CRANFIELD UNIVERSITY

K. PAXTON-FEAR

UNDERSTANDING INSIDER THREATS USING NATURAL LANGUAGE PROCESSING

SCHOOL OF DEFENCE AND SECURITY
PhD in Defence and Security

PhD
Academic Year: 2020-2021

Supervisors: Dr D. Hodges, Dr O. Buckley
February 2021

CRANFIELD UNIVERSITY

SCHOOL OF DEFENCE AND SECURITY
PhD in Defence and Security

PhD

Academic Year: 2020-2021

K. PAXTON-FEAR

Understanding Insider Threats Using Natural Language Processing

Supervisors: Dr D. Hodges, Dr O. Buckley
February 2021

This thesis is submitted in partial fulfilment of the requirements for the degree of PhD.

Abstract

Insider threats are security incidents committed not by outsiders, such as malicious hackers or advanced persistent threat groups, but instead an organisation's employees or other trusted individuals. These attacks are often more impactful than incidents committed by outsiders. Insiders may have valid security credentials, knowledge relating to the organisation they work for (such as competitors), knowledge of security controls in place and potentially how to bypass those controls. This activity could be unintentional, such as an employee leaving a laptop on public transport, or malicious, when an insider purposefully chooses to attack for some gain, such as selling IP to a competitor. When an outsider chooses to attack, they may leave digital breadcrumbs as they perform various stages of the cyber kill-chain. These breadcrumbs can allow organisations to detect and respond to an incident, flagging suspicious behaviour or access. Comparatively, an insider may be able to continue their attack for years for being caught. Therefore, insider threat activity can be considered co-spatial and co-temporal with legitimate activity; an insider conducts their attack during their work or very soon after leaving their jobs.

There are three fundamental approaches to control the risk of malicious insider threats: organisational, technical, and psychological. More recently, insider threat models attempt to encapsulate all these factors into one approach, combining all these into a single framework or model. However, one issue with these models is their static nature; models cannot adapt as insider threat changes. For example, during the COVID-19 Pandemic, many organisations had to support remote working, increasing the risk of attacks. This work attempts to address this flaw of models directly. Instead of attempting to supplant existing practices in these three domains, this work will support them, providing new techniques

for exploring an insider threat attack to better understand the attack through the lens of strategic and tactical decision making. This dynamic, custom insider threat model can be constructed by leveraging natural language processing techniques, a type of machine learning completed on text, and a large corpus (body of documents) of news articles describing insider threat incidents. This model can then be applied to a new, previously unseen corpus of witness reports to offer an overview of the attack. The core technique this work uses is topic modelling, which uses word association to identify key themes across a document, similar to grounded theory approaches. By identifying themes across many different insider threat incidents, the core attributes of insider threat are recognised, such as methodologies, motivations, information about the insider's role in an organisation or the weakness they exploited. These topics can be further enriched by identifying temporal, casual and narrative clues to place events on a graph and create a timeline or causal chain.

The final output of this process is a collection of visualisations of the incident; this visualisation then aims to support the investigator as they ask critical questions about an incident, such as "What was the motivation of the insider?" "What assets did they target and how?" "Were there any security controls in place?" "Did they bypass those?" allowing for the full exploration of the attack. Informed organisations can make changes using the answers to these questions combined with existing controls, policies, and procedures.

The work presented in this thesis has many implications for both insider threat specifically and the broader domains of sociology and cyber security. Primarily this work introduces a new approach to incident response, supporting the reflection stage of incident response. While this work represents a proof of concept for NLP to be used in this way, due to the technical nature of this work, it could be improved to produce an implementable and deployable piece of software, generating further impact, while there would be some necessary training required, this could offer a new tool for handling insider threat within an organisation. Aside from this direct impact in the insider threat domain, the methods developed and designed during this work will have a broader impact on cyber

security, mainly due to its interdisciplinary nature within social science. With the ability to leverage witness reports or organic narratives and map these automatically to an existing framework, rather than ask a witness to adapt their narrative to a framework directly. Reports can then be collected on a large scale and analysed. These techniques provide a holistic view of an attack, considering many aspects of an insider threat attack by using reports already collected after an incident to create a better understanding of insider threat which leads to more techniques in prevention and detection.

Keywords

Insider threat; natural language processing; reports; organic narratives; topic modelling;

Papers Published

- K. Paxton-Fear, D. Hodges and O. Buckley (2018a). 'Connected events and malicious insiders: Investigating patterns of insider threat using natural language processing'. In: *Behavioural and Social Sciences in Security*
- K. Paxton-Fear, D. Hodges and O. Buckley (2018b). 'Increasing the accessibility of NLP techniques for Defence and Security using a web-based tool'. In: DOI: 10.17862/cranfield.rd.10066229.v1. URL: https://cord.cranfield.ac.uk/articles/poster/Increasing_the_accessibility_of_NLP_techniques_for_Defence_and_Security_using_a_web-based_tool/10066229/1
- K. Paxton-Fear, D. Hodges and O. Buckley (2020). 'Understanding Insider Threat Attacks Using Natural Language Processing: Automatically Mapping Organic Narrative Reports to Existing Insider Threat Frameworks'. In: *HCI for Cybersecurity, Privacy, and Trust*. Springer International Publishing

- K. Paxton-Fear, D. Hodges and O. Buckley (2019). 'Using Topic Distribution to Classify Fuzzy Topics'. In: *In Review Human-centric Computing and Information Sciences*

- K Paxton-Fear, D. Hodges and O. Buckley (2021). 'Visualizing an insider threat incident from witness reports using natural language processing'. In: *Conference on Applied Machine Learning for Information Security 2021*

- A system for creating custom insider threat models from witness reports (in preparation)

Contents

List of Figures

List of Tables

List of Abbreviations

NLP	Natural Language Processing
ML	Machine Learning
AI	Artifical Intelligence
IDS	Intrusion Detection System
IP	Intellectual Property

Acknowledgements

I would first like to thank my supervisors Duncan Hodges and Oliver Buckley for their mentorship and guidance throughout my PhD. DSTL for their generous funding which allowed me to perform this research. Alex Gibb who has been a mentor for many years, and whos guidance and sage advice led me down this path.

I would like to thank my family for their support. Especially partner Richard for his unwavering support throughout my PhD, during the difficult time that 2020 presented us with. My mum who has no idea what my PhD is in, but who listens and tries to help anyway. My dad and my brother who both know slightly more and are always cheering me on from the slide lines

My friends who have had double lives as shoulders to cry on, people to relax with and cheerleaders. Pierre, Aniket, Michael, Michael, Liz, Penny, Benjy, Tyler and Daryl.

Chapter 1

Introduction

Insider threats are security threats that are not perpetrated by external actors such as Advanced Persistent Threat groups (APTs) but instead by internal actors, such as employees or contractors. These individuals often have valid credentials, are experts in the computer systems at their employers and may be aware of valuable assets or systems to competitors. Therefore these insiders (employees and trusted contractors) can be one of the most impactful security threats an organisation may face (Randazzo et al. 2005). However, this is not always malicious in the case of accidental insider threats; this may be that an insider may be forgetful or unaware of security practices and accidentally bypassing security controls. Furthermore, while a malicious actor may attempt many stages of the Cyber Kill Chain such as reconnaissance techniques (Yadav and Rao 2015), which may leave a digital fingerprint as they attempt to gain access or information about a system. In contrast, insiders commit their attacks during their employment, where even legitimate behaviour can mimic malicious behaviour. For these reasons, insider threats can often be more challenging to investigate. The challenge is particularly visible by examining reports of data breaches; data breaches involving insiders are prevalent, accounting for between 10-20% of data breaches, both accidental and malicious (Verizon 2019; Verizon 2020).

When organisations consider security threats, security teams often focus on external

threats, using formal threat models, models of threat actors or threat intelligence briefings (Mohamed and Belaton 2021) and their behaviour in the hopes of identifying likely malicious actors and exploitation paths. However, trusted individuals, an organisation's own trusted employees or contractors, are not usually included in these threat models. This oversight can also lead to more damage over an extended period. Unlike the external actors leaving breadcrumbs as they perform their attack, which are detected by typical intrusion detection systems allowing security teams to catch an intruder easily and quickly. Insiders have valid credentials, may also have in-depth knowledge of the business or IT systems and can hide their activity (incidentally or maliciously) alongside legitimate behaviour and this behaviour may also be present for long periods of time.

Many approaches have been investigated, developed and deployed to manage the threat posed by insiders, particularly in technical controls identifying anomalous behaviour (mirroring the advances in defences for external threats). However, these controls are often not enough, as insider threats continue to rise (Verizon 2019), with insiders sometimes able to circumvent these tools (either knowingly or not) or a potential organisational mistrust in the tools due to over alerting leading them unimplemented or when implemented, ignored (Forte 2019). In addition to technical solutions, there are various linguistic, sociological and psychological approaches to understanding the 'why' and who of insider threat. Revolving around the question of 'why did that attack occur' and, more broadly 'What are the traits of an insider?' 'How could they have been detected sooner?' This lack of knowledge and understanding has led to many organisations citing lack of information as a barrier to managing insider threat attacks (Forte 2019), particularly in preventing and ultimately detecting an incident. There is a gap between the research such as Cappelli, Moore and Trzeciak (2015), and the incident response and IT governance practitioners (Forte 2019). Although this gap has been sought to be addressed using advanced technical controls in research and software (Gavai et al. 2015; Agrafiotis et al. 2015; Young et al. 2013; Senator et al. 2013), even with these technical controls, the number of incidents is still high, suggesting that insider threat is still complex and challenging

to detect, even when using state of the art technical controls such as machine learning (ML) systems. One common approach for managing insider threat revolves around the use of models; these encapsulate many elements of insider threat into a single model capturing the psychological, social, technical and organisational elements, such as Nurse et al. (2014b), Trzeciak (2011) and Randazzo et al. (2005)

This PhD presents three technical capabilities for a system that leverages NLP to understand a large corpus of witness reports after an insider threat incident. These reports can then be mapped to the existing state of the art models. These reports are routinely collected after an incident but are usually considered atomic. This system will consider the reports as a whole instead of considering each statement an atomic document, aggregating these together and visualising the result. By applying Natural language processing techniques the large volume of reports can be analysed and adapted. The NLP model can be more dynamic by training a model on a general corpus (body of documents) of many different incidents, with different insiders, organisations, motivations, and methodologies adding to this corpus as insider threat attacks evolve and change. Then the model can be applied to the previously unseen corpus of witness reports, in this case, an example insider fraud case from the existing literature. The ultimate aim of this is to support the creation of a custom, dynamic insider threat framework, which builds upon the existing work on insider threat models such as Nurse et al. (2014b). This model was built using a grounded theory approach, however this thesis instead uses topic modelling instead finding the thematic characteristics computationally. Individual reports can be aggregated, analysed, visualised and mapped to an existing insider threat model, even if those reports differ in detail, language use, formality and writing style. Using these visualisations, investigators can pinpoint specific methodologies, critical events, assets or conspirators and implement managerial, IT governance or policy changes to prevent the next attack. To demonstrate the effectiveness, this has initially been applied to a case of insider fraud from the existing literature; however, this work represents an initial proof of concept, and this could be expanded further in the future with further case studies from other archetypes and the

literature. The system has three core capabilities, representing the technical objectives:

1. Map a corpus of organic narratives automatically to an insider threat model. Creating a 360- degree view of an attack whilst also reducing cognitive bias and cognitive load. Using topic modelling and topic segmentation to segment and classify text by topic. Allowing sentences from different reports, but that describe the same event to be labelled.

2. Identify causal, temporal or incidental relationships between these topics. Showing how the characteristics of an attack relate to each other. Creating a custom insider threat framework, which is then visualised using several graphs.

3. Deconstruct a single topic to visualise and identify the themes and core ideas that the computational model has found. Creating a single visualisation for each characteristic by aggregating the sentences in a single topic.

This approach blends the technical and sociological approaches considering the attack as a whole rather than focusing on a specific characteristic or approach. The use of organic narrative reports (those written in a natural narrative style rather than formal reports) and natural language processing allows for a 360-degree view of an attack. This process considers all the elements that may have influenced an attacker, from technical methodologies to external factors, to allow investigators to react to an incident implementing policy, technical or procedural, changes to mitigate the next attack. In contrast to many state of the art approaches which focus on technical approaches to detecting an attack in progress by analysing technical artefacts. While this work focuses on understanding an attack rather than detection, understanding insider threat as a whole leads to improved detection and mitigation capabilities. Therefore, this work's specifically targets incident response practices and supports the reflection stage, providing a retrospective look at an incident evidence to support organisational or IT governance changes within an organisation. This approach's significant benefit and novelty is the use of NLP to analyse witness reports. Allowing a witness to freely write about their experience without

needing to adapt their report to a formal model reduces cognitive load and bias towards the model. In addition, this process would allow for the collection of more reports from different individuals than typical technical reports, which may contain details that would be missing otherwise. In contrast to reports with a prescribed style, this approach is used to map reports into the formal model, rather than to ask someone to fit their narrative to a model. The impacti of which is significant for insider threat but also any field that uses formal models and organic narratives.

This thesis presents the following chapters describing the development and integration of these approaches. First, the background section presents insider threat research and the core issues facing insider threat research, identifying the core literature gap. Next, the aim of this project will be discussed, and the technical objectives set. The methodology section will then introduce natural language processing and the tools and techniques used. The following four chapters represent each technical objective. First, the data collection process detailing the datasets used and how they were collected. Next, chapter 6 concerns the creation of topic models and the mapping process to an insider threat model. Next, the causality and temporality chapter demonstrates how each computational topic can be contextualised with others to produce a custom insider threat model, creating a narrative, causative and temporal model. Finally, the topic analysis objective shows how a single topic can be analysed to understand the core themes. Throughout all of these objectives, both the technical methodology, visualisations and tooling are discussed. While these represent a proof of concept for using NLP for these problems and are an initial first step, the technical objectives demonstrate the effectiveness of this approach. Next, the Discussion chapter presents a critical analysis of this work and details the implication and impact of the research. Next, the Future Work chapter describes the future research that this project could produce, acknowledging the issues and suggesting routes to improve them and the impacts these improvements could have on other fields. Finally, a conclusion and statement of contribution describe this PhD's academic contribution and this research's novelty.

Insider threat remains challenging to detect, this work presents new methods and approaches to support natural language processing in the insider threat domain. Although this remains in its infancy, it represents small steps in exploring NLP within the insider threat domain and focusing on supporting expertise within organisations. Several improvements could be made to this research, allowing the techniques to be operationalised, such as developing a software package and training investigators to use it.

Chapter 2

Background

The traditional image of an insider threat is that of a vengeful employee who maliciously steals insider information for their own gain while irrationally angry at a small slight committed by their employer. While malicious insider threats are often caused by a breakdown in the relationship between an employer and employee, the reasons and motivations are a large complex web of interacting characteristics. Examples include the building of resentment, insiders' historical behaviour, tardiness or other negative attitudes toward work and personality. While they may be considered normal, individually combining these characteristics in a single individual may ultimately lead to them committing an insider attack. There is a large volume of work that attempts to understand insider threat attacks from different perspectives, including psychological, technical and sociological approaches (Elmrabit, Yang and Yang 2015; Greitzer et al. 2012). Technical approaches often use complex tools to monitor anomalous behaviour, sociological approaches attempt to monitor subtle change when an insider intends to commit an attack, and psychological approaches look into an insider's past and personality. Insider threat models encapsulate all these aspects of an attack (organisational, psychological, technical and sociological) in the hopes of understanding the attack as a whole while recognising the importance of each characteristic. However, models can often be particular and difficult to implement in an organisation, requiring expertise and experience to aid in their use. The work in this

thesis builds upon the existing state of the art models and considers a holistic approach to addressing insider threat with organisation focused techniques and tools.

2.1 Introduction

Insider threat is a security threat perpetrated by internal actors, such as an organisation's employees, contractors or other trusted individuals, in contrast to the traditional image of security threats as external actors. Defined as 'a current or former employee, contractor, or business partner who has or had authorised access to an organisation's network, system, or data and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organisation's information or information systems' (Cappelli, Moore and Trzeciak 2015). Insider threats can be broken into two types: A malicious insider, who purposefully attacks their employer or former employer and an unintentional insider, whose actions lead to an attack on their employer or former employer, but this is not the intended outcome (visualised in Figure 2.1). Many insiders have access to valid security credentials, and even if they may not be technical, can have in-depth knowledge of a system and potential technical controls, and potentially, how to bypass them. In other cases, the security or IT team may not follow security principles like the principle of least privilege, compartmentalisation of separate systems or the Zero-Trust Security Model. The management of insider threat often lies with all members of an organisation rather than just IT, with managers and colleagues able to report to handle disgruntled employees before they attack, but it can be challenging to recognise and act (Cappelli, Moore and Trzeciak 2015). Insiders can vary in role, motivation and technical skills; therefore their attacks vary in methodology, outcome and the assets compromised. Because of this variation and the difficulty in managing the risk posed by the variation, it is clear that insider threat attacks are a nuanced security threat (Hunker and Probst 2011). A 2019 report compiled by Verizon (Verizon 2019) suggests that 34% of all data breaches involved an internal actor and this continued in the 2020 report with 30%

of breaches involving internal actors (Verizon 2020).

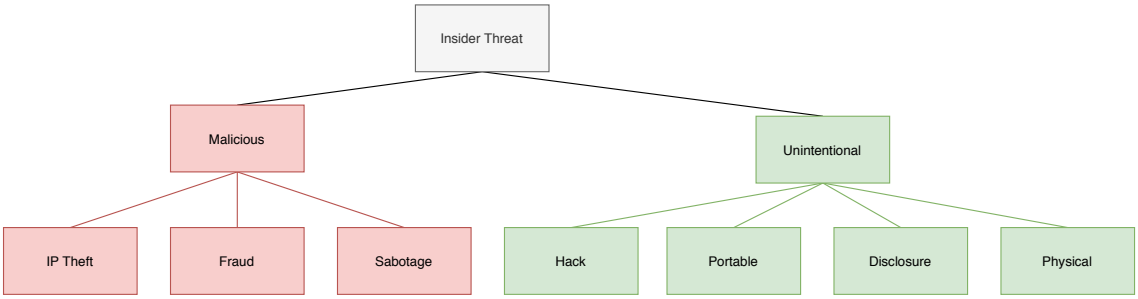


Figure 2.1: Types of Insider Threat

When discussing insider threats, is it important to understand the difference between the two types, malicious and unintentional, each with their archetypes. An unintentional insider does not deliberately attempt to compromise the security of their employer. The definition of unintentional insider threat 'An unintentional insider threat is (1) a current or former employee, contractor, or business partner (2) who has or had authorised access to an organisation's network, system, or data and who, (3) through action or inaction without malicious intent, (4) unwittingly causes harm or substantially increases the probability of future serious harm to the confidentiality, integrity, or availability of the organisation's resources or assets, including information, information systems, or financial systems.' (Greitzer et al. 2014). However, this harm can be caused by several factors, including inaction, action, or knowingly breaking the rules (Greitzer et al. 2014). Greitzer highlights several potential concerns and mitigations for managing unintentional insider threats, demonstrating that often intentional and malicious insider threats have the same root causes, such as drug use, risk-taking, stress and anxiety.

These two types can then be broken down further into the malicious archetypes, IP Theft, Fraud, Sabotage and the unintentional incident types Hack, Physical, Portable and Disclosure. Each of these requires a different approach to defence and mitigation. The work presented in this thesis primarily focuses on the malicious insider threat; however, it is crucial to understand both malicious and unintentional insider threat and the archetypes within these broad definitions. Insider threat is nuanced at its core, and these boundaries may not be as distinct as they may seem, especially for cases that may fall between arche-

types or involve a mix of malicious and unintentional actions. Therefore, it is crucial to consider both types and their subtypes to fully understand the risk that insiders pose.

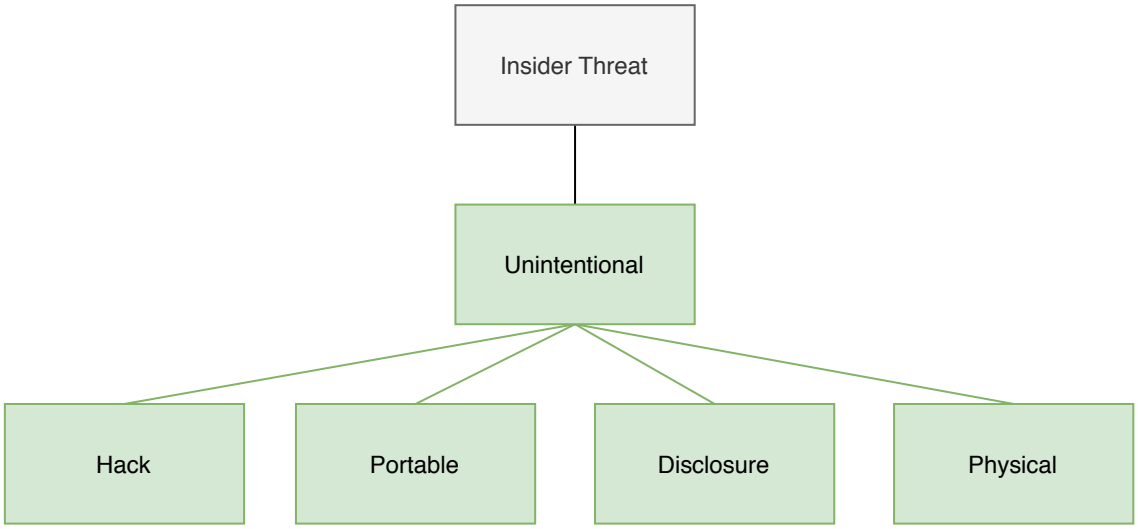


Figure 2.2: Unintentional Insider Threat

2.1.1 Unintentional Insider Threat

Unintentional insider threat is usually broken into several types: Hack, Physical, Portable and Disclosure seen in Figure 2.2. Hack refers to an attack where an insider's accounts or privileges are targeted; this is usually through social engineering or phishing attacks or devices containing malicious code which is executed when plugged in (Greitzer et al. 2014). Hack type attacks primarily target users with limited technical knowledge and can potentially be mitigated by providing training and support to staff members (Greitzer et al. 2014). Physical attacks are those where physical data is lost, stolen or improperly discarded; these attacks refer to physical data such as printed documents rather than electronic (Greitzer et al. 2014). The next type of attack, Portable, is similar. This involves lost, improperly disposed and stolen devices rather than documents; these include phones, laptops, PDAs, as well as hard drives and memory devices (Greitzer et al. 2014). Finally, unintentional attacks related to the Disclosure archetype are when sensitive data is accidentally leaked, such as when information is sent to an incorrect email or posted publicly (Greitzer et al. 2014).

The causes of unintentional insider threat are varied and combine human and organisational factors with broader psychological and social issues (Greitzer et al. 2014). These include data flow, work setting, work planning and control and employee readiness. if these core issues are mismanaged, there are risks of these becoming contributing factors to insider threat, such as high levels of stress or time pressure, fatigue and high workloads of employees or lack of organisation or technical controls on the organisation's critical systems. Greitzer et al. (2014) suggests that these issues can then lead to severe consequences, such as reduced working memory or lack of attention. These consequences can then increase the risk of accidental disclosure or increase susceptibility, which may increase the risk of a Hack incident (Greitzer et al. 2014). In addition, increased risky behaviour and decision making can also increase the risk of many types of incidents due to intentional disregard for rules and procedures, including intentional (Cappelli, Moore and Trzeciak 2015) and unintentional insider threat (Greitzer et al. 2014). Finally, personality types and gender may further increase the risk of some attack types (e.g. Hack), as some demographics may also be more susceptible to attacks (Greitzer et al. 2014).

Human Factors and Training	High-Level Organizational Best Practices	Automated Defense
• Enhance awareness of insider threat and UIT. • Heighten motivation to be wary of UIT risks. • Train employees to recognize phishing and other social media threat vectors. • Engender process discipline to encourage following of policies and guidelines. • Train continuously to maintain proper level of knowledge, skills, and ability. • Conduct training on and improve awareness of risk perception and cognitive biases that affect decision making. • Improve usability of security tools. • Improve usability of software to reduce likelihood of system-induced human error.	• Review and improve management practices to align resources with tasks. • Improve data flow by enhancing communication and maintaining accurate procedures. • Maintain productive work setting by minimizing distractions. • Provide effective security practices (e.g., two-factor authentication for access). • Implement effective work planning and control to reduce job pressure and manage time. • Maintain employee readiness. • Maintain staff values and attitudes that align with organizational mission and ethics. • Implement security best practices throughout the organization.	• Deploy better software to recognize bogus emails. • Deploy data loss prevention software to recognize potentially harmful sites and email practices. • Use firewalls. • Use virus and malware protection software. • Enable remote memory wipe for lost equipment.

Figure 2.3: Mitigations for unintentional insider threat(Greitzer et al. 2014)

The mitigation for unintentional insider threat attacks can be broken down into Human factors and training, organisational best practices and automated defences. These are shown in Figure 2.3, from (Greitzer et al. 2014). For defensive measures, the focus is on awareness and risk, offering employees training to ensure that best practices are followed and risk perception is not impaired. For organisations, communication, work

planning, and best practices ensure employees are not experiencing issues from being overworked and any guidelines are communicated clearly to employees. Finally, automated defences help reduce the risks of unintentional insider threat without intervention, blocking harmful software, virus and malware scanning and email scanning to help find malicious and phishing emails. These organisational changes can help reduce the risk of unintentional insider threat. However, insider threat is nuanced, and the stress that these controls and training apply to employees and the ethical considerations of implementing them can cause unintended side effects, e.g. making insider threat more likely (Greitzer et al. 2014; Reeves, Parsons and Calic 2020; Palm 2009; Greitzer, Frincke and Zabriskie 2011). For example, employees being sanctioned for clicking on phishing links, or being overwhelmed by training (Reeves, Parsons and Calic 2020). These controls, therefore, may even increase the risk of breaches rather than decrease them if not implemented correctly.

2.1.2 Malicious Insider Threat

The majority of the literature on insider threat discusses malicious insider threat, as opposed to unintentional insider threat. In contrast malicious insider threat is committed by an employee purposefully to damage their employer or gain something. Malicious insider threat can include a range of impacts to employers, from direct financial to national security, including a loss of reputation for the employer. The CERT MERIT models (Cappelli, Moore and Trzeciak 2015) describe three archetypes of insider threat, Sabotage, IP Theft and Fraud; these models represent different insiders, targets, methodologies. The MERIT Models (Cappelli, Moore and Trzeciak 2015) are based on a dataset of real insider threat cases across multiple industries and form the backbone of all state-of-the-art insider threat research. Therefore, it is essential to understand each archetype and answer the key questions: Who are the insiders? How do they commit their crimes? Why do they commit them? What impact do these attacks have? Moreover, what steps can be taken to prevent them? By answering these questions, a further understanding of insider threat

can be gained.

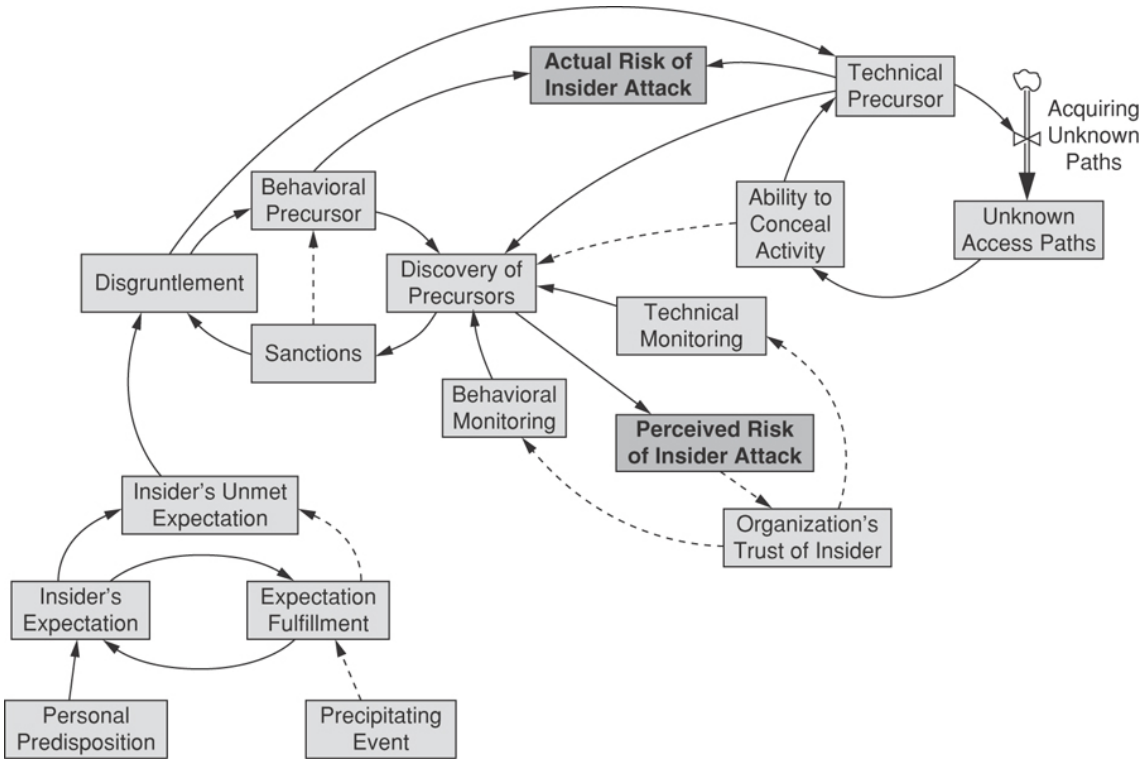


Figure 2.4: MERIT Model Insider Sabotage (Cappelli, Moore and Trzeciak 2015)

Insider Sabotage shown in Figure 2.4 is an insider threat attack that targets an organisation's technical infrastructure; an insider with privileged access to the IT infrastructure may plant a logic bomb, malicious code designed to intentionally damage systems. These insiders are often technical individuals, programmers or system administrators due to the complex methodologies needed to carry out these attacks (Cappelli, Moore and Trzeciak 2015; Keeney et al. 2005). Usually, these insiders can commit their attacks using their existing privileges, creating additional accounts that the organisation is unaware of giving them a backdoor. These insiders may have a predisposition to attack, with previous issues with law enforcement or their employer, but these attacks are often motivated by unmet expectations and disgruntlement (Cappelli, Moore and Trzeciak 2015). Sabotage can significantly impact the organisation, both financial (forming a significant motivation of many insider attacks) and business losses, specifically in losing customer trust. Cappelli, Moore and Trzeciak (2015) calls explicit attention to the customer loyalty lost

in a bank when their customers cannot access ATMs and the personal stress this would cause to the customers. Preventions primarily suggested for insider sabotage are the management of both the hiring and departure of employees, ensuring background checks are completed and that accounts an employee has made during their job are removed from IT systems and that an insider does not have any access after they leave their job. Finally, the MERIT models propose various controls to manage disgruntlement, consistent consequences for rule-breaking, open communication between employer and employees, identifying concerning behaviour early and ensuring that these controls are reflected on at regular intervals. These steps manage insider threat behaviour via prevention, stopping insider threat before an attack occurs. In contrast to other types of insider threat, insider sabotage is challenging to manage with technical controls, as insiders who commit these attacks are usually the individuals who install and control technical controls and may bypass any form of monitoring. Therefore, insider sabotage remains an incredibly impactful insider threat to combat and protect against, with lasting financial and reputational impact (Cappelli, Moore and Trzeciak 2015).

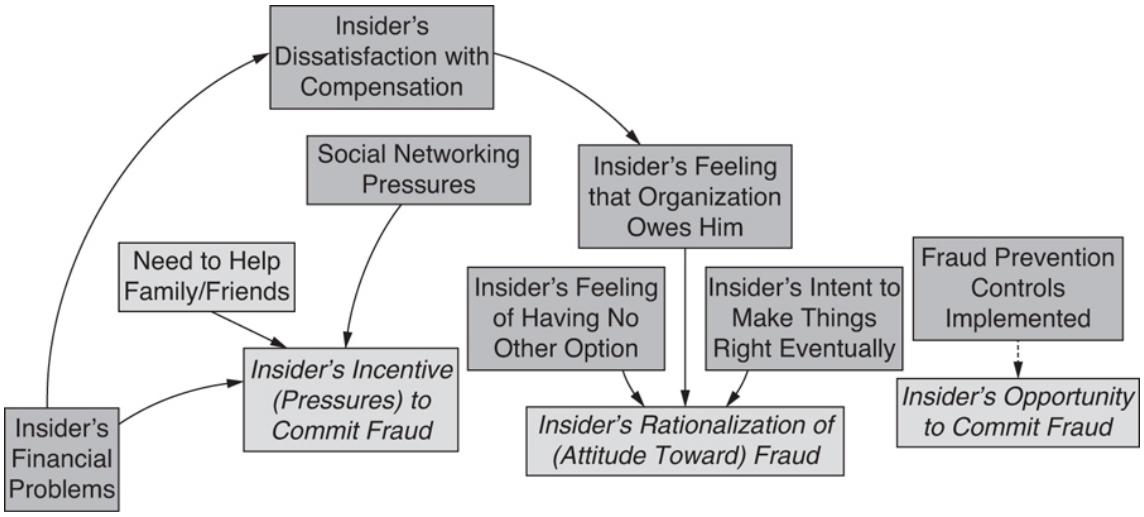


Figure 2.5: MERIT Model Insider Fraud (Cappelli, Moore and Trzeciak 2015)

Insider fraud occurs when an insider uses their position to commit an act of fraud, shown in Figure 2.5. Unlike other cases of insider threat, this often occurs during an insider's employment rather than after an employee leaves or plans to leave. An example

of this archetype is a utility company employee who allowed fraudulent meter readings costing their employer $325,000. This type of insider threat can also involve outsiders, including organised crime; in these cases, the insider is either recruited or targeted by the outsider to relay, change, add or delete information, for example, Tidy and Molloy (2020). These insiders are usually not technical members of staff but may have, due to their job, a large amount of access to privileged information. These insiders primary motivations are to benefit themselves; the entire motivation to commit fraud is best represented by the Fraud triangle shown in Figure 2.6

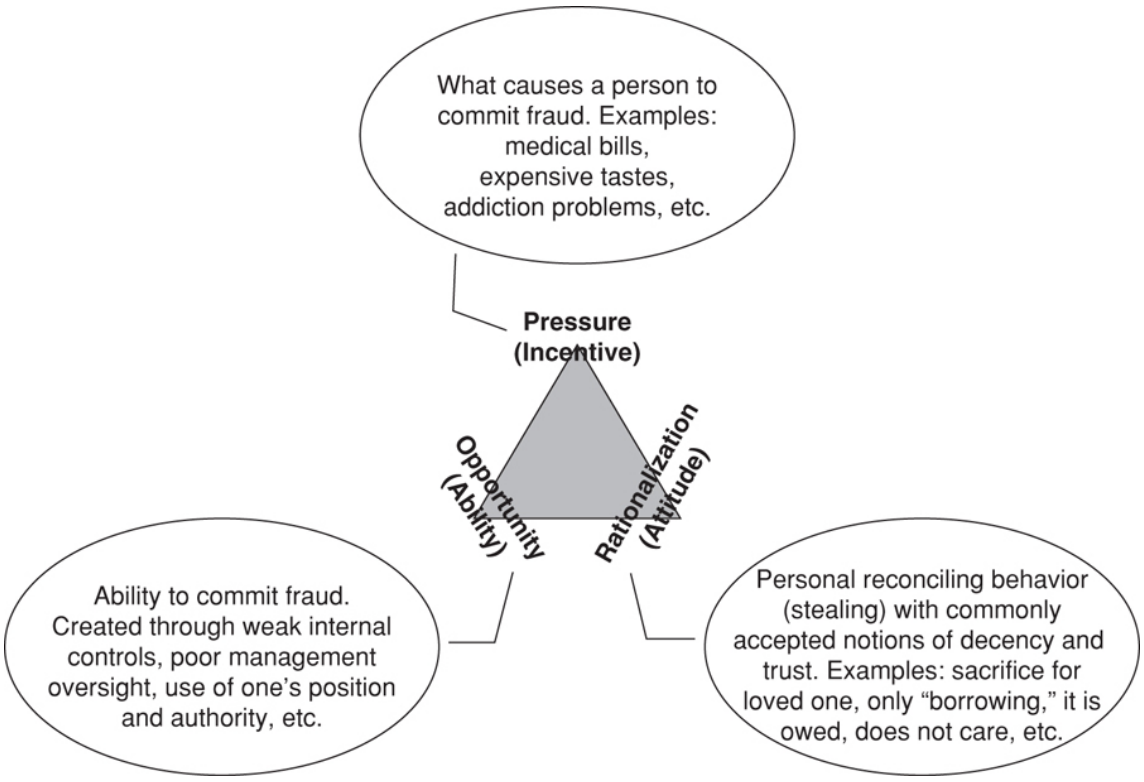


Figure 2.6: Fraud Triangle applied to Insider Fraud (Cappelli, Moore and Trzeciak 2015)

The primary motivation for insiders who commit fraud is financial pressure, usually brought on by debts or bills or sometimes pressures such as addiction. They can commit their attacks due to: opportunity, the ability to commit crime and rationalisation successfully, that it is 'okay' to commit these crimes, as a sacrifice for a loved one, a view that they are borrowing or getting what is owed to them. These core elements of the fraud triangle create the psychological environment where an insider can commit fraud. This

is comparable to other types of fraud, which can be analysed using Maslow's Hierarchy of Needs, for example, Cendrowski et al. (2007) discusses a link between an employees lack of recognition and the rationalisation of fraud. In addition, lower needs, for example, the need for shelter may also drive an individual towards fraud. Insider Fraud has a range of pressures and impacts, direct financial impacts, data protection, such as GDPR (*Data Protection Act* 2018) and in some cases, national security. These attacks can cause a significant reputational loss for an organisation or country. These attacks are not technically sophisticated and usually involve insiders physically or electronically copying or modifying data, for example, through email, removable media, telephone, or download. In preventing these attacks, organisations should attempt to disrupt the fraud triangle. For example, adding security controls can limit an insider's opportunity to attack. Recognising financial problems early and mitigating with a financial assistance program to manage the financial pressure that leads insiders to attack. Regular auditing and background checks during recruitment can further recognise potential insider threats. Auditing, in particular, can recognise modification of data. Insider fraud can be an incredibly damaging and impactful attack, as these can occur for years without detection.

Insider intellectual property theft shown in Figures 2.7 and 2.8 is the final malicious insider threat archetype, this type of insider threat primarily involves insiders stealing trade secrets from their employer. IP theft can also involve foreign governments, with insiders having competing interests between their organisation and another country, e.g. country of birth or family residence. These insider threat cases are usually committed to benefit the insider's future employment, stealing intellectual property (IP) either for a new job or to start their own company. They are committed by the individuals who create the IP, such as programmers, engineers, scientists. However, sometimes it can be committed by others such as sales staff. There are two key types of insider IP theft, the ambitious leader, who recruits others intending to develop a competing product or relay to a foreign organisation or government and the entitled independent who plans to move to a new job and sees IP theft as an advantage. This type of insider threat can also be directed

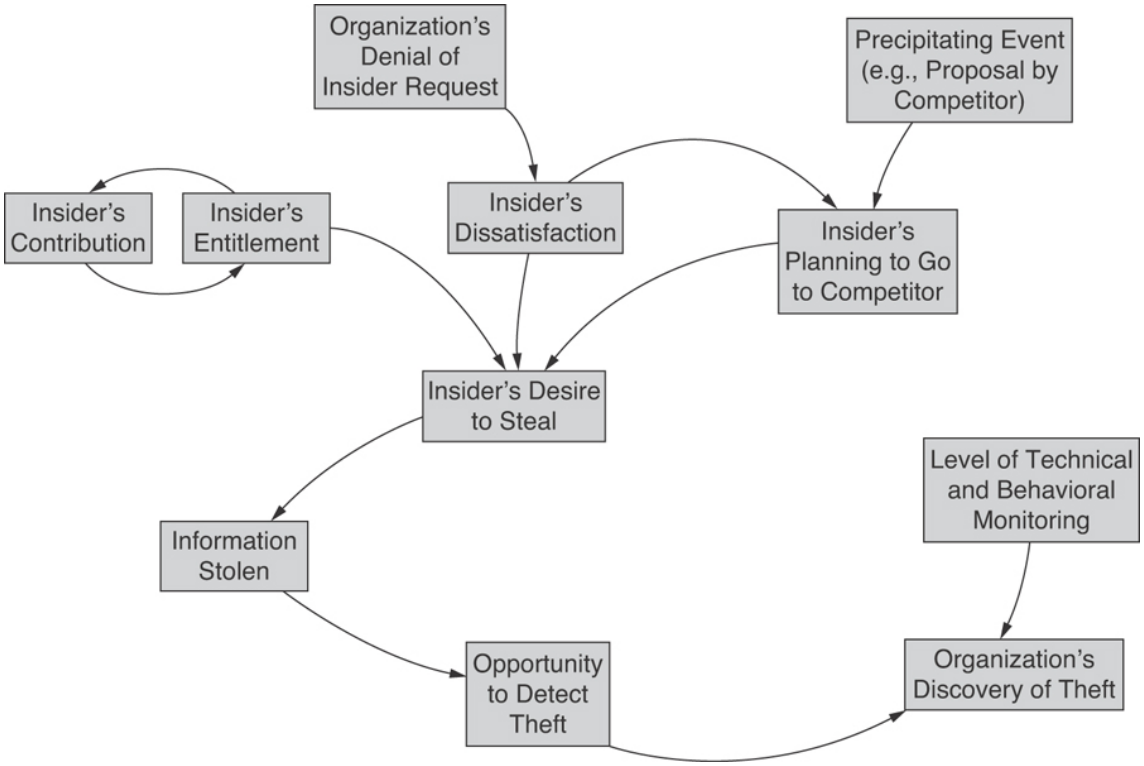


Figure 2.7: MERIT Model Insider IP Theft: Entitled Independent (Cappelli, Moore and Trzeciak 2015)

by another company that purposefully attracts talent from a competitor. These insiders' motivations are a mix of dissatisfaction with their current job and the insider feeling entitled to the IP because of their high contribution level. Ambitious leaders' motivation is usually linked to the desire to create a competing product or foreign interference loyalty to another country. The theft is usually carried out using removable media, email, leaking data, printing/copying and removing documents from the building. The prevention of this insider threat attack is primarily technical, using watermarks or DRM (Digital Rights Management) and continually scanning network traffic. Another important control that can be implemented is the management of dissatisfaction and the termination process, similar to other archetypes. These attacks can have a wide range of impacts, financial, loss of trade secrets to competitors or foreign powers and intelligence.

These are the core archetypes; however, insider threat is changing, and some cases fit between archetypes. Two such issues are Insider Threat from Trusted Business Partners and Insider Threat on the Internet Underground. Insider Threat from Trusted Business

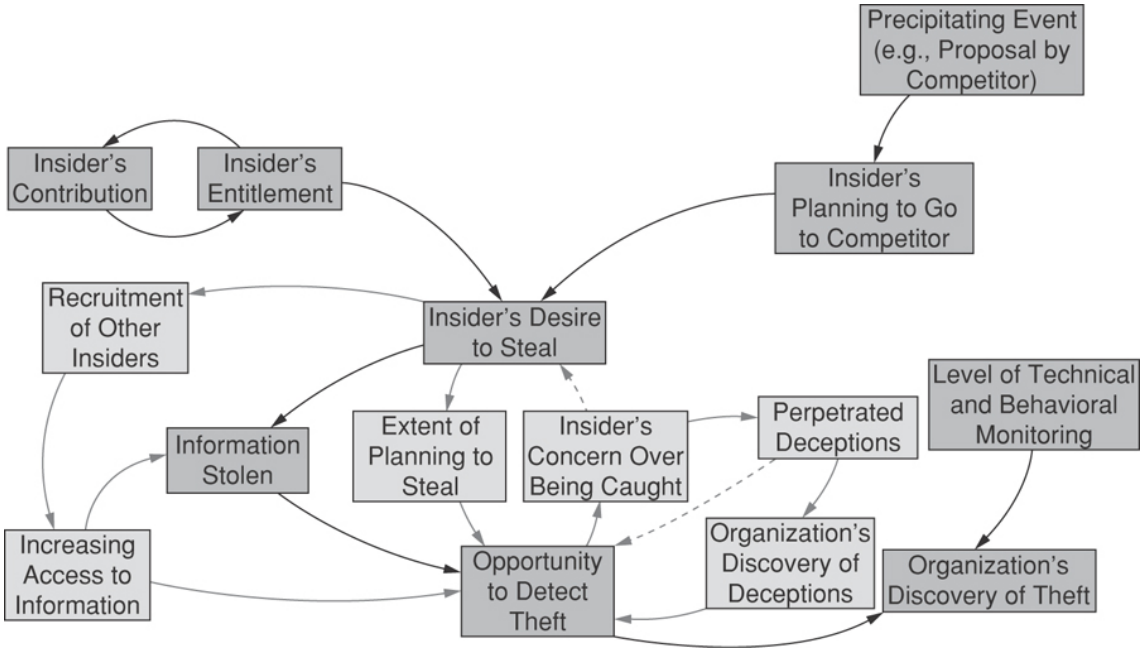


Figure 2.8: MERIT Model Insider IP Theft: Ambitious Leader (Cappelli, Moore and Trzeciak 2015)

Partners includes the risk posed by services and companies an organisation uses in a supply chain, such as accountants. These external companies may have access to large amounts of proprietary business information or trade secrets, and there is a trust relationship between both organisations. The external organisation's employees can become an insider threat; therefore, ensuring that any external organisation has procedures for insider threat in place is particularly important. Insider Threat on the Internet Underground involves malicious insiders with ties to the Internet Underground, such as hacking forums or other websites associated with the deep and dark web. These insiders may sell the data or, in some cases, distribute and brag. This behaviour can escalate to asking outsiders to attack the organisation when disgruntled.

2.1.3 Conclusion

Insider threat detection is often cited as a major issue and barrier to managing insider threats. Insider threat activity can be challenging to detect for the following reasons:

1. Insider threat activity is often co-temporal and co-'spatial' with legitimate activity,

taking place in the same systems and the same time, especially for cases of insider fraud.

2. The behaviour associated with insider threat activity is often similar to legitimate activity (e.g. copying files, IP theft).

3. Insiders may be aware of the security protocols, processes and technical controls and therefore able to bypass them, especially for sabotage, IP theft and fraud

The risk and the difficulty in detection make insiders threats very impactful and, if successful, incredibly costly to the organisation affected and can cause damages such as 'financial, reputation, organisational disruption, long term impact on an organisations culture' (Liang, Biros and Luse 2016). Much of the research in insider attacks is tasked with mitigating the insider threat by understanding how they attack and why they attack (Homoliak et al. 2019). Despite the vast literature of insider threat work, insider threat remains a complex problem, with breaches throughout 2019 and 2020 caused by insider threat (Verizon 2019; Verizon 2020). Breaches are both frequent and impactful, with financial, reputational and data protection issues arising from them. In particular, it seems clear that it is challenging to manage potential insiders and intervene before an attack to prevent an attack, focusing on containing, responding and recovering after a breach.

Although each archetype recognises the differing attributes of an attack, both of the insider and their target, the insider differs in their job, method and asset attacked. The organisations differ in technical protections and mitigations, industry, and there are many similarities between attacks. These elements, the insider's disgruntlement, the access to IT systems and the behavioural and technical controls, are similar across each. However, much of the research considers insider threat as a whole, combining archetypes and sometimes both accidental and malicious insider threats.

For organisations, the vital issue of insider threat is detecting, preventing, and managing the threat. There are several different views on this issue, organisational policy, technical solutions and psychological understanding, each offering a different lens and

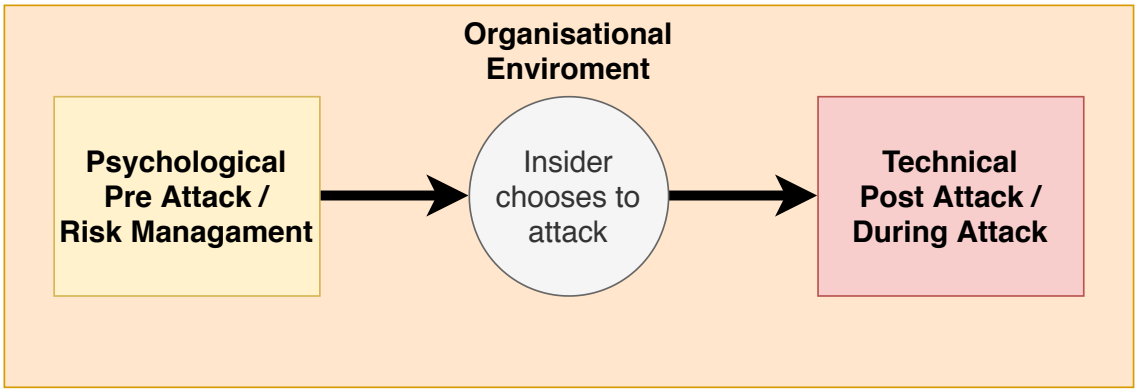


Figure 2.9: Where do the approaches to insider threat take place?

solutions to managing insider threat. Insider threat models, such as the MERIT models, attempt to encapsulate all of these views, providing a method for retrospection and policy changes.

2.2 Managing Insider Threat

In mitigating and managing insider threats, organisations and organisation structure play a huge role. The approaches an organisation takes represents the environment of the attack. Organisations play a role in managing employees, either reducing or mitigating disgruntlement, creating trust and loyalty. Their overall goal is to recognise an insider threat before the insider attacks and intervening, therefore preventing an attack or investigating insider attacks to prevent the next. The role members have in an organisation are extremely important in managing insider threat attacks, especially in preventing attacks. If a manager can step in and provide positive intervention, for example, if an insider does not get a promotion, offering a raise or additional responsibilities can reduce the escalation of disgruntlement. However, these interventions can be challenging to identify and implement. Complex organisations will have staff members with conflicting roles and responsibilities with a complex web of trust relationships that can mean technical approaches are favoured.

The most critical issue in managing insider threat within a workplace or organisation

is identifying insiders before an attack. Organisations and teams can have complex social and hierarchical structures; managers and colleagues are the first line of defence. As discussed in the previous section, although insider threat attacks differ in many ways, especially by archetype, all possess an escalation of the insider's disgruntlement, leading to either apathy or anger, ultimately impacting the decision to attack. Therefore, recognising disgruntlement or taking steps to avoid or manage disgruntlement is critical, especially for line managers of potential insiders (Cappelli, Moore and Trzeciak 2015). In addition, recognising anger and disgruntlement at work and external factors such as financial or personal issues is also crucial. Allowing insiders to have open and transparent communication with line managers can allow these issues to be discussed and dealt with openly. Therefore, the goal for organisations and specifically managers should be to recognise disgruntlement and financial or personal issues.

One of the most significant issues that organisations face is trust (Nichols, Danford and Tasiran 2009), both the insider trusting the organisation and excessive trust placed on the insider by the organisation. Particularly in the case of insider sabotage as noted by Cappelli, Moore and Trzeciak (2015), colleagues do not wish to report insider threat activity. However, it is essential to note that trust is two way, between insiders and organisations and the organisation's trust of insiders. This disconnect between an attack's actual risk and the perceived risk of an attack can be a significant factor in employees becoming distrustful and ultimately deciding to attack.

Too much monitoring, either technical or behavioural, can erode trust, however too little or allowing insiders to bypass monitoring reduces the ability to identify and act on insider threat activity. Managing this trust between the organisation and its employees is a critical activity that line managers perform. If behavioural and technical monitoring remains in place this trust relationship can be damaged and other approaches should be used to develop this trust (Yerby 2013; Martin and Freeman 2003). Managers must take on the responsibility of understanding insider threat, creating a trusting environment without reducing monitoring capacity. If trust is reduced, such as when an employee faces adverse

events, it is essential to manage and handle disgruntlement in the corporate structure.

Expectation management takes a significant role in handling disgruntlement in employees and fits into the broader issue of negative workplace issues. For many insider threat incidents, an insider's unmet expectations are critical to their decision to attack in many cases of insider threat, with some insider threat cases having additional motivations that aid in the rationalising of an attack (Cappelli, Moore and Trzeciak 2015). This unmet expectation may be a missed promotion, a salary raise that an insider feels entitled to or disciplinary measures. The core method to manage these expectations is communication, specifically creating an environment where employees feel they can communicate negative issues and know management will address them. In addition, a consistent application of workplace rules and consequences fosters this environment. Employees should be offered regular reviews where concerns will be listened to and managed, with an organisation accepting feedback following unmet expectations. Of particular concern should be employee termination, with clear procedures to remove IT equipment and all potential access points.

If disgruntlement is likely to occur due to unmet expectations and specifically the insider's entitlement to outcomes (such as promotions), a positive intervention, if appropriate, can be implemented. Positive interventions can help manage expectations and give an individual a feeling of progression, therefore avoiding the reduction of trust and disgruntlement. This is challenging but taining can be offered to new managers (e.g. technical managers are more likely to be promoted to the position and likely have not had this training previously) (Cappelli, Moore and Trzeciak 2015). A focus on employees' wellbeing can aid in both developing trust and offering positive intervention; these can take the form of counselling or additional training (Kirk and Brown 2003). Those who feel entitled to promotions offering additional responsibilities or professional development whilst communicating can mitigate disgruntlement as a compensation measure (Cappelli, Moore and Trzeciak 2015). However, it is essential to note that many people in an organisation can help recognise and prevent insider threat, including HR departments.

Particularly in both the start and end of employment, co-workers and colleagues are often witnesses to insider threat behaviour but may be disincentivised to report it to line managers or upper management. Other models attempt to understand the role deterrence plays in reducing insider threat such as Safa et al. (2019), which adds additional features such as the perceived sanction, increased effort, increased risk, reducing rewards, reducing provocations, removing excuses, managing intention and by placing their actions in behavioural norms. Each of these attributes deters a potential insider threat by making the act more difficult or less desirable. For example, a perceived sanction certainty might be ”I believe that if I sell organisational information my organisation will discover it”, which can be combined with a perceived sanction severity such as ”I think the punishment will be high if I sell or transfer organisational information outside” this deters potential insider threat as it disrupts the ability for the activity to go undetected for an extended period of time, increasing the perceived risk an insider must take. This perceived risk can be increased further by technical controls such as monitoring or the identification of behaviour by managers or other staff members. In particular, deterrence models highlight the role of managers and other staff members by positively enforcing security policy and helping an employee understand encryption or data deletion (which reduces the reward should an employee choose to become an insider threat.

Human Resource departments are vital to the management and prevention of insider threats, specifically as their work occurs at the start or end of an employee's job. Practical options involve background checks and other pre-employment checks, such as references. While in some countries (such as the UK), this may be unfeasible due to legislation, including data protection laws, e.g. GDPR (*Data Protection Act* 2018), this is discussed in detail in the ethics section of this literature review. Much of the literature on insider threat within the context of HR is specific to US laws and cites previous records with other employers or with the police, such as tardiness at work or previous criminal conviction, can be used to indicate a potential insider threat (as insider threat activity usually involves a negative attitude towards work (Greitzer et al. 2012)). However, in all countries employ-

ment contracts and employee handbooks ensure rules are applied across the organisation and lay out potential punishments (Greitzer et al. 2012; Safa et al. 2019). During the onboarding process, documentation and clear legal documents are essential; such documents as Non-Disclosure Agreements provide insiders with consequences if they were to attack. These documents should also set out policy clearly, and show that internal consequences will happen for misbehaviour. Applying these policies throughout the organisation can help reduce disgruntlement among employees, creating a sense of fairness in applying rules.

Aside from the onboarding process for new employees, HR plays a crucial role in dealing with individuals leaving the company. Clear policies are vital, especially in ensuring that no equipment or back doors into servers are left. For example, ensuring equipment is returned and wiped, and that any accounts former employees created or used have been deactivated and no longer have access. These crucial steps form a collaboration between HR and IT and ensure that each account is sufficiently deactivated. The 30 days after an insider leaves have been shown to be the most likely window of an insider attack, therefore it can be essential to manage resignation well, including additional controls such as gardening leave (Sullivan 2016). Managing concerning workplace behaviour from an insider must be identified early on and appropriately managed to avoid conflict or the escalation of insider threat activity. In particular, managers are adept at recognising missed expectations, for example, not receiving a promotion, concerning behaviour such as being consistently late for work or being unprofessional. Colleagues of an insider may also approach their manager to report insider threat activity or other concerning behaviours such as bullying; this allows a manager to identify and manage an insider before they attack. However, this can be extremely difficult, especially when asking a line manager to become a psychologist or insider threat expert. These are specialist skills and is an enormous responsibility on a single person, which may further increase the risk of insider threat activity (Cappelli, Moore and Trzeciak 2015; Greitzer et al. 2014). Instead, all management levels must understand insider threat and have some knowledge on how

to identify and manage a potential insider before they choose to attack. However, as discussed above, this can be particularly difficult with conflicting priorities, projects and difficulty in recognising insider threat activity.

Colleagues of the insider can also play a large part in the investigation of insider threats. This is particularly true for noticing and reporting concerning, but difficult to identify, behaviours within the organisation, although sometimes this may be indirectly communicating with the insider as a support network. Many colleagues may be reluctant to report, out of loyalty to the insider or because they are uncertain of the insider's motivations, so they choose not to report. Even when colleagues are willing, often they are not sure what to report or what is relevant to an investigation (Forte 2019). IT staff are more likely to report due to their knowledge of technical controls which detect anomalous activity.

Technical solutions are often employed to manage the insider's risk at an organisational level; while these are usually tools, IT policies such as an acceptable use policy and IT governance can also form some part of the technical controls. Due to the overall difficulty in managing insiders using only organisational controls and the level of training and experience required to identify and prevent insider threat activity, technical controls can be a desirable option. However, these technical controls come with many caveats, and despite vendors promising accurate results, can often misidentify legitimate activity as insider threat activity, particularly as many insider attacks are subtle (Homoliak et al. 2019; Yuan and Wu 2020). Often these controls present the core defence of an organisation to insider threat attacks and provide an easy bypass for technical insiders. This rush to implement technical solutions, although attractive, may not work at best, and may actively make insider threat activity easier and undetectable.

2.3 Technical Approaches

Technical approaches are often favoured in organisations and often provided as software packages or part of existing controls for intrusion detection such as Thompson, Whittaker and Andrews (2004). Technical controls vary widely in complexity, from simple, low-interaction honey pots aimed at would-be insiders, to machine learning solutions that continually monitor and flag anomalous traffic. As discussed in the previous section, these are particularly favoured due to their ability to detect insider threat attacks before an insider can complete them (Yuan and Wu 2020). In particular, they monitor network traffic or other computer artefacts, detecting insiders who access data that their job does not require, or transferring files from internal networks to removable media such as USB drives (Noever 2019). Unlike relying on management, these systems do not require expert knowledge and instead focus on insider threat signs. This section will discuss a range of these solutions, from simple honeypots to anomaly detection systems, and discuss why these technical controls do not offer complete protection from insider threats. Despite these controls, data breaches resulting from insiders are still extremely common (Homoliak et al. 2019).

These techniques often focus on detection either after an attack has taken place or during an attack. These techniques often require specialist knowledge of computer forensics or insider threat to install or use. These can be broken down into two categories, passive controls and active controls. Passive controls do not need a large amount of additional monitoring and use a flag system. An example passive control is machine learning approaches, such as anomaly detection, where user behaviour is flagged when it differs from the 'usual' network activity. Active controls use honeypots and decoys to lure an insider into attacking, as well as technical restrictions to discourage insider attacks. These require much more monitoring to passive controls as well as upkeep.

The first approaches to technically understanding the insider threat through detection used honeypots (Bowen et al. 2008; Spitzner 2003). These files tempt a possible insider with promises of classified information such as trade secrets. If a user does not need

access or this information, they should ignore it and not take any action. However, an insider threat or an employee considering insider threat will access the file and ultimately not be met with the information promised. The honeypot will then alert IT staff, HR or their management chain.

Although this approach is relatively simple, it can serve as an early warning system and help find employees who may use this opportunity to commit an attack. (Bowen et al. 2008) expanded upon this creating more sophisticated decoy documents by increasing the believability of documents and increasing the file's perceived value lure potential insiders. The honeypot technique's simplicity makes the approach accessible to many organisations and allows organisations to test many types of files to lure different insiders into attacking; therefore, an organisation can understand insider threats within their organisation. However, it is hard to generalise these mitigations for different insider threat attacks across different organisations. These approaches can only act as an early warning system and can cause many false positives (such as curious individuals who are not interested in committing an insider attack) and must be used with other techniques, such as behavioural or psychological indicators to identify an active insider threat. Therefore although these honeypots can aid in detecting insider threat, they have significant drawbacks, in the form of curious employees as false positives and having to be very tailored to likely attacks to ensure an insider is tempted.

Much of the early technical literature outside of honeypots attempts to define an insider's technical activity. This definition stage is an essential step for the evolution of technical approaches and begins defining an insider's behaviour within the cyber domain and links the behavioural to the technical. An early example is the development of a domain-specific language for the insider threat domain (Magklaras, Furnell and Brooke 2006), this work focuses on taking existing understanding from external threats and develops it to apply to insider threats. Building upon this work was then creating knowledge bases (Althebyan and Panda 2007), and a formal model to show how an insider escalates their permissions or accesses data outside of their job description. These early models

and applications are discussed in more depth within the next section. However, these early technical approaches are essential as they led to more advanced anomaly detection and graph-based approaches. The ability to define what insider threat activity consists of underpins all future technical approaches; this critical step is not only the first step to detection but provides IT departments with actionable intelligence.

Graph-based approaches representing insider knowledge and insider acquisition were the next iteration for technical understanding. Applying the modelling techniques into a deployable detection system. One of the graph-based approaches involves using a directed graph to represent an insider's knowledge and detect when an insider attempts to increase their knowledge before an attack. These approaches rely on an analyst using them to detect insider attacks, unlike approaches which monitor the network (Althebyan and Panda 2007; Mathew et al. 2008; Yaseen and Panda 2012). These approaches have since been usurped by the anomaly detection work and machine learning detection systems due to the technical and modelling requirements to create and exploit these knowledge graphs. This early work in combining a technical approach with an insider threat model and creating actionable intelligence was key in understanding the insider threat. In addition, this work takes a different approach to earlier and later work, uniquely representing an insider threat as a person trying to increase their knowledge. These approaches are not the only graph-based technical understanding to insider threat; however, representation is the first step in moving from definition to detection.

Other graph-based approaches predate the machine learning approaches to anomaly detection, finding anomalous graphs based on deviations from typical patterns (Eberle, Graves and Holder 2010). In their work Eberle, Graves and Holder (2010) create a graph based on observed traffic and examine this for outliers within that graph, particularly structural issues which suggest that the activity observed is not like others. This approach avoids manual analysts by utilising machine learning. As a result, anomaly detection systems can be built more quickly and efficiently using this initial approach.

Anomaly detection techniques and user behaviour flagging are the most common

forms of modern technical approaches to understanding insider threat activity. These detection systems work by analysing network traffic logs and attempting to take a normal baseline reading of legitimate activity. Once this normal reading has taken place, any anomalous activity which could suggest an insider attack is discovered using machine learning techniques (Agrafiotis et al. 2015; Gavai et al. 2015; Legg et al. 2015; Young et al. 2013). Although often anomaly detection is applied to network traffic (Legg et al. 2015), there are others such as graph-based (Eberle, Graves and Holder 2010), network traffic monitoring (Legg et al. 2015). There are disadvantages to these approaches because anomaly detection techniques must take a baseline reading. If an insider attacks while this baseline is recorded, any system will assume that behaviour is 'normal'. Like other approaches, a certain amount of specialist technical knowledge is needed to install and utilise thhse effectively with a general push towards tools set up once and record constantly similar to an antivirus or firewall, which reduces the ongoing technical knowledge required. These are extremely common, often included in software packages.

Software packages which promise insider threat detection are usually built from the anomaly detection research, offering detection using traffic management such as (Splunk 2020; Forcepoint 2015; Securonix 2020), other commercial approaches offer traditional engagements with a focus on insider threat such as (Redscan 2020). These common software packages focus on providing insights to IT teams, flagging suspicious traffic on the network, much like intrusion detection systems (Legg et al. 2013). Although these recognise the different types of insider threat, each primarily focuses on attempting to recognise traffic where an insider is accessing files or increasing their access to files, copying large amounts of data or simply breaking from their usual habits. This may not catch some types of insider, who already may possess a high level of access or technical skill such as insider IP theft or insider sabotage (Cappelli, Moore and Trzeciak 2015). In the case of insider sabotage where the insider is often a member of IT staff, they may even have access to monitoring systems and be able to avoid detection. A different approach to insider threat detection is provided by (Redscan 2020), combining the detection system

with a more traditional penetration test (Redscan 2020) and phishing campaigns. This more varied approach attempts to uncover not just malicious insiders but also accidental insiders and recognise entry points, not just reacting, as suggested by Cappelli, Moore and Trzeciak (2015). In general, detection software is often deployed onto a network, with each software package offering different anomaly detection systems, whether machine learning models or traditional approaches. These packages proclaim a complete solution to insider threat detection, with state of the art system's focus on machine learning and work to improve accuracy, reducing false positives and false negatives and detection.

Other approaches are not yet developed into deployable software and focus on improving machine learning systems for insider threat detection. One technique could be deep learning(Yuan and Wu 2020). Although there are still many challenges; however, there is a lot of potential for improved detection systems (Yuan and Wu 2020). Deep learning has been used to analyse mouse dynamics for insider threat; by analysing the user's mouse movement, insider threat could be detected as a biobehavioural marker (Hu et al. 2019). Another use of deep learning is presented in (Lu and Wong 2019), which uses LSTM (Long Short Term Memory), a type of RNN (Recurrent Neural Network) to improve anomaly detection techniques. Other uses of machine learning in the detection process are varied, such as (Le, Zincir-Heywood and Heywood 2020; Elmrabit et al. 2020), and whilst this work does not focus on detection directly, the specific issues this research targets, such as risk, are key to detection efforts. One such use is an approach that uses natural language processing to analyse logs for detection purposes (Hu et al. 2019).

However, as noted by (Yuan and Wu 2020), there are significant challenges when using machine learning in the insider threat domain, such as issues with data, lack of explainability, and the difficulty in detection in general, such as finding new attack patterns. These are vital issues in insider threat as often insider threat is not purely a problem that can be solved with technical solutions; these technical approaches are missing significant features. Explainability, for example, is essential for managing insider threats in an or-

ganisation. If the concerning behaviour can be explained by a system, suitable mitigations can be put in place. Although outside of machine learning, another approach has been to categorise insider threat and focus on the key elements and technical solutions to each, such as (Homoliak et al. 2019), this work shows that insider threat is nuanced, and each technical aspect needs to be addressed with different technical controls. In general, the technical aspects of insider threat focus on better approaches to insider threat detection, using machine learning or better-defining insider threat activity.

However, data breaches still occur, and a large number of data breaches have been attributed to insiders. This shows that these technical controls that organisations often employ are not enough; in particular, the lack of explainability hinders organisations by reducing the intervention techniques. Despite the use of new technology such as machine learning and state-of-the-art deep learning, insider threat activity is complicated to detect Yuan and Wu (2020) noting issues with data, including lack of data, class balance, lack of temporal information as well as the detection of insider threat activity in general, specifically the difficulty in adapting these controls for new attacks. Anomaly detection using machine learning is particularly common in both academic and software packages (Michael, Fusco and Michael 2008; Noble and Cook 2003). However, as discussed, this may not capture insider threat activity. Anomaly detection requires a large amount of data, usually by creating a baseline of 'normal' network activity (Legg et al. 2013), and if the organisation is already a victim of insider threat activity, there will be no anomaly. Anomaly detection also may flag false positives, making an organisation less likely to address issues (Greitzer, Frincke and Zabriskie 2011). Technical controls, in particular, are used widely in organisations, and often they are the only tool used to manage insider threat. The technical controls for insider threat management have evolved and form part of a defensive toolset to manage, mitigate and prevent insider threat.

The technical approaches of understanding and detecting insider threat have become far more sophisticated over time, from simple honeypots to flag potential insiders to complex anomaly detection methods. As insider threat has moved from a niche area of se-

curity into its own domain, the difficulties of detecting and understanding the technical aspects of insider threat have become apparent. Due to the similarity of insider threat activity to regular activity (e.g. Copying high-risk files) and that it can be committed simultaneously as typical activity, it is clear that a purely technical approach cannot solve the problem. However, the technical understanding of insider threats, where insiders will, for example, increase their security level or access high-risk files, can be a useful technical indication, requires additional information to fully understand and intervene in an attack. Considering additional approaches, such as psychological and social approaches to an attack, can form the core of the organisational and managerial approach to insider threat understanding and provide insights into the underlying reasons. It is not limited to just the technical aspects but also the insider's motivation and the catalyst that caused them to attack enabling a full root cause analysis. This is extremely important as, as discussed above, the technical, detection-driven lens of these approaches often fails to capture an attack's human element. A more thorough understanding of the human element can often lead to prevention rather than detection, providing actionable tasks to find, intervene and reduce the risk of attack rather than aim for only detection and remediation.

2.4 Psychological and Social Approaches

Psychological and social approaches primarily focus on the pre-attack phase, identifying those who may be a high risk for becoming an insider threat or those on the path to a higher risk. In identifying these individuals, they can be flagged for additional technical monitoring or the intervention techniques discussed in the organisational section. Psychological approaches aim to understand the insider and primarily focus on motivations and behavioural changes when an insider attacks. These approaches often use techniques from sociology and psychology to do this. This approach offers a different perspective than the technical, instead of considering just the technical aspects of attack to consider the insider themselves, what motivates them, what turns an employee into an insider threat. These

questions are best understood by considering the psychology of insiders and considering the social aspects of insider threat. This different approach to insider threat understanding does not focus on detection in the sense of automatic detection, such as the technical approaches. Instead, it aims to identify insiders before they attack and, similarly to the technical controls, allow an organisation to mitigate insider threat before an attack occurs.

For psychological and social approaches, these controls can take many forms, including identifying precursors to insider threat activity, e.g. (Greitzer et al. 2012) or indicators that may increase the risk of an employee becoming an insider threat, e.g. (Shaw and Stock 2011), identifying other risk factors such as a drug, alcohol or gambling problem, the motivations behind insider threat attacks. Social techniques attempt to understand the difference between an individual who will attack compared to one who will not, e.g. (Brown, Watkins and Greitzer 2013).

Although the majority of work of psychological approaches has been in identifying risk factors for insiders, another major approach has been language use. Identifying subtle changes in language use when an individual chooses to attack (Brown, Watkins and Greitzer 2013; Greitzer et al. 2013; Ho et al. 2016). Brown, Watkins and Greitzer (2013) relates word use to the 'Big 5' personality model (Goldberg 1993), identifying patterns which might predict insider threat. Greitzer et al. (2013) approaches the issue similarly, examining word use and mapping these onto a model of personality traits, adding weights to each word. This builds from Greitzer et al. (2012), where scores of words from these same personality traits are used to predict words that insider threats use. These approaches hope to build an early warning system, where an individual's communication could be monitored, and those who use certain words or score highly on these measures could then be subjected to additional security screens. However these approaches come with ethical concerns similarly to employee monitoring. Another different approach was completed by Ho et al. (2016) which focuses more on the tipping point when an insider chooses to attack, looking for subtle communication changes, which the authors note is hard to detect. Although the factors which are discussed next focus on the characteristics an insider may or may not

possess, these focus on observable changes, aiming to bridge the gap between prevention and detection by using this behavioural manifestation of an insider's personality or state.

Insider threat factors, by contrast, focus on these personality characteristics and attempt to answer the question of why different individuals who may be put under similar pressures at work decide to become an insider threat or why certain events may become the 'straw that breaks the camel's back'. Organisations then attempt to develop mitigation strategies that reduce an individual's likelihood of transitioning to an insider using this fundamental psychological understanding. These strategies focus on the pre-attack phase, which can be managed at an organisation level. For example, if an insider attacks because of financial problems caused by a gambling addiction, a possible mitigation strategy may be to screen for potential issues during the hiring process (Nurse et al. 2014b; Cappelli, Moore and Trzeciak 2015); however, this may not be possible or desirable due to legal and ethical concerns and these are discussed in detail. Many of these factors are encompassed in the insider threat archetypes discussed in the previous sections; however, there are general factors for malicious and accidental insider threats.

Greitzer et al. (2012) investigated more general insider threat indicators, Figure 2.10 shows various psychological indicators associated with insider threat activity. These include disgruntlement, rejection of feedback and stress. Each factor is expanded upon to describe the set of behaviours that the factor encompasses. This expansion ensures that this research is more easily actioned in an organisation, with clear examples of the behaviour and potential issues. Although many factors may be work specific, for example, difficulty accepting feedback and disengagement, others are more easily observed by closer colleagues than managerial staff, and further demonstrate that to fully manage insider threat, many different members of an organisation must collaborate. It is worth noting that only some of these directly result from pressures from work; others are related to the insider's general personality, for example, being detached from others or holding grudges. Using a Bayesian network model the authors assigned weights to indicators and compared them to observed behaviours of insiders, creating a method for detecting in-

sider threats. Creating a method for investigating insider threats in combination with a technical approach in order to move from a reactive approach to insider threats to a more proactive approach, identifying high-risk employees before they attack.

Indicator	Description
Disgruntlement	Employee observed to be dissatisfied in current position; chronic indications of discontent, such as strong negative feelings about being passed over for a promotion or being underpaid, undervalued; may have a poor fit with current job.
Not Accepting Feedback	The employee is observed to have a difficult time accepting criticism, tends to take criticism personally or becomes defensive when message is delivered. Employee has been observed being unwilling to acknowledge errors; or admitting to mistakes; may attempt to cover up errors through lying or deceit.
Anger Management Issues	The employee often allows anger to get pent up inside; employee has trouble managing lingering emotional feelings of anger or rage. Holds strong grudges.
Disengagement	The employee keeps to self, is detached, withdrawn and tends not to interact with individuals or groups; avoids meetings.
Disregard for Authority	The employee disregards rules, authority or policies. Employee feels above the rules or that they only apply to others.
Performance	The employee has received a corrective action (below expectation performance review, verbal warning, written reprimand, suspension, termination) based on poor performance.
Stress	The employee appears to be under physical, mental, or emotional strain or tension that he/she has difficulty handling.
Confrontational Behavior	Employee exhibits argumentative or aggressive behavior or is involved in bullying or intimidation.
Personal Issues	Employee has difficulty keeping personal issues separate from work, and these issues interfere with work.
Self-Centeredness	The employee disregards needs or wishes of others, concerned primarily with own interests and welfare.
Lack of Dependability	Employee is unable to keep commitments /promises; unworthy of trust.
Absenteeism	Employee has exhibited chronic unexplained absenteeism.

Figure 2.10: General Insider Threat Factors (Greitzer et al. 2012)

The factors that contribute to unintentional insider threat are similar, including lack of attention, workload, risk tolerance, cognitive limitation and influence of physical state, drugs or hormone imbalances (Greitzer et al. 2012). However, work regarding unintentional insider threat focuses on the impact of these factors on decision making. Although, for example, stress plays a role in both malicious and unintentional insider threat attacks, in unintentional insider threat, the insider commits their attack by bypassing security controls in order to manage their workload, rather than not feeling appreciated by their employer. In addition to these human factors, Greitzer et al. (2012) suggest more general psychosocial, socio-cultural and other factors, including personality predispositions, concerning behaviour and demographic factors. These factors include the ‘Big 5’ personality traits (extraversion, agreeableness, openness to experience, conscientiousness and neur-

oticism) (Goldberg 1993), and how they may interact with each other to make individuals more susceptible to becoming an insider threat, for example, links between openness, extraversion and agreeableness scores can lead to greater trust (Greitzer et al. 2014) and therefore lead someone to be more trusting and susceptible to social engineering and phishing attacks. These factors can be managed, and rather than methods that react to attacks, such as technical approaches, identify those high-risk employees preemptively (Greitzer et al. 2014).

Despite the differences between malicious and unintentional insider threat, there are significant similarities between these factors. Specifically, the behaviours which may suggest a risk of becoming an insider threat, include anger, disgruntlement and stress (Greitzer et al. 2014), which are also noted as malicious factors such as anger management issues and grudge-holding, disgruntlement and stress. For unintentional insider threat, this often leads to impaired decision making and issues in risk perception (Greitzer et al. 2014). This demonstrates that although malicious insider threat and unintentional insider threat are considered different, their differences are nuanced and what may be considered different types of insider threat may not be as different as they might appear. The literature on unintentional and malicious insider threat suggest methods of managing each factor, which are very similar; opening up communication, offering programs for employees to reach out for mental health issues, and employment assistance programs to aid in reducing outside stressors. Although personality characteristics differ between the two, all the 'Big 5' (Goldberg 1993) have been shown to interact with each other to make someone more likely to become a malicious or unintentional insider threat. This shows that under similar pressures, different individuals can commit insider threat attacks, but these may differ only in whether or not this was malicious or unintentional. Although there are differences, specifically in human factors, for example, workload and workplace stress, and lack of attention or knowledge on security risks, which do not appear in malicious insider threat cases, where an insider often chooses to attack for differing reasons, the similarities between two types are striking. Insider threat models attempt to encapsulate

these, generalising over insider threat cases, combining not just the environment of an organisation, pre-attack psychological and post-attack technical, but also these boundaries of insider threat.

Another psychological approach includes addiction theory to understand why insider attacks (Maasberg and Beebe 2014), CMU's insider threat division (Cappelli, Moore and Trzeciak 2015) refers to insiders being motivated to attack due to addiction issues such as gambling or alcoholism. In their work Maasberg and Beebe (2014), use addiction theory in order to better understand these motivations behind attacks where addiction has led to an insider attack, similar to the personality factors—examining the insider's motivation, mainly focusing on addiction, rather than the insider's personality.

These and other approaches have been key to understanding insider threat from a psychological point of view. The psychological aspects are seen as a warning sign before the attack, while the technical approaches look to catch an insider in the act of attacking. These approaches, therefore, cannot exist alone and must be combined with other techniques to understand and prevent insider attacks. By combining both anomaly detection with psychological indicators, an organisation can begin to better detect insider threats (Brdiczka et al. 2012), creating a harmony between a reactive and proactive approach to managing insider threats.

The combination of technical and psychological understanding is the start of the use of models. These models do not limit the understanding of insider threats to one specific viewpoint but attempt synthesis across all aspects of an attack, from the organisation to the technical to the psychological. These focus not on detection but instead on understanding insider threat, with the intention that from a better understanding of the problem space and all elements of an attack, many different mitigations or detection systems could be introduced, rather than just one.

2.5 Insider Threat Models

As previously discussed, these three approaches, managerial and organisational, technical and psychological and social, cannot be considered separately. Indeed each attack contains many of these aspects, and whilst some insiders may use an extremely complex technical attack, especially for some archetypes such as insider IT Sabotage, whilst others such as insider fraud are commonly perpetrated by non-technical members of staff (Cappelli, Moore and Trzeciak 2015), which may have fewer digital artefacts, but there may be more psychological indicators. Hence, viewing an incident through a combination of lenses is important to understand the phenomena.

To achieve this one major approach has been developing models of insider threat, these include the insider threat archetypes considered by Cappelli, Moore and Trzeciak (2015), as well as others such as Nurse et al. (2014b). These models also allow insider threat to be visualised and show how organisational characteristics influence technical characteristics which are in term influenced by psychological characteristics. Modelling can take many approaches using processing analysis (Bishop et al. 2014), agents (Sokolowski and Banks 2016), information-centric (Ha et al. 2007) and assessment focused (Chinchani et al. 2005). In addition models can be exploited for prediction, anticipation and detection (Legg et al. 2013; Greitzer and Hohimer 2011; Kandias et al. 2010; Althebyan and Panda 2007). These models attempt to encapsulate many elements of an insider threat attack, and although each model uses different methods, all focus on creating actionable insight that could be used to understand an attack.

By far the most well-known insider threat models are by CERT (Cappelli, Moore and Trzeciak 2015), their MERIT models are used widely in both industry and academia. Using group modelling and a database of insider threat incidents, CERT creates four models and identifies three archetypes of insider threat within the broader malicious insider type of insider threat. These are insider IT sabotage, insider IP theft and insider fraud, with insider IP theft split into two models, ambitious leader and entitled independent. These models represent insider threat as a range of events, motivations and psychological states

with connections between each one to show how one can influence another and where an organisation can prevent attacks. For example, in Figure 2.11, the insiders desire to steal can be mitigated by an insider being concerned about being caught but can be increased by an insider's entitlement which is increased by the insider's contribution to the IP. These visualise many elements of an attack, from the technical to the psychological, showing how each influences another.

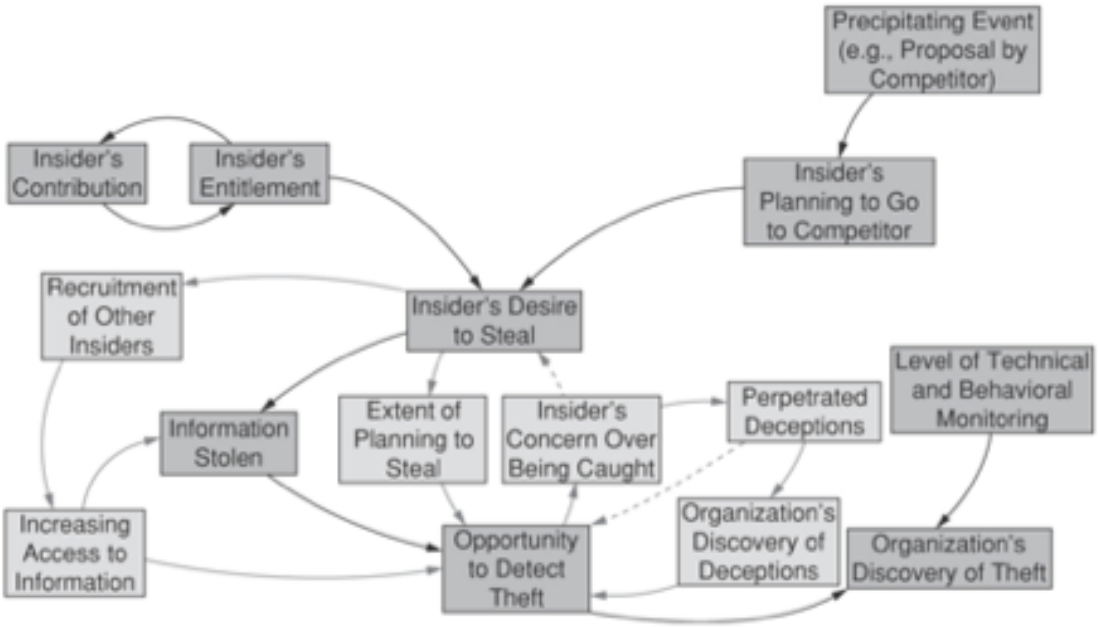


Figure 2.11: MERIT model of insider theft of IP: Ambitious Leader (Cappelli, Moore and Trzeciak 2015)

Figure 2.11 shows the MERIT model for insider theft of IP: Ambitious leader. In each CERT model, general themes are identified, such as 'information stolen' or 'opportunity to detect theft'. These are represented on the visualisation as text within boxes. These boxes are then linked with arrows showing relationships with both casual and causative relationships. By tracing the causative relationships, an investigator can suggest mitigation strategies, and CERT provides some mitigation strategies for each of its insider threat models.

The MERIT models provide a very easy way for an investigator to understand an insider threat attack and mitigate them. MERIT models are also more specific, separated into types of insider threat they can easily describe many types of insider threat. This

allows the framework to be specific, with each model being considerably different, with different actions, relationships and mitigation strategies. The MERIT models are widely used in insider threat research, with these models being built upon to improve them, using them to understand new cases of insider threat and develop new solutions to insider threat.

The MERIT models do have disadvantages due to the separation of models into individual archetypes of insider threat, for example, sabotage and IP theft. They cannot be compared directly even if they are similar for a particular organisation. This makes the models hard to use in an organisational setting but can help organisations understand and mitigate insider threats by considering each archetype individually.

Nurse et al. (2014b) created a framework that characterises many different types of insider threat using a grounded theory approach. Grounded theory is a methodology adopted from the social sciences which involve the categorisation of text into codes or themes. The codes are then validated using multiple contributors, and relationships between concepts are discovered and validated against each other. Nurse et al. (2014b) identified key events and characteristics which are shared and can be generalised across many incidents of insider attacks. From these key events and characteristics, the relationships between them are theorised and finalised. The final completed framework is shown in Figure 2.12.

The framework combines the psychological aspects (Actor Characteristics) with technical aspects (Attack Characteristics) and events surrounding the attack (Catalyst); in addition to these aspects, which are investigated in previous work, the framework includes Organisation Characteristics. This builds upon work done by CERT (Cappelli, Moore and Trzeciak 2015) however, where the CERT MERIT models choose to separate different types of insider threat, insider fraud, insider IT sabotage and insider IP theft into separate models and in the case of insider IP theft into two models, Nurse et al. chooses to create a framework for characterising all kinds of insider attacks.

The grounded theory approach adopted by Nurse et al. is a unique methodology for developing insider threat models. It is similar to how other models are developed with a library of insider threat cases.

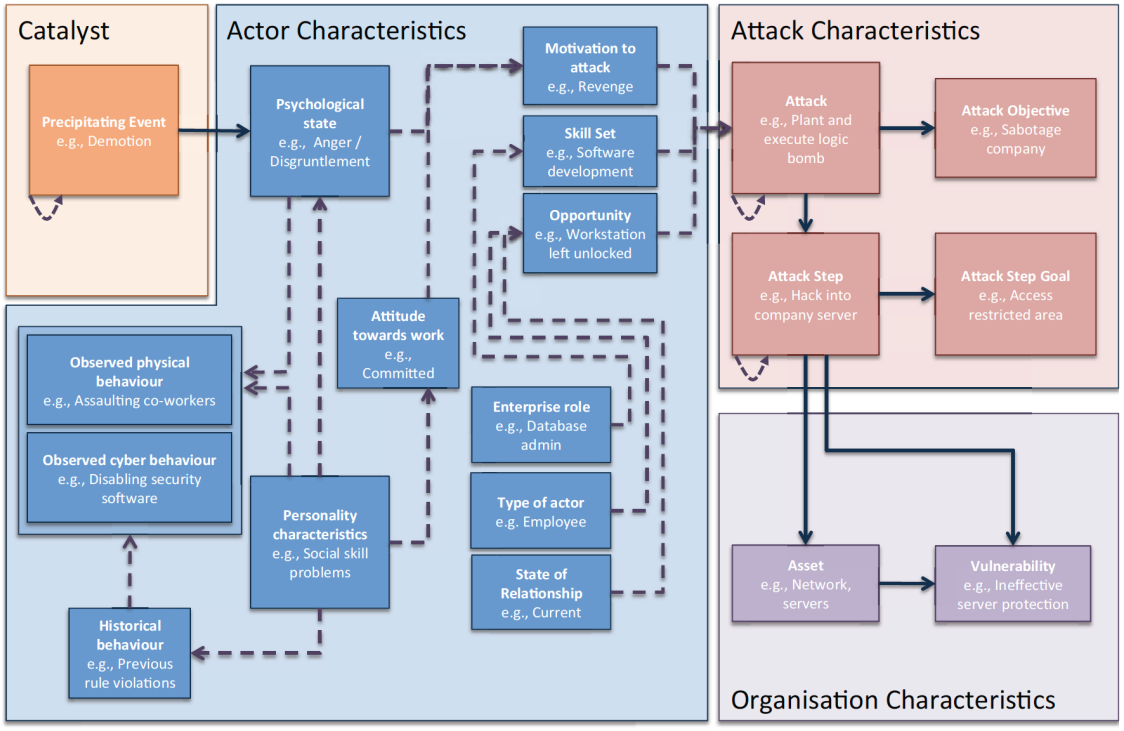


Figure 2.12: Cappelli, Moore and Trzeciak (2015) Framwork

This library is then analysed by humans to discover themes in the text, these themes are themn mapped as relationships between them are developed and finally they are validated by exprts, this approach is called Grounded Theory. Although many approaches follow this development approach, the first step of the analysis is completed using different approaches; for example, the MERIT models were created using group modelling. Grounded theory as an approach has a nmber of advantages with less bias, a clear focus on the literature (Nurse et al. 2014b). Codes or themes are taken from sample cases, constantly developed to ensure that it is guided by the literature, this can take many forms in grounded theory, such as coding each word individually, coding sentences based on these words, and paragraphs based on the sentences, creating a layered approach to understanding the textual reports. An example theme may be 'motivation to attack'. These themes is similar to natural language processing; however, instead of a computer processing the text, a human takes on the role. The validation stage of the grounded theory involves applying the themes or codes to new case studies, and by comparing the codes across multiple people, a validated code should have a 70% match between experts (Nurse et al.

2014b).

Nurse et al. (2014b) shows how this approach can be used to great advantage by applying their framework to different types of insider threat case studies to aid in understanding why and how they happened. In their case studies, the authors apply their framework to different insider threat cases using the same model to analyse both. This is a clear advantage of the model by using a single framework to understand many different cases of insider threat. This is because similarities across different cases can be noted and mitigated. The MERIT models, by comparison, only allow, attacks of a single archetype (e.g. insider IT sabotage) to be compared and understood rather than in contrast to the other archetypes of insider threat. In addition, the grounded theory approach allows the framework to be solidly guided by the insider threat literature, and while difficult to validate, the authors used a cross-validation method where a code can be changed as necessary if it is not applicable in the literature. This allows the model, which was primarily created by human discussion and understanding to be critically evaluated as it is being developed and led by the literature. Representing the nuances of organisations and cases is important to understanding insider threat, as these can make a framework unsuitable for understanding cases outside of the original cases used to develop the framework. As some organisations choose not to report certain incidents, they may not have the tools to analyse these cases as the authors were not aware of their existence.

Other models have also encapsulated both psychological and technical approaches, these however focus on detection and prediction of insider threats rather than mitigation and understanding. These include prediction models (Althebyan and Panda 2007; Kandias et al. 2010; Thompson, Whittaker and Andrews 2004), agent-based models (Sokolowski and Banks 2016), process analysis (Bishop et al. 2014) and graph-based (Chinchani et al. 2005). These models worked to confirm earlier psychological and technical work by using their approaches to assess an insider threat attack.

It is clear that although there have been many models that hope to describe the insider threat, they share similar disadvantages. In order to apply to the greatest number

of cases, they must be general; however, the more general a model, the more difficult it is to capture aspects that may be unique to an organisation or an attack. Understanding the insider threat is clearly an important factor for organisations and research with the work being used to create more sophisticated detection systems, create organisational policies, aid in psychologically understanding why an insider chooses to commit an attack. However, frameworks need to be adaptable and should contain a feedback loop; as organisations deal with attacks, the organisation should able to develop the framework to meet its organisational needs.

2.6 State of the Art

From each lens of insider threat, there are various state of the art techniques, with each lens supporting the management of insider threat attacks. This difference is pronounced when considering the goals of each approach; technical approaches tend to focus on detection, psychological approaches on recognising potential insiders, managerial approaches with practical methods of managing insider threats and finally, models hope to encapsulate each element and provide a method of examining a wide range of aspects of an attack. Each of these approaches is valuable to understanding insider threats. This section will summarise the current state of the art for each lens and finally discuss the future trends of insider threat research and the next steps in managing insider threat before introducing the literature gap that this work attempts to fill, why this gap, and the potential impact that addressing this gap will have.

Managerial techniques of insider threat management, those that involve the organisational structure and members of an organisation, often focus on mitigation. These approaches involve many members of an organisation, including IT staff and IT governance, line managers as well as c-suite, co-workers and HR departments. These techniques blend psychological approaches to recognising likely insider threat attacks with clear actions that an organisation can take to mitigate attacks. State of the art in this do-

main involves distilling existing research from the psychological or insider threat models into practical advice. Primarily this includes engaging with the potential insider before they attack, managing risk on an organisational level, ensuring that an insider does not become disgruntled. In addition, this work engages colleagues and all levels of management in insider threat reporting. When concerning behaviour is reported, all levels of management to be aware. However, organisations tend to support technical solutions to managing insider threats, automatically flagging concerning behaviour.

While managerial approaches aim to mitigate insider threats, technical approaches generally aim to detect insider threat. For state of the art research, this often involves automatically flagging suspicious behaviour using machine learning algorithms. These approaches are often sold as software to organisations as a solution to insider threat. These approaches use machine learning techniques, specifically classification or anomaly detection, to continuously scan networks and highlight activity that may be classified as insider threat or anomalous. These may include an insider increasing their level of access, opening high-risk files, transferring large amounts of data onto external media or email. Traditionally these approaches include a dataset with many different types of network activity, both malicious and incidental, from which an algorithm is then trained. However, most of the research recognises the difficulty in detecting insider threats using this approach. Although these approaches may offer high accuracy, often, this is done in a controlled, experimental setting rather than in an organisation. However, to combat this, some insider threat software vendors offer alternative engagements to help manage insider threat, including phishing campaigns and digital forensics. These approaches focus on artefacts visible by computer systems, such as opening files, and therefore are limited in the ability to detect all types of insider threat (Noever 2019).

To detect insider threat, other behaviour should be considered, especially those which may be strong indicators of likely insider threat activity, such as drug abuse or other addictions such as gambling. Psychological approaches attempt to identify these indications, giving organisations the power to identify those at risk of becoming an insider threat. Al-

though these do not focus on detection or mitigation specifically, this work can be applied to both identifying insider threats so the threat they post can be mitigated or detected. Some of these indications include financial issues, addiction, previous behaviour or rule-breaking, and mental health problems, which may make an employee attack or be coerced into attacking on behalf of an outsider. These indications are gathered by speaking to employees and attempting to understand the underlying psychology of insiders and what led to them attacking (Greitzer et al. 2012). However, other state of the art approaches attempt to identify the change in an insider once they have chosen to attack as an attempt at detection using social or linguistic clues (Greitzer et al. 2013).

State of the art insider threat models aid in the investigation and understanding of insider threat. These represent a different approach to insider threat research by combining many aspects of insider threat and attempt to understand the issue of insider threat as a holistic whole. The state of the art models are the CERT models (Cappelli, Moore and Trzeciak 2015) and Nurse et al. (2014b), with the difference between them being the Nurse framework aiming to encapsulate many different types of insider threat incidents into the model, while CERT approaches the different archetypes as separate models. These models aid in the issue of insider threat by aiding researchers in understanding an attack, the psychological, organisational and technical aspects that make up an attack. These models look at many different attacks and attempt to find related themes across them, the Nurse et al. (2014b) model uses the approach of grounded theory from sociology to analyse attacks. In this approach, natural themes of the text are drawn out by labelling documents and coming to an agreement across multiple participants. Models show how important understanding insider threat is to greater goals such as detection or prevention.

It is clear that there are several different threads to insider threat research, and each has a different aim, such as detection, prevention and mitigation. Insider threat remains a problem for security teams and organisations, still causing a large number of breaches, data loss and financial loss. While the domain has evolved dramatically over the last

decade, there are still gaps in both our understanding of the nuances of insider threat and tool support for potential protections.

2.7 Ethics

An important issue to discuss is the ethics of insider threat research. Although not widely discussed in the literature, aside from the approval of ethics committees for ongoing research, the implications of behavioural and technical monitoring can increase issues between employees and their employers. Particularly from a privacy point of view and workplace surveillance, employees and employers have an asymmetric power relationship; the employer is able to withdraw a job at any time while employees are reliant on it. Other research asks if intervening in a potential insider threat attack may worsen a trust relationship between an employee and employer, removing the element of consent to be monitored (Palm 2009). In particular recent legislation such as GDPR (*Data Protection Act* 2018) protects an employees right to privacy at work.

A particular concern is that the majority of the work completed on insider threat, such as Cappelli, Moore and Trzeciak (2015), has a US focus. This is a particular concern for the ethics of insider threat as data protection legislation is not as widespread currently only California has some of the same protections as EU countries (Pardau 2018). Therefore much of the literature introduces themes of workplace surveillance as a method of controlling the threat of insider threat, including monitoring email and website access, storing keystrokes and resource access, conducting psychometric or drug testing and health data(Ball 2010). This is of particular concern when these are aggregated to create a profile such as Greitzer et al. (2013) and Greitzer et al. (2012), and then used in a predictive context such as is the case with software packages, e.g. Splunk (2020), Forcepoint (2015) and Securonix (2020), in legislation such as General Data Protection Legislation (*Data Protection Act* 2018) the consent given to an employer to hold this data would not extend to the further processing of it. Privacy is a recurring issue in insider threat ethics; often,

for interventions to be a success, an organisation must use both internal data about an employee (performance, timecards, attitude towards work) but also external data (life events, background checks, health, use of assistance programs) (Greitzer, Frincke and Zabriskie 2011). The use of this data has not just ethical implications of privacy but also legal requirements such as General Data Protection Legislation (*Data Protection Act* 2018). Furthermore other legislation such as the Equality Act (*Equality Act* 2010), Rehabilitation of Offenders Act (*Rehabilitation of Offenders Act* 1974), Gender Recognition Act (*Gender Recognition Act* 2004), Human Rights Act (*Human Rights Act* 1990), Regulation of Investigatory Powers Act (*Regulation of Investigatory Powers Act* 2000) (RIPA), Computer Misuse Act (*Computer Misuse Act* 1990) (Fafinski 2013) and rights such as the right to be forgotten, right to respect for private and family life all impact the legal and ethical issues surrounding insider threat. Unfortunately this has not been discussed more widely in the literature, with this wider context often being a footnote (Hanson, Thorsen and Hunstad 2021; Dounis 2017; Taylor et al. 2011).

Much of the work in the ethics of insider threat involves which data to include in any model or disciplinary action and how to interpret any results. Particularly in punishing before an individual has committed an attack, which can be caused by misinterpreting a false positive, punishment can escalate insider threat behaviour (Greitzer, Frincke and Zabriskie 2011). Predictive models can be extremely appealing for managing insider threat. However, the creation and interpretation should always be considered, particularly for the legal and ethical issues (Greitzer, Frincke and Zabriskie 2011). However, this does not extend to wider contexts than just the US. This research focuses on not detecting insider threat but instead to understand a particular attack and empower organisations to manage insider threat attacks or to understand their particular risks (assets, processes, people). The hope of this research is to build upon existing management processes and threat models to secure organisations, rather than to focus on monitoring, technical or behavioural. It is clear that more work must be done to better understand the legal and ethical implications of insider threat research, particularly in a UK or EU frame of reference.

2.8 Literature Gap

Although the state of the art includes various tools, indicators and models of insider threat, there is a very clear disconnect between academic research and practitioners. Organisations prefer software solutions; however, they cite lack of training as a major reason they feel unprepared for an insider attack. Even then, insider threat attacks are extremely common and impactful to the organisation, despite the large body of research. One reason for this may be the difficulty in implementing this research. In general, the primary insider threat investigation tools tend to be automated tools. As discussed, the interpretation and data used for these tools are of ethical concern and may also cause an organisation to break data protection laws (Greitzer, Frincke and Zabriskie 2011). Another key issue is the need for specialist skills, where insider threat training is not commonly offered by organisations to staff. There is a real issue in getting research to practitioners and having deployable solutions. It is clear that the state of the art is not enough to reduce insider threat attacks and often remain undeployed in organisations due to the number of breaches and the damage these breaches can cause.

One barrier for insider threat research is the difficulty in engaging with practitioners. Therefore offering a new solution that engages with practitioners and ensures that they are considered is a major factor. For insider threat research, the core research appears to be software solutions. These offer key advantages over more difficult to use advice, focusing on a deployable, easy to understand solution. While there are many approaches to managing insider threat, almost all require an understanding of the risk, approaches and mitigation of insider threat. This understanding is limited due to the nuanced nature of insider threat, and the current solutions do not offer a structured, holistic understanding of insider threat. This lack of understanding, therefore, causes issues when controls are implemented. Models attempt to bridge this lack of understanding. However, they have disadvantages. Firstly models require a high amount of domain knowledge, and combined with their static nature, it can be challenging to implement meaningful changes, to either adapt to new attacks or become tailored to a particular organisation or domain. Finally, the

confirmation bias and limited view, that is, they are dependent on the data used to create them and the inherent bias that the researchers may include, with a focus on technical or psychological characteristics rather than showing a full view of an attack. This data may actually not improve predictions, and many existing models are biased (Greitzer, Frincke and Zabriskie 2011).

Any system which attempts to remedy these issues and also offer a deployable solution using technical controls. Firstly a system must limit the amount of domain knowledge required to use it; this will enable policymakers and experts within an organisation to use any tool. Organisations often cite lack of training as a reason that they struggle to implement lasting changes to manage insider threat (Greitzer, Frincke and Zabriskie 2011; Forte 2019; Cappelli, Moore and Trzeciak 2015). Models are also static and cannot adjust to changing business requirements and organisation, nor react to potential controls and how they may impact an attacker or how future technology may allow an insider to attack.

Any system must also support existing practices within organisations rather than attempt to replace policymakers in IT governance or other valuable decision-makers. Any system should support this tactical and strategic decision making, providing both macro and micro-level decision making. Fully supporting an organisation to understand the insider's threats they face, as an increased understanding can be key for managing insider threats within an organisation. A better understanding feeds into existing incident response, preventing, detecting and responding to a threat. These requirements and the gap in the literature will form the aims and objectives of the system. These technical requirements serve as objectives, and the literature gap serves as the aim. In the next section, the specific objectives are discussed, which is followed by the technical approaches that this work will use.

Chapter 3

Aims and Objectives

The core aim of this research is to create a toolset that can be used to investigate an insider attack, supporting existing practice. Specifically, this toolset will use existing reports, regardless of writing style, to help a practitioner map these to an existing model to better understand the individual characteristics that lead to an attack. This is done by completing the following technical objectives. First, several datasets must be collected, one containing many different insider threat attacks and another that best represents the data that investigators may use. Next, individual characteristics of the attack, such as the methodology or motivation, must be found, regardless of which report may contain them. These characteristics must then be linked, creating a custom insider threat framework unique to the attack and organisation. Finally, specific details must be visible, allowing investigators to suggest organisational changes to protect against the next attack. Each of these forms the system, and each of the technical objectives must interact with each other.

3.1 Introduction

The overall aim of this project is create capabilities that can aid in the investigation of insider threat cases. Creating a toolset to aid in the investigation of insider threat attacks. This will allow for the further understanding of an insider threat attack. Although this does not aim to solve or detect insider threats, understanding is an essential step in insider threat

research. A focus on practitioners can be used by approaching the problem with the lens of understanding. With an ultimate aim to improve and work with the existing practices rather than supplant or change a practitioner's existing processes. This is particularly important as one of the critical barriers to insider threat detection and management is often a lack of training, specifically on tools, and a lack of understanding of the issues. Staff members who witness an attack are reluctant to report, not necessarily because they do not want to report, but they do not know what to report and when to report it.

Therefore, with a focus on understanding insider threat, this research focuses on giving organisations and their staff the power to manage insider threat attacks by improving investigation tooling. By using organic narratives, a type of narrative where there is no direction given to witnesses and they are free to write in as much or as little detail or in any format, and natural language processing, this work aims to work with these existing practices. Reports after insider threat attacks are already written after an incident. However, often lack a range of perspectives due to the difficulty of writing more formal reports.

Understanding the attack is extremely important to the mitigation process, often knowing which assets an insider targeted, their motivations, how they may have manipulated people or processes or how they were caught, leading to critical organisational changes. In particular, strengthening mitigations that may have been bypassed or removed altogether or adding additional auditing processes can act as barriers to employees' ability to become an insider threat. Therefore, understanding the insider's methodology is key to creating lasting organisational change.

The fundamental aim of this system is: is create capabilities that can aid in the investigation of insider threat cases, by leveraging reports already written after an incident and written in an organic style without a prescribed format. When these individual tools are combined, the full system should allow for the full exploration of an attack, from the precise details about the methodology the insider used and the overview of an attack.

3.2 The System

When considering the individual technical objectives that must be completed to meet this research's aim, it is important to consider the system as a whole. Shown in Figure 3.1, is a visualisation of the system and an example of how it can be used. This diagram shows how the system will function and how an investigator would use these capabilities together. By breaking apart this functionality, the individual objectives can be understood. Data collection and the insider threat model are further broken down into the three technical objectives. Each objective has a task within so that the results can be visualised to allow an investigator to interact with them. These represent the core objectives of this research, each one aiding in the entire system.

1. Creating a corpus of insider threat: To perform any natural language processing first example cases of insider threat will have to be collected.

2. Detecting characteristics: This process finds commonalities between insider threat attacks and uses topic modelling to segement witness reports not by document but by insider threat characteristic.

3. Finding relationships between the characteristic: Similarly to other insider threat models this static analysis of characteristics must be connected together.

4. Finally, the specific details of a characteristic: While the previous two provide an overview of an attack, it is difficult for an investigator to be able to use these to make organisational changes.

The data collection stage is the first significant objective that must be completed; two datasets must be collected. The first is a general corpus of insider threat cases which should represent a wide range of cases of insider threats, with different organisations, insider threat archetypes, outcomes and methodologies. This will be used to train the final model and ensures that the model will be trained on a range of different features of insider threat. The second is an example of a single case of insider threat but written by

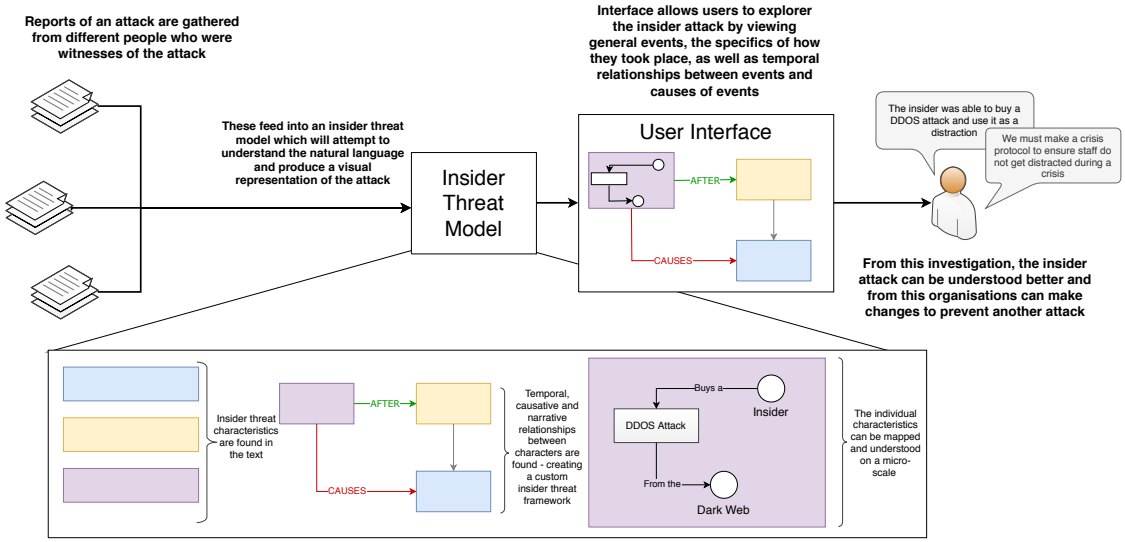


Figure 3.1: Overview of the system

different individuals. This replicates the potential input to the system and provides an opportunity to tune and improve the model. These must be separate corpora, as the model must be trained on many different insider threat cases rather than a single archetype or case. Another consideration is that for the topic modelling to be effective, words must be repeated in differing contexts; therefore, using a corpus written in a similar style for the training would be preferable. However, for the model to meet the aim of this work, it must be applied to different accounts of an incident that may differ in style.

First is a large amount of insider threat documents, and the second is a sample case. The insider threat documents are used to train the insider threat model. This corpus (collection documents) must contain many different insider threat cases with different methodologies, different insiders, different goals, and different motivations to allow the model to be applied to any case. Therefore, it was identified that these requirements could be met by gathering news articles describing insider threat attacks. The sample case allows the research to be tested and validated. This approach has many advantages, including the use of similar language choices and the ease of collecting this data. They were chosen from the literature, which an existing insider threat framework has already analysed. If the model created by the machine can produce similar results, the machine model is validated. This was done by asking participants to retell the story of an insider threat attack

told to them in a series of sound clips, encouraging them to write in any format with as many details as they liked, producing organic narratives. This process is fully described in Section 3.3.

The insider threat model represents the core technical objectives of this research, broken into three stages which are described in Sections 3.4.1, 3.4.2 and 3.4.3. These represent the major technical tools that must be developed and implemented. First is the ability to detect insider threat characteristics. These represent characteristics that any insider threat cases may have, such as the motivation or the technical methodology of the insiders. Second, the ability to gain further insight into these characteristics by finding the connections between them. Finally, to inspect a single characteristic and understand the specific details of that characteristic. For example, if a characteristic is a technical methodology, this will show the exact assets targeted and how. These three tools represent major stages in the project and represent the technical approach to this research. Each of these uses a combination of machine learning and natural language processing techniques and approaches.

The final task of each technical objective is to visualise the work. While these technical objectives are fundamental to this approach, without an accessible approach to view and understand this data, this cannot be communicated to practitioners, and therefore, the overall aim of this work will not be met. This is extremely important and touches each technical objective. By combining these visualisations, an investigator can explore the attack from both an overview perspective and a detail-oriented perspective. In the future, this visualisation task would be considered a final objective, and each visualisation would be implemented in a piece of software; however, because this work represents an initial proof of concept, this will be using other freely available tools. The critical piece of software in use is the Gephi software package. This provides an interface for interacting with graphs and offering graph-specific algorithms to explore and understand the underlying graph structure. This is extremely useful as many graphs can be large, and the ability to zoom or rearrange the graph physically emphasises the most important aspects. In addi-

tion, the graphs can be coloured to highlight additional information, such as how causal a single topic is. In addition, some support tools have been created during this thesis which helps manage different corpora and trains models. While this is designed as a research tool, this visualisation helps meet the overall aim of this work to support an investigator in investigating an attack.

With the aim of this research being to improve understanding of insider attacks using technical tools, it is imperative to keep the human and the insider threat investigator in mind at all stages in this process. Therefore, this research's final objective is to show how these tools could be used in an investigatory context, showing that these individual technical objectives can be used together as a system to fully understand insider threat attacks and specifically highlight important information such as the technical approach or the motivation.

3.3 Data Collection

The datasets objective involves creating two different datasets; these are used for two different purposes, and it is extremely important that they are not confused. The first is a collection of insider threat reports collected in the form of news articles, and the second is a sample case comprising of organic narratives written by participants in an experiment. This work requires two datasets; the first is only used to train the model, and the second is not. This ensures that the model is not overfitted to a specific type of insider threat attack. The second dataset is a sample Insider Threat case and allows for the validation of the models, specifically by choosing a case that has been analysed using an existing insider threat model. The similarities and differences between the insider threat framework and the natural language processing models can be found.

The insider threat news dataset is collected using a web scraper on an insider threat aggregator that reports insider threat activity from a range of new sources. This corpus is then enlarged by using the initial reports to seed an automatic classification tool, classify-

ing new insider threat documents. The advantage of this is that these reports are publicly available, although additional documents may be added to tune a model to a specific organisation. This process produces a corpus of insider threat news reports, which report different attacks in different levels of detail but use a similar writing style.

The organic narrative dataset is created using an experiment named 'The Perspectives Experiment'. In this experiment, participants are asked to listen to 3 audio recordings from different witnesses of an insider threat incident and then retell the story in their own words. This encourages participants to write freely and not be concerned with the level of detail, encouraging them to write down what they remember. This creates a corpus that is similar to the expected input for the final system. The case chosen for this experiment is from the literature that has already been analysed using an existing insider threat framework. This allows the machine models to be compared and contrasted with the human models, creating validation and reflection opportunities. As the Insider Threat domain has not yet used natural language processing to understand incidents better, this provides the only corpus that represents insider threat narratives.

3.4 Technical Objectives

3.4.1 Extracting Insider Attack Characteristics

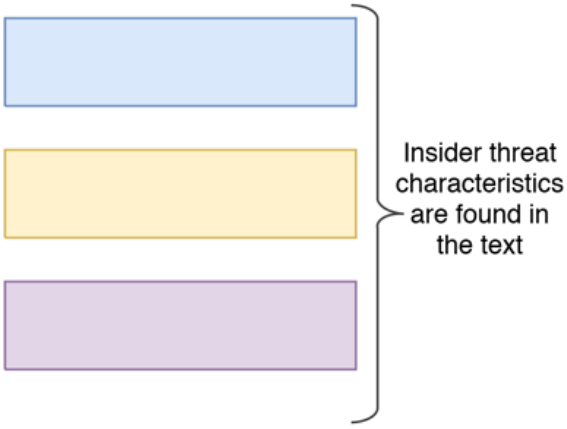


Figure 3.2: Attack Language Objective

The next objective and the first technical objective is to automatically find insider threat characteristics within the text shown in Figure 3.2. This uses 'topic modelling', a similar concept to 'coding' and grounded theory in the social sciences, but uses NLP to do this automatically. A topic model attempts to automatically discover topics in the text by finding words in similar contexts; this can then be reapplied over new text and assign sentences, paragraphs, or whole documents a topic. In this research, topic segmentation (Riedl and Biemann 2012) is applied to sentences in the perspectives experiment. This allows sentences and paragraphs from documents to be aggregated by document, or importantly by topic creating a similar output to grounded theory (Baumer et al. 2017). The comparison to grounded theory is important here, as this method has been used to create insider threat models, notably Nurse et al. (2014b), of which will be used during the mapping process. The topic model is only trained on the initial insider threat corpus but applied to the perspectives experiment corpus, this is important as it means that the model will not be linked to the specific sample, and therefore it is likely that this model could then generalise to new cases of insider threat. Initially, these topics can only place sentences in a topic with no additional contextual information. However, this is then achieved by linking topics using casual, temporal and narrative information. This enhances the initial model by capturing these details even from different reports from different individuals.

The next stage is to add this context to each topic, mapping the topics to an existing insider threat model categorising each machine topic by a characteristic from the insider threat model. This process uses a portion of labelled perspectives experiment data to achieve this labelling of the entire model. The result from this objective is a list of topics, with each labelled with an insider threat characteristic and sentences categorised by document and by topic or characteristic.

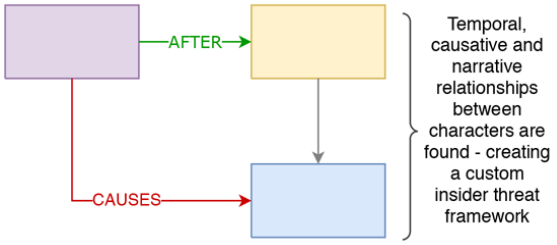


Figure 3.3: Causality Objective

3.4.2 Creating a Custom Insider Threat Framework

These topics do not have any indication of the connections between them; therefore, the next step and objective are to find these links between topics and characteristics shown in Figure 3.3. This creates a type of custom insider threat framework specifically for a case. This is a significant advantage for NLP, as the relationships between characteristics will come naturally as people write. Importantly causal information is key for summarising information, particularly on a timeline (Mirza 2016; Girju, Moldovan et al. 2002), as these can summarise events from news reports, it is likely that this summarisation approach would also be helpful when considering insider threat case summarisation, providing the further context missing from work such as Jacobi, Atteveldt and Welbers (2016). To note is that previously the vast majority of this work was completed on smaller scales by individual words (Mirza 2016; Girju, Moldovan et al. 2002), however this work instead aggregates each sentence into topics and then find casual clues, providing a middleground in terms of accuracy. This is done by first finding all the connections between topics; these may be narrative - one topic follows another, causal - one topic causes another, temporal - One topic occurs after another or coincidental - two topics just happened to appear in sequence. This is done using Markov chains, a statistical data structure, which models can be represented as a state change, and the topic can be viewed as the probability that one state becomes another state. Therefore, this represents a participant describing one insider threat characteristic to speak about another. By examining the individual words in use between topic changes, this link can be assumed to be temporal or causal.

This objective shows the high-level overview of the insider threat attack focusing on

the links between characteristics rather than each characteristic's details. However, these details can be vital to creating lasting organisational change, empowering both strategic and tactical decisions. The output of this objective is a series of graphs that represent these temporal causal narrative relationships.

3.4.3 Examining Specific Details

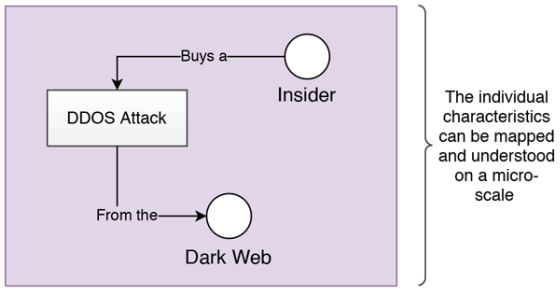


Figure 3.4: Topic Analysis

The final technical objective is the micro-level view of a particular topic shown in Figure 3.4. Whilst the overview is given by the previous objective. This objective represents the details of a topic, this is vital as without the key details no defensive measures can be put into place. Much of the early work in the domain of insider threat was in defining insider threat activity (Costa et al. 2014), these definitions then impacted models (Trzeciak 2011) and in particular advice on handling specific incidents such as by contextualising them within IR frameworks such as Cichonski et al. (2012). The precise details for IR are very important and this is similarly true for Insider Threat, with reports on specific cases being published to present lessons learned (Randazzo et al. 2005). Therefore allowing a single topic to be understood, extracting specific details such as turning a topic related to a technical method into the precise details. This objective informs an investigator of the specific technical method used and should highlight details such as the asset targeted, how it was targeted, the weakness exploited, and any tools the insider used. This method visualises the many sentences within a topic, merging them and creating a graph representing the core topics. This process brings out these concepts, highlighting the con-

nections between actors and actions. This objective's output is a graph representing a particular topic and a method for generating these graphs efficiently.

3.5 Using the Tools

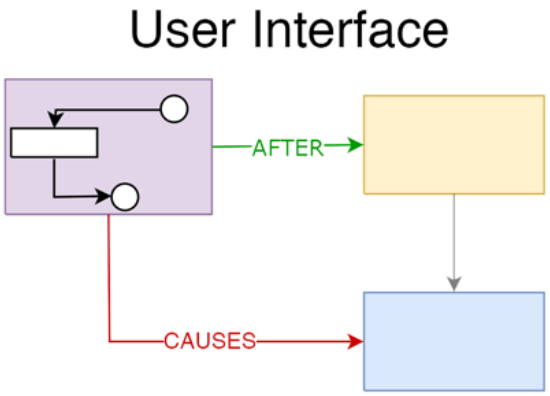


Figure 3.5: Visualising an Attack

Finally, the last objective shows how these tools work together to investigate an attack, showing that these tools are effective and can be used together by using the visualisation tasks in each technical objective (Shown in Figure 3.5). Visualisation is very important for understanding large amounts of data and being able to process and act upon it (Iliinsky 2010). Many insider threat cases are presented without visualisations and this may limit their uses such as Randazzo et al. (2005), as frameworks provide easier to understand and read such as Nurse et al. (2014b) and Trzeciak (2011) the state of the art models. These tools work by first mapping the data to an existing framework, allowing insider threat models to be more easily used, next by identifying key topics that are highly causal and examining these for the key details (Paxton-Fear, Hodges and Buckley 2020). The first level is helping to understand the experience of an insider threat attack, the second understanding the case as a whole giving a 360-degree view of an attack and finally the specific details which may get lost in the above steps, in particular, those details which allow mitigation strategies to be developed. This objective's output is an investigation into the insider threat attack and showing how an investigator can gather motivation, a

technical overview, an overview of the attack, and an outcome (Paxton-Fear, Hodges and Buckley 2021).

3.6 Conclusion

In conclusion, this research aims to create a toolset that allows people to investigate and understand an insider threat attack. Leveraging organic narratives and existing reports, using natural language processing to process and visualise the data. The objectives represent the technical stages of the research, first to collect two insider threat corpora, next to use topic modelling to automatically code these documents and assign them an insider threat characteristic, next to create a custom framework for this attack, finding the narrative, causal and temporal links between topics and then to examine the details of each topic. Each objective is a chapter is in this thesis, finally ending with an example of how the system can be used and how key details in the discussion chapter can be highlighted using the methods developed.

Chapter 4

Methodology

This research will leverage Natural Language Processing, a form of machine learning used to process text. NLP is used in various systems, fields and domains and includes tasks such as machine translation, question answering, information extraction and retrieval. This project uses topic modelling, particularly the LDA (Latent Dirichlet Allocation) algorithm and Standford CoreNLP suite (Manning et al. 2014). Topic modelling is a technique that uses unsupervised learning to find topics within the text, automatically finding related words and phrases. These topics can then be applied over the same or new text to allocate sentences, paragraphs or documents into a topic. The Stanford CoreNLP suite (Manning et al. 2014) contains a range of out-of-the-box tools and NLP pipelines that can be used to understand small units of text. In addition, some bespoke algorithms and techniques have been developed to add additional understanding.

These techniques are then applied to several datasets, first a corpus of known insider threat news articles and then a corpus of a single insider threat incident written in natural language. These two separate datasets are vital as the news articles contain many different insider threat cases but are the same in terms of language, making the removal of stopwords or general topics not related to insider threat easier. The models are trained on this corpus, and while this thesis only demonstrates the efficacy of the approach on Insider Fraud, this general corpus should allow this model to generalise to different cases. While

the second corpus provides a sample input, no model is trained on this data, reducing the likelihood of bias. First, these two datasets are gathered from news articles for the first and collected using participants in an experiment for the second. These techniques will first segment the documents by insider threat characteristics, add contextual information regarding the causality and temporality, and finally explore individual topics combining information from different witnesses into a single visualisation. To segment the documents, topic modelling will be used on the corpus of general insider threat cases and then applied to the documents for a single insider threat attack. This should capture similarities between insider threat attacks and therefore produce topics similar to the insider threat characteristics. This allows the sentences within documents to be ordered within a document starting from the first sentence to the last or combined by topic where different sentences from different documents appear together due to their topic. These two approaches to segmentation are important as often within a document contextual clues which suggest temporal data ”after that”, ”before this” or temporal data ”because of”. To capture this, the topics in a single document is considered a state and modelled as a Markov chain using a list of causal and temporal patterns (Mirza 2016) to include specifically causal or temporal information. This provides a timeline of an incident but is still limited by topic rather than individual sentences, this is to say that the topics represent characteristics of an attack, but there is no specific method to see the details of that characteristics. Finally, CoreNLP (Manning et al. 2014) provides out-of-the-box tools which are used to recognise the distinct actors, actions and assets from different witnesses and merges and combines them.

4.1 Natural Language Processing

Natural language processing (NLP) is a subfield of machine learning and involves the process of teaching a computer to process and ‘understand’ natural language (Manning and Schütze 1999). This natural language could be in the form of human speech or the written

word. Common NLP tasks are machine translation, spell checking, summarisation, voice recognition, and question/answer systems. Machine learning systems are taught human language using many different methodologies; however, they generally rely on statistical methods. Statistical NLP, in contrast to rule-based NLP, allows a system to be flexible and adapt and learn the underlying rules of a language based on a large number of examples, called a corpus (Manning and Schütze 1999). These examples can be used in many ways, supervised learning uses labelled examples to demonstrate some text features, and an NLP system can apply this to an unseen example. Semi-supervised is similar but allows the NLP system to incorporate a feedback loop improving the rules it has learned, and unsupervised relies on a completely unlabelled dataset and requires the system to discover its own rules (Manning and Schütze 1999)

Natural language processing has been developed since the 1950s, with the primary focus of the research in tasks such as machine translation, information retrieval, text summarisation, question answering, information extraction, topic modelling and opinion mining (Cambria and White 2014). Machine translation has evolved from a simple dictionary system for English to Russian in 1954 (Hutchins 2004) to Google Translate and sophisticated neural machine translation systems (Wu et al. 2016) which have been used widely to translate to and from many languages instantly. Information retrieval has made huge developments due to the modern Internet requirements, where searching for relevant information has become extremely important to search engines, and the techniques used have become industry secrets. In addition to search engines, information retrieval is vital as more records become digitised in order to find relevant information and act upon it (Manning, Ragahvan and Schutze 2009). Text summarisation algorithms have also found wide use, with APIs that can summarise pieces of text (Smmry 2018) implemented into bots on social media (autotldr 2017). Question answering has similarly been under major development as Watson, a question-answering system competed with humans on a gameshow (Loftus 2009), the technology has now evolved into other fields, including medical, marketing, finance and engineering (Loftus 2009). NLP is widely used

in many fields as the amount of text information increases, with businesses, organisations and governments eager to generate actionable intelligence from the text that is generated every day.

Information extraction (IE) is closely related to NLP and uses many techniques; however, it differs in the end goal and use. Information extraction is limited to the extraction of structured representations from unstructured text or other data (Manning, Ragahvan and Schutze 2009). this might be domain-specific or open domain. It is considered related to both NLP and machine learning (ML); however, information extraction is not limited to text and includes many different types of unstructured or semi-structured data such as images, video or audio, however, this work will focus on text only. There are many approaches to represent relationships, aiming to increase the potential use of a tool. This usually begins with the extraction of tuples: object, subject and verb (Angeli, Johnson Premkumar and Manning 2015), this is expanded to include relationships with nouns and adjectives, rather than just verbs, sentence fragments rather than full sentences and inter-proposition relationships. IE's primary use is often searching (Manning, Ragahvan and Schutze 2009), for example, searching across the web for search engines or as part of a pipeline for further NLP tasks.

4.1.1 Ethics of Natural Language Processing

When applying NLP, it is important to consider the context of the problem. Traditionally NLP systems aim to learn 'meaning', with a focus on machine comprehension and understanding. However, this machine representation is far from human-level comprehension or understanding (De Vault, Oved and Stone 2006). Insider threat, however, is a human-focused problem, and the implications of using technology have profound real-world outcomes. For example, someone deemed an insider threat is at risk of being fired or, for those who commit larger crimes, arrested and sentenced. However, this issue has been discussed at length in the academic literature, with a focus on bias, such as Bolukbasi et al. (2016). With these human consequences, it is extremely important to discuss

the limitations of NLP and how this technology can be used to support human decision making rather than replace it.

A widely discussed issue in NLP is regarding different biases such as gender bias (Bolukbasi et al. 2016), as NLP uses large amounts of textual data, any bias evident in the text will create biased results, and models must specifically be developed not to replicate this bias. This is a fairly well-known issue in all statistical fields; however, the text is often treated as data in other fields. For example, word embeddings, which is a process used to replicate meaning and understanding in text. Words with similar semantic meanings have similar vectors, and the vector difference between two words can show relationships between them. For example: 'man is to king as woman is to queen', the vector between 'man' and 'king' should be equal to the vector between 'woman' and 'queen'. These word vectors are used in various applications, including for web search. These can also be used in secondary NLP tasks such as topic modelling (Zhao, Du and Buntine 2017). However, these embeddings are often trained on data that enforce gender or racial stereotyping: 'man is to computer programmer as woman is to homemaker' (Bolukbasi et al. 2016). Although there are various critiques of this particular and suggestions on how to debias NLP such as Bolukbasi et al. (2016), these debias systems focus on mitigation approaches all rely on understanding the sociological and societal context of language.

Relate to these issues is the issue of how an NLP model understands meaning(De Vault, Oved and Stone 2006). Often NLP systems are said to understand and learn the same way a child might. By being exposed to language naturally, the rules and understanding are learned. However, a child's understanding is grounded in the physical world. This grounding is extremely important to developing meaning but difficult to replicate for NLP. This grounding usually takes place using additional real-world items such as images or producing various artificial tests (De Vault, Oved and Stone 2006).

These two issues seem separate but are closely linked. The lack of meaning combined with the lack of understanding of sociological and societal contexts can produce biased NLP models that are not grounded for real-world applications. Although they may per-

form well on certain tests, these tests are artificial, and on more realistic tests (Ribeiro et al. 2020) these models do not produce the same accuracy.

Therefore, it is extremely important to consider this context and ensure that NLP alone cannot solve these issues. We need to ground each model using existing insider threat models, providing this richer context. Although these models, as with all things, express a bias, primarily towards public attacks or exist in large databases. This bias is understood and known; in fact, NLP in this instance provides an advantage. For example, when new attack patterns emerge, the more flexible NLP models are likely to recognise these, and the model can be trained with new data to understand future attacks better. This process lets the model adjust and change with NLP, supporting the understanding and aiding in decision making.

4.1.2 Domains using NLP

As described previously, machine learning is already used within the insider threat domain. In the detection of insider threats, anomaly detection is widely used. These approaches use machine learning techniques to learn what normal, legitimate activity is on a network or in user actions and then flags activity that deviates from this norm. However, natural language processing is not widely used in the insider threat domain due to the labour-intensive labelling and model creation. This is perhaps not surprising as creating tools in a single small domain can be time-consuming. This is a problem in several domains, and various methodologies and frameworks have been developed to solve this problem from an NLP perspective (Huang and Riloff 2010). As unsupervised and semi-supervised approaches have become more widely used within NLP, NLP has been used within the insider-threat domain.

One use in the domain has been validating insider threat characteristics (Liang, Biros and Luse 2016), the authors describe NLP as an empirical approach to understanding language and used NLP techniques to evaluate insider threat models empirically. Although NLP was not used to understand insider threats, it shows how NLP is adopted within the

domain. For example, Kandias et al. (2010) used a dictionary-based approach to classify user comments on YouTube videos as being negative towards law enforcement in order to categorise people who are likely to become insider threats. Although NLP is not widely used Linguistic techniques such as discourse analysis, language-action cues and word use analysis have been used by researchers to help to understand the psychosocial side of insider threat (Brown, Watkins and Greitzer 2013; Greitzer et al. 2013; Ho et al. 2016). Much of this work is manual, however, and does not exploit machine learning techniques, focusing on a linguistic study to provide evidence.

Although not natural language processing, many approaches for creating insider threat models such as Nurse et al. (2014b) or the linguistic analysis discussed above use textual data in the form of reports in order to analyse a large number of insider threat cases. This suggests that insider threat as a domain could benefit from natural language processing as it allows the mass analysis of these reports, which can be found internally within an organisation, on news articles, within press releases and discussed on forums such as Hacker News. Furthermore, this volume is likely to increase with the GDPR legislation in force; organisations will be legally bound to report data breaches as they occur to their users. This seems likely to reveal even more insider threat attacks where data has been lost (Information Commissioner's Office 2018).

Although NLP is not yet widely used in the insider threat domain, NLP to understand domain-specific reports in other domains NLP is often used to classify pieces of text within the security domain, such as detecting terrorism-related articles (Choi et al. 2014). Causality mining is used to understand what events caused other events, such as a poor company policy causing stock prices to fall (Radinsky, Davidovich and Markovitch 2012). In the medical domain, text mining of reports has been used to find hidden relationships between diseases and medicine (Bruijn and Martin 2002). In archaeology, NLP has been used to process reports to help archaeologists sort their work using named entity recognition (NER) and relation extraction to find places and people, and relationships between them such as person A lived-in location B (Richards, Tudhope and Vlachidis 2015). In

Engineering NLP, it has been used to detect where two reports describe the same defect and minimise report duplication (Runeson, Alexandersson and Nyholm 2007). In the medical domain, NLP has been used to create search engines for medical records designed to deal with domain-specific requests; this has been met with praise from the medical community allowing doctors and health care professionals to better interact with the medical documentation (Hanauer et al. 2015). The wide range of uses and success of NLP demonstrates how powerful it has been in understanding reports from many different fields. In order to leverage NLP for report understanding in insider threat, a clear pipeline must be developed, creating the steps for the tools that will be used and techniques that can be implemented.

4.2 Understanding Insider Threats

To meet the aim and objectives, this work will use NLP, creating a pipeline for understanding an insider threat case. The first step is to create a series of domain-specific datasets; this is a well-known stage in the creation of NLP tools, especially for unique domains. Next, a technique called topic modelling functions similar to grounded theory, which has already been used to create insider threat models; however, the NLP approach creates a computational model. This alone does not encapsulate all the context; therefore, additional layers are added, temporal, narrative and causal information can be layered to create a richer model. These give an overview of an incident; however, an attack's characteristics must be studied in detail to fully understand an incident and make effective policy changes. For example, while an insider threat characteristic may be 'motivation', knowing the exact motivation (e.g. 'financial due to a large unexpected medical bill') can allow an organisation to make effective, meaningful changes, for example, an employee assistance program to reduce the risk of an insider attack.

To achieve this, several techniques from NLP will be used. Topic modelling is widely used throughout this research; this creates a model using the Latent Dirichlet allocation

(LDA) algorithm, an unsupervised technique that finds topics automatically by examining word frequencies. Many of the other tools are provided by existing toolkits such as the Natural Language Toolkit (Bird, Klein and Loper 2009) and CoreNLP (Manning et al. 2014). These provide open-source NLP tools to reduce development and training times. Additional algorithms have been developed for this project specifically, which provide additional functionality such as training new models or processing custom corpora. Finally, a bespoke web interface was created to aid in the visualisation and management of this research gephi. Visualisation is extremely important as it can aid in understanding data, particularly in supporting: the efficiency of data (showing only what matters), and the information from the data. Therefore by ensuring that the data is visualised, this work can create a greater impact as it can be more easily understood (Iliinsky 2010).

4.2.1 Topic Modelling

Topic modelling is an NLP technique that creates topics of related concepts and words. It is a generally unsupervised technique that uses probability to assign words to a k-number of topics. This is used for many domains and fields, including journalistic texts, scientific journals and Twitter conversations, and is a common technique for categorisation tasks with early working demonstrating this potential such as Blei, Ng and Jordan (2003), specifically for information retrieval tasks such as Lafferty and Blei (2006). This approach is common because topic modelling is not limited to creating topics; these topics can then be applied as a model to previous unseen or the original corpus to create posterior probabilities. Over time topic modelling has been expanded to include more tasks, including automatic labelling (Lau et al. 2011) and performance measures (Arun et al. 2010). In addition to these tasks the most common technique, the LDA algorithm has been improved: dynamic topic models (Blei and Lafferty 2006), supervised topic models (Blei and McAuliffe 2010), correlated topic models (Lafferty and Blei 2006), semi-supervised topic models (Jagarlamudi, Daumé and Udupa 2012), building new functionality into topic models, allowing them to be used for more tasks or more development. Specifically,

Lafferty and Blei (2006) allows topic models to be improved by including a layer of temporal information and can be used to understand how topics change over time. This has been used to analyse scientific topics and journalistic texts (Jacobi, Atteveldt and Welbers 2016).

Topic modelling is similar to grounded theory, both aim to discover common topics or themes within a piece of text, with grounded theory discovering themes, called codes and topic modelling discovering topics automatically using a compatible algorithm. However, most topic modelling algorithms cannot label and give meaning to topics; a human must interpret them. A comparison between topic modelling and grounded theory is shown in Figure 4.1, here the topics are represented in the first column labelled 1-8 with a human interpretation, and the themes from grounded theory in the first row, the matrix represents when they overlap. Many of the themes are also represented as topics. For example, 'Positive Response from Friends' and 'Friends' Reactions' has a strong overlap, as did 'no reaction'. However, other topics show a relationship to several codes from topic modelling, for example, Topic 2 'Necessary for communication' and 'need to communicate as a type of trigger' and 'Reasons for communicative necessity', with the latter also relating to Topic 4 and 7. Topic modelling, however, cannot completely replace this human labour but can aid in: 'selecting what to read and organising documents into groups that are likely to be thematically coherent' (Baumer et al. 2017).

Interpreting topic model results is a recurring issue limiting topic modelling in traditional social science. One such use is Jacobi, Atteveldt and Welbers (2016), the authors use topic modelling to understand journalistic texts. The first experiment considered how news articles have framed issues relating to nuclear technology over time. Creating a topic model over ten topics and identifying those that change over time, e.g. research, cold war, proliferation and accidents/danger, continuous topics, e.g. weapons, power and US politics, and finally, irrelevant topics, e.g. summaries, book reviews and films and music. Using the temporal topics, the topic model found an increase of the accidents/danger topic correlated with Chernobyl and Three Mile Island, with peaks in mentions of this

	Triggers for returning	Communicative necessity	Morality	Renegotiated	Social reconnection	Friends' reactions
1. Positive response from friends						Friends had positive reactions
2. Necessary for communication (distance, tragedy, etc.)	Need to communicate as a type of trigger	Reasons for communicative necessity				
3. No reaction						Friends showed no reaction
4. Brief, focused, guilty return	Utilitarian (e.g., info seeking)		Qualified guilt			
5. Negative emotions (guilt, disappointment, addiction)			Guilt, let myself down	Addiction implies limited control		
7. Stories, obliged return, immediate reaction	Major life events as triggers				Welcomed back	Mixed reactions
8. Positive emotions, changed use				Increased self control	Positive about reconnecting	

Note. Cells describe how each topic resonates with one or more themes.

Figure 4.1: Topic Modelling and Grounded Theory, the table shows example topics (rows) and example grounded theory model (columns) comparing each using the cell

topic in 1979 and 1986. When the topic model is expanded into 25 topics, a hierarchy can be seen, where a topic such as accidents/danger becomes more specific. While originally Three Mile Island and Chernobyl were on the same topic, 25 topics became split into their topics, following their peaks. The authors suggest topic modelling and grounded theory working together during the interpretation, suggesting new codes, highlighting codes that are too vague (and need to be split up) or too specific (and need to be combined). This work demonstrates the potential use of topic modelling with grounded theory to improve existing models or create new ones.

Topic modelling can be used for various tasks, and in particular, this is useful for combining NLP with grounded theory. This will allow the project to meet the objective of creating a custom insider threat framework. However, topic modelling can also be used for classification tasks, and this will allow the project to create a large dataset suitable for NLP, classifying a large number of public reports of insider threat. Therefore topic modelling will be used in two tasks. First, it is used to automatically categorise insider threat and non-insider threat news articles in the Datasets chapter and second to create an automatic, machine-generated model similar to a grounded theory model such as the one

in Nurse et al. (2014b) creating a custom insider threat model.

4.2.2 NLP Tools and Pipelines

Before more specialised tools like topic models can be applied to a corpus, the corpus usually has to go through some pre-processing. This is usually done by existing NLP tools and pipelines, as these are often not domain-specific. The existing tools which will be used during this project are 'NLTK' (Python) (Bird, Klein and Loper 2009) and 'CoreNLP' (Java and API) (Manning et al. 2014) and 'tm' (R) (Feinerer 2013), which tool is used is dependent on the language used. These often provide pre-processing and general NLP tools such as stemming and lemmatising, tokenisation and stopword removal, with some having more specialist tools such as Information Extraction tools. Stemming and lemmatising helps normalise the text, collapsing word tenses into a single token. Tokenisation splits sentences into individual words; this allows NLP tasks to process individual words in a sentence. Finally, stopword removal removes very common words that do not impact a document or sentence's overall meaning. These tasks are usually placed in a pipeline, with each document being processed in the same way for NLP tasks.

The most common NLP pipeline is CoreNLP (Manning et al. 2014), created by Stanford in Java, run as a web API (Application Programming Interface), this includes the core features described above but has some additions. The pipeline is shown in Figure 4.2, each additional tool uses input from the previous, creating rich annotations. These include POS (Part of Speech) tagging, which labels each token by their part of speech, NER (Named Entity Recognition), which labels tokens that refer to names, in particular people and places, tagging and dependency parsing, which allows researchers to view underlying tree structure of a sentence. These additional tools allow the CoreNLP pipeline to add additional annotations; these can be extremely useful for recognising people or places names in text and allowing for more tools that may require POS tags. In addition to the pipeline, CoreNLP also includes additional tools such as OpenIE, which splits sentences into IE triples, sentiment analysis and coreference resolution, which can resolve rela-

tionships between names and 'she'. CoreNLP features several options for this pipeline, including state of the art neural networks. However, this is limited due to the time to train models with state-of-the-art tools, which may not increase performance in some tasks. While there are some limitations, these tools offer prebuilt models and support for many NLP tasks formalising the tools into a pipeline that can be applied to any text. This further has the advantage that as all these tools have already been built, it is appropriate to apply them directly to text without needing additional training sets or additional models, therefore reducing the requirements.

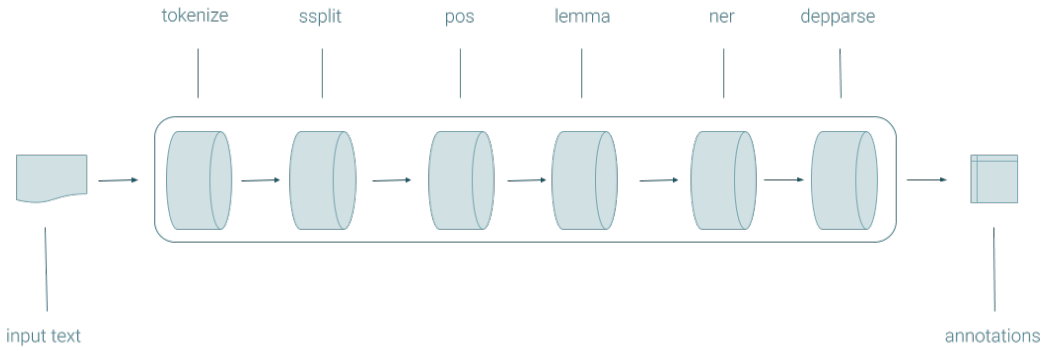


Figure 4.2: CoreNLP pipeline(Stanford NLP Group 2020)

NLTK (Bird, Klein and Loper 2009) offers a simplified NLP pipeline in Python, with many of the additional tools being built open source. NLTK does offer POS tagging, similarity measures, classification and information extraction tools. Additional, third-party tools offer more functionality, including machine translation and sentiment analysis. The final text mining package that was used is the 'tm' package in R (Feinerer 2013), this provides similar tools. However, it adds some statistical analysis, including frequency matrices and transformation tools. Due to the flexible nature of many of these tools, features from different toolkits can be used for the same project, and additional, bespoke tools can be created to supplement the functionality of these pipelines. Additional tools will expand these pipelines during this project to create a tool and domain-specific algorithms.

4.3 Conclusion

NLP can analyse a large corpus much faster than a human and then easily visualise and present the results. NLP is widely used for a variety of tasks in other domains, including machine translation, speech recognition and spell check, as well as being used in other domains for the analysis of large corpora of reports including archaeology, engineering and medicine, which demonstrates the ability of NLP to reduce human labour. Although NLP is not yet widely used in the insider threat domain, there has been a push for its use in cybersecurity, which has been a success in the field. NLP allows organisations to analyse existing data in reports, leveraging these algorithms' analytical power to explore and understand events. This process allows for a custom insider threat model, built similarly to grounded theory models, but with significantly reduced human labour and a focus on dynamic models which change alongside the organisation. Model visualisations and bespoke software are created to open up these tools to non-programmers. This allows organisations to make policy changes and prevent future attacks. This work will use these existing techniques and existing tools to analyse a large volume of text, in this case, witness reports of an insider threat incident varying in detail and language use.

The specific NLP tools will be topic modelling with the LDA algorithm, CoreNLP and NLTK to provide an NLP pipeline. On this pipeline, bespoke algorithms enable additional domain-specific functionality, and finally, the results are visualised and managed using gephi and a bespoke web interface. However, to create these tools, first corpora must be created, which allows for initial models and an insider threat example case. NLP requires large corpora, collections of documents, in order for models to represent the data, with a dataset that is not sparse and can be easily cleaned. The datasets that will be created are a sample of insider threat cases taken from real news articles, providing a breadth of different attacks, a sample insider threat case where different witnesses to an attack provide a report of an incident, and a grounded example, linking the machine topics to human codes from grounded theory. The next section will discuss the creation of these datasets and provide a sample from each.

Chapter 5

Creating the Datasets

NLP requires a large amount of relevant textual data to build models; these are usually domain-specific; however, there are open-domain datasets. In NLP, datasets are called corpora and contain a collection of documents. For this project, two main corpora have been created: First, a dataset of insider threat news reports provides a range of different insider threat attacks and captures language features that insider threat reports have; second, a dataset of organic narratives, where participants were asked to retell the story of an insider threat incident. These two data sets have two uses; the first allows the NLP models to be trained on publicly available data and an example dataset to which these models can be applied. In addition to the general application of models, this second use can aid in the refinement of models. Creating a corpus suitable for topic modelling requires an extensive collection of news reports, and simply extracting articles from known insider threat aggregators is not enough data due to the sparsity (a word may only appear in a single document). Therefore this section also presents a methodology for increasing these corpora of news articles automatically; this is shown in Figure 5.1. Finally, this chapter presents the creation of organic narratives using the 'perspectives experiment' study and labelling a portion of this data using an existing insider threat model.

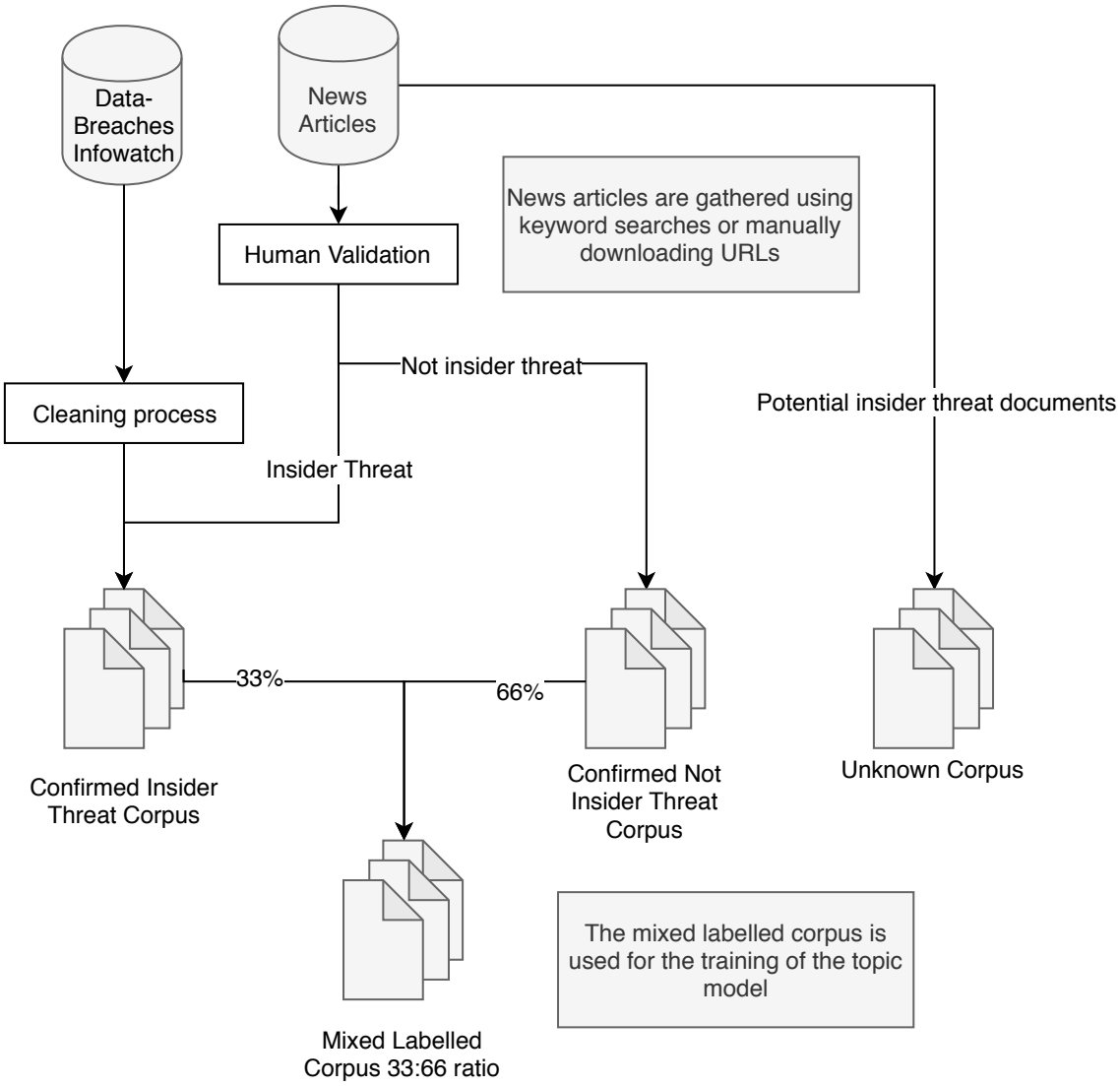


Figure 5.1: The datasets used in the automatic classification system

5.1 Introduction

In order to use NLP, a large body of relevant textual data must be created; this allows for models to be trained, verified and provides potential sample input; these datasets are referred to as 'corpora'. Typically, once corpora are gathered, they are manually annotated with additional information to provide additional context for the documents. Corpora can be domain-specific or open-domain. Due to the nature of insider threat, domain-specific corpora must be created, although open-domain corpora can be used for more general tasks, such as identifying causal verbs and patterns. Existing corpora will be preferred as

the addition of manual annotation often improves the performance of NLP models on particular tasks; hence we can exploit a large volume of existing labelled corpora. However, as there is little work on NLP in the insider threat domain, specific insider threat corpora must be created. In this work, two insider threat specific corpora have been created, first a dataset containing many different insider threat cases, which should capture features that are general to insider threat. The second corpus provides a sample input containing reports with varying levels of details about a single incident. This chapter discusses these corpora, specifically how the data was gathered, augmented, and transformed.

It is important to first discuss the differences between the two proposed corpora collected during this chapter. The general corpus contains many different insider threat cases, and the sample corpus contains many different reports of the same incident. It is essential to gather both as these allow for different operations on the data. The general insider threat corpora will allow topic models to be built on a range of cases that potentially use the same language style. This is important as topic modelling requires the data to be cleaned, with a document written in a similar style, the stopwords are words that are incredibly common such as 'the', 'a' or more specific words from a specific domain, are more likely to be consistent and therefore easily removed Manning and Schütze (1999). However, considering the overall aim of this work, a specific style will not be consistent over multiple witnesses all giving their account of an insider threat incident; therefore, a sample corpus that can be used to develop and test these tools is extremely important. This second corpus will have multiple accounts of the same insider threat incident; these accounts are organic narratives, allowing a witness to write freely in their own voice, with as much or as little detail as the author wishes to include. This corpus ensures that this work is consistent with the ultimate aim of the research and that these models are not specific to one specific style of writing. The following sections will discuss the creation of each corpus.

5.2 General Corpus of Insider Threat

The general insider threat corpus contains a wide range of insider threat cases; these vary in methodology, target industry, attacker motivation and insider archetype of attack. This corpus represents the features that any attack may have, with a wide range of different attacks, allowing for models to be flexible in the attacks it can recognise. This corpus will then be used to train the topic models discussed in the previous section, an unsupervised process similar to thematic coding used in the social sciences. This work uses known insider threat news articles to create this corpus. News articles are written in a similar style; therefore, these articles are ideal for topic modelling, as the stopwords will be similar. Although this is not vital, this general insider threat corpus could be replaced or improved with another, for example, a corpus of cases specific to an organisation or a time-bound corpus that will better recognise emerging insider threats. This section will discuss how this corpus was created and processed in preparation for topic modelling. First, the initial corpus was downloaded from known insider threat news aggregators and then the process by which more documents were automatically classified to increase the size and variety of the corpus.

5.2.1 Inital Corpus

The initial corpus was created from known insider threat cases using an insider threat aggregator. The aggregators chosen were Data breaches (Johnstone 2022) and Infowatch (InfoWatch 2019), these aggregate known breaches for malicious insider threat, accidental insider threat and cybercrime. There are many advantages to using these sources: first is that these aggregators source breaches worldwide, allowing for a greater variety of reports, second these aggregators are kept up to date, and new breaches are reported on very soon after they were published, ensuring that new methodologies or archetypes for insider threat can be found, third these aggregators differentiate between a breach caused by an insider and other data breach types (such as a malicious, external individual), finally, the

same case can be discussed multiple times as new information becomes public, allowing for a wide range of potential insider threat features to be reported. Therefore these sources offered a broad range of types of insider threat cases while not including non-insider threat cases, in addition to being a trusted aggregator. Finally, a custom extractor was created; this extractor reads these URLs and removes those referring to other incident types. For Data Breaches, each article is tagged- only those with the 'insider' tag were downloaded; for Infowatch, any article with the words 'hacker' or 'hack' were ignored. This simple extractor then visited each new article and extracted large bodies of text by identifying those portions over 100 characters within HTML paragraph tags. This corpus was then extended by downloading likely insider threat articles by searching for specific keywords, such as 'employee stole' generated from the insider threat literature. The keywords were chosen by examining the CERT dataset published by Trzeciak (2011), for the incident and subject data. These are then manually labelled as insider threats ensuring that the initial corpus is varied but still contains known insider threat cases.

This corpus was then cleaned and normalised; this is standard practice in natural language processing and was completed using the CoreNLP library (Manning et al. 2014) a set of industry and academic standard for text pre-processing. First, the text is transformed into lower case and stopwords are removed. Stopwords are extremely common words such as 'the' 'a' etc. (Manning and Schütze 1999). In addition, punctuation is removed. Finally, the document is stemmed; this process collapses multiple forms of a word into one removing -ing -ed endings. This process normalises the text; this is important as text can often contain many different endings to the same word, which may be important for a potential reader, however for topic modelling, this does not need to be preserved and is a typical process (Jacobi, Atteveldt and Welbers 2016). Stopword removal uses three specific lists. First common English words are removed, next, a list of stopwords specifically for news articles (Singh 2020), these are the most common words taken from a corpus of news articles and finally a custom list which contains the names of the sources used 'infowatch' (InfoWatch 2019). These sources aggregate insider threat articles from

Figure 5.2: Inital Text : ‘President Donald Trump’s decision to impose tariffs on another $200 billion worth of Chinese goods drew a swift rebuke from lawmakers on both sides of the aisle and business groups. Trump announced Monday that the US Trade Representative would begin to impose a 10% tariffs on Chinese goods ranging from food to fabrics to industrial chemicals. The tariffs will increase to 25% on January 1, 2019, unless the US and China agree on a trade deal. The escalation of the US-China trade war was quickly criticised by both Republican and Democratic lawmakers along with many business groups. All of the critiques centred on economists warnings that the tariffs would ultimately harm US business and consumers by raising prices on imported goods from China.’

Figure 5.3: Normalised text : ‘impos tariff good drew swift rebuk lawmak side aisl repres begin impos tariff good rang fabric industri chemic tariff 2019 unless agre escal critic republican democrat lawmak critiqu center economist warn tariff ultim harm consum rais import good presid donald trump decis impos tariff anoth 200 billion worth chines good drew swift rebuk lawmak side aisl busi group trump announc monday us trade repres begin impos 10 tariff chines good rang food fabric industri chemic tariff will increas 25 januari 1 2019 unless us china agre trade deal escal uschina trade war quick critic republican democrat lawmak along mani busi group critiqu center economist warn tariff ultim harm us busi consum rais price import good china’

other news sources and are well known in the industry. The stemming process collapses word endings, removing tense but still capturing the meaning, normalising the text. This process output is shown in Figure 5.3

When the corpus was examined, it exhibited a high sparsity, with some terms only appearing in one document. This is not optimal for topic models, as the algorithm can more easily generalise if certain words are repeated in multiple documents, giving a wider range of context. When considering mitigation approaches, one method that could be used is to remove sparse terms. Sparse terms are those which only appear in a single document, as topic modelling relies on words appearing in differing contexts to build the model (Blei, Ng and Jordan 2003), if the corpus contains a high amount of sparsity, this model cannot generalise over many terms. This would imply that many insider threat incidents are unique in some way, such as industry or threat archetype. This work aims to create a general model for all insider threat cases, and this dataset must have a wide

variety of cases. However, these cases should all share some attributes. Simply removing sparse terms, however, is likely to reduce large amounts of context, which will likely impede the model further. Another approach is to increase the corpus, adding additional insider threat cases. Increasing the corpus is not necessarily ideal as it introduces more sparsity if the documents still do not share terms (Naveed et al. 2011). However, it is likely that both Data Breaches and Infowatch are not reporting all insider threat cases but do report a wide range and that there are likely other insider threat cases that would be similar enough to improve the sparsity.

5.2.2 Expanding the Corpus

Due to the sparsity of the original corpus, it was clear that more documents would be required for topic modelling to be effective. Therefore, it would require additional articles of insider threat representing new cases. This is done by further searching news articles to find likely insider threat documents. However, due to the number retrieved, manual labelling was considered inappropriate. This would involve labelling approximately 10,000 further articles, where a small percentage would be insider threat-related; any human labeller would also have to be familiar with insider threat as a research field. Therefore, an automated approach was first investigated due to time constraints, and a methodology was developed.

This system uses topic modelling (Blei, Ng and Jordan 2003), an unsupervised method of finding related terms in a document. Topic modelling approaches are commonly based upon the LDA algorithm, and developments have been made to improve performance in particular domains and particular applications, such as information retrieval (Lafferty and Blei 2006) The LDA algorithm creates a statistical model of a corpus-based on the appearance of surrounding words within a document. This creates a cohesive list of keywords that represent, in a statistical manner, concepts crucial to that topic. A topic model can then be reapplied to the data upon which it was created to achieve the posterior probabilities (Blei, Ng and Jordan 2003), these are used as a measure of the underlying topical

decomposition of documents within a corpus (Manning, Ragahvan and Schutze 2009). Alternatively, the model can be applied to previously unseen data to help categorise new documents.

The nuanced nature of insider threat is challenging in this context; many topic modelling approaches perform poorly when the topics for understanding the context of a document are not statistically significant due to the nature of the LDA algorithm and document classification (Blei, Ng and Jordan 2003; Jacobi, Atteveldt and Welbers 2016). Any document is rarely contextualised around a few significant topics; for example, a news article that reports a scientific breakthrough adds additional context by talking about a similar breakthrough, the scientists earlier work or comments from the public or government. In many domains, these small topics are not contextually important enough to limit the understanding of a document; however, in some, they are key, such as insider threat. Insider threat requires the actor to be a trusted individual. The use of traditional, more discriminative topic modelling may classify insider IP theft, fraud and sabotage into external theft, fraud and cybercrime, committed by someone outside of removing the element of the employee.

This approach uses the full topic membership distribution rather than the maximal membership to manage this nuanced problem. This is to say that each document is categorised not by whichever topic appears the most commonly in a document but instead by the individual proportions of each topic in the document. This allows for the consideration of the nuanced topics which may be salient to the overall relevance of a document but yet may themselves have a relatively small contribution to the overall document. The assumption is that insider threat documents contain smaller topics at different ratios than non-insider threat documents. For example, using the traditional document classification approach, both insider fraud and fraud may be classified together; they may share an overwhelming number of terms between them, the only different factor may be that the perpetrator of one was an employee. Using the topic model results (topic distribution), a classification system may recognise what proportions of topics may imply that the doc-

ument is regarding insider threat. An example of this is found in (Blei, Ng and Jordan 2003). This method can be expanded for insider threat specifically due to its nuanced nature, such as its use by Trzeciak (2011).

This section will describe the method used to create the insider threat corpus, which will extract key insider threat characteristics in the next chapter. It will consider how the data was collected, how these documents are classified, and finally, the corpus is analysed to reflect on both the method and the new corpus. This experiment was originally piloted by choosing a small subset of queries and manual classification; however, given the requirements on corpus sparsity and relative rarity of insider threat articles, this was later used as labelled datasets for the categorisation. However, analysis on topic proportion of a single insider threat document and prediction can be seen in Figure 5.4. Using topic proportion rather than membership is discussed in Blei, Ng and Jordan (2003) and Jacobi, Atteveldt and Welbers (2016), however this was not used with a document classification system.

The first stage is to increase the number of potential insider threat articles; this is done using the bespoke downloader and combining more potential insider threat keywords. These keywords were initially gathered from the CERT dataset (Trzeciak 2011), by extracting insider threat attributes such as ‘disgruntled’ and combining these with other keywords from the insider threat literature, such as the insider threat modes, by examining the attributes of both attacks and insiders specifically. These are then expanded, finding synonyms ensuring more coverage, creating a total of 13,197 keywords. Each keyword is combined into a two-word query; this allows for the automated discovery of many different news articles but complies with the APIs existing rate limiting. As these queries are generated automatically from the existing insider threat literature, such as ‘applications + cybercrime’, ‘advisor + disdain’, ‘roots + unemployment’ and ‘crime + employee’, the list is very broad, and many may not be related to insider threat. However, this is less restrictive and, therefore, should find new reports that were not extracted by the first stage of searches or in the insider threat aggregators. However, the broad nature of this search

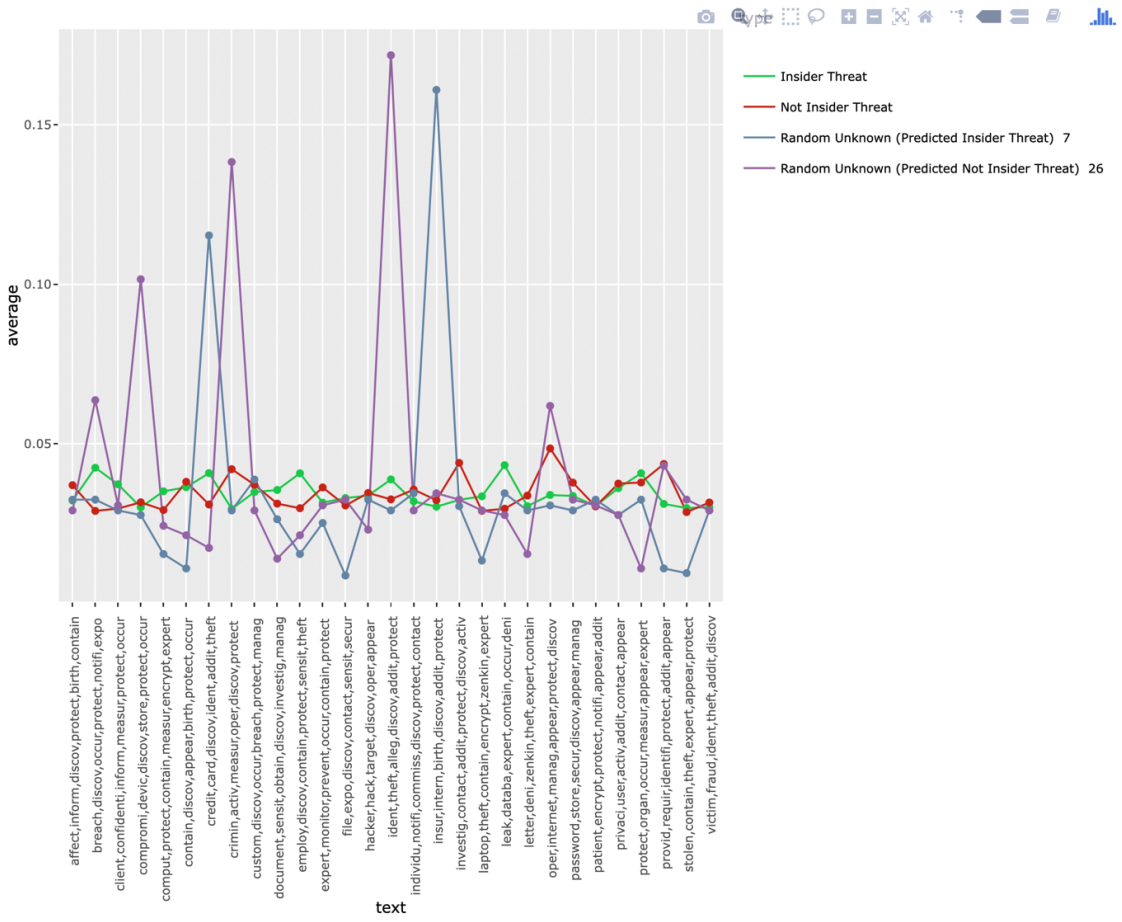


Figure 5.4: Results of the Pilot study

includes both insider threat and non-insider threat documents; therefore, this must be filtered further. This process creates the target corpus, the 'Unknown Corpus', that then must be labelled as insider threat or not insider threat during the classification process.

During the first stage, a similar method was used, with a much smaller wordlist; these were selected from the insider threat literature and specifically included insider keywords such as 'employee', these documents were then classified manually as insider threat or not insider threat. Again, this smaller wordlist was created manually, choosing the most likely insider threat words from the initial list, primarily these involved words relating to an incident combined with words relating to employment (employee, contractor).

This dataset contains 2,096 labelled insider threat cases and has a representative class imbalance of 33% insider threat and 66% not insider threat; this was an arbitrary pro-

portion but chosen due to the rarer nature of insider threat in the corpus. This creates the following corpora: A known, confirmed insider threat corpus, a known confirmed not insider threat corpus, and a mix labelled corpus of a 33:66 ratio. In all the corpora, each document has minor pre-processing to normalise the text, as discussed above, stopwords (domain-specific, English and custom) are removed, and the text is stemmed. These corpora are shown visually in Figure 5.1.

The next stage involves training the topic model. This topic model is created from the mixed labelled corpus and represents both insider threat and non-insider threat documents. The posterior probabilities are then reapplied onto each document in this training set. This generates a numerical representation of the topic membership of each document. Traditional topic models would, at this stage, typically assign the document to the most dominant topic. However, since identifying attacks from internal actors (rather than external actors) requires a nuanced understanding of the text, this approach considers the entire membership distribution. At this stage, each document is assigned a value for each topic; this represents the amount that this document is 'about' this topic. This final topic model contains 100 topics; this was chosen using the Cao et al. (2009), Arun et al. (2010) and Griffiths and Steyvers (2004) algorithms who suggests that for a corpus of this size a large number of topics (100 topics) is appropriate.

These topic membership distributions, having been labelled in the previous steps as insider threat, are then used to build a classification model. In this work, a random forest (Breiman 2001) classifier from the R package 'randomForest' is applied (Liaw and Wiener 2018); however, this classifier could be replaced with any suitable classification algorithm, random forests were chosen as they have been shown to be comparable to other classifiers in NLP tasks such as (Palomino-Garibay et al. 2015) and can estimate the importance of each feature, in this case, the topic proportions.

A classifier built with 500 trees delivered the best performance. The confusion matrix for this classifier is shown in Table 5.1, when measured using cross-validation, splitting the training corpus into a test/train set at 33% test / 66% train and a topic model containing

	FALSE	TRUE	Error
FALSE	545	154	0.22
TRUE	96	1096	0.08

Table 5.1: The confusion matrix for the cross-validation classification

100 topics. The results are presented in a confusion matrix, this table shows the number of true positives and true negatives when compared to the number of false positives and false negatives (B. 1997). While the misclassification is high in the case of true negatives, this is small when compared to the true positive rate, and due to the process of topic modelling these documents can be filtered further on and these outliers may not matter as any irrelevant documents will tend to form their own topics. The recall and precision trade-off will always be a concern in machine learning (Gordon and Kochen 1989), so it is important to consider alternative methods for handling false positives.

The classifier can then be applied to the corpus of documents from the wider news and blog posts to create two corpora, one of 'predicted insider threat' and one of 'predicted non-insider threat'. This evaluation process is shown visually in Figure 5.5. The predicted insider threat corpus will be the corpus that will be used in future experiments. Therefore, this corpus must be evaluated to ensure that the new documents are relevant to insider threat and are not miscategorised. Due to the size of the corpora, is it not feasible to manually examine each document to confirm the classification. The documents that are not predicted as insider threats can be discarded and not used in the process further; while these could be examined for accuracy, there are many more discarded documents than predicted documents, which would be too time-consuming.

To enable the investigation of these documents, a topic model is used to summarise the corpus using a total of 20 topics. This was chosen arbitrarily to enable a summary; this method has the additional advantage of finding documents with a low proportion of relevant topics and removing them. In this case, the remaining topics were manually categorised. This aims to capture new insider threat incidents rather than reports of the same incident from a different news source. Classifying new, unseen insider attacks is

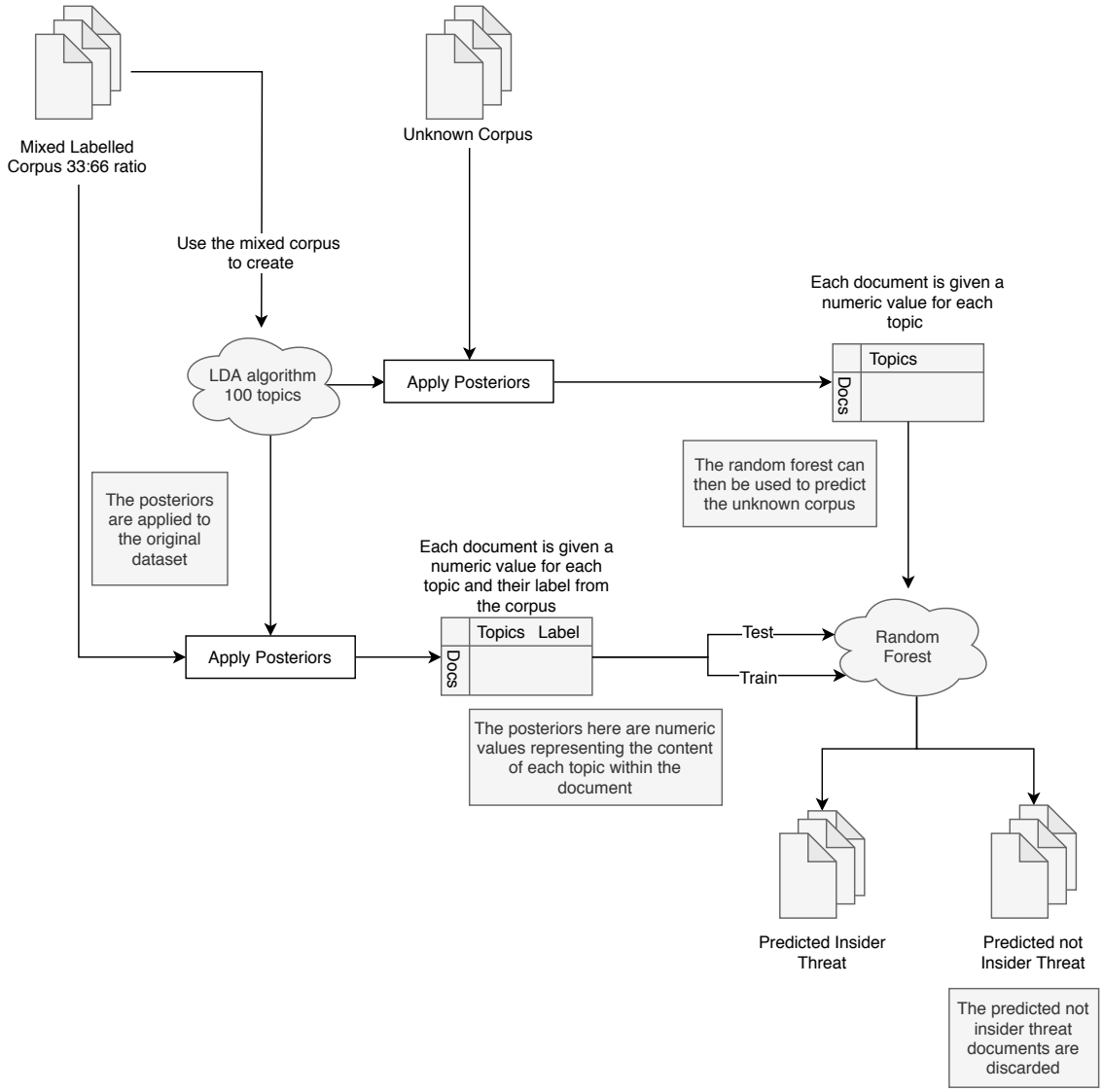


Figure 5.5: The training process

done using topic models to summarise the new corpus. If the model has been successful, some topics will be shared, others will have words in new contexts, and new topics should emerge. The final corpus is then manually evaluated by insider threat experts.

There are a small number of documents that have been incorrectly classified, and this is usually due to specific themes which may be common in related text; these can be removed using the traditional approach of discriminative models to filter and remove documents with high proportions irrelevant topics, a common technique for topic modelling (Jacobi, Atteveldt and Welbers 2016; Baumer et al. 2017). For example, several news articles discussing Donald Trump's presidency tend to exhibit similar language to

insider-threat attacks, with language such as 'whistleblower' and 'Whitehouse insider', which could imply insider threat. While these misclassifications could be reduced further by adding more examples of these documents, the data collection represents a small stage of this research and would be too time-consuming. These documents form a unique topic at this final stage, making it easy to identify the documents associated with these large systemic errors. This evaluation process is shown visually in Figure 5.6.

The results of the topic modelling summarisation are shown in Figure 5.7 and Figure 5.8, these were created in R (due to familiarity and availability of implementations of algorithms such as LDA, as well as visualisation tools such as ggplot2), and these results use the text mining, topic model, tidy text and wordcloud packages. Figure 5.7 shows the manually categorised documents either by downloading these from an insider threat aggregator or manually labelling them. Figure 5.8 shows the automatically labelled documents. Of note is that these words are stemmed as part of the pre-processing, so for example, the word employ covers words such as 'employed', 'employee'. These visualisations show the ten most discriminative terms for each topic, and the size represents the beta coefficient.

This is considered successful if these two visualisations from Figure 5.7 and Figure 5.8 are similar but not identical, this shows that the automatically categorised corpora is relevant whilst finding new, unseen insider threat cases. These two figures are similar, with 'breach' and 'employ' appearing in many topics across both corpora. This shows that similar concepts are being discussed, specifically related to breaches and employees. However, there are differences in the smaller, manually verified corpus; there are several topics which related to particular incidents from the insider threat literature, for example, the allegations of Russian money laundering using an insider at Danske bank (topic 10) and an insider who leaked many damaging emails from Royal Dutch Shell (topic 4). The larger corpus has fewer examples and instead shows overall concepts or 'classes' of threat; these are more general and suggest that the cases are new and more varied across the insider threat archetypes, rather than showing the same cases from the existing ones

literature. This demonstrates that this model is not overfitted and can generalise on insider threat cases. Interestingly there are a number of topics that are related through sharing a few words, such as the topics containing legal terms, e.g. topic 9 in the predictions and topic 14 in Figure 5.7, the visualisation of the manually verified corpus. These topics share some words but not others, suggesting that similar words appear in the documents but in differing contexts; this is further evidence that the system is appropriately trained and able to generalise the concepts seen in the training dataset.

Within Figure 5.8, the automatically categorised corpus, there are some potentially irrelevant topics, for example, topic 16, which appears to relate to compromised cryptocurrency exchanges. However, there may still be an insider element to this topic (although the most discriminative terms involve the cryptocurrency exchanges). While there are measures of similarity such as Keeney et al. (2005), these all require a dimension of human analysis of each topic; therefore, a purely human decision analysis was chosen, particularly as insider threat is a nuanced field. Depending on the application of the corpus, these could be filtered out using typical discriminative topic models, as discussed previously, or by manual inspection of documents that have a large amount of these topics (a significantly smaller task than verifying the entire corpus).

The predicted, automatically categorised corpus has met the expectations, with it being similar to the manually labelled corpus. In addition, the two corpora share some topics suggesting shared concepts; some topics share some words but not all words suggesting that these words have appeared in new contexts, and finally, the model has shown some emerging topics demonstrating new cases of insider threat.

Hence, we can conclude that the method allows the identification of the nuanced differences between attacks originating from internal actors and those originating from external actors. Furthermore, this allows the automated creation of a large corpus of reports of insider threat attacks for an ongoing stream of news articles. In order to measure the effectiveness of this approach, summary topic models are then used to explore the predictions from the classifier and a corpus of labelled documents. Preliminary heuristic

validation by human experts suggests that this approach effectively classifies a large corpus where the documents include smaller periphery topics that are conceptually important but not statistically significant.

5.2.3 The Final Corpus

The final output of this process is a large corpus of 3,500 documents that represent a wide range of insider threat documents, from specific, well-publicised cases to the general concepts surrounding insider threat and its archetypes. This process does not ensure that all archetypes are included, particularly in equal numbers. However, some insider threat archetypes are less common than others, or some archetypes become less common over time as controls are put into place. Therefore, it is likely that there will be a natural distribution of insider threat cases with a sufficiently large enough corpus, with the limitation that these are publicised cases. As there is no existing corpus of insider threat cases, this corpus becomes the general insider threat corpus discussed at the start of this section. Therefore, this process has created a domain-specific general insider threat corpus that represents a large variety of cases and therefore contain elements that all insider threat attacks may have, containing cases that vary in methodology, insider, attack target, and attacker motivation. Although this corpus was created using news articles, this corpus could be modified with more specific incidents of insider threat within an organisation. The presented method could expand other corpora, where there is a nuanced problem domain. This classifier could potentially also be tied to a news feed, automatically classifying new insider threat attacks. Allowing the model to be updated with new insider threat attacks continually.

This corpus will be used in the following chapter, where general topic models are created. However, in addition to this general insider threat corpus, it is important to have a corpus of organic narratives; several documents discuss the same insider attack. It is important to collect the data for both corpora as the next will represent the sample input for the system and allow the topic models to be tested developed and ensure that they are

fit for purpose. The next section discusses the collection of this data in detail.

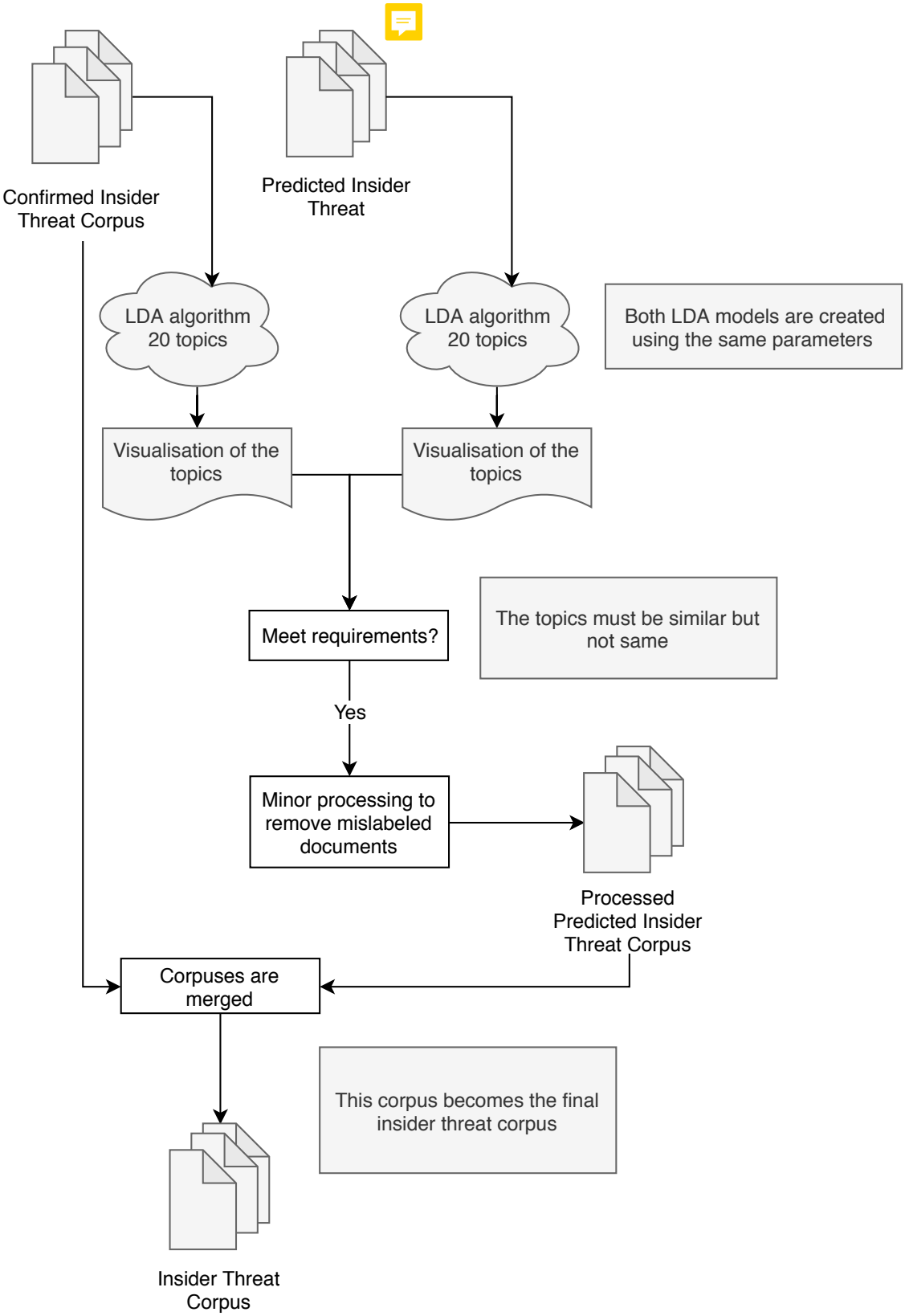


Figure 5.6: The evaluation process

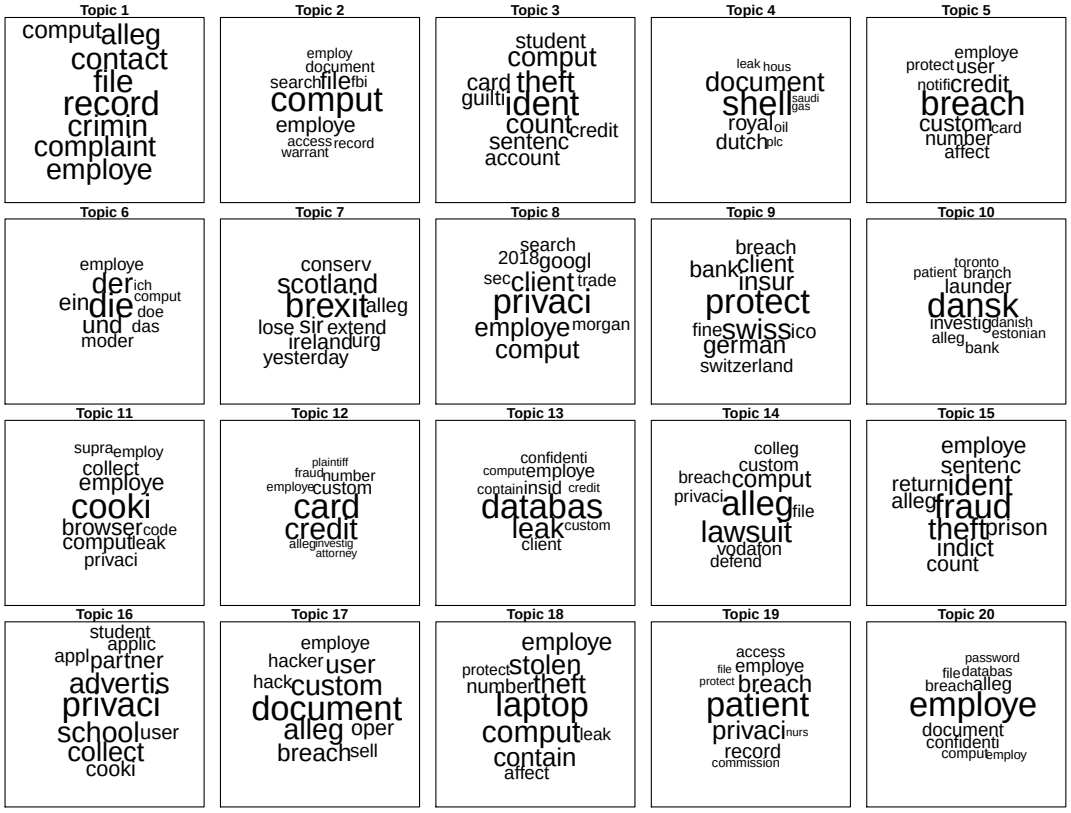


Figure 5.7: The Manually categorised corpus summarised using topic modelling

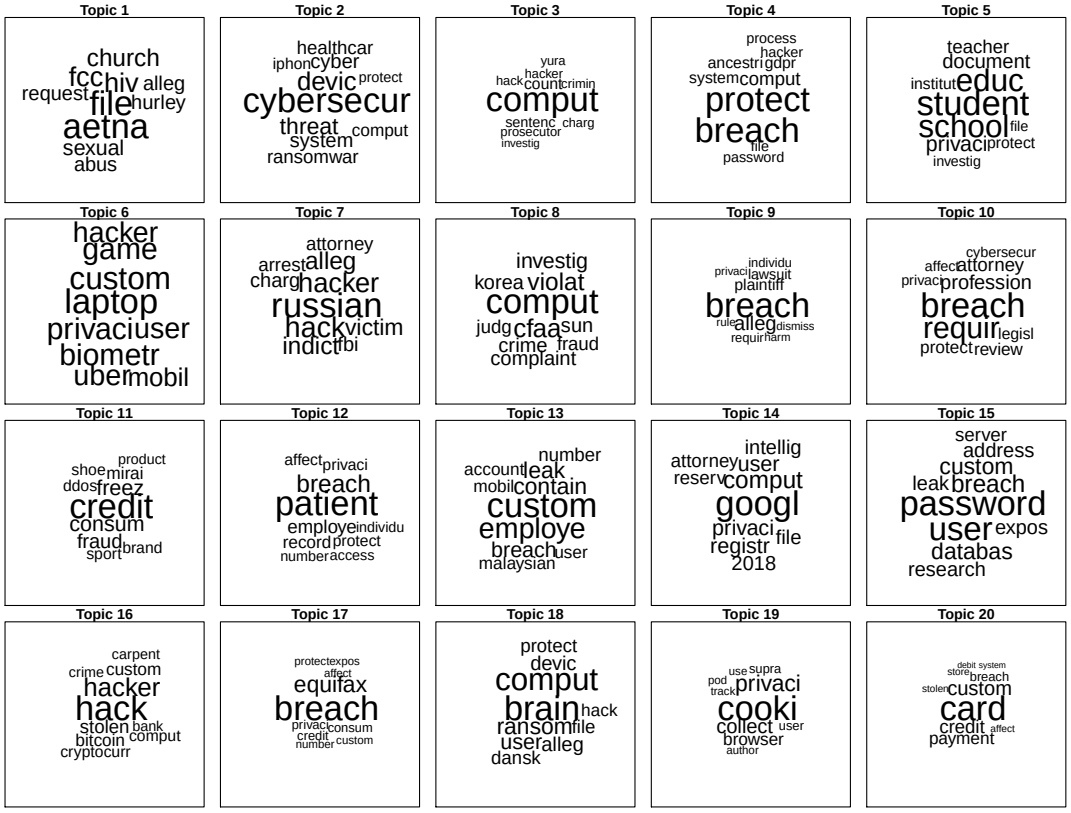


Figure 5.8: The automatically categorised corpus summarised using topic modelling

5.3 Corpus of Organic Narratives

The perspectives experiment will produce a dataset that closely resembles the eventual input data for the system. Creating this dataset allows the tools developed during this research to be grounded by the eventual input data. This corpus should be representative of the input data and therefore should contain multiple reports of the same incident written by different individuals, therefore differing in style, level of detail, format, formality and tone. Contrasted with the general insider threat corpus, this will contain the same incident written in these different styles, compared to separate incidents written in a very similar style.

This corpus of organic narratives can be used throughout the research as a sample insider threat case. In addition, it is also used as a validation corpus, and as a grounding corpus and as a development tool. Although initially, it provides a simple sample insider threat case, throughout this research, by labelling a portion of the data, it becomes multiuse. In this section, the creation of this corpus will be discussed, first the rationale behind creating the corpus, next the ethics and data collection methodology and finally in labelling the data to ground the NLP models in an existing insider threat model.

This corpus will consist of organic narratives; these are written narratives written in any style, allowing for an individual to be unconstrained. Particularly these are witness accounts where style, format, tone or formality is not dictated to an individual, and instead, individuals are asked to write in whichever form they feel is natural. Therefore, this will create a corpus of the same incidents written by multiple people with these key differences in style. This is particularly important as not every member of an organisation will have the same educational, socio-economic, technical or employment background; however, they are all important for understanding the incident. In addition, not every member of an organisation will have witnessed the same chain of events. For example, those outside of an IT department will not have necessarily witnessed any technical exploitation but may have instead witnessed the insider's declining mood or attitude towards work. Ensuring that this data represents these key differences in both style, format, and detail will en-

sure that format and levels of detail, this work remains relevant and implementable in an organisation.

Organic narratives are particularly important; they allow participants to write in their own way and do not force a particular style on them. This presents the initial system with insider threat reports in a similar style to witness reports that may be collected as part of an investigation. There are three key advantages of this approach, first the reduction of bias specifically towards a bias towards technical descriptions of insider threat incidents; second, the reduction of cognitive requirements when individuals are asked to formalise their experiences into a report; and finally, reduce the expertise needed to contextualise a report into an existing insider threat framework. To reduce bias, many existing insider threat solutions focus on technical details, such as insiders accessing files or bypassing restrictions. Although this can be extremely important for IT governance, many of the recommendations for managing insider threat at an organisational level focus on human factors rather than technical elements, for example, processes for employees leaving or joining an organisation (Cappelli, Moore and Trzeciak 2015; Greitzer et al. 2013). Therefore, it is extremely important that many aspects of insider threat attacker considered rather than strictly those which may be visible on a network level.

By reducing the cognitive load of asking individuals to formalise their experiences, reports are more easily gathered from a wide range of witnesses. However, there is still a barrier to asking people to report incidents; research has shown that there are many witnesses to an insider threat attack who would be willing to provide reports but worry about whether or not behaviour may be report-worthy (Forte 2019). With this organic narrative approach, individuals are asked to report in however much detail they would like with the assurances that pertinent details are more likely to be considered with other reports. Finally, this approach reduces the need for experts on insider threat, specifically and those able to use and exploit existing insider threat models. This can also allow people to feel more comfortable reporting incidents as if they are party to information that may be considered privileged and do not feel comfortable reporting it.

In particular, to gather these organic narratives, individuals are asked to listen to three perspectives, a member of IT staff, a colleague, and a news report, as a representation of likely insider threat perspectives an individual may hear from, but that will also vary in particular details such as technical understanding, and then are given a 'free-text' box and encouraged to retell the story in their own words. This ensured that participants were not pressured to remember all the details, nor were they asked to consider one perspective over the others. Therefore, producing representative organic narratives without the need to gather data specifically from an insider threat case. This experiment was given full ethical approval by CURES (Cranfield University Research Ethics) and was run on a custom-built platform hosted on Google cloud platform. After the data has been collected, some of the data was manually labelled, and this allows for both supervised and semi-supervised topic models by aiming to create models which closely follow the known insider threat framework. By connecting the existing framework to an NLP model in this way, the model can be grounded to the framework, creating a kind of ground truth for NLP, rooting it into a societal context (De Vault, Oved and Stone 2006). However, it is possible to encode bias in this way, and this has been mitigated by using people familiar with insider threats from different backgrounds to mitigate the bias as much as possible. Each participant had different levels of familiarity with insider threat, ranging from knowledgeable to expert and offering space to justify their choices. These mitigations may not be enough, and by not training the topic model directly on this data and instead using it as part of the tuning process, if there are significant biases encoded with a poor output, the tuning can be discarded. By the end of this process, the final product is a corpus that represents real data, one insider threat case written by multiple people, each with differing levels of detail, language use, formality, writing style or other linguistic features.

5.3.1 Data Collection Methodology

This section will discuss how the data was collected using the perspectives experiment. This experiment forms the basis of the organic narratives and provides the corpus of or-

ganic narratives, which will be used in future work. This section discusses the experimental design, ethical approval, the results of the pilot, the results of the experiment and how the data was analysed. This experiment was created using a custom platform launched via Google Cloud Platform. Participants are asked to listen to three audio recordings (a transcript is also provided); each recording is a report of the same insider threat attack from different perspectives, a co-worker, a member of IT staff and a news report. The task then asks participants to retell the story with the prompt: 'Who was the person? Why did the person do it? How did they do it? What happened to the person after the events?', these were chosen arbitrarily but mapped to the existing insider threat model literature. This approach was chosen to ensure that participants do not just repeat a story and that they may forget certain details to represent realistic data better, as well as encourage different tellings of the same story. While the data gathering from new reports gives different events written in the same way, this experiment explores the same event written in different ways.

NLP models often perform better when grounded in an existing model (De Vault, Oved and Stone 2006), especially for those with human factor considerations. Therefore, it is extremely important to ground this experiment in an existing model that can be compared and contrasted. The case chosen for this experiment is analysed in (Nurse et al. 2014b); this was chosen as this model is particularly flexible and can apply to many insider threat cases. The adaptability of the model is important at this stage as the aim of the research presented is to create an adaptable model.

The incident chosen for the participants is a dramatisation of a sample insider threat case from the existing literature using audio. The case presented concerns an insider fraud case from Cappelli, Moore and Trzeciak (2015), which the Nurse et al. (2014b) model was applied to, by the authors (providing the grounding): where a bank manager was able to write fraudulent cheques, hiding the evidence by taking advantage of a paper-based system, eventually caught when a cashier spotted the cheque. The catalyst for the insider attack was the introduction of a computer-based system, something the insider

helped design but ultimately could not exploit. The insider was motivated by money, with a history of gambling and treating others in the office, although colleagues were not suspicious as she informed them she had inherited the money. This is very typical of insider threat and is encapsulated in the insider fraud archetype. However, because the model will not be directly trained on this data, instead trained on a wide variety of insider threat cases from the general insider threat corpus and applied to this corpus.

This case was chosen due to the mix of perspectives involved, as it includes both colleagues (who may have noticed her spending or general mood) and IT staff (who observed the reluctance to use the new system), as well as having a general outcome (caught and fined). This case offers a large range of details; however, these details would have been noticed by a particular member of the team. Therefore, the following perspectives were chosen, a news broadcast describing the overall details of the case, a colleague who witnessed some concerning behaviour and a member of IT who only witnessed the technical aspects of the attack. It is likely that those with certain backgrounds may only notice some of these details, which is likely to be realistic to reports written after an attack; for example, someone with a technical background may focus on the technical aspects. As participants are just asked to retell the story, this is not important, as, with a wide enough range of participants, all details will be recorded in some way.

Each perspective was written by a professional in the field to allow the experiment to seem realistic. For example, the news perspective was written by a journalist who currently works at the Daily Mail, the IT staff perspective was written by a former IT support manager, and the general perspective was written by someone with a performance background. These perspectives were also recorded by the authors. Using audio instead of text for this task was done to ensure the participant had to transform the information, hopefully, to ensure that the writing was original and did not borrow elements from the perspective, allowing the data generated to be representative of insider threat reports. A video was not used for time and equipment considerations.

The experiment was deemed a low risk for ethics and data management. The ex-

periment collects no personal or demographic information and responses are completely anonymised. Each participant is given a participant identification which allows them to ask for their response to be deleted. Responses are stored in plain text files, stored on an encrypted drive with a backup provided on the University file server. This data will be documented and made available for reproducible research.

After piloting this experiment with 10 participants, feedback was collected to ensure that participants understood what they were being asked to do, and various changes were made as a result of this. The majority of the respondents found the task challenging and commented that the task made them feel concerned about not remembering details or that the task felt too much like a test of memory. The addition of allowing participants to take notes was piloted, but this caused participants to copy more, and the organic narrative was lost - with participants focusing on the precise details, both short notes and longer notes were trialled. It was decided that notes would not be appropriate, and instead, a prompt was used ('Who was the person? Why did the person do it? How did they do it?') along with different language explaining that it is normal for them to forget details. This change was liked by participants who found the experiment less intimidating. Participants reported that they would also like the task to be changed to question/answer format for the prompts. However, this would not be appropriate for the experiment as prescribing the format would also lose the organic narrative, prompting the participant to answer in a certain way.

From this initial release, 41 responses were gathered from the general public and became the first half of the experiment; however, it was clear that ideally, 100+ responses would be preferable. The task was added to Amazon's Mechanical Turk; this offers a platform to pay people a small reward to complete tasks that only humans can do, such as transcription, allowing for the increase of responses. Using Mechanical Turk, a further 66 responses were gathered, creating a total corpus of 107 documents. While the Mechanical Turk experiment phase was ongoing, the existing data was analysed and processed. An example output is shown in Table 5.2. This data was analysed and did not differ in Flesch

(Flesch 1948) reading ease score.

Table 5.2: Inital Experiment when compared to Mechnical Turk

Initial experiment	A middle manager at a company committed fraud by using an old paper-based system to report false numbers instead of a new electronic system, which would have caught her out. Afterwards she was required to pay back the money (over $60 million), another 9 conspirators may also be charged.
Mechanical Turk	–>A manager at a tax office was able to commit fraud for 18 years and steal from the company because her company allowed her to use paper system for her department. –>When the company introduced a new computer system to replace the paper based system so as to avoid any frauds, she denied to adopt the computer system. Her department was surprisingly given exemption from adopting this new computer system. This helped her to continue her fraudulent activities. –>Nobody suspected her and her fraudulent activities went undetected until one day a bank teller reported a suspicious cheque. It was then that her crime was uncovered. –>It is suspected that she also had accomplices who helped her in covering up the paper-based records. –>Her office was surprised when they found out this news. The manager is described as kind and generous by her colleagues. –>The manager was fined more than 60 million dollars for her crime. –>Taking a lesson from this incident, the company no longer exempts any department from adopting the computer-based system.

Once the data has been collected and cleaned (e.g. correcting general spelling mistakes, normalising the text as discussed previously), the corpus can be analysed. Unlike the general insider threat corpus, this corpus is widely used throughout this research, as this provides a simulated version of the eventual input to an operational system. The analysis of the data will be used in both the attack language, where key insider threat characteristics are extracted and topic analysis, where an individual characteristic can be explored objectives. The first stage in analysing this data is the labelling of a portion of this data. This allows a topic model to be grounded by an existing insider threat model, an important step to using NLP in this domain in particular. The next section will discuss this process, the methodology used, and the output of the labelling process.

5.4 Labelling the Organic Narratives

The next stage in creating the corpora to be used in the experiments detailed in the following chapters is the labelled organic narrative corpus. Using a grounded theory approach, each sentence from a subsection of the previous experiment is coded according to an existing insider threat model (Nurse et al. 2014b). This provides a labelled organic narrative corpus which can be used for supervised or semi-supervised NLP techniques; these require some information to be known about the data; in this case, this is used as a form of human topic model. Instead of machine-generated topics, human judged codes are used, with the codes originally generated from models. Topic modelling is often considered analogous to grounded theory (Baumer et al. 2017), making this process ideal for validation and labelling of topics. In particular, this process allows for the model to be grounded in an existing social sciences research methodology, in this case, topic modelling, improving the accuracy and relevance of the later NLP experiments.

As discussed in the ethics section, having a grounding in a social science model can be extremely important, especially for a socio-technical problem such as insider threat, where the impact of investigating has far-reaching social implications for individuals. Grounding the work in an existing model, the results from the NLP techniques can be relevant and take into consideration multiple aspects. This grounded approach will primarily be used during Chapter 6 as a method of validation and interpretation. Specifically, during topic modelling, much of the context can be lost during the unsupervised approach. By layering this additional labelled data, more context can be gained.

The result of this experiment will be similar to the result of topic modelling but with labelled topics. Each document is broken down into individual sentences and assigned a code depending on an existing insider threat model. Sentences over different documents can then be organised by document, the structure or by code, the meaning. Therefore, related sentences, which discuss the same insider threat characteristics, can be organised together. Due to the similarities between coding and topic modelling, this creates an ideal validation dataset, providing a ground truth.

During the data collection process, an initial experiment subset was completed prior to the release of the Mechanical Turk experiment. Of this initial release, there were 41 documents in total, with varying levels of information and detail as expected for the final corpus. This data is then labelled by people knowledgeable in insider threat, but not necessarily experts, as they were simply applying a model and choosing which characteristic each sentence represented; doing this ensured that the data collection could be completed in a timely manner and ensure the final thesis could be delivered. While this may have an effect on the data, this piece of research represents a first step in using NLP within the domain of insider threat, and as the field develops further, it is likely that more robust datasets will be collected. The experiment received ethical approval from CURES, and six researchers were chosen to code the data were chosen. These individuals are familiar with insider threat and the insider threat literature. Each human coder was given a spreadsheet and a drop-down list of all the codes, and an instruction document. The instructions prompted each participant to examine each sentence and choose an appropriate category from the list. If a participant was not sure, they were asked to choose one and make a note of additional categories. This was completed quickly by most participants, and each gave feedback explaining their decisions.

5.4.1 Human-generated Topic Model

The Human-generated topic model experiment, which forms the labelled perspective experiment, creates a dataset that is similar to machine topic modelling; however, as it was labelled by a human creates an 'ideal' topic model. This experiment uses a subset of the perspectives experiment, split into individual sentences, and assigns each an insider threat characteristic from the Nurse et al. (2014b) model. This is done by asking participants, who are familiar with insider threat, but who may not be experts to apply characteristics to each sentence. These opinions are then collated, creating an ideal characteristic for each sentence and therefore an ideal 'topic'. This concept is similar to grounded theory, which was the approach used to create the Nurse et al. (2014b) model that was chosen.

The Nurse et al. (2014b) model was chosen for several reasons, as the initial dramatisation for the perspectives experiment used a case study, with an example of the model applied, this provided a useful reference for individuals completing the tasks. In addition, the Nurse et al. (2014b) model was created using a grounded theory approach for a general insider threat model; this is similar to topic modelling and consistent with the overall aim of the research. The Nurse et al. (2014b) model identifies insider threat characteristics, and these represent attributes that all insider threat attacks may possess. In this experiment, these are also called 'codes' to distinguish them from the machine created 'topics', as both represent insider threat characteristics.

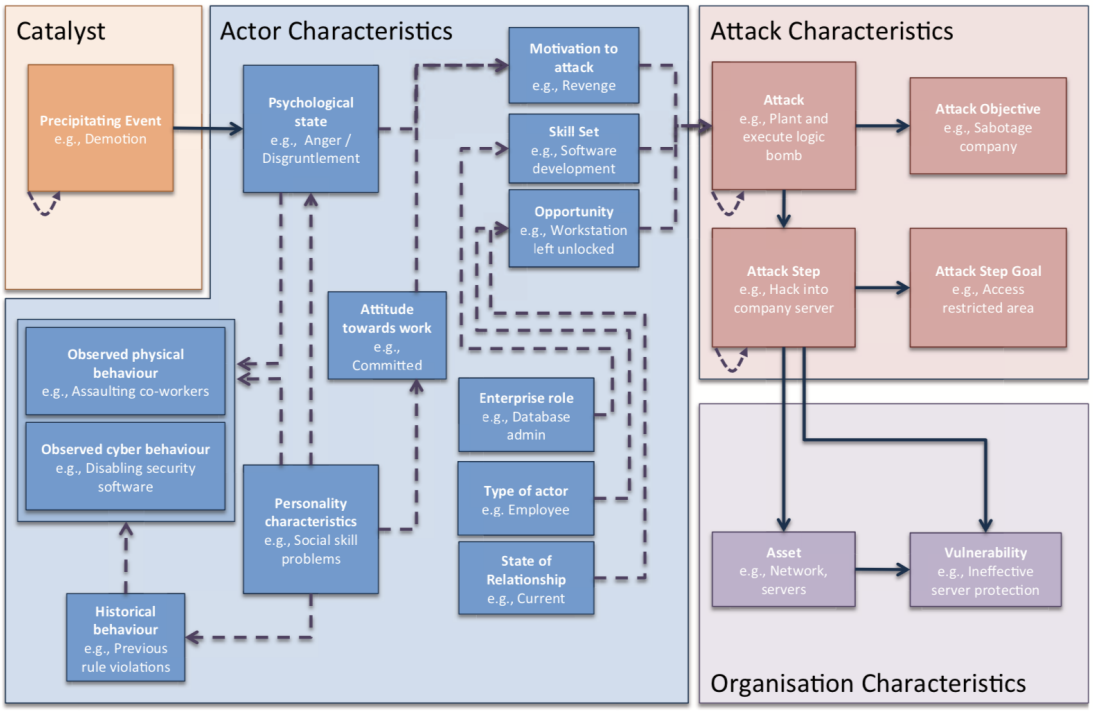


Figure 5.9: Insider threat framework by Nurse et al. (2014b)

The participants were asked to use the Nurse et al. (2014b) model, both the initial model and the specifically labelled model shown in Figure 5.9 and 5.10. These participants ranged from experts to casual familiarity to ensure that they were diverse and that results represented the model rather than an individual's knowledge of insider threat. Each participant was given the choice of how to answer, with each sentence being given a code and a location to write notes. This was primarily used during the piloting process where

individuals would give feedback, especially if they were conflicted between two potential codes. This was completed using Excel, using formulas to allow participants to confirm their code selections with examples and the full characteristic presented in the original Nurse et al. (2014b) model.

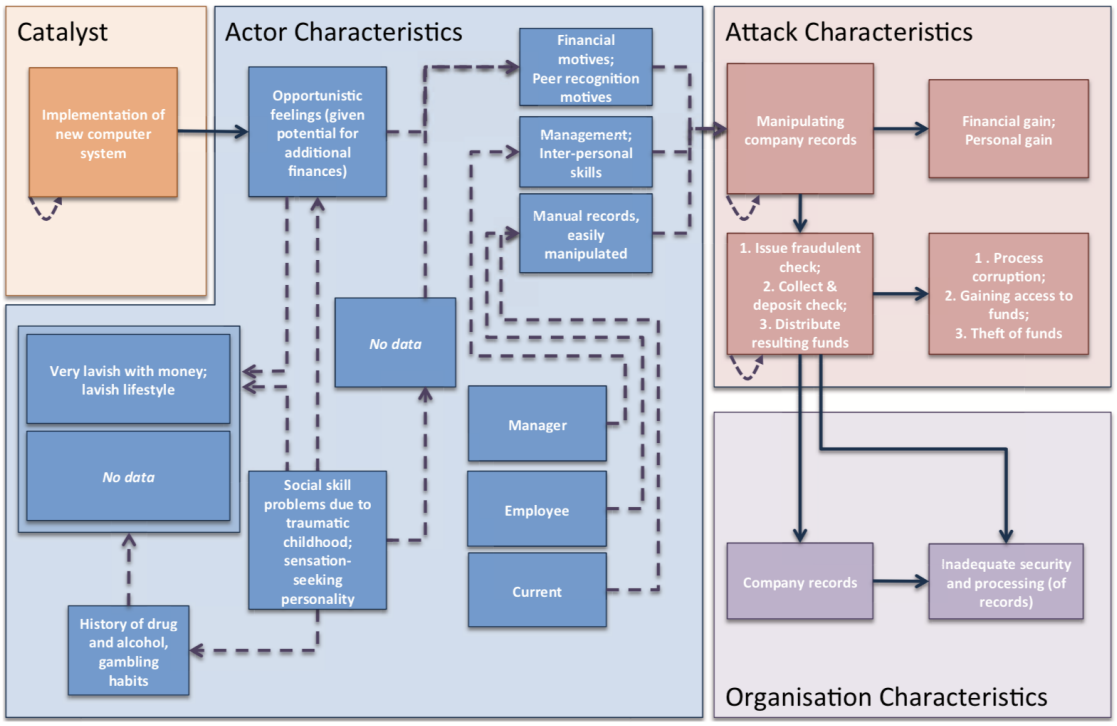


Figure 5.10: Insider threat framework applied to a known case by Nurse et al. (2014b)

As this task was ongoing with the Mechanical Turk expansion of the prospective experiment, the initial run of 41 documents was segmented for this experiment. These documents were written by a mix of individuals from different backgrounds and therefore differed in specific details, level of overall detail, length and writing style. This produced a total of 270 sentences.

The participants ranged in backgrounds; however, all were familiar with insider threat, whether an expert or casual familiarity level. If a participant was unsure, they were encouraged to label a sentence in the closest code and write any thoughts in the notes section. Although this was not used during the final experiment, this feedback was considered when piloting. Each was given a choice in the form of a drop-down of the general char-

acteristic from Nurse et al. (2014b), e.g. organisational, attack, actor, and the specific, e.g. attacks step, asset, historical behaviour. The drop-down then gave an example of what may constitute this, e.g. 'Actor Characteristics - Motivation to Attack' would give a definition of 'Financial motives; peer recognition.

5.4.2 The Dataset

These six participants' answers were then merged together to create a single agreed code or, in some cases, two agreed codes. This was done by choosing any code where a majority of participants agreed; if there was a tie, both would be chosen. For example, if the participants were split 3/6 and 3/6 or 2/6, both codes of 3/6 would be chosen. For those that were more ambiguous where there was no clear majority, the rules shown in Table 5.3 were applied.

Table 5.3: Rules used to code sentences

Average code	Clear majority (3+) or tied with the second code (3/3 2/2)
Second Average Code	Second majority (2+) or tied with first code (3/3 or 2/2)
Ambigous	When there is no clear majority (2/6)
No Agreement Met	All respondents had different responses

Table 5.4: Number of sentences per rule

Average code	198
Second Average Code	29
Ambigous	35
No Agreement Met	6

In the majority of cases, the modal topic was assigned the first code, with the second modal topic the second, the full results are available in Table 5.4. In some cases, the sentences were highly ambiguous, and no code could be chosen. However, when reading the individual comments, it was clear that a few codes were being used interchangeably or when the meaning of the model was not clear on the differences between them. This may have been because the participants were not necessarily experts in insider threat, however many of these characteristics describe the same element from another actors

point of view, for example, the opportunity the insider had and the vulnerability that they exploited. These codes were combined in the hopes of reaching a majority. This was done twice and can be seen in Table 5.11. In addition, some characteristics were missing from the model due to the nature of the model as it was not trained on this specific insider threat case or for this specific purpose, specifically the organisational and the actor's outcome after the attack. An EDA was completed to ensure that there were no outliers; this found no issues with the results.

Actor Characteristics - Enterprise Role	Actor Characteristics - Job Info
Actor Characteristics - Type of actor	
Actor Characteristics - State of relationship	
Organisation Characteristics – Vulnerability	Actor-Organisation - Vulnerability/Opportunity
Actor Characteristics – Opportunity	

Figure 5.11: Merged characteristics

This produces a result shown in Table 5.12. Each document is identified using a unique ID to ensure that individuals remain anonymous but that their data can be removed when requested. Each document is then broken into sentences; these are ordered so for future experiments, the sentence order is not lost. Finally, the two codes are chosen by the respondents and the rules, and a note of the overall uncertainty in the coding.

Document	Sentence Number	Response	Average Code	Second Average Code
5c6d17f2c581e	1	Tax office manager was writing fraudulent cheque's, 200 of them to be exact.	Attack Charateristics - Attack Step	
	2	She had done so over 18 years.	Actor Characteristics - Historical behaviour	Attack Charateristics - Attack
	3	Her department were made except from the new IT system that did auditing and accounting.	Catayst - Precipitating Event	
	4	She was ordered to pay back $48 million in restitution, $12 million in federal taxes and $3.2 million in state taxes	Outcome - Actor	
5c6d2abb29bf9	1	A manager in a company was seen as generous and helpful, however she insisted that her department should not use a new computerised system for managing financial transactions because it was deemed as too complicated.	Catayst - Precipitating Event	
	2	She was well liked and generous with her staff, although some of them thought she was a little bit odd and had some history of alcohol abuse.	Actor Characteristics - Personality characteristics	Actor Characteristics - Historical behaviour
	3	When a bank teller raised a concern about a cheque the manager tried to cash, it alerted the company to further discrepancies, and an internal investigation discovered that the manager in question had been defrauding the company for the past 18 years, to the tune of millions of dollars.	Attack Charateristics - Attack	Outcome - Organisation
	4	Staff did notice that the manager always seemed to have money, but generally put it down to her having an inheritance of some kind.	Actor Characteristics - Observed physical behaviour	
	5	The IT department had not noticed the suspicious activity due to the fact that the manager had managed to persuade her superiors that the computerised system was too complicated for her team to use (even though she helped to design it), and as such there was only a paper based record that could easily be manipulated.	Organisation Characteristics - Vunerability	
	6	The manager was found guilty and ordered to pay back 60 million dollars, and as a result all of the departments within the company now use the computerised system without exception.	Outcome - Organisation	
	7	Nobody suspected that such a well liked member of staff could be guilty of such a huge amount of fraud.	Actor Characteristics - Personality characteristics	

Figure 5.12: Sample Coded Data

This experiment has therefore created a labelled subset of the perspectives experiment data. This dataset contains the opinion of six participants who labelled a total of 41 documents according to an existing insider threat model by Nurse et al. (2014b) From this, up to two codes were chosen to represent the viewpoints of the participants, creating a 'human topic model'. This process merged some characteristics or codes from the Nurse et al. (2014b) framework and added additional ones. Specifically, Actor Characteristics - Enterprise Role, Actor Characteristics - Type of actor, Actor Characteristics - State of relationship were merged into Actor Characteristics - Job Info and Organisation Characteristics – Vulnerability and Actor Characteristics – Opportunity were merged into Actor-Organisation - Vulnerability/Opportunity. Despite this, some sentences will have a degree of ambiguity. These were noted in the final dataset so that future experiments could weight this ambiguity in the future. This creates a final dataset with each sentence in the document given a code or characteristic, similar to a machine created topic, and a note on the overall ambiguity.

5.5 Conclusion

This chapter has discussed the creation of the three datasets that will be used in the course of this research. These three datasets vary in use and overall aim, but each will allow this work to deliver on each objective. These three are a corpus of general insider threat, organic narrative reports written about an insider threat incident, and a portion of the narrative reports which were grounded to an existing insider threat model. These will be used throughout the research and be used for NLP techniques such as topic modelling.

The general insider threat corpus is created from publicly accessible news articles of insider threat incidents. This gives a general corpus of many different types of insider threat incidents, which are written in the same style. Originally this corpus was created using known insider threat aggregates such as data breaches and Infowatch, which share articles of insider threat attacks from across the internet, both major attacks and smaller,

local attacks. This corpus was then expanded by downloading potential insider threat articles, using insider threat keywords, creating a topic model and comparing the topic proportions to the original, known insider threat corpus. This creates a varied corpus of insider threat news articles covering many different attacks in many different industries, not limited by insider threat archetype, methodology, motivation or other case characteristics. This corpus is not varied by language use, and the similar language allows for NLP techniques to more easily normalise the text for future use. One such use will be the attack language objective; during this objective, the corpus is used to create topic models. These topic models represent features that any attack may have and can be thought of as the characteristics that make up an insider threat attack.

The organic narratives corpus was created from the results of the perspectives experiment, during this experiment, volunteers were asked to listen to three individuals tell the story of an insider threat attack from their perspective and then asked to retell the story in their own words. Participants were asked to retell the story however they liked, with formal or informal language, bullet points or paragraphs and as much or as little information as they liked or remembered. This was completed by a range of individuals, including Amazon Mechanical Turk members, although this data was anonymised so that demographical information was not stored. Thus creating a final corpus of 107 documents, these documents indeed ranged in writing style, level of detail and language use, while the general insider threat corpus aimed to find different insider threat attacks, but with a similar style, this corpus aims to show the same insider threat attack with different writing styles. This corpus will be used as a sample corpus and therefore play a crucial role in invalidation and development activities.

However, as discussed previously in the methodology, with a nuanced issue like insider threat, it is important not just to consider this a solely technical problem that can be solved with NLP. Instead, the problem of insider threat is rooted in sociology, and therefore it is important that any NLP model is also rooted in the social domain. The organic narrative corpus is grounded to an existing insider threat model, ensuring that the model

exists without the social domain. This model, created by Nurse et al. (2014b) was created with no specific insider threat archetype and was created using grounded theory, a similar technique to topic modelling, making it ideal to ground the organic narrative corpus. A section of 41 documents was chosen from the initial batch of the perspectives experiment, broken down into sentences and assigned a label. This was done by asking 6 participants with varying levels of expertise in insider threat to consider each sentence and label each with the characteristic which seemed to fit the sentence best. These 6 participants responses were then merged and the modal code or codes assigned to each sentence. This creates a 'human' topic model, using a similar data structure to topic modelling, but instead of using machine-generated topics and an algorithm to assign them to the text, it was done by a collection of people. This corpus can then be used for validation, ensuring that the machine-generated model is comparable to existing social models.

Each of these corpora is different and is used for different tasks within the research. Every effort has been made to ensure data gathering is limited to only necessary data; therefore, these corpora are refused for different tasks, within reason. However, this process is open, and the pipelines that have been developed could be used to create new datasets. For example, the general insider threat corpus contains many different types of insider threat activity; however, these are not relevant to all organisations, and to create a more bespoke model, an organisation may want to tailor the insider threat model to cases they are more likely to experience so that additional documents can be added, or a new corpus created.

As all the data has now been gathered, the first objective of this research has been completed. This data can then be used for all future objectives, with the first of these being the Attack Language objective. This objective will use topic modelling to create a custom insider threat model by analysing the datasets that have been created. Using the general insider threat corpus to create the initial topic models and then to apply these to the corpus of organic narratives. This then allows the organic narratives to be organised by document, with a single narrative having a collection of topics for each sentence or

be organised by topic, with each topic having a collection of sentences from different documents. Therefore creating the initial stage of an insider threat model, where the key characteristics of an insider threat attack are identified.

Chapter 6

Attack Language

Topic models form the core of this work and allow the text to be segmented by topic, a process called Topic Segmentation (Riedl and Biemann 2012). This technique uses the LDA, Latent Dirichlet Allocation algorithm to statistically find related words and place these in separate topics (Blei, Ng and Jordan 2003). In this section, first, the creation of topic models is discussed, then the model selection and tuning process. Finally, it can map the topics onto an existing insider threat model. The final output of this project stage is to automatically find related sentences, place them in the same topic, and then label this topic with an insider threat characteristic from a known insider threat model. This output allows the corpus to be organised by document or by topic, finding related pieces of text across the documents.

6.1 Introduction

The goal of the attack language stage is to re-organise the corpora not by document but by topic. Finding sentences across different documents that refer to the same topic allows the sentences to be not just organised by document, in order, but by meaning. A topic model was trained on the general corpus of insider threat discussed previously and then applied to the perspective experiment corpus, which then uses the labelled data to refine and label these topics. This chapter aims to demonstrate that a topic model can be created that

can capture a range of features of an insider threat attack. Topic modelling, as discussed previously, is remarkably similar to grounded theory (Baumer et al. 2017); however, in grounded theory, the 'codes' are created and analysed/shortlisted by humans. In topic modelling, a statistical algorithm called LDA (Latent Dirichlet allocation) is used. This statistical algorithm places each word in one topic by examining the words surrounding it. The final output is a model that can be applied to a piece of text and then describe the text in terms of a model. During this stage, all three corpora will be used, the general corpus of insider threat will be used to train the topic model, the labelled perspectives experiment will be used for tuning the parameters, and finally, the final topic model will be applied to the perspectives experiment in preparation for the next stages.

6.2 Methodology

The full methodology for this technical objective is shown visually in Figure 6.1. To accomplish this, a three-step process is employed: however, before this can begin first, the datasets must be collected, this technical objective uses two datasets, first the collection of news articles reporting cases of insider threat and the second the individual organic narratives of a single case of insider threat. These two are used in different stages; the corpus of news articles is used to train the topic model, while the individual narratives are analysed using the topic model and a known insider threat model. These can then be used to tune the model further. The general insider threat news articles were created using a custom web archiving tool combined with an automatic classification tool, while the organic narratives were created with the perspectives experiment. These are the two core corpora used during this technical objective; however, this methodology is not limited to insider threat, and any organic narratives can be understood using the methodology developed.

This system was developed primarily in R using the 'tm' package. This could also be done using another language, and implementation, R was chosen due to familiarity.

The tools created were scripted, automating the process of training and applying models. A simple web interface was created to analyse and explore the output. Although this particular combination of tooling is not required, this methodology can be applied to other programming languages such as Python. Many of the tools were developed with APIs, so the tools developed can be implemented in another interface.

The first step in this technical objective involves labelling part of the organic narratives; although this was discussed in the data gathering chapter, this forms a crucial step in developing the Attack Language models. Next, an existing insider threat model is chosen, and several participants are asked to label each sentence according to the corresponding insider threat characteristic from the model. These are then merged, and the most common one or two topics were chosen, creating a similar output to a topic model. This experiment replicated the process of grounded theory, and therefore creating this similar output was necessary for the results of this experiment mirror the results of the topic model, creating a 'human topic model'. Creating this 'human topic model' is particularly important as topic modelling requires the number of topics to be selected, which provides an evaluation mechanism.

The following assumption is made to evaluate the topic model: that sentences that appear in the same insider threat characteristic or code for the 'human topic model' will also appear in the topic model's same topic. That is to say that a human has marked these sentences as being related by insider threat characteristic, so any topic model that can evaluate based on insider threat characteristic should also place these sentences in the same topic. The second assumption made is that this is not necessarily a 1:1 ratio between topics and grounded theory codes and that a machine topic model may be more specific and therefore require more topics to achieve the same effect, which is noted by the literature (Baumer et al. 2017).

Stage 2 involves creating the topic model; the topic model is trained on the general corpus of insider threat. The general corpus was created using news articles. This is important, as the similar language style allows some aspects of the articles to be normalised.

The normalisation process reduces the number of words while ensuring that the remaining words are related to insider threat. The normalisation process removed punctuation and stopwords and stemmed the text, collapsing features such as tense into a single word. This process removes two sets of stopwords, a general English list and a more specialised news list; this is a standard process that ensures that the remaining text is domain-relevant. The stemming of text reduces the overall number of words, merging different contexts of the same word. This process optimised the final topic model by using a grid search and scoring mechanism, the grid search methodology is widely used for the tuning of hyper-parameters of NLP models and machine learning more broadly, such as in Ghawi and Pfeffer (2019). The highest scoring model is then chosen as the final topic model in step three.

Step 3 describes the mapping process; this is important as grounded the topic model to the existing insider threat framework ensures that the model meets the overall aim of the research and this technical objective. First, the full, unlabelled corpus of organic narratives is broken down into documents, and each document is broken into many sentences. Then, the posterior probabilities of the model are applied to this corpus of sentences. Although in the data collection phase, the proportion of topics was used to aid in classification, the topic with the highest beta was chosen. This represents the best 'match' between a word or sentence to a topic, using the topic model more discriminatively. If desired, the corpus could be classified by paragraph, instead of sentences within a document, and in the literature, this is common, e.g. Riedl and Biemann (2012). Finally, sentences were chosen as these can be better represented as Markov chains and more easily merged. The result of this is a list of sentences and a topic number representing each sentence's topic. However, the topic number does not provide additional context and is arbitrary; therefore, the labelled sentences are also used to map these topics to an insider threat model characteristic. This process is simple; for each topic with sentences that were manually categorised, this topic is assigned the code that the majority of sentences are labelled with by the system.

The next section will detail how the final topic model was chosen. As topic modelling requires various parameters to be selected before a model can be created, and specifically, a topic number must be chosen, a metric for evaluating the models was developed, based on Lund et al. (2019). Creating an evaluation method was challenging, as parameters may interact with each other, as no single parameter can be considered the independent variable. To address these issues, a grid search method was employed. Many models are created during a grid search, which are then scored and investigated, examining those that produce the highest score. This process allows for the discovery of the most impactful parameters, and these parameters will then be selected for in the final model.

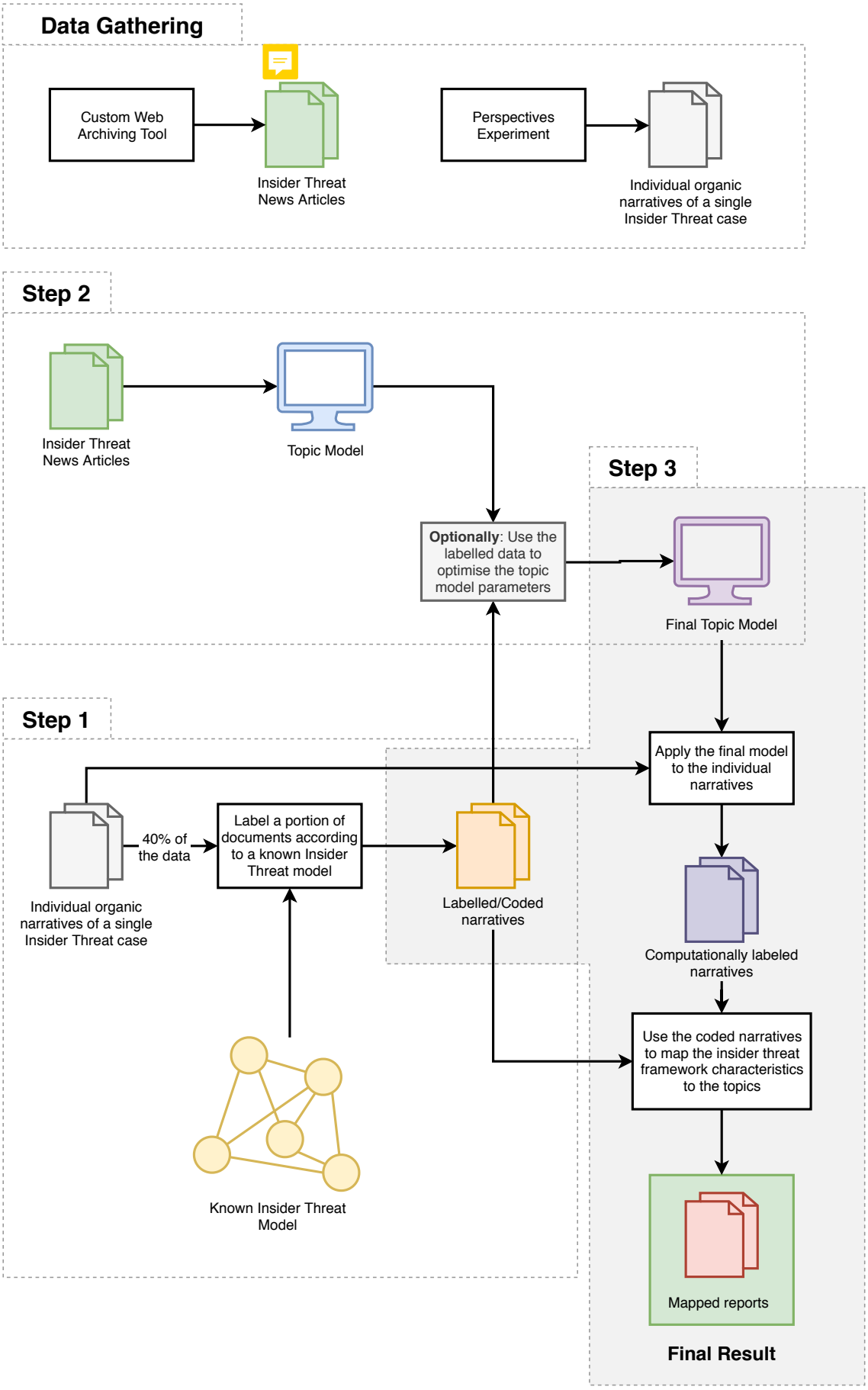


Figure 6.1: Full method for the attack language

6.3 Choosing the final topic model

This experiment aims to find the best topic model, the topic model with the qualities that the event extraction task will require, specifically that correctly places the sentences that the humans have coded, without overfitting to the training data allowing it to be used on the organic narratives. Therefore, first potential topic models must be generated, adjusting the parameters used to create them. These can then be scored against the Human Topic Model experiment, finding the most similar model to human performance. This process encapsulates several experiments, first reducing the total dimensions and finally selecting the best model. This section will discuss this process, first the scoring, next, the generation of models and dimensionality reduction, the experiments examining the final dimensions, choosing the final model and finally applying this model to the data.

The scoring system uses a simple algorithm and reward structure based on work by Lund et al. (2019). We expect a topic model that performs well would place similar sentences into the same topic. We consider the results of the human topic model experiment as our ground truth. Therefore, if a computer model places sentences in the same code into the same topic, this is considered a success. The model will score +1 per matching sentence. During the human-coding process, we identified that some codes covered multiple interpretations; therefore, the codes may be more general. However, the computer model may classify these more general codes into separate topics due to the mechanism of topic modelling. For example, the attack code was often used for any sentence, which generally explains the attack. However, a topic model would likely distinguish sentences with a high proportion of technical words from those with a high proportion of summary words, even if these would appear in the same code. To account for this, if a topic model correctly encapsulates a subset of the code, it is rewarded with +1 to the score for each sentence. If the topic model cannot correctly identify any matching sentences, this model is punished with a -1 score per sentence. If a model has some matching sentences and others that do not match, it is merely not rewarded.

A worked example is shown below in Figure 6.2. In summary, topic models which

rank highly on this scoring mechanism will demonstrate the following characteristics: They will be close to human performance with similar sentences appearing in the same topic. However, sometimes these may appear in subsets of other topics. Conversely, a topic model which does not rank highly or will have a negative score will primarily have a few related sentences that appear together.

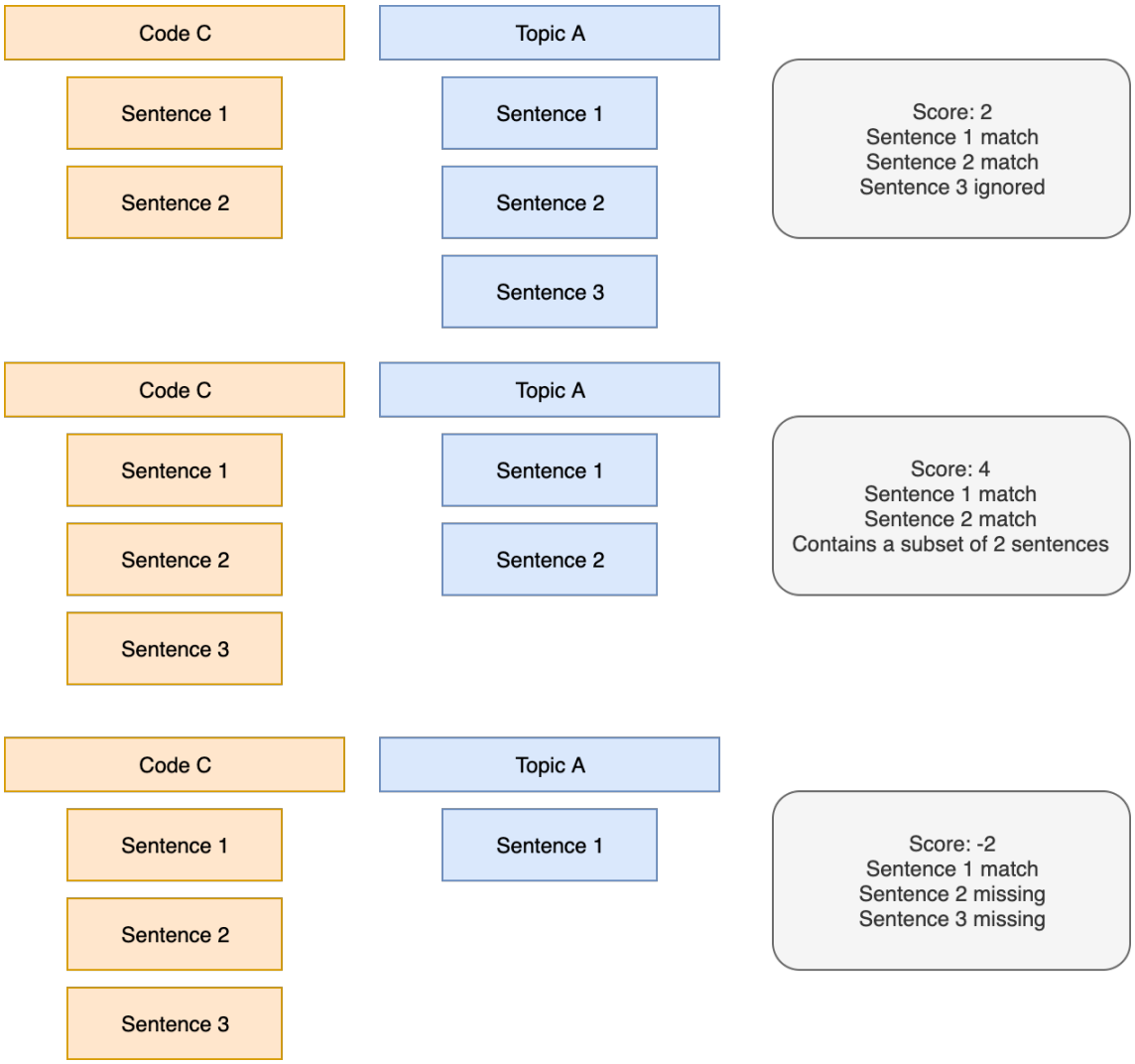


Figure 6.2: Worked scoring example

A grid search approach was taken to choose the model parameters, as the relationship between parameters and final score may not be clear, and a brute force approach would not be appropriate given the time constraints. For the model's generation, the following assumption is taken: That not one of the parameters can be considered an independent variable. Instead, each parameter's exact values, including those that have not been

Table 6.1: Grid search values used

Experiment	Model	Value
Data	dimensionality_reduction_data_1	Data=100%
	dimensionality_reduction_data_2	Data=50%
	dimensionality_reduction_data_3	Data=75%
K	dimensionality_reduction_k_1	K=25
	dimensionality_reduction_k_2	K=50
	dimensionality_reduction_k_3	K=400
	dimensionality_reduction_k_4	K=450
	dimensionality_reduction_k_5	K=500
	dimensionality_reduction_k_part_2_2	K=550
	dimensionality_reduction_k_part_2_3	K=600
Method	dimensionality_reduction_method_1	Method=Gibbs
	dimensionality_reduction_method_2	Method=VEM
Stopwords	dimensionality_reduction_stopwords_1	Stopwords=en+news10k
	dimensionality_reduction_stopwords_2	Stopwords=en+news1k
	dimensionality_reduction_stopwords_3	Stopwords=en+news100k
	dimensionality_reduction_stopwords_4	Stopwords=news1k
	dimensionality_reduction_stopwords_5	Stopwords=news10k
	dimensionality_reduction_stopwords_6	Stopwords=en
	dimensionality_reduction_stopwords_7	Stopwords=news100k

changed, are the independent variable. Due to the large number of potential parameters and parameter values, the number of dimensions must be reduced, as the total amount of parameters would cause a combinatorial explosion. The dimensionality reduction experiment looks at four potential final parameters which can be adjusted in the topic modelling library and scores them as independent to understand how each might affect the final score; initially, each value was chosen arbitrarily to examine the effect of each. Note that this does not mean that a change in this value will always produce a higher scoring final model, just that it may and can therefore be removed if it has a small influence. For the dimensionality reduction, the parameters chosen were: how much of the data to use, what the final value of K (number of topics) will be, which sampling methods to use, and how many stopwords there should be, the full values of which are shown in Table 6.1. The stopwords lists that will be tested are English only, 1,000, 10,000 and 100,000 news specific stopword lists. The sample methods include Gibbs and VEM. The data represents a random subset of the dataset. Figure 6.3 below shows the results of this experiment.

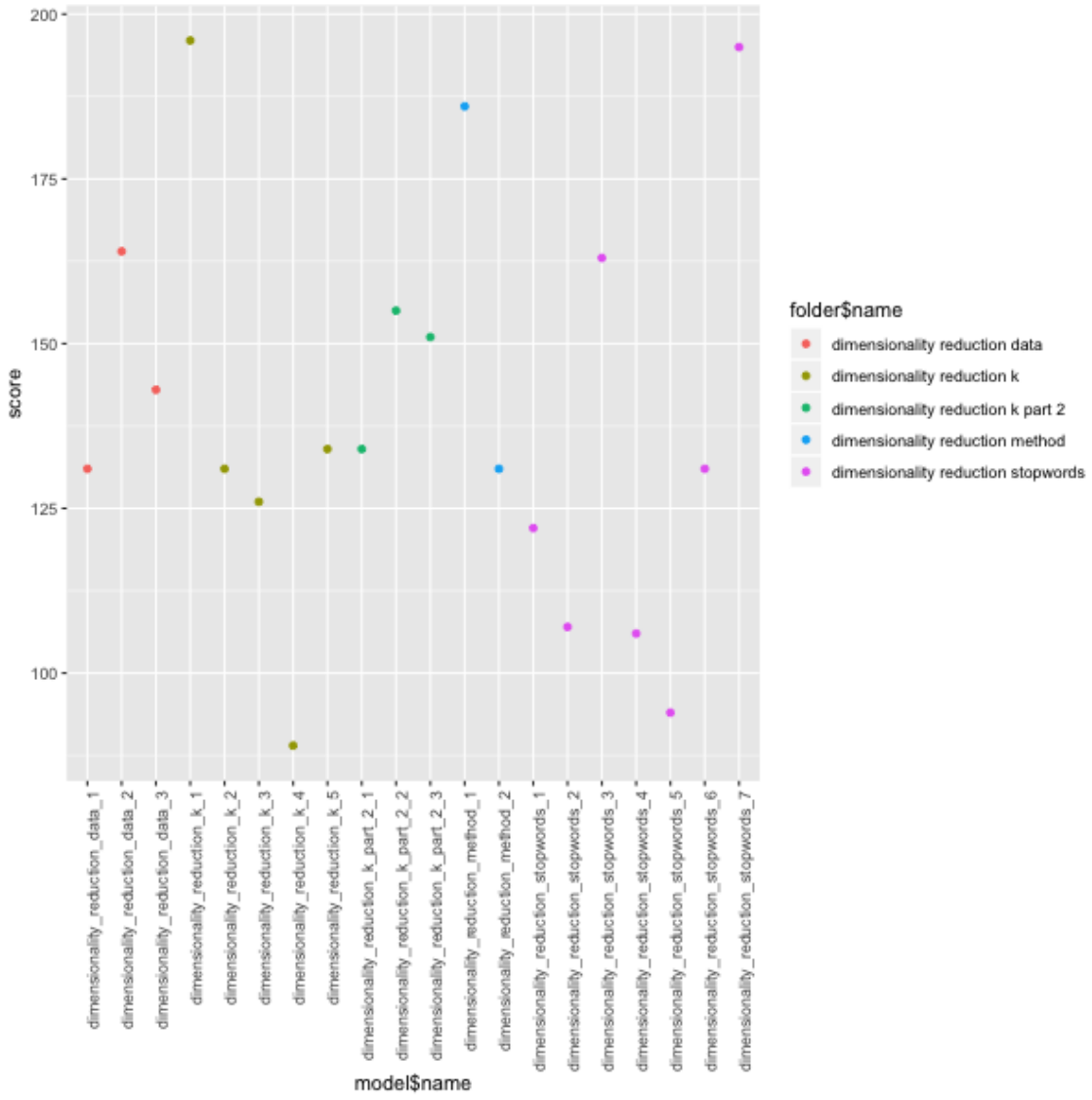


Figure 6.3: Inital grid search results

From the results of this experiment, various important features were noted. First, after the initial experimentation, it was concluded that the scoring algorithm might be greedy, giving high scores to those models with a high and a low number of topics. These scoring issues are because this scoring algorithm does not punish larger or small topics. This decision was made because of the assumption that any characteristic can be represented, potentially, as multiple topics, similar to hierarchical topic modelling (Jacobi, Atteveldt and Welbers 2016).

First, it was likely that the scoring algorithm was greedy when dealing with a small

amount of data. Therefore, the amount of data should not be considered in the final experiment. Next, it was shown that the value of K dramatically improves the performance, and therefore it would be essential to keep K as a final parameter and potentially explore a large number of K values in a future experiment. The sampling method was shown to change the score dramatically, which may indicate better performance; however, it is not clear why; therefore, this was also kept. Finally, the number of stopwords was also a great indicator, with a large number of stopwords and a small number of stopwords outperforming a medium amount of stopwords. After this experiment, it was then decided that a further experiment should be done to investigate the link between stopwords and K. We theorise that with a small number of stopwords, a larger number of topics must be used as some topics will encapsulate new specific stop words, with a large number of stopwords a lower K value can be used to achieve the same result. In addition, the scoring algorithm was shown to be greedy, when a model had a low K value, the models performed better on the scoring mechanism. This is expected as the model is not punished for having a small number of very large topics. For example, if the model were to correctly place three related sentences in one topic, this would award the model a score of three. However, if this same topic also placed three additional sentences, even if they belonged to a different characteristic, the scoring algorithm would award a total score of six, despite the fact that these two groups of sentences may not be related. Finally, we would further expect that the model would score better with a larger number of topics as the models are rewarded for subsets of existing topics. Therefore it is essential to consider these bounds when selecting the final topic; these bounds can be directly impacted by either a low number of stopwords or a high number of stopwords.

Figure 6.4 stops a sample distribution demonstrating the relationship between scoring and the value of K. The red series shows the effect of the scoring; it is expected that with a small number of topics or a large number of topics, the scoring will be higher because of the greediness of the algorithm. The blue and green series both show the actual scores but with differing numbers of stopwords (English only and English with 100,000 news

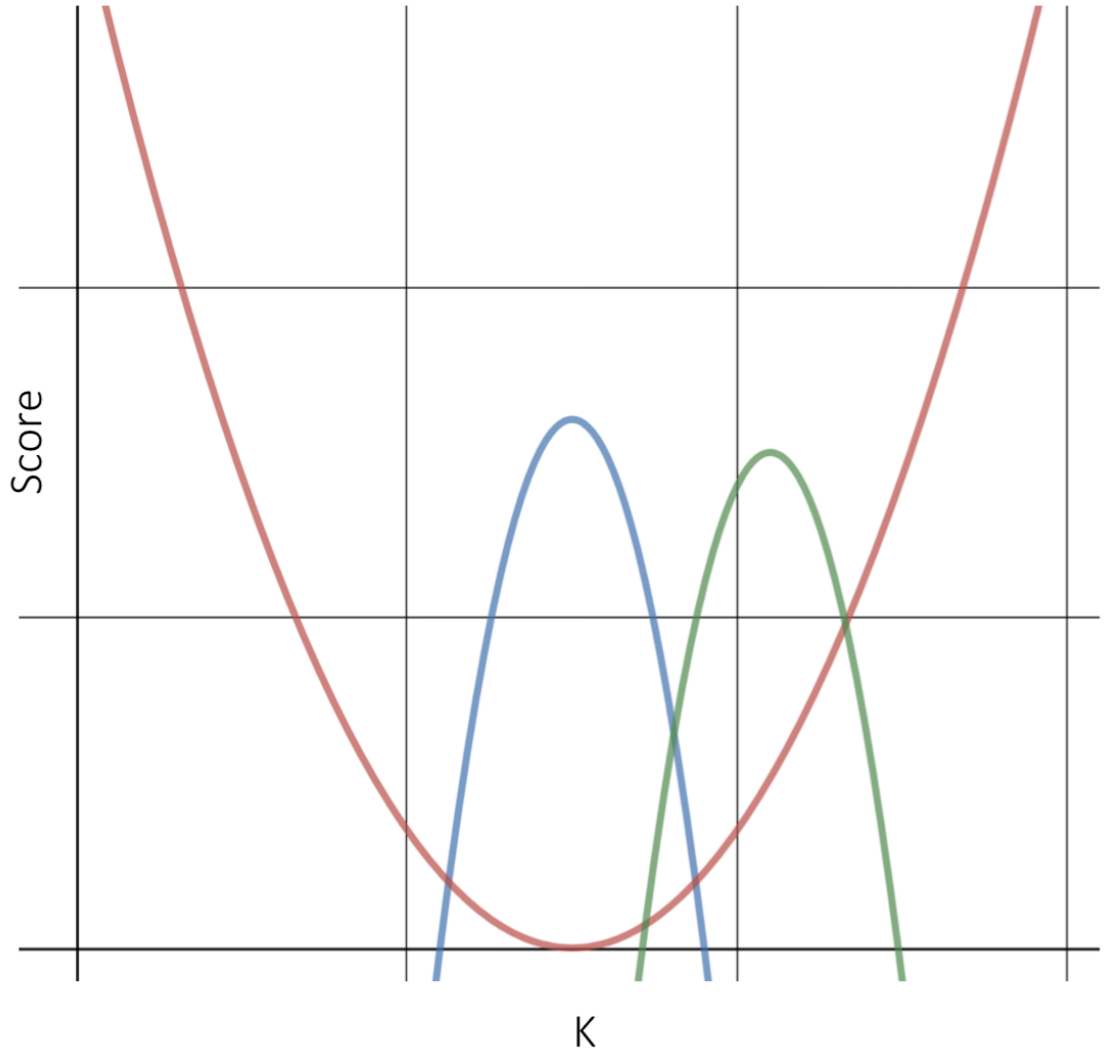


Figure 6.4: Expected distribution of scoring

specific stopwords). It is expected that the fewer stopwords that are used, the higher K will need to be to see the peak of the score. This is because some news specific stopwords will be captured in another topic, so for the insider threat topics to show without the words associated with news interfering. Therefore, the best performing stopword list will be whichever peak, blue or green, is highest overall, without considering the score's effect at a high or low number of topics. This peak can then be examined to find the appropriate K value for the final model.

It was essential to investigate this link thoroughly to ensure the model was correctly tuned. Although this is not necessary for the topic modelling process, by investigating this

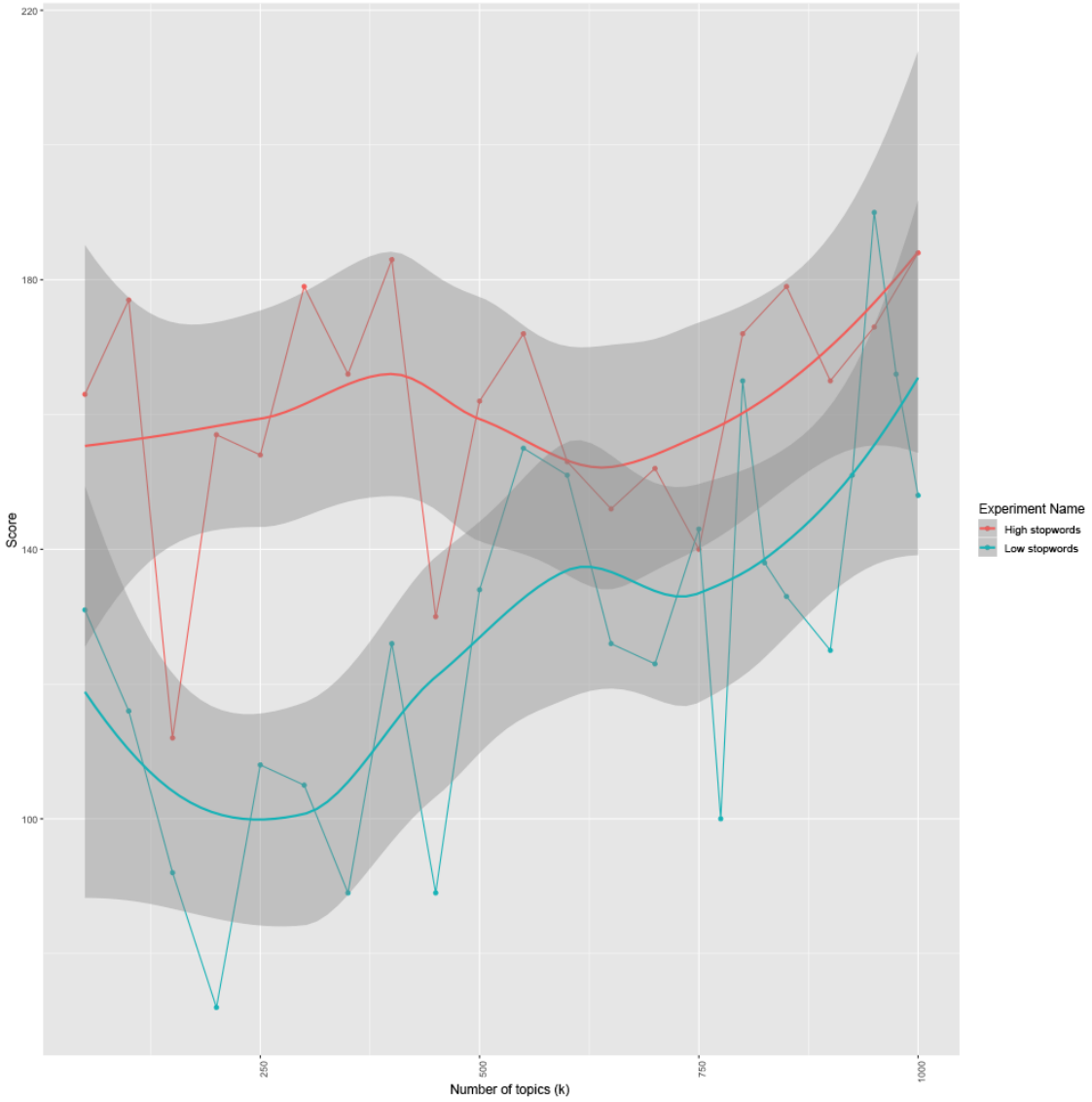


Figure 6.5: Final results of the experiment

link, issues such as overfitting of the model to the training data can be found and avoided. Therefore an experiment was devised to examine a range of K between 50 and 1000 topics in 50 topic increments, the number of stopwords English-only and English and 100,000 news specific stopwords. We would expect both the models with a smaller K value and a high K value to score highly on the scoring mechanism. Moreover, each would have a peak between these two values. Figure 6.5 shows the results. The peak appears sooner on the red series (the higher number of stopwords), approximately at 350 topics, and later in the blue series (the smaller number of stopwords) and approximately 650 topics.

However, there appears to be an outlier at 800 topics; this can be caused by random chance. Therefore, another experiment was run to ensure that this is an outlier rather than a more suitable model. It is also essential to discuss that it takes an increasingly longer time to train models with a small number of stopwords. This process should be quick, allowing users to adapt models. For this tool to be operationalised effectively, we believe it is essential to maximise the ease of use and therefore prefer models that train quickly and have a high score. The best performing models (shown in Figure 6.5) had a high number of stop words; this appears at the peak of the red series; the next experiment can expand the range of the peak. The next experiment can now introduce the method from the previous experiment and compare the results of using the VEM and Gibbs sampling methods.

From the results of this experiment shown in figure 6.6, the best sampling method is the VEM method (red series), with Gibbs scoring significantly lower (blue series). The difference in these results is likely because the Gibbs method scores higher with a smaller number of topics, as the dimensionality reduction tested only 50 topics, demonstrating that each parameter does not represent an independent variable. Therefore, the highest score within the peak can be chosen at k=370 topics. This model will become the final topic model and can be applied to the perspectives dataset. Although this model has been tuned using the perspectives dataset, it was not directly trained on this data. Instead, the model was trained on the general insider threat corpus. Therefore, this can likely apply to other types of insider threat attacks; however, this was beyond the scope of this work, and if it does exhibit overfitting, a similar tuning process can be employed again to reduce this. However, as this has not been more thoroughly investigated, the results of this particular tuned model can be considered as only representing insider fraud.

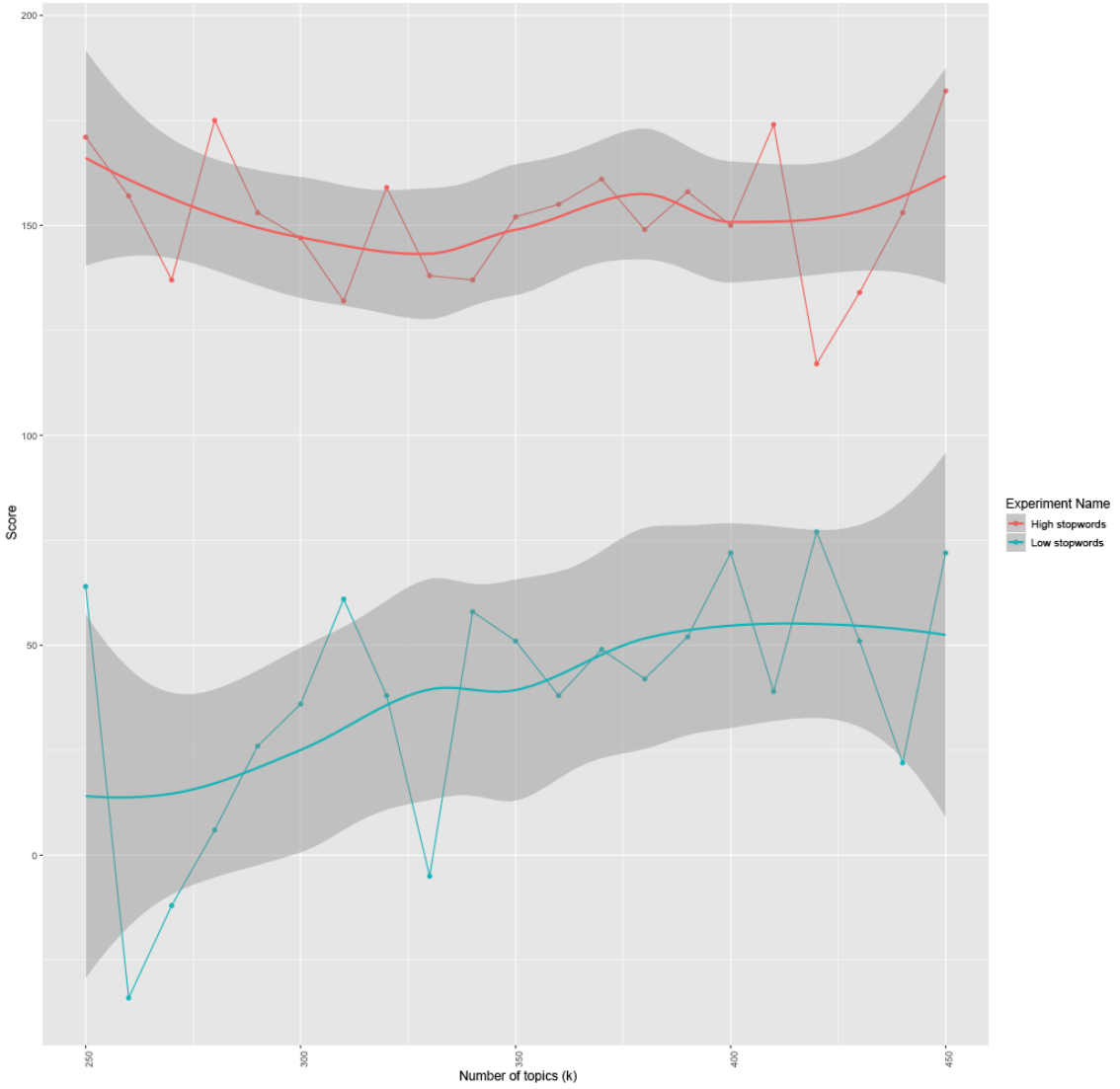
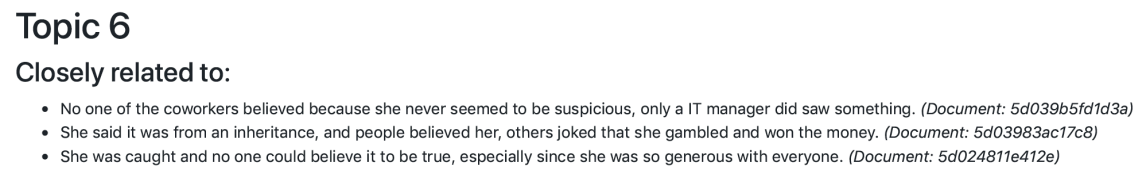


Figure 6.6: Expanding the peak of the models

6.4 Results

This section will demonstrate the topic model, as there are 370 total topics, it is not feasible to include each. However, this section will show some selected results. The model will be applied to the full corpus from the perspectives experiment. Although a subset of this data was used to score the models, the final model has only been trained on the general corpus of insider threat news articles. Therefore, the topics generated do not rely on data from the perspectives experiment, and additional documents which were not labelled have also been included. Each topic is labelled, and each sentence categorised

into that topic is shown. A sample of these are shown in Figures 6.7—6.13.

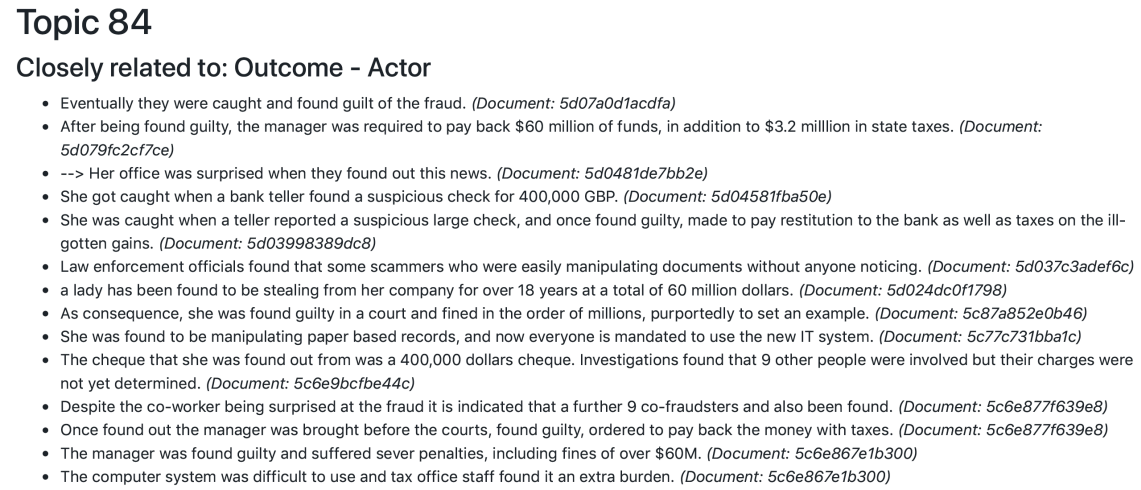


Topic 6

Closely related to:

- No one of the coworkers believed because she never seemed to be suspicious, only a IT manager did saw something. *(Document: 5d039b5fd1d3a)*
- She said it was from an inheritance, and people believed her, others joked that she gambled and won the money. *(Document: 5d03983ac17c8)*
- She was caught and no one could believe it to be true, especially since she was so generous with everyone. *(Document: 5d024811e412e)*

Figure 6.7: Topic 6

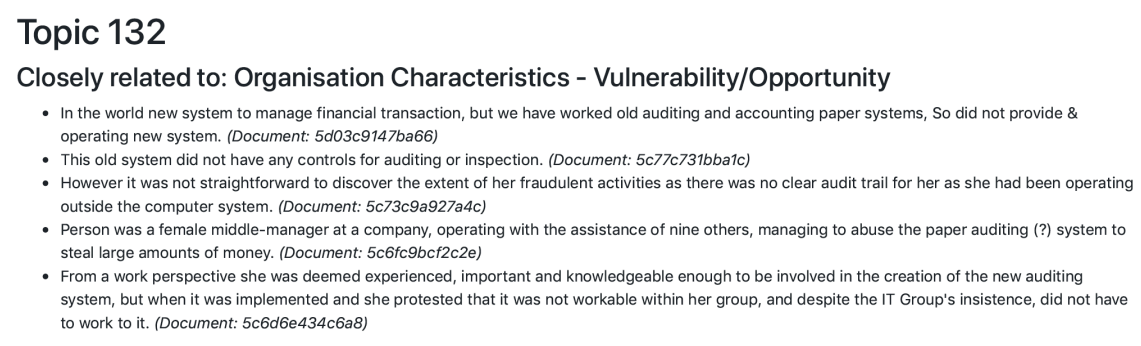


Topic 84

Closely related to: Outcome - Actor

- Eventually they were caught and found guilt of the fraud. *(Document: 5d07a0d1acdfa)*
- After being found guilty, the manager was required to pay back $60 million of funds, in addition to $3.2 milllion in state taxes. *(Document: 5d079fc2cf7ce)*
- --> Her office was surprised when they found out this news. *(Document: 5d0481de7bb2e)*
- She got caught when a bank teller found a suspicious check for 400,000 GBP. *(Document: 5d04581fba50e)*
- She was caught when a teller reported a suspicious large check, and once found guilty, made to pay restitution to the bank as well as taxes on the ill-gotten gains. *(Document: 5d03998389dc8)*
- Law enforcement officials found that some scammers who were easily manipulating documents without anyone noticing. *(Document: 5d037c3adef6c)*
- a lady has been found to be stealing from her company for over 18 years at a total of 60 million dollars. *(Document: 5d024dc0f1798)*
- As consequence, she was found guilty in a court and fined in the order of millions, purportedly to set an example. *(Document: 5c87a852e0b46)*
- She was found to be manipulating paper based records, and now everyone is mandated to use the new IT system. *(Document: 5c77c731bba1c)*
- The cheque that she was found out from was a 400,000 dollars cheque. Investigations found that 9 other people were involved but their charges were not yet determined. *(Document: 5c6e9bcfbe44c)*
- Despite the co-worker being surprised at the fraud it is indicated that a further 9 co-fraudsters and also been found. *(Document: 5c6e877f639e8)*
- Once found out the manager was brought before the courts, found guilty, ordered to pay back the money with taxes. *(Document: 5c6e877f639e8)*
- The manager was found guilty and suffered sever penalties, including fines of over $60M. *(Document: 5c6e867e1b300)*
- The computer system was difficult to use and tax office staff found it an extra burden. *(Document: 5c6e867e1b300)*

Figure 6.8: Topic 84



Topic 132

Closely related to: Organisation Characteristics - Vulnerability/Opportunity

- In the world new system to manage financial transaction, but we have worked old auditing and accounting paper systems, So did not provide & operating new system. *(Document: 5d03c9147ba66)*
- This old system did not have any controls for auditing or inspection. *(Document: 5c77c731bba1c)*
- However it was not straightforward to discover the extent of her fraudulent activities as there was no clear audit trail for her as she had been operating outside the computer system. *(Document: 5c73c9a927a4c)*
- Person was a female middle-manager at a company, operating with the assistance of nine others, managing to abuse the paper auditing (?) system to steal large amounts of money. *(Document: 5c6fc9bcf2c2e)*
- From a work perspective she was deemed experienced, important and knowledgeable enough to be involved in the creation of the new auditing system, but when it was implemented and she protested that it was not workable within her group, and despite the IT Group's insistence, did not have to work to it. *(Document: 5c6d6e434c6a8)*

Figure 6.9: Topic 132

Each topic model is shown mapped to the existing insider threat characteristic or, in the case of the unlabelled data, the most representative words in the topic model. The mapping process examines the sentences within the topic and how the participants labelled these during the ‘human topic model’ experiments, this gives each topic additional context and aids in understanding. The aim of this work focused on supporting decision

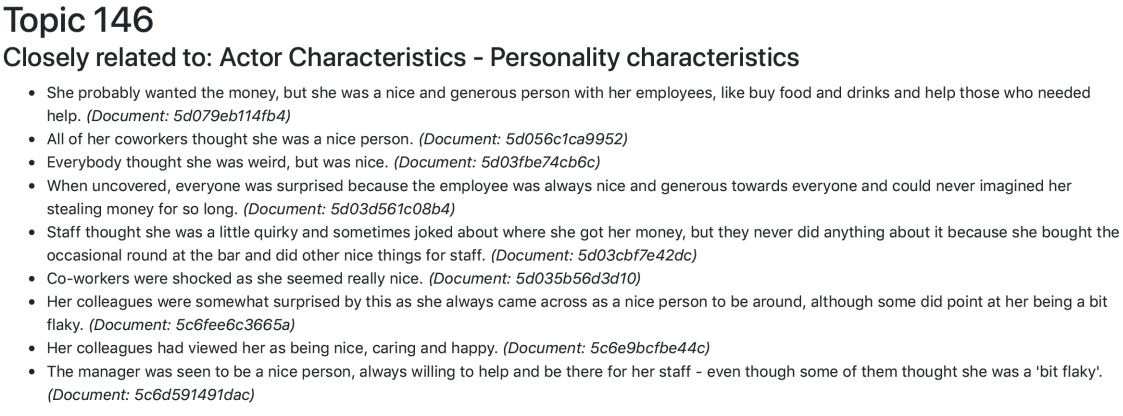


Figure 6.10: Topic 146

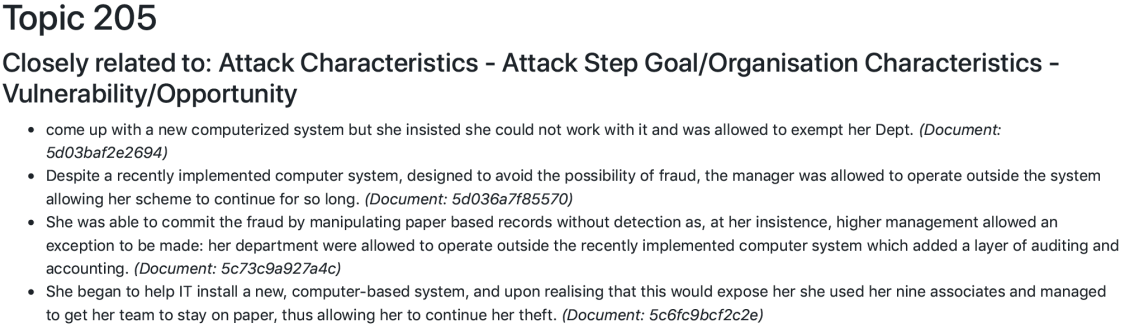


Figure 6.11: Topic 205

making. Therefore it was essential to add some additional information to the machine-generated topics. For example, each topic may have one label, with others having several or no labels. This is due to the approach used to map the labels onto the topics by examining how the participants categorised them. This process was designed to better understand the topic model and its decisions.

The topic model is much more specific than the human coders; during the perspectives experiment, the coders expressed how general they considered the characteristics. This phenomenon is expected as there are fewer characteristics in human-based models to be flexible to account for many different types of insider threat attacks. However, by comparison, topic models are specific to word use and, therefore, more sensitive than a human coder. The final model has many topics (370), and the method used to generate the topics relies on the proportion of specific words, creating much more specific topics, often revolving around a few keywords that the sentences may share. Therefore, many topics

Topic 265

Closely related to: Actor Characteristics - Personality characteristics

- The employee, a middle level manager, was said to be kind and generous. *(Document: 5d07afae00e24)*
- The news came as a surprise to many who say that she was kind and generous. *(Document: 5d07ad29c4338)*
- The manager is described as kind and generous by her colleagues. *(Document: 5d0481de7bb2e)*
- The news was a shock to the office because the manager had always behaved in a kind and generous way with everyone. *(Document: 5d037c3adef6c)*
- The news came as a shock to the office who had found the manager to be a kind and generous co-worker who frequently supported those in need. *(Document: 5d036a7f85570)*
- Reports say that the manager was a kind and friendly person and no one suspected her of being a thief, despite her oddities. *(Document: 5d02dc8e97cb6)*
- The woman spun a web of lies to hide where the money was coming from and was kind to her friends and staff, enabling her to keep her secret. *(Document: 5c7f9270c1494)*
- But she was kind and generous to her colleagues and was known to frequently support those in need. *(Document: 5c73c9a927a4c)*
- Therefore due to her kindness and generosity, the fraud seemed out of character and came as a shock to her colleagues. *(Document: 5c73c9a927a4c)*
- A new IT system was brought in which would have made it harder to manipulate the records, therefore making it harder to commit fraud, but the manager used her influence to give her department exemption from using the new system. *(Document: 5c6ecc9f2513e)*
- The manager appeared kind, generous and supportive of people in her group. *(Document: 5c6e877f639e8)*
- The actions of the manager, who was described as kind and generous by one of her colleagues, was discovered after a teller at the bank questioned a cheque she had written for over $400,000. *(Document: 5c6e80c32e77a)*
- Colleagues perceived the manager as kind and generous, but there were some jokes and rumours about where they got their money from. *(Document: 5c6d857487398)*
- A middle manager, female, known as kind and generous/nice and understanding committed financial fraud. *(Document: 5c6d530129acc)*

Figure 6.12: Topic 265

Topic 340

Closely related to: Attack Characteristics - Attack

- She was only caught because a bank teller questioned a cheque she wrote. *(Document: 5d07bded75b16)*
- There was a person who was a manager at the bank that some would describe as a nice woman who had been stealing millions from the bank overtime. *(Document: 5d03c707b6dfa)*
- She ended up getting caught when a bank teller marked a check as suspicious. *(Document: 5d039950a0ffa)*
- the bank manager was stealing checks she got away with it by telling everyone it was a family inheritance she was able to hide the transactions becasue the bank used a paper based system she was caught by a bank teller who thought a check looked suspicious the bank manager was caught and fined 60 million dollars *(Document: 5d028986a93c6)*
- A female bank manager was caught stealing money from the bank. *(Document: 5c7939498b4b6)*
- generally well-liked and popular female tax office manager had been stealing from taxes over 18 years by exploiting loopholes in paperwork systems and was very against an electronic system which she probably knew would make her theft harder to carry out and easier to detect, and was caught by a bank teller who noticed/questioned a suspicious cheque for $400,000; there were at least nine other accomplices; required to pay at least $45 million in restitution/taxes and other costs *(Document: 5c6fc0970bcac)*
- A bank teller helped catch her when they spotted a suspicious cheque for $400'000. *(Document: 5c6ecc9f2513e)*

Figure 6.13: Topic 340

have the same label, and those that revolve around the critical attack steps have many topics associated with a single code. However, the mapping approach allows these to be aggregated with each topic existing under a super-category of insider threat characteristic.

Similar sentences are successfully categorised together however, some topics may need to be joined together or outliers removed. This can easily be done during the next phases of the project either using graph-based techniques during the event specifics stage, where the sentences inside the topic will be visualised using a graph, or during the causality phase where topics are linked, if a topic has similar links, it can be considered the same ‘event’.

6.5 Discussion

This technical objective has shown that is it possible to segment organic narrative reports by sentence and then classify each sentence by topic; topics can then be mapped to an existing insider threat framework. This process creates a list of topics and then, within each topic, a list of sentences that belong to this topic. Although this technical objective is the first stage in this wider project, however, the implications of this technical objective should be discussed. Primarily the implications of the ability to map organic narratives to existing frameworks, in this work, the Nurse et al. (2014b) insider threat framework. However, this could be applied to any formal model.

The techniques developed are not domain-specific rather, the corpora used to train, score and map the models were specifically developed for insider threat. If this was to be applied to a different domain, first, a general corpus of domain-specific texts must be created. These do not have to be news articles. However, this was the easiest approach as it could be done using a bespoke web archiving tool. Next, a corpus of organic narratives to apply the topic model to would have to be gathered and a portion labelled according to an existing framework or formal model. However, these are the only changes that would have to be made; the underlying methodology is unsupervised and does not require vast amounts of training data. This could be key to increasing the use of organic narratives, particularly as they produce less bias overall and allow individuals to report what they feel comfortable with rather than pushing them to explain all characteristics of an incident.

The mapping approach allowed each topic to be labelled according to the framework. However, this is not the only approach to do this. There are several alternative approaches for topic labelling including Lau et al. (2010) and Lau et al. (2011). These aim to add additional context to the machine-generated topics, allowing them to be more easily used; however, these approaches would have been suitable due to the time constraints on this research. With this mapping approach, some outliers are present; these topics do not have a code associated with them due to the small number of sentences. While other topics such as Attack Characteristics - Attack are overrepresented, with this used to represent a

general overview of an incident.

Although the topic labelling does aid in providing context, by mapping each topic to a characteristic, relationships between topics, a key advantage of models, is not recreated. This is extremely important for insider threat as digital forensics processes often look back at an attack to determine how their response could have been improved. This is represented in an insider threat model as arrows, connecting each characteristic with each other. This will be the next stage by attempting to draw some meaning between topics and improve this context. This creates a 'custom' insider threat framework based on the existing literature, allowing models to adapt to a business without interfering with existing processes.

Other improvements that could be made if this approach is developed further, addressing alternative methods of labelling topics, the scoring algorithm and the number of documents used could also produce overall performance increases—for alternative topic labelling, using an automated approach to label topics which could not be labelled during the mapping process. This could be done using the contextual clues discussed in the next chapter, specifically the narrative chains, or potentially introducing some human verification or labelling. This could allow for an investigator to more deeply understand topics that are borderline between two characteristics or which cannot be labelled. Improving the scoring algorithm could lead to more relevant topics, therefore reducing much of the labour involved in the human verification presently the scoring algorithm more favourably scores topic models with very few or too many topics, although this did not impact the performance of the model during this experimentation, this could improve the model further in other domains. The issue of choosing a number of topics is not new and is ongoing in topic modelling recent work such as Lund et al. (2019) builds upon the work of Griffiths and Steyvers (2004), Cao et al. (2009) and Arora et al. (2012). Finally, adding additional documents in the labelling stage could also improve the mapping. However, there is a trade-off between human labour and the model performance that must be considered.

In conclusion, this technical objective has shown how a corpus of organic narratives

can be understood within the context of a formal model by segmenting the corpus by sentence and then applying a topic model trained on a corpus of general cases. This has been shown using insider threat. However, it is likely that this work could impact additional fields beyond insider threat and into any domain where there are sufficient training data and a formal model. Although there are improvements that could be made, the key advantage of this approach is that each topic can be labelled, giving an investigator additional context. This system presently allows an investigator to organise the documents by the report, viewing each sentence in a report in order, or by topic, with sentences from many different reports that concern the same topic. In the following chapters, additional context will be added by investigating the structure of the text, adding causal and temporal clues and creating a custom insider threat model from the reports.

6.6 Conclusion

This methodology has successfully extracted and mapped the insider threat characteristics from the perspectives experiment corpus of organic narratives. The corpus can be viewed either by document or by topic, where sentences from different topics but which discuss the same insider threat characteristic can be viewed. Topics can then be labelled according to the insider threat characteristic in order to make viewing these documents together easier. This can allow investigators to identify reports that must be prioritised quickly. For example, if an investigator is primarily interested in the insider's motivation, all the sentences regarding motivation can be found, and if necessary, the documents that the sentences belong to can be highlighted and read later.

One of the key advantages of insider threat frameworks is the ability to visually understand not just a single characteristic but how each individual characteristic plays a part in the larger whole of an attack. Although frameworks struggle to capture all views of an insider threat attack, due to the limitations discussed in the literature review, connecting individual aspects of an attack, even if the creation of the model only considered

the technical or sociological, is extremely important. Particularly in developing effective mitigations, root cause analysis can be challenging for investigators.

Therefore, the next stage of the system must create these links between each topic, effectively re-creating an insider threat framework. This allows the model not just to prioritise those organic narratives which the investigator is already interested in but to allow an investigator to find emerging details. The next chapter will discuss the creation of a simple insider threat model from the results presented in this chapter and then add the additional context of temporal and causative information from the text. This will transform individual topics into a graph of topics, with each connected to the others. This then allows an investigator to move from highlighting topics to a strategic view, visualising the entire attack using topics.

Chapter 7

Causality and Temporality

To understand the relationships surrounding each topic, they must be linked, recreating an insider threat model specific to an attack. These topics can initially be linked using Markov chains, a statistical model which shows the likelihood of a state transitioning. Considering a topic in a document as a state, a Markov model can be created showing the probability of one topic following another. However, from these initial links, it is not clear which of these links are causal, temporal, narrative or simply coincidental in nature. The next stage is to examine these links and attempt to recognise these connections and highlight those topics which may be of great interest to investigators. For example, those topics that show a timeline of events or a chain of causality can allow an investigator to highlight activity that can mitigate a causal chain that results in an attack. This then highlights the topics that must be examined further and potentially the reports that may be necessary to understand an attack.

The initial topic models and mapping allow the organic narratives to be segmented by topic; however, each topic exists independent of others. To use the models to successfully to understand an attack there must be additional context between characteristics and this is reflected in insider threat models such as the Nurse et al. (2014b) and the Cappelli, Moore and Trzeciak (2015) models. In these, each characteristic of an attack is given additional context by providing links between them. These connections are key to understanding

and influencing the organisational culture, and this is a particularly challenging problem (Coles-Kemp and Theoharidou 2010). Following an incident, it is even more important to understand the relationships between the individuals concerned, their peers and the organisation; the policies and processes in operation within the organisation as well as the technical details associated with the attack.

Exploring an incident within an organisation, it is clear that elucidating a 360-degree report of the internal actor, their relationship to other staff and the organisation, in addition to other factors outside of the workplace can help better understand the attack from both a pre-event motivational perspective and the attack methodology itself. A better understanding of these factors will help create more effective mitigations and improve the overall security posture of the organisation when defending against insider threats. One of the keys to extracting this 360-degree view of the incident is to collect a wide view of the incident from many different perspectives, to enable this it is important the approach has a low cognitive load to enable as many 'bystanders' as possible to contribute. It is also important that, during the data-gathering phase, we do not require individuals to contextualise their observations within a model or an existing understanding of how insider threat is understood — this may result in valuable information being discarded as it does not 'fit' with our existing understanding of insider threat. The 'organic narratives' of the incidents are a very typical way of writing about an incident where the narrative of the event naturally unfolds in a roughly temporally ordered manner. If an individual is asked to write about an event, this is typically the style of narrative that will be used.

Gathering organic narratives is a good way of ensuring a 360-degree view of an incident, there is a much-reduced cognitive-barrier to contributing, ensuring everyone, not just those with security backgrounds, can contribute meaningfully. It also ensures that narratives or observations are fully captured, without being manipulated to fit a formal understanding of insider threat, or details discarded because they do not fit in the current understanding of how insider attacks are committed. These often contain language describing the connections; this chapter will exploit these words and explore temporal,

causative and narrative relationships.

7.1 Method

This section discusses the key steps taken to create the causal and temporal relationships between topics contained within reports of insider threat attacks, in general, we will represent these as directed graphs where the nodes represent the topics and the edges the relationships. This builds upon the previous work, using the results from Chapter 6. In this approach we take the topic representation of the reports associated with the attack and these topics form the vertices of a directed graph. The graph is then enriched in layers; first, the narrative layer is created using Markov chains. Next, a causal layer is added to encapsulate the causal relationships within the corpus. Finally, a temporal layer is added, identifying the temporal relationships encoded within the document.

Initially, we create a model of the narrative structure of the report; this encapsulates an initial understanding of the relationships between the topics. Organic narratives are particularly advantageous for this approach as they describe an incident naturally, with casual or temporal words to join together separate events relating to an incident. If we assume these are organic narratives then the narrative and temporal structures are likely to be similar (although not identical), but there is little context — we cannot identify causal structures within the report, and indeed some relationships may not have any semantic meaning outside of the individual's narrative structure. However, this narrative structure provides a base for discovering more semantically-rich connections. Through analysing the entire corpus related to the single incident, we can construct a Markov chain (Ching et al. 2013), where each edge is a transition probability between two topics. This is founded on the assumption that the documents, to some degree, follow a structured, overarching narrative flow.

For the next layers, we consider the problem of extracting causal and temporal relationships. The temporal relationships allow the automated construction of a timeline

of events, whilst the causal relationship is particularly important for understanding the efficacy of defensive interventions, whether technical, behavioural or organisational.

To identify the causal and temporal links, we employ a pattern-based approach to managing the relationship extraction, the process for extracting causal relationships is shown in Figure 7.1 (the method for building temporal graphs being very similar). The causal verbs, causal patterns, temporal verbs, and event patterns were sourced from a dataset generated using a mixture of rule-based and classification approaches (Mirza 2016). These patterns (Mirza 2016) use Regular Expressions Aho 1991 and therefore are easily to apply to existing documents. The dataset provided by Mirza (2016) contains both causal phrases and causal verbs. This dataset was then improved with a clue on sentence struture, as in English there are mnay ways to express information and this does not always follow the second sentence following from the start. An analysis of the location of causal words in a sentence is shown in Figure 7.2, as can be seen, the initial peak in causality is within the first five words of a sentence. The algorithm only considers these if they appear within the first five words of a sentence, to ensure that there exists a link between two topics rather than between two noun phrases within a sentence. If a sentence matches a causal pattern an edge is drawn between the previous sentence's topic, this sentence's topic and the next sentence's topic, and the count of total causality in a topic is recorded.

Similarly to causality, the temporal graph is created by searching for temporal patterns at the start of the sentence; however, with temporality the problem becomes more difficult. Within the English language it is common to write 'out of order' exploiting different temporal words as noted by Girju, Moldovan et al. (2002), for example, *'Before x, y happened'* suggests that the next sentence happened before the previous, rather than after. Each temporal pattern was labelled using before/after/during, managing this problem. These labels instruct the algorithm on the correct approach to temporally linking topics, including when two topics happen simultaneously, this is represented graphically in Figure 7.3.

The final directed-graphs were visualised using the Yifan-Hu graph layout (Hu 2005).

If a topic in the graph visualisation appears near another, it implies some close relationship between the topic. Each topic records the number of sentences within the topic that exhibit causal or temporal patterns; this normalised value represents the causal or temporal impact of that topic. This will become clearer in the following results section. Following the validation exercise performed in Chapter 6, it is also possible to label the topics with the closest code from the from an existing model of insider threat; this label provides more context to the machine-generated topics.

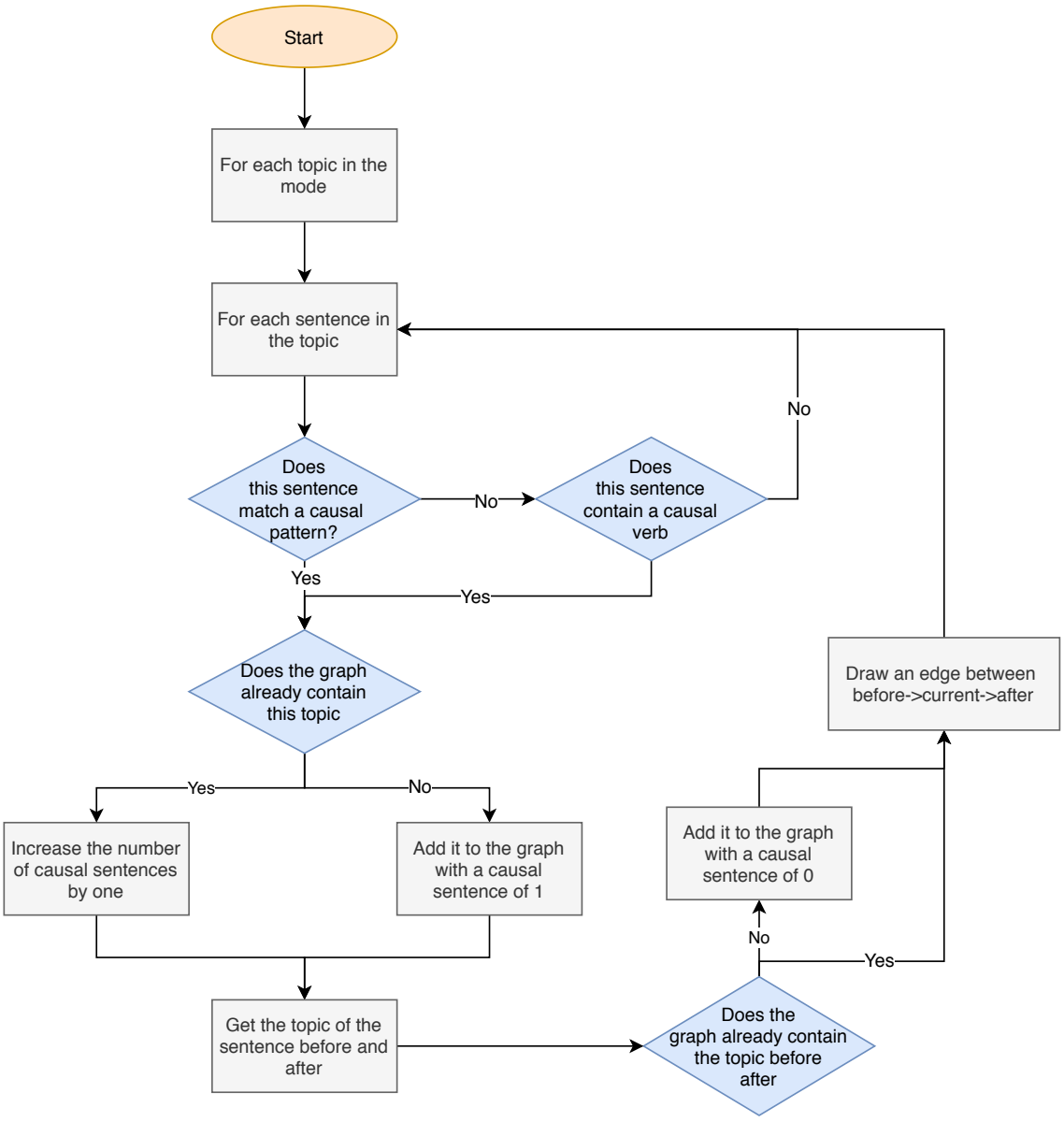


Figure 7.1: Building the causative graph

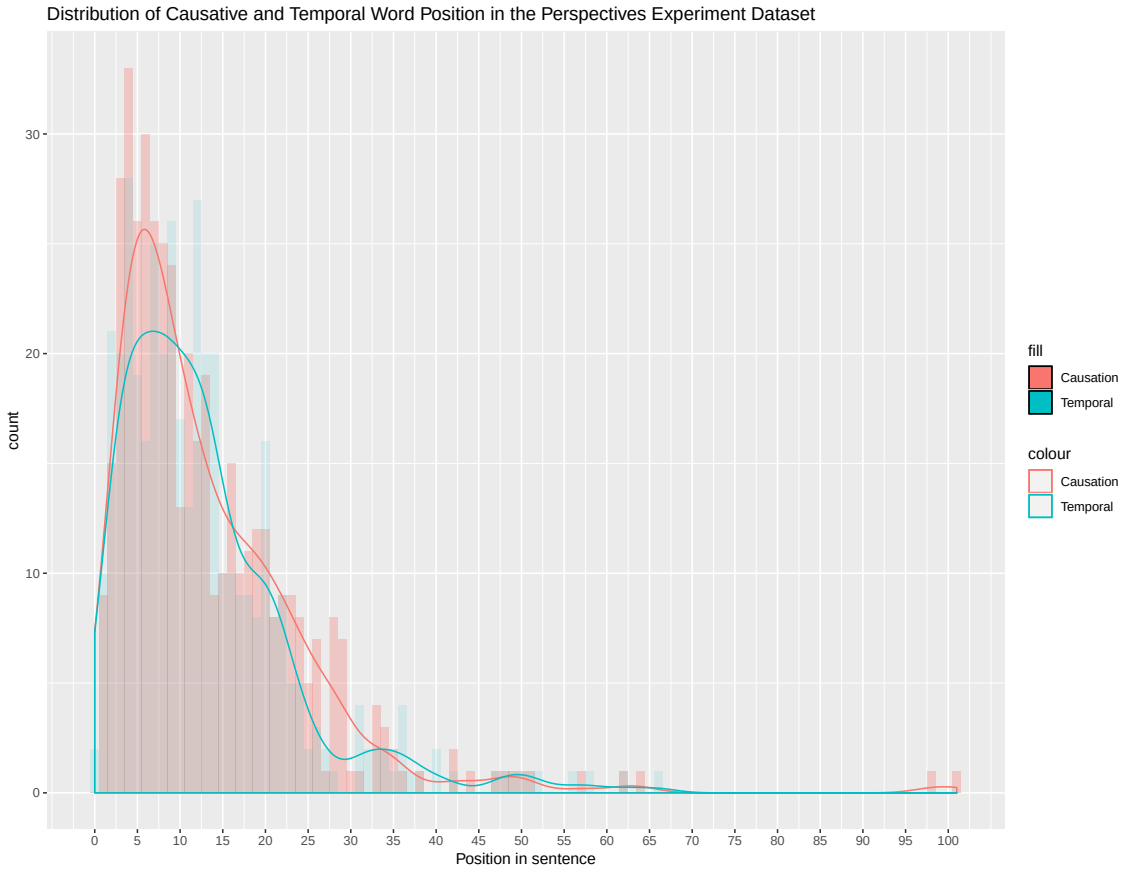
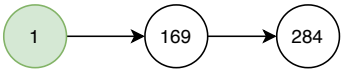


Figure 7.2: In which position do causal words appear in a sentence

Before that...

Previous Sentence: A manager at a Tax Office .....
Before that...
Next Sentence: When an IT system, that included fraud detection measures, was introduced...

Previous Topic: 169
This Topic: 1
Next Topic: 284



After that...

Previous Sentence: A manager at a Tax Office .....
After that...
Next Sentence: When an IT system, that included fraud detection measures, was introduced...

Previous Topic: 169
This Topic: 1
Next Topic: 284

During that...

Previous-1 Sentence: This story...
Previous Sentence: A manager at a Tax Office .....
During that...
Next Sentence: When an IT system, that included fraud detection measures, was introduced...

Previous-1 Topic: 34
Previous Topic: 169
This Topic: 1
Next Topic: 284

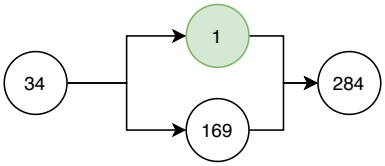


Figure 7.3: Example construction of the temporal graph (example taken from the data)

7.2 Results

In this section, we present the results from layering the relevant relationships on top of the underlying narrative structure. First, the narrative layer is created using Markov chains, the second layer is the causative layer created using the algorithm in Figure 7.1 and finally the temporal layer created with an adapted algorithm shown in Figure 7.3. This has been slightly modified as the temporal clues require the correct positioning of events.

The narrative links between topics are shown in Figure 7.4 there are clear structures that many documents follow, with the edge weight showing the probability of one topic following another.

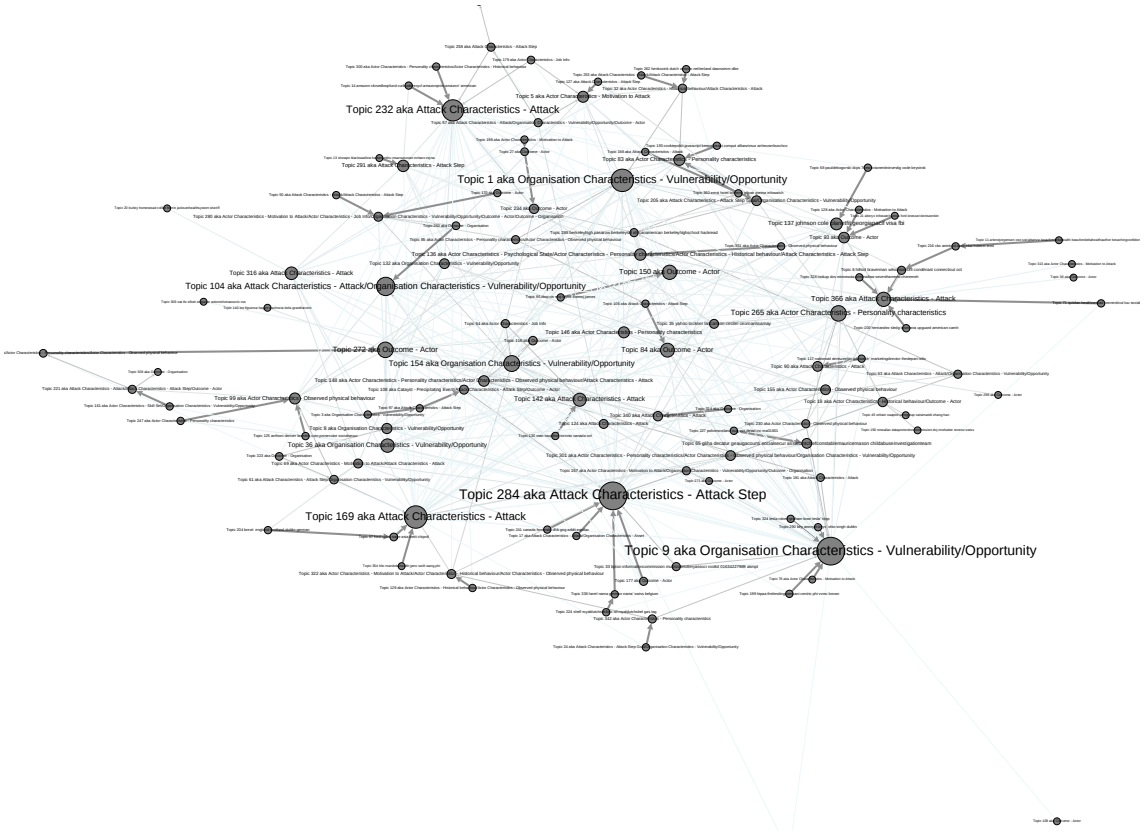


Figure 7.4: Narrative model of the topics using Markov chains

The computational nature of the LDA algorithm may mean that a description of an insider attack with more technical words, would be separate from a more general description — this is interesting from an analytical perspective as a more specific representation of a topic is more likely to have specific causal links than a generalised view. However, as

discussed previously, we can exploit the previous validation exercise to map the machine-generated topics to human-validated codes. At the narrative level, we can merge topics which are allocated the same human-generated codes to make this graph more intuitive (with the codes taken from a model of insider threat, this graph is shown in Figure 7.5.

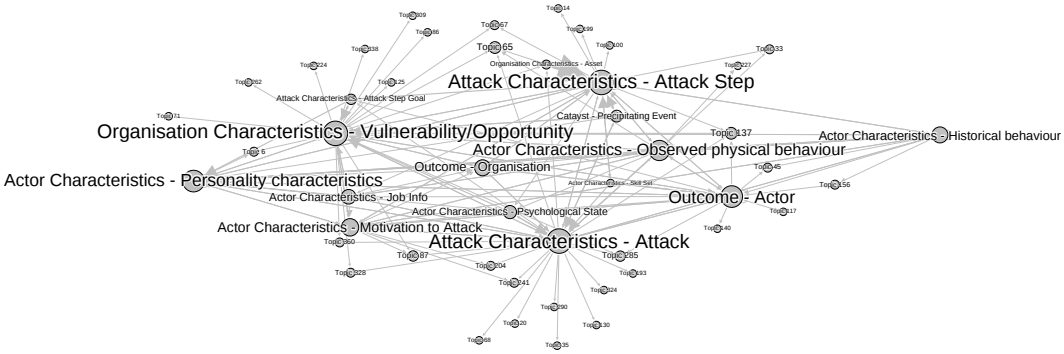


Figure 7.5: Narrative model of the data with the individual topics mapped to codes

These two graphs provide the initial layer that the causal and temporal graphs will build upon, specifically to bring additional context to identify key relationships between topics. This further demonstrates the close connections that some characteristics present, as well as being smaller. For example, the motivation for an attack, the personality characteristics of the insider and the insider's job are all close together in the graph in Figure 7.5. This smaller graph also shows the unlabelled topics; these are similarly close to one code, which could suggest that these also belong to that code.

The causal relationships are shown in Figure 7.6 showing the data coloured by amount of causal words/phrases; the node size shows the number of causal sentences that have been found in the topic and edge weight shows the number of links between the topics. For example, Topics 25 and 234 are smaller topics which have a smaller amount of causal sentences, but these make up the majority of the sentences within the topic. Similarly, Topic 284 and Topic 1 contain a large number of casual sentences; these represent a smaller proportion of the total sentences within a topic. Examining these topics shows there are clear, strong causal links and chains of causality, where one event causes another.

Figure 7.7 represents the same directed graph; the edges are coloured to represent the

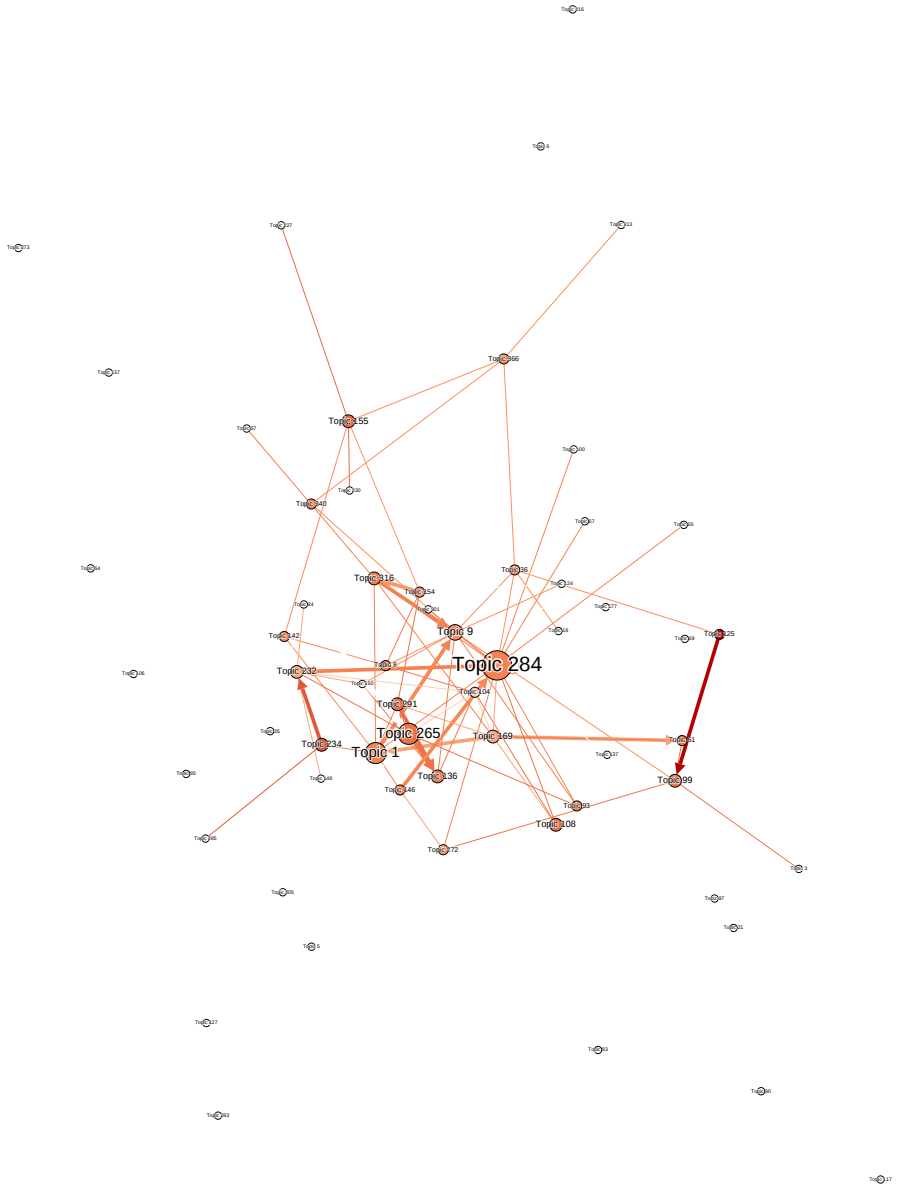


Figure 7.6: The causal layer coloured by causality

closest human-generated code to the topic. It is clear that causality occurs between different codes, demonstrating that there exist causal chains of topics that represent distinct insider threat characteristics. However, causality does appear within some codes; this may be because causality is an inherent property of this code, for example, when describing a single Attack Step.

The temporal layer is shown in Figure 7.8 and Figure 7.9, similar to the causality with each node size indicating the number of temporal sentences and edge weight showing how often these topics are linked. With Figure 7.8 coloured by temporality normalised by the number of sentences, and Figure 7.9 coloured by code. Figure 7.8 shows there are similarities between causality and temporality, with smaller topics with a proportionally high amount of temporal sentences such as Topic 32, 108 and 86 and some larger topics with a proportionally lower amount of temporal sentences such as Topic 9, 284 and 104. Unlike causality, there are far more temporal links, this is likely due to the rarity of causal links, in general people write in a very narrative manner, which is likely to demonstrate a larger number of narrative temporal structures. Figure 7.8 and Figure 7.9 show there are very clear 'stories' made from temporally-linked chains of topics, which can be seen to move across different codes.

These directed graphs across both narrative, temporal and causal layers clearly do exhibit structure, indicating that there is potential insight within the automatically generated relationships; this is explored in the following section.

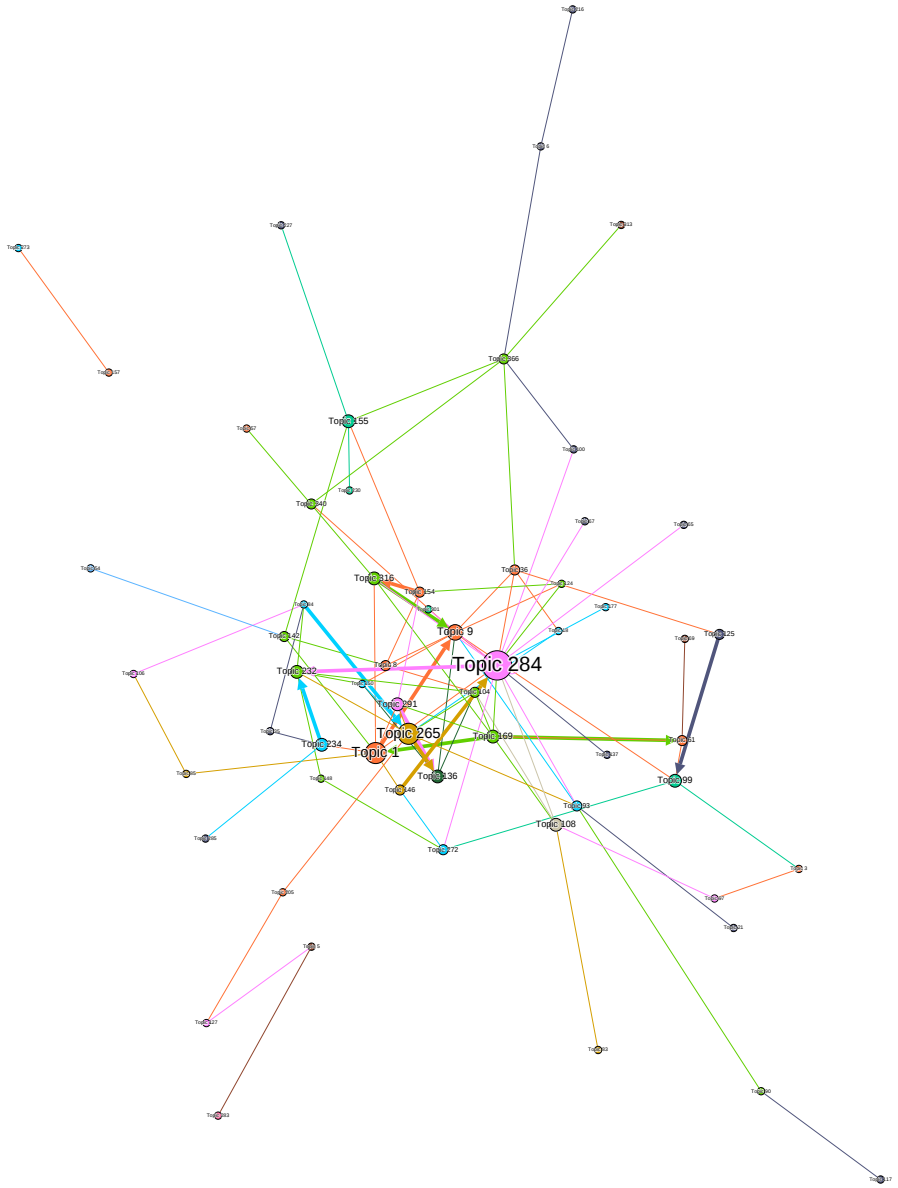


Figure 7.7: The causal layer coloured by code

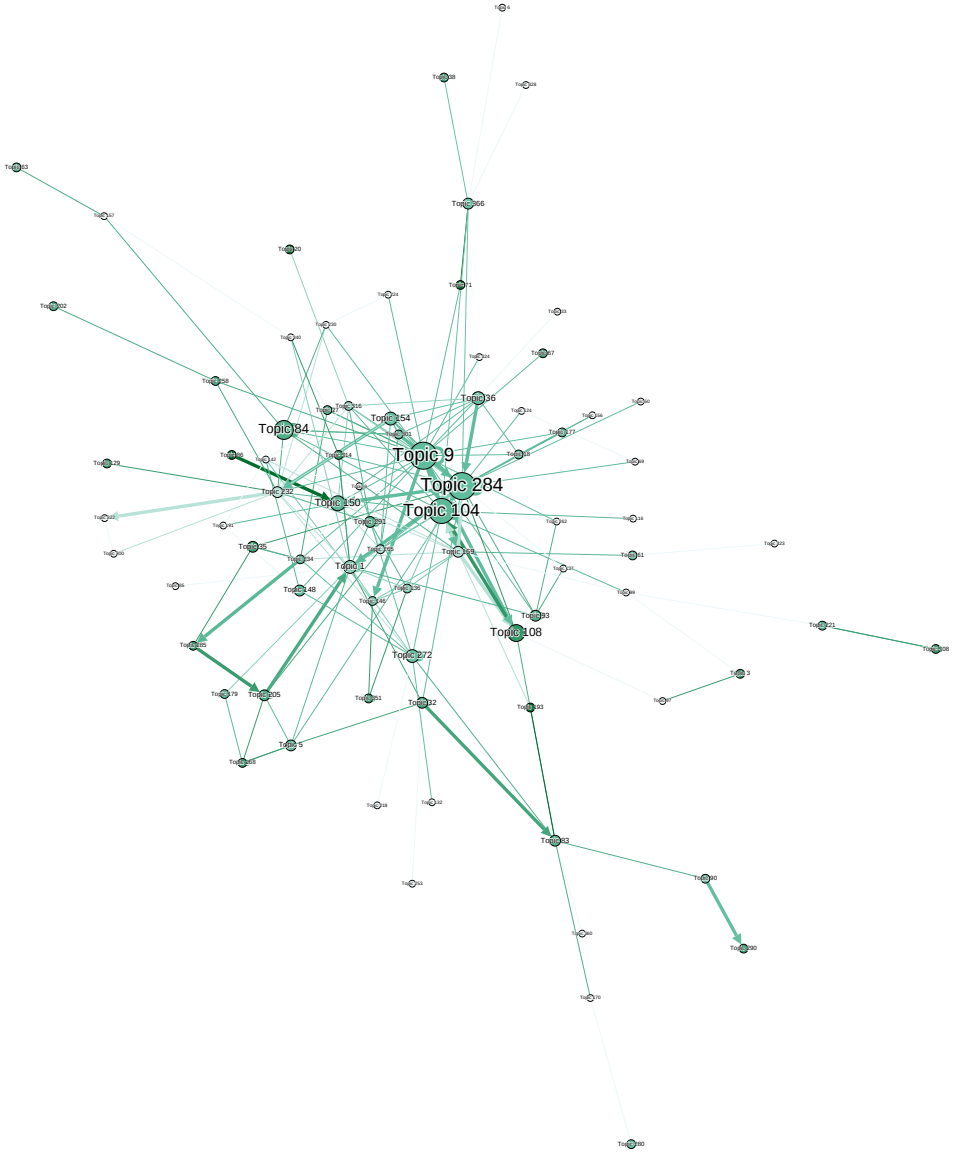


Figure 7.8: The temporal layer coloured by temporality

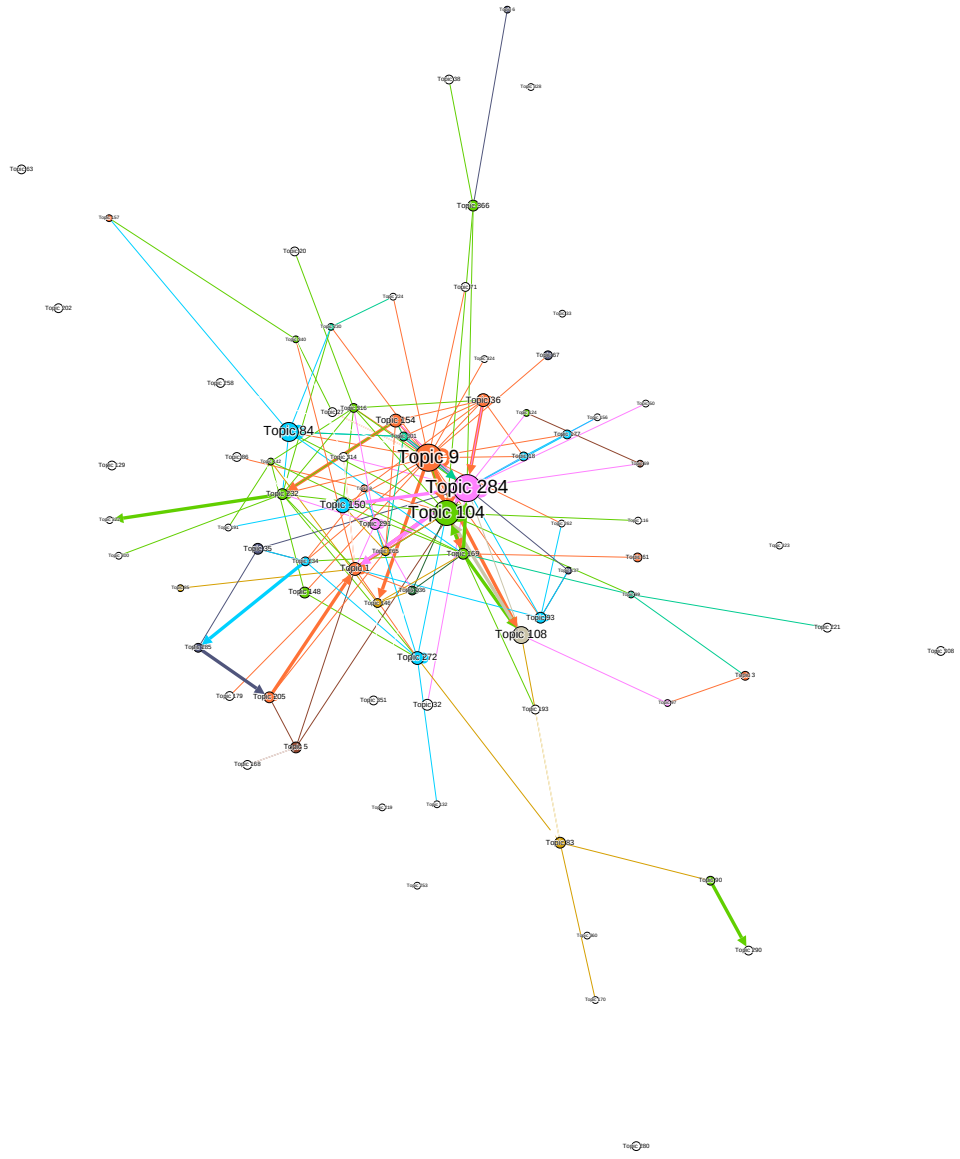


Figure 7.9: The temporal layer coloured by code

7.3 Discussion

This section of the chapter is broken down by the different types of relationships that are identified from the text corpus. Initially, the implications of the narrative structure are discussed; this is followed by the causative structure and finally, the temporal structure is explored. The section closes with a discussion of the implications of the entire visualisation and how it can be exploited to improve understanding.

7.3.1 Narrative Layer

The first layer that will be discussed is the narrative layer built from the Markov chains. This layer provides the foundation, showing the links between each topic as participants have written them. Of note is the various clusters which appear in Figure 7.4, these show common narratives used to describe an attack, for example, one such chain is shown in Table 7.1.

Table 7.1: An example narrative chain: Each topic is annotated by the nearest code from the human validation exercise - this code is from a common insider threat model (Nurse et al. 2014b)

Topic Number	Code
Topic 24	Attack Characteristics - Attack Step Goal / Organisation Characteristics - Vulnerability / Opportunity
Topic 342	Actor Characteristics - Personality characteristics
Topic 322	Actor Characteristics - Motivation to Attack / Actor Characteristics - Historical behaviour / Actor Characteristics - Observed physical behaviour
Topic 284	Attack Characteristics - Attack Step
Topic 63	Attack Characteristics - Attack / Organisation Characteristics - Vulnerability/Opportunity
Topic 285	unknown code
Topic 234	Outcome - Actor

From this example chain, we can observe the narrative account: first a discussion of the vulnerability, then information about the insider, followed by discussing the attack in detail and finally discussing the outcome. It is important to note that this chain does not represent an attack observed in one document, but many throughout the corpus. For example, one document may have started a chain: $24 \rightarrow 342 \rightarrow 322$ but then discuss 169 rather than 284, whilst another document may have a chain such as $322 \rightarrow 284 \rightarrow 63$ before ending their account, in essence, our approach synthesises narrative chains from the entire corpus of documents.

To add more context to this discussion, Figure 7.5 shows the same narrative relationships but this time each topic is merged with those with the same code from the validation exercise, so each vertex in the graph, e.g. the single vertex labelled 'Actor Characteristics - Personality characteristics' in Figure 7.5 is represented by multiple topics (vertices) in Figure 7.4.

This graph demonstrates that the automated approach is able to generate a coherent structure with clear narrative paths, e.g. the 'Actor Characteristics' appear close to each other, and 'Attack' is closely linked to 'Motivation to Attack' and 'Outcome Actor', 'Outcome Actor' is also linked to both physical and historical behaviour, whilst 'Attack Step' is more closely linked to the outcome of the organisation and the catalyst.

Finally, there are some topics which do not have codes; this is because these topics do not have any sentences that were in the labelled validation set due to the mapping described in the previous chapter. However, when investigated in this way, it is clear that these tend to cluster around a code, suggesting that these could also belong to the same code. Some are more ambiguous, appearing between two codes, suggests some link between both such as 'Attack' and 'Attack Step', which are contextually similar and these between topics may represent transitional sentences between the two. To assign a topic to these they could be investigated further and labelled manually, or a technique to computationally apply labels could have been created. However, not all sentences must be assigned a code and it is not necessary for the later stages, given this and the time constraints, this was not completed however could be a insightful piece of future work.

At this stage, the relationships that have been identified are purely narrative in structure, i.e. this is the way individuals tend to write about the incident. The next stage is to give each relationship a contextual meaning, specifically in organising the event temporally or identifying which chains may be causative in origin. Alternatively, there may be relationships which are entirely narrative or coincidental and arise from the organic narrative rather than having any 'meaning' within the incident itself.

7.3.2 Causative Layer

The first stage of adding additional context is to show causal relationships between topics, and therefore allow investigators to identify causal chains and consider mitigation strategies specific for the relevant chains. There are three key outcomes of the causal layer, firstly the overall level of causality both in the corpus and in individual topics, secondly the causality within and between different codes and finally the existence of specific causal chains.

Figure 7.6 shows the causal layer coloured by causality, this graph is notably smaller than the temporal graph (shown in Figure 7.8). This is likely due to the rarity of causal links, with some reports using temporal language to imply a causative relationship. These topics are usually smaller topics but with stronger causative links. Topic 125 $\rightarrow$ 99, 125 is a smaller topic, but with a number of causal sentences (i.e. it has a high edge weight) and a high percentage of casual sentences (i.e. the vertex colour), we can also see there are strong causative chains between Topics 169 $\rightarrow$ 1 $\rightarrow$ 9 or 146 $\rightarrow$ 284 $\rightarrow$ 232. If the smaller less weighted links are considered (i.e. those that are causal but not as strong), there are many more chains possible.

When examining causal chains, it is useful to understand the code of each topic, for example, the chain: Topic 169 $\rightarrow$ 1 $\rightarrow$ 9 becomes 'Attack Characteristics - Attack' $\rightarrow$ 'Organisation Characteristics - Vulnerability/Opportunity' $\rightarrow$ 'Organisation Characteristics - Vulnerability/Opportunity' which is indicative of multiple vulnerabilities resulting in the attack being made possible. Whilst the chain of Topic 84 $\rightarrow$ 265 $\rightarrow$ 136 becomes 'Outcome Actor' $\rightarrow$ 'Actor Characteristics Personality characteristics' $\rightarrow$ ' Actor Characteristics Historical behaviour', indicating that the insider's personality characteristics are related the later behaviour by the insider. A final example chain may be Topic 146 $\rightarrow$ 284 $\rightarrow$ 232 which becomes 'Actor Characteristics - Personality characteristics' $\rightarrow$ 'Attack Characteristics - Attack Step' $\rightarrow$ 'Attack Characteristics - Attack', in this chain the actors' personality characteristics has a direct causative link to the attack step which has a causative link to the overall attack.

When considering the causality between codes shown in Figure 7.7, it is noteworthy that some codes are highly causal appearing multiple times within the graph additionally there are causal links between codes as well as within certain codes. The best-represented codes in the graph with many nodes these are topics which would be mapped to Attack Characteristics - Attack (shown in green in Figure 7.7), Attack Characteristics - Attack Step (shown in pink) and Organisation Characteristics - Vulnerability/Opportunity (light blue). These three clearly map to the core elements of the attack (the factual 'what happened') and are the easiest to join as a clear causative thread. In addition, causality can link topics within a code, such as Attack Characteristics Attack (green), this will be because the automated topic discovery is likely to generalise less than a human, with humans able to '*capture the high-level conceptual patterns*' (Baumer et al. 2017) and therefore to fully express the complexity of an attack is likely to require a number of causally linked sentences, within a topic.

There is a clear link between causal and temporal relationships in that causality tends to imply a temporal relationship (the cause will precede the effect); however, there is value in explicitly identifying temporal links. This links may not encapsulate a causal relationship, but the reconstruction of a timeline (even if not causal) is of significant use to investigators.

7.3.3 Temporal Layer

The final relationships we can automatically discover are the temporal relationships; there are three key findings from this relationship, first the degree of temporality expressed in the corpus, second the relationship between temporality and code (i.e. the human coded validation exercise) and lastly specific temporal chains of interest.

Firstly the amount of temporality over the entire corpus, this is shown in Figure 7.8, in contrast with the causal relationships the graph of the temporal relationships has far more vertices and, generally, topics have a lower number of temporal sentences. This suggests that temporality is not a feature of a small number of topics, but instead, any topic is

likely to have some temporal sentences. The ability to find and understand this temporal information is extremely useful, showing how one event can lead to another and telling the story of an attack. This can allow investigators to understand each event in context to the previous events as well as to create a timeline.

Considering the codes associated with the temporality graph (shown in Figure 7.9), it is clear that there are temporal links between different codes as well as the same code, as was seen in the causal relationships. However, unlike the causative graph there is no strong very common code, but Organisation Characteristics Vulnerability/Opportunity (orange) and Attack Characteristics Attack (green) have a large number of temporal links, likely due to this forming the core story. This suggests a kind of timeline, with one core event such as Attack Characteristics - Attack Step happening and then Attack Characteristics - Attack following it. This supports the notion of there being a timeline of events, specifically from Topics 234, 285, 205, 1 or 36, 284, 150.

When looking at these specific temporal chains, it can be useful to compare to their code; this can help provide a comparison with the original model. For the temporal layer a common chain involves 'Organisation Characteristics - Vulnerability / Opportunity' $\rightarrow$ 'Attack Characteristics - Attack Step' $\rightarrow$ 'Actor Characteristics - Historical behaviour' or 'Organisation Characteristics Vulnerability / Opportunity' $\rightarrow$ 'Attack Characteristics Attack' $\rightarrow$ 'Actor Characteristics - Historical behaviour'. These links suggest a discussion of the insider finding a vulnerability, to exploiting it, and, as the code 'Historical Behaviour' was used in the initial labelling by participants to describe 'behaviour identified before the insider was caught', finally the insider's behaviour before they committed the attack.

In summary, the temporal relationships add more additional context to the narrative and causal structures of the directed graph, specifically creating a timeline that an investigator could follow. In addition, the structure of the relationship shows that temporality is fairly common when describing an attack, but with a focus on the events preceding and following certain codes. Specific, strong, temporal chains can also be discovered using

this method, forming a backbone of a timeline of events.

7.4 Conclusion

This section demonstrates how topics from the topic model created the previous chapter can be understood contextually using additional linguistic information, primarily using temporal and causative words. This section also demonstrates how computationally created 'topics' (or themes) can be enriched by building narrative Markov chains. These chains can then have additional contextual information added by identifying causal and temporal structures. This layered approach allows investigators to understand an attack and specifically highlight those areas that are of potential interest by finding chains of causality, where one event causes another or temporal chains, where one event follows another, or by the narrative relationships where one event is related to another.

Considering each layer separately, the underlying 'backbone' of an incident emerges. The causal layer is showing direct cause-effect relationships between distinct events, the temporal showing a timeline and the narrative layer showing how attacks are reported. This separation of the layers allows investigators to highlight key topics or chains of interest, before combining a much smaller graph to understand the relationships between each topic. This technique shows how through the consideration of each relationship (the narrative, temporal and causal) a full picture of the attack emerges, and key details can be recognised.

The causal layer highlights those topics with direct causal links between them. For example, the insiders' motivation caused an attack to happen; interestingly, we also see other rich causal relationships such as the vulnerability and the insiders' personality characteristics being linked to the attack. The temporal layer shows a timeline of events, these may not be directly related to one another, but show the attack 'as it happened' to those telling the story, placing each event in temporal context with others. Finally, the narrative layer shows how witnesses write about attacks, linking together important information

such as the personality characteristics to the attack. Each layer adds additional context to the attack, giving a high-level overview, but importantly shows which topics, and therefore which sections of reports should be prioritised for triage. When all relationships are combined into a single graph it can be difficult to interpret, however, by creating a subgraph focusing on the topics that are of most interest (for example those which form highly causal chains) allows an investigator to understand the details of a chain.

These first two technical objectives form a similar visualisation to those provided by existing insider threat models. This gives an overview of an attack, and this research can allow an organisation to understand how characteristics impact each other. However, this does not allow an organisation to see the fine details. Fine details can be vital to making organisational changes, such as IT governance. This next chapter will seek to address this issue and analyse individual topics.

Chapter 8

Topic analysis

While the attack language and causality technical objectives offer a high-level overview of an attack, it can be useful to explore an attack's details. This approach analyses each sentence within a topic, understanding the topic as a whole at a micro-scale. In the context of an insider threat framework or model, the first stage of this research, described in Chapter 6, collects each characteristic of an attack. The second, described in Chapter 7, automatically finds relationships between these characteristics using causal or temporal clues. These technical objectives create the initial custom insider threat framework from the organic narratives. This final technical objective aims to describe a topic in detail. For example, rather than labelling a topic as 'motivation', this offers an approach to visualise the sentences and allow the topic to be fully understood. Gathering these key details is vital as it enables an organisation to make lasting organisational changes that may prevent the next attack. For topics such as the methodology, this process may also be key in identifying how an insider bypassed technical controls or insight into how technical controls could be strengthened in the future. While the previous technical capabilities created, covered in Chapters 6 and 7, this Chapter enables tangible controls to be implemented and actions to be taken.

To analyse each topic, off-the-shelf tools were employed using CoreNLP (Manning et al. 2014) and Python. CoreNLP offers a suite of tools designed for common NLP tasks,

such as part-of-speech (POS) tagging and named entity recognition (NER). This is designed as a pipeline, allowing a developer to get results from different tools, enriching the results from a single process. In our approach we use the OpenIE (Open Information Extraction) module, (Angeli, Johnson Premkumar and Manning 2015) to extract IE triples. IE triples represent relationships in a piece of text, labelling each element as subject, object and relations. An example is provided by (Angeli, Johnson Premkumar and Manning 2015) where 'Born in a small town, she took the midnight train going anywhere.' becomes 'she; born in; small town'. Extracting these triples allows sentences to be simplified and visualised, merging several sentences with the same triples. This research maps these triples visually and forms a directed graph. A merging algorithm was designed and employed to aid in the simplification. Finally, these are visualised using the software package Gephi 0.9.2 (Bastian, Heymann and Jacomy 2009).

This Chapter presents the method to analyse topics, showing the development of the technical methods. Then selected results will be shared from the 300 topics with sentences assigned to them (70 topics did not have any sentences assigned to them). These 70 topics likely represent other characteristics not included in this scenario or potentially other archetypes of insider threat attacks. Finally, this Chapter ends with a discussion on how this stage can be used in the investigation processes and specific recommendations for practitioners.

8.1 Method

To perform this analysis, three steps were completed. First, the sentences in each topic and analysed individually to extract the OpenIE triples. Second, the sentences are merged, linking sentences that would usually appear in different documents together. Finally, these are plotted on the graph and visualised. Importantly the Hu (2005) layout is used, which pushes nodes with many connections together, allowing the visualisation to cluster related concepts. This creates a final graph for each topic, showing the individual concepts within

a topic. Although this is limited to the original language used in the organic narratives, this produces intuitive graphs, varying in word use rather than concepts.

As discussed in the previous section, this work uses the CoreNLP suite of tools, specifically the OpenIE triple extraction. This open-domain information extraction tool analyses sentences to extract the subject, relation, and object. This produces the following output 'Born in a small town, she took the midnight train going anywhere.' To 'she; took; midnight train' and 'she; born in; small town'. As both sentence fragments use the subject 'she', they could be merged, assuming that each represents the same 'she'. In the case of this work in particular and the subjects, sentences can share many words if they are in the same topic. Although some details are not captured, the core ideas of the sentence are. However, this only applies to a single sentence and does not consider the other sentences in a topic. This process and an example can be seen in Figure 8.1

This initial process has many repeated nodes that can be viewed in Figure 8.2, this requires the use of the merging as mentioned earlier algorithm. The merging is done using a bespoke algorithm, identifying where two sentences share common relations, subjects and objects and joining these together. This algorithm also merges similar triples to attempt to reduce the overall size of the graph. This was created in Python and outputs to a Gephi file. This algorithm is in Figure 8.3 and the merged version of Figure 8.2 can be seen in Figure 8.4. Importantly this merging preserved many features of the sentences, such as causality.

After the merging, the graph is drawn, drawing a directed edge between the subject and the relation and the object and the relation. Keeping note of whether or not the word appeared as an object, subject or relation. This was then visualised using Gephi. In Gephi the nodes are colourised and visualised with the Hu (2005) graph layout. This layout attempts to model the graph as a collection of springs, allowing similar concepts to appear closer together. This layout allows for the discovery of distinct concepts, which is important to understand the case fully. This graph can then be explored or exported for further analysis.

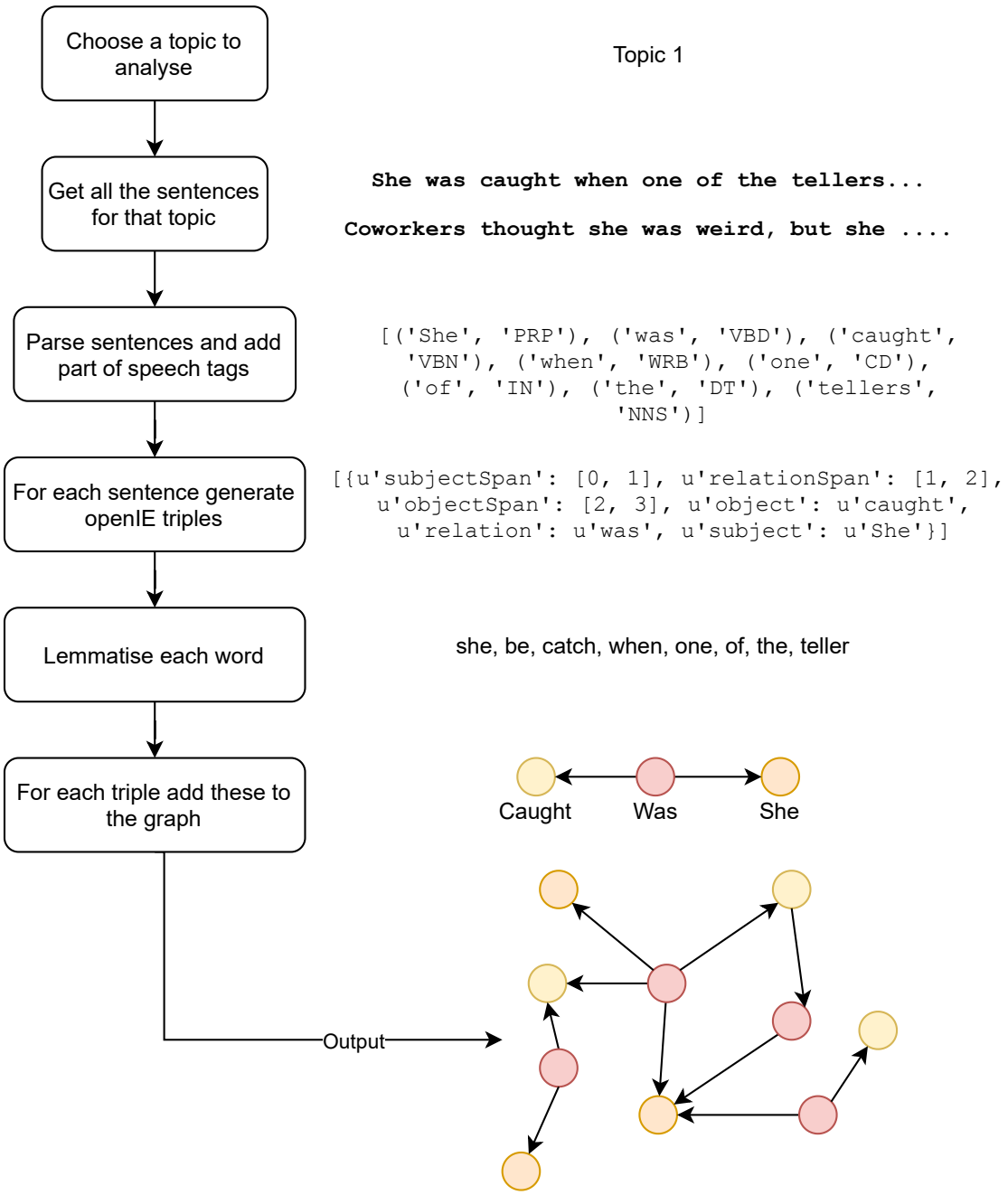


Figure 8.1: The inital pipline

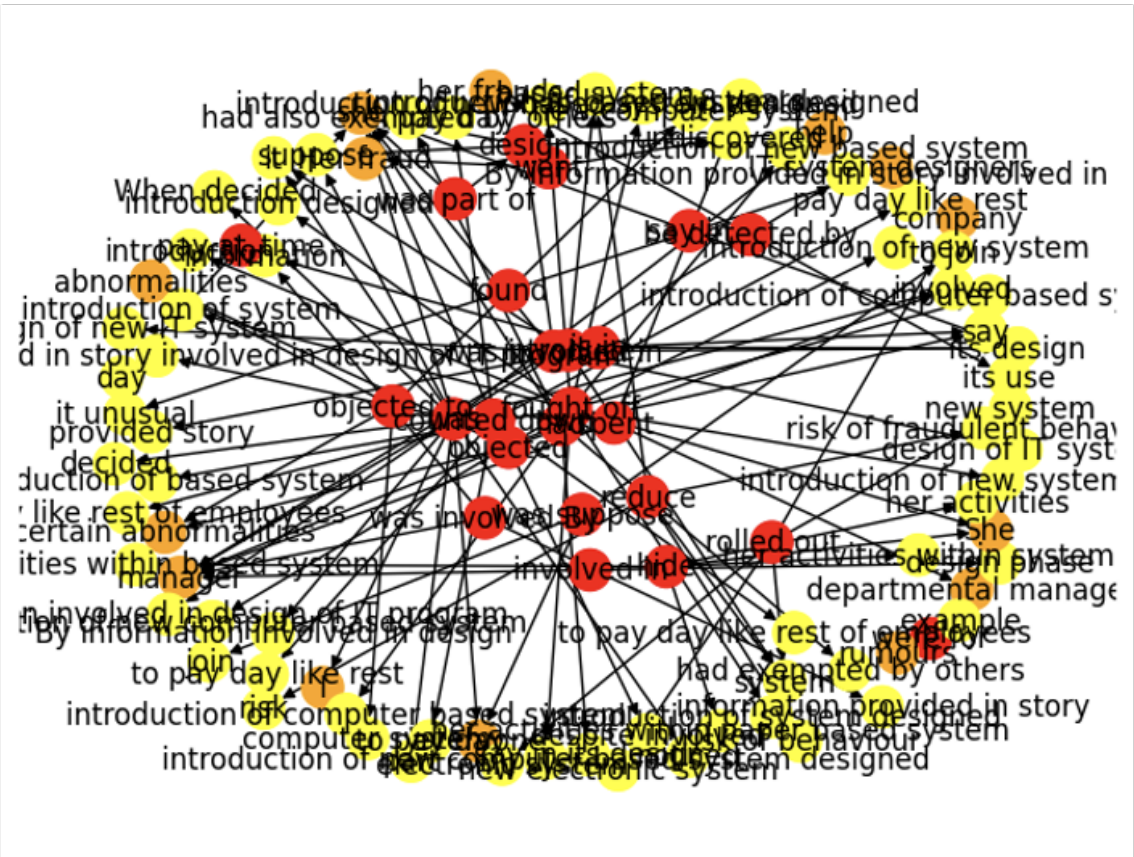


Figure 8.2: An unmerged topic

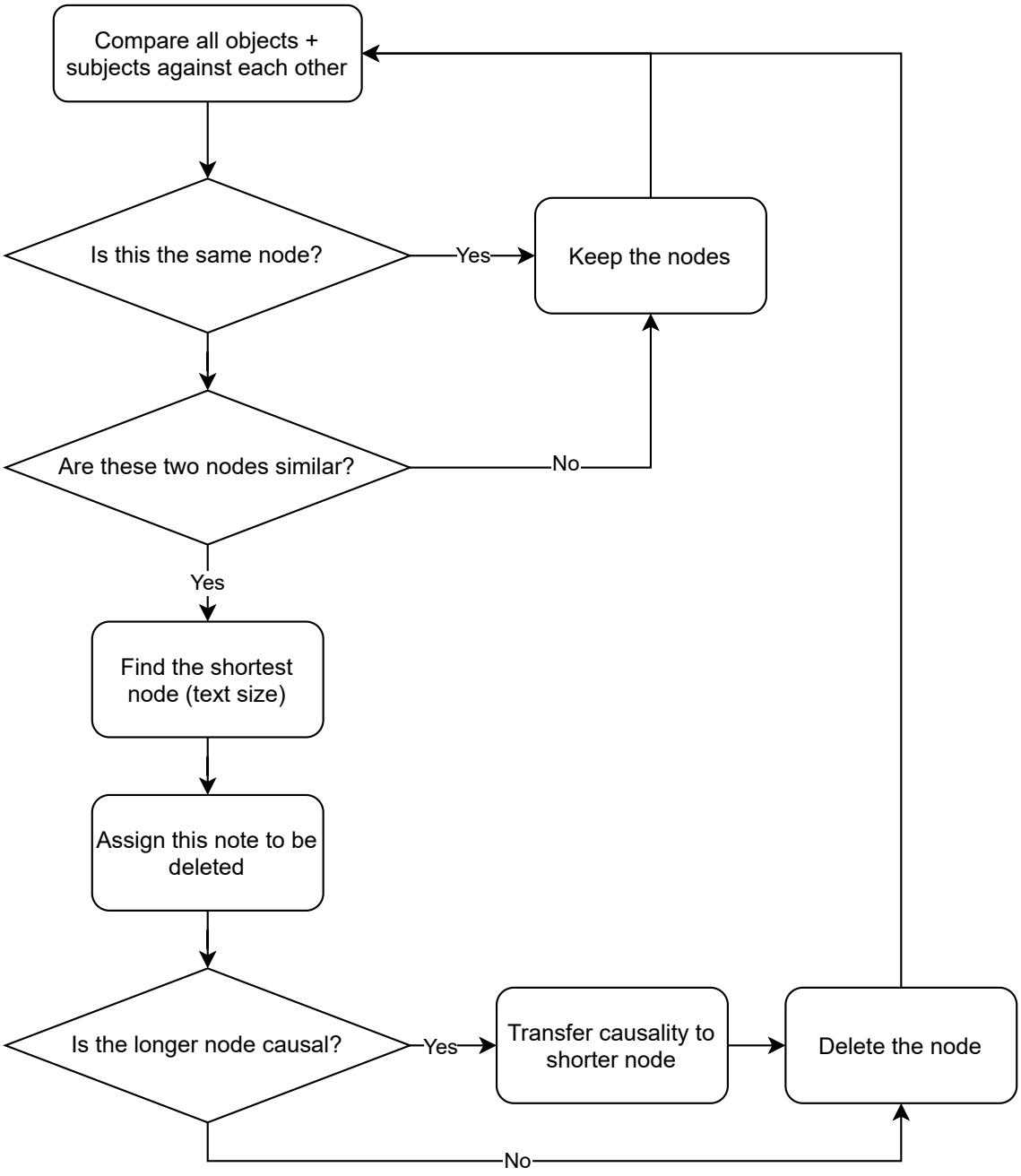


Figure 8.3: The merging algorithm

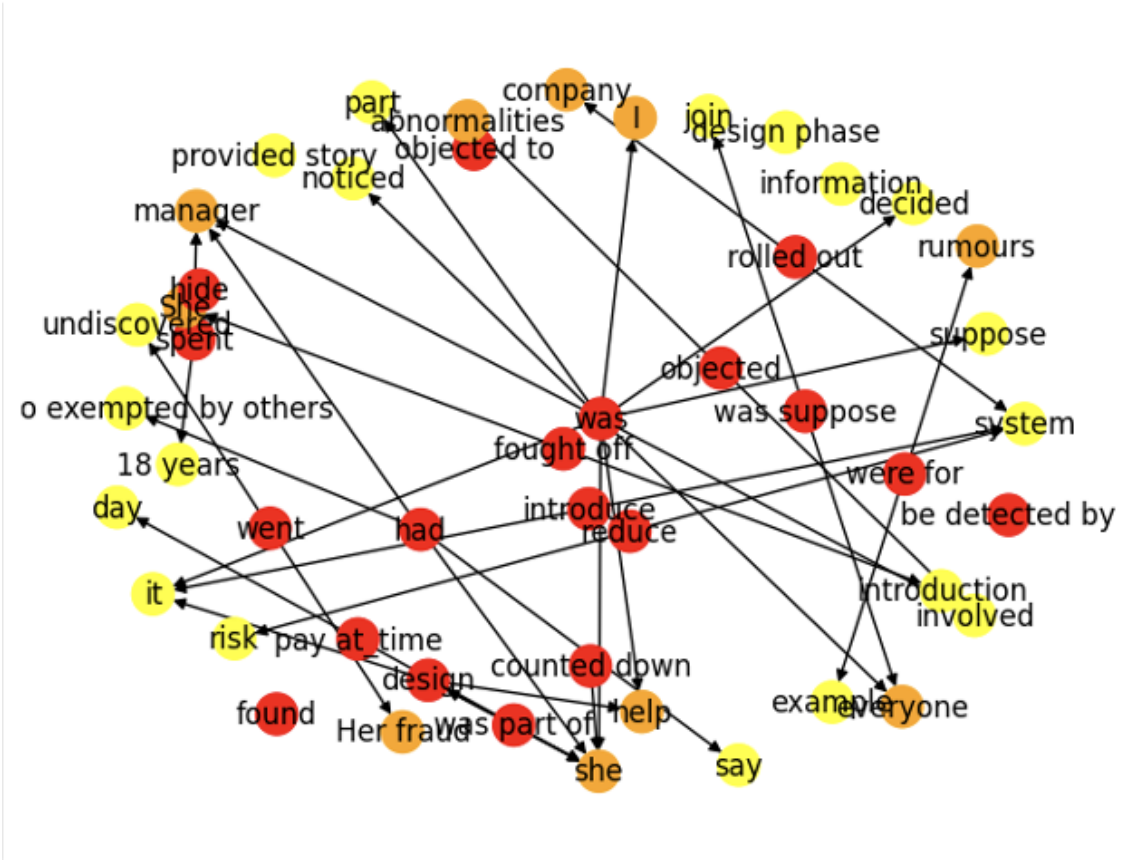


Figure 8.4: A merged topic

8.2 Results

This section will discuss the results from this approach to explore the data from the perspective experiment case. Each graph has been created using Gephi and represents a single topic, containing one or more sentences from one or more documents. Each graph node is coloured by either the subject, object or relation while node size represents the total number of connections to that node; the graph layout uses the Hu (2005) graph layout, which models the graph as a series of springs, pulling together nodes which have many connections while those that do not have many connections are forced apart. Therefore, the location of the nodes shows how closely related a node is to another. Finally, nodes with red text are causal in nature; these contain either a causal verb or a causal phrase. It is important to preserve this causality as this can show why a topic may be causal overall. The approach for detecting and mapping causality is discussed in detail in Chapter 7. The results below show a selection of representative topics. The first shows a typical topic of medium size, the next shows a much smaller topic, then a larger topic and finally a typically sized topic.

Figure 8.5: Topic 36

Figure 8.5 shows the final graph for Topic 6. This topic is of medium size and contains a small amount of causality. This topic clearly shows three core ideas, represented by the three linear structures. First, a new computer system was implemented. Second, she was made exempt from using it, and third, it reduced the risk of fraudulent behaviour and could

have detected her. There is also an outlier, the mention of 'waiting to payday'. This has significantly less detail than the three concepts discussed previously. From Figure 8.5 we can observe that topic primarily revolves around the insider and her relationship to the new computer system. From this, the following can be gained: first, there is a new computer system, and second that it was to reduce risk and third that the insider's methodology was to become exempt from this system, despite system designers finding it unusual. This might be of particular use to inside a threat investigation. This describes the insider's core methodology; future policy changes can be enacted, making the computer system mandatory or ensuring that employees do not become exempt from auditing measures.

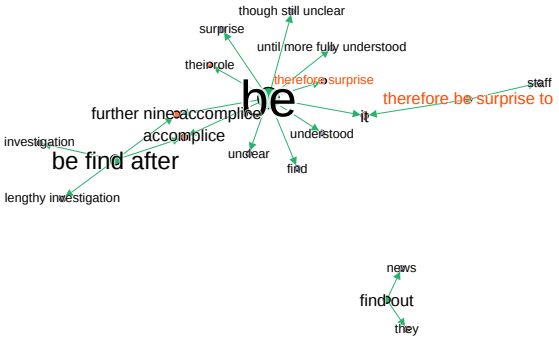


Figure 8.6: Topic 285 Graphed

Figure 8.6 shows Topic 285. This topic is much smaller, containing only a few sentences. However, even though this topic is smaller, it is evident what this topic discusses: staff were surprised, and there were nine further accomplices. This topic primarily involves the others in the insider threat attack, the accomplices, and other staff members. There is more noise on this topic due to the difficulty in merging the nodes because of the smaller number of sentences.

Figure 8.7 shows Ropic 9, this is a much larger topic with many causal sentences, and the graph reflects this complexity. This graph shows several details; however, they all relate to the process of committing the attack, with details regarding creating discrepancies, the inside of being caught, and the use of a new system, although this does not have the same amount of detail regarding the inside exemption, it does show an overview of the entire attack. This topic revolves around the insider and the details regarding the attack,

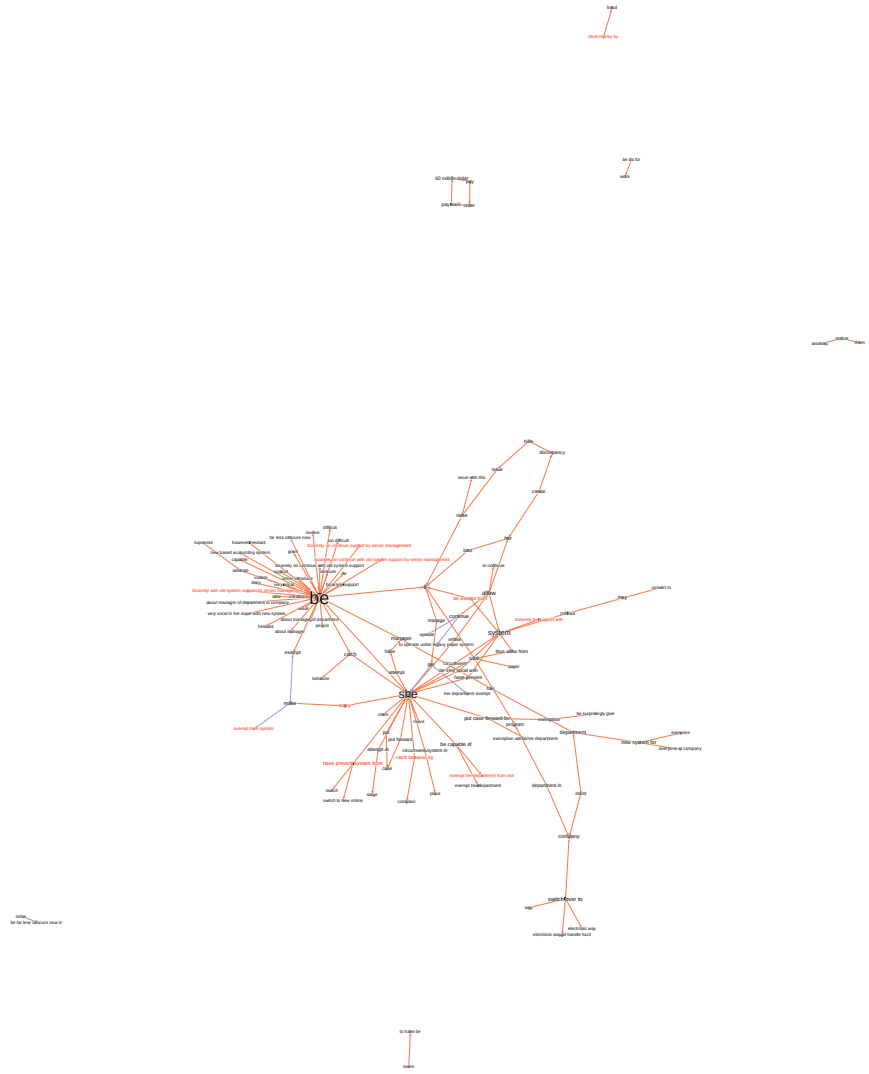


Figure 8.7: Topic 9 Graphed

shown as ‘she’ appears in the centre. There are many nodes under ‘be’ due to the overall common nature of the ‘to be’ verb structure ‘she is/was...’. However, these are better represented in the rest of the graph with further details. Finally, this shows a large amount of causal information, and it is clear that these primarily involve prevention or exemption. This shows how this method can be used on a larger topic to understand it, despite a large number of sentences classified into that topic.

Figure 8.8 above shows Topic 1. This topic is an average size for the dataset and shows how a typical topic may have several ideas relating to the insider and not related to the insider. For example, many of the details revolve around the word ‘she’, with the additional graphs showing additional information from a different point of view, for

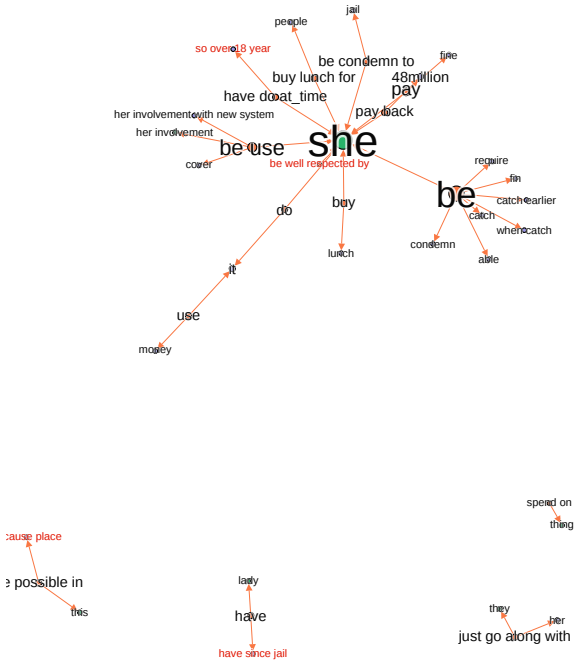


Figure 8.8: Topic 1 Graphed

example, that of the co-workers with ‘they’. This shows several core ideas firstly, the insider’s relationship with the co-workers, she was well respected by them and bought them lunch, secondly content related to the insider being caught, and having to pay a fine. The subgraphs show the insider being jailed and that the co-workers just went along with her. This is a fairly narrative topic telling the story of the insider being caught and how the co-workers did not suspect her. A human would not necessarily categorise this as being related. However, the machine identifies that these sentences tell ‘the story’ and are related despite being two different characteristics of an insider threat attack.

The selected results in this section show several graphs representing the most common topic types. However, with differing levels of causality and sizes and differing levels of detail in the sentences, these results are also shown across different insider threat characteristics. Each result shows a core idea surrounding the topic, and an investigator can analyse important details surrounding the attack. As these results are likely to be used by highlighting topics of interest from the causality and temporality graphs and ‘drill down’ for further investigation and gain specific details about the attack such as the methodology, each graph can show this information without overwhelming a potential investigator with

a large amount of detail. These graphs also show the efficacy of this approach, demonstrating that it is possible to gain information from the text without reading each sentence. The following section will discuss how an investigator is likely to use this information and recommendations for this approach.

8.3 Discussion

In this section, the results will be discussed in further detail and suggestions for use by a practitioner. First, a short discussion of how to interpret these graphs, next, a discussion of the similarity between topics and the differences, then discusses how a practitioner may use this tool and finally, a discussion on the tool itself and its strengths and weaknesses in an investigation. It is important to note that the analysis stage aims not to replace a human investigator but to help one understand the insider threat attack from several points of view and without having to read individual reports to do so. Therefore, this is considered successful if it is possible to understand the topic without the investigator reading each sentence from every report.

The most important thing about reading these graphs is to understand the core ideas from each topic. However, each graph represents a single topic. Due to how topic modelling works, this may read to humans as being about several different topics, whilst the machine links these together a human may not. When humans are given coding tasks, they are often more general than a machine; this is shown both in the literature (Baumer et al. 2017) and in our experiment in Chapter 5. However, when dealing with small units (in this case, sentences), they may be thematically related and contain more than one core idea. The best way to read these graphs is to look for long interconnected sections of the graph. These subgraphs show the concepts with the most significant detail, such as if several sentences discussed the same key idea but used different words to express it. For example, in Figure 8.5 it is clear that there are three core ideas, shown by the long and interconnected subgraphs from the central node ‘she’. Although this topic contained

different words, as individuals chose to use different words, the similar context of these sentences creates a clear graph with the three core concepts.

Although these figures represent different topics, there are some apparent similarities between them. First, each graph contains one key node, for example, 'she' — referring to the insider, and from this key node, we see each graph grow, adding new details with new nodes. Even the most interconnected graph Figure 8.7 still has two key nodes, 'she' and 'be' representing the insider and each verb form of 'to be', for example 'she was/is'. This shows that this approach can focus on the key details surrounding specifically the insider, and the graph grows naturally from this key concept.

In addition to this key node, the graph grows organically from this idea when placed using the Hu (2005) graph layout. This visually demonstrates pulling out key ideas related to each other. For example, in Figure 8.5 there are three core ideas in this topic that all relate to the introduction of a new computer system. However, in this figure, one subgraph focuses on the insider's exemption to the new system while another discusses its design. These are visually separated from one another on the graph showing that although these are related thematically, they are still separate concepts and may have to be investigated separately. This is even true of Figure 8.7, which is significantly larger and shows one concept–the overall attack details, which is highly interconnected. From this figure, some clear topics are visible: the manager was exempt from the new system, that the insider circumvented the system, committed the attack, and now everyone must use a new system. This shows the same characteristic when discussing the vulnerability/opportunity the insider had. However, it does it in less detail and includes additional information regarding the overall attack. However, the size of this topic is still clear that this key information can be visualised.

Each graph does possess a level of noise; this is due to the difficulty in merging these concepts. These usually appear as disjointed subgraphs, outliers, rather than in one of the complex chains to visualise the details. This noise is challenging to remove using a computational approach due to the nature of human language. This approach allows for an

investigation to be completed with less bias and quicker. However, reports can be noisy, especially when analysed on a smaller scale, such as individual sentences. However, human analysts can identify these outliers, and the core details can be focused on, as is clear from the figure shown above.

The primary difference between the Figures is that larger topics have far more nodes and relationships than the smaller topics. This is not necessarily a disadvantage of this approach since smaller topics are more focused, and it can be easier to gain important information about the incident from them. Although the larger topics may give more of an overview, they often do not contain the same detail as the others. For example, in Figure 8.6 and Figure 8.6, both contain details about the attack but Figure 8.5 contains the specifics about the new system, being made exempt as well as being involved in the design of the system.

Several steps must be followed; this is discussed in detail in the next Chapter. First, topics must be selected for investigation, selecting those topics based on the causal, temporal or narrative graphs, or by choosing a specific characteristic. This technical objective then allows for the specific details, for example, the assets targeted or how security controls were bypassed. This is extremely important when creating insider threat precautions for IT governance or policy. If this research were to become a piece of software, an investigator would narrow down the choice of topics ahead of time, so it would not be necessary to view all the topics in the corpus.

This tool allows investigators to understand the topic better. As topics are generated automatically, this stage allows investigators to analyse a particular topic and understand what the topic is about and partially how the machine has decided to place the sentences together. Therefore, this stage improves the initial understanding of the computational models whilst not overloading an investigator with extra details. In addition, this tool offers a way to inspect each topic and decide which reports might need further investigation, enabling the maximum benefit of human expertise and machine.

This is a critical stage for understanding insider threats as it allows an investigator

to understand the particular topic fully. Initially, an investigator understands individual topics by observing the insider threat characteristic label associated with the topic, then understanding the links from this topic to others using causal, narrative and temporal links. Finally, this stage allows an investigator to dig deeper into particular topics and fully understand the details surrounding them.

One of the key improvements that can be made would be to develop the merging algorithm further to remove the noise generated on some graphs. For example, measures could be created to measure the decrease in size relative to the amount the graph was merged presently, the merged sentences are merged on the stem of words, comparing walked to walking to walk, rather than additional contextual information. Specifically, the merging of sentences where writing style differs greatly between sentences would require additional tools and would not necessarily be possible with out-of-the-box techniques. In addition, adding further information onto this graph may make the graph overwhelming and therefore unusable by an investigator. Therefore, the decision was made to limit this particular stage to these graphs. However, in the future, it would be possible to improve the merging algorithm using techniques such as Word2Vec (Mikolov et al. 2013) to calculate the vector distance and, therefore, the similarity between two words. However, this requires a large amount of data, specifically with the less formal organic narrative style and would be considered out of scope for this project.

This section has shown how this algorithm's results and the graphs visualised can be used to fully understand the insider threat attack, specifically by allowing investigators to understand the more complex characteristics of an attack, such as the vulnerability opportunity. We discussed in detail the implications of this task and staging the project and how investigators will use this to ameliorate their existing practice. In particular, how this can be used to visualise a large number of sentences from different reports that may use slightly different writing styles but fundamentally discuss the same incident, and therefore the connections within a topic are visible.

8.4 Conclusion

This section demonstrates that it is possible to combine sentences from different reports grouped by topic and then visualise each sentence to obtain key details about an insider threat characteristic. Also shown is the ability to gain important information from different topics and also the value of the smaller machine-generated topics. These topics are more specific but contain key details that an insider threat characteristic may not fully represent. In addition, recommendations for use by practitioners has also been given with an explanation of how this can be used in the future. Finally, potential approaches for improving this approach were suggested. Although out of scope for this project, improvements could be made in merging these sentences, specifically, using an insider threat word vector representation, which could be used to find similar words from different sentences but requires a large amount of training data from insider threat organic narratives.

This stage forms the final part of the project, building upon the earlier work. Firstly insider threat characteristics are mapped to a topic model, automatically assigning insider threat characteristics to text. Then organise these using Markov chains to find relationships where one topic likely follows another. This gives an overall view of an insider threat attack. Next, from these Markov chains, the connections between both are analysed to find temporal or causative links between topics to highlight those machine-generated topics likely to contain pertinent information for an insider threat investigation. This stage of the project builds upon this, allowing an investigator to look at highlighted topics and understand the text. For example, examining individual sentences and linking all sentences with a topic together can discover important details. In addition, by examining different topics, it is possible to get additional details such as specific technical approaches the insider took rather than discussions of the entire incident, which may overlook specific details.

This work is of particular importance to understand organic narratives that may be numerous, differ in levels of detail, expressed using different language formality, and use different writing styles. By their nature, organic narratives are difficult to analyse and

understand as they are personal to the writer. However, this approach shows that it is possible to combine many organic narratives to understand an insider threat characteristic fully. Therefore, this work sits alongside existing practice and provides a method of quickly understanding and insight into an incident to allow an investigator to find important information about an attack without reading larger amounts of text, where some details may be easily skipped or ignored.

This concludes the technical aspects of this project; the next section will therefore discuss the uses in practice and how this system can be used as a whole to understand a particular incident—providing useful information and highlighting some information that may be easily skipped or ignored simply by accident, or by an overwhelming amount of details.

Chapter 9

Discussion

While these three capabilities have been discussed in technical detail, much like insider threat, it is essential to take a holistic view of the work focusing on how it may provide the tools and techniques to understand insider threat reports. Furthermore, this section explores and advises on the implementation of these tools to the work of practitioners. In addition, this chapter considers the broader implications on insider threat and the future of insider threat research, particularly for understanding insider threat as a necessary step to manage the insider threat, particularly as a stage before predicting future threats can occur. Finally, this section discusses the system's limitations and where potential flaws may lie, the ethical implications of this system if it were in use, and how these issues may be mitigated and improved.

The previous sections present each stage of the project and the results from the technical objectives; this section will discuss how these complement each other, providing an understanding of the attack from the investigators perspective. By examining the case from the experiment but from the investigator's point of view, this section presents how an insider threat incident could be examined. This section demonstrates the practical application of the work and how this research could provide a tool to aid in investigating insider threat attacks and the impact these capabilities have. First, the overall case will be examined from a top-down perspective, exploring codes of particular interest. Then these

codes will be examined for their casual and temporal relationships to each other to highlight different topics of interest. Then each topic can be considered atomically, providing individual and specific details such as exact methodologies and motives. Finally, showing how these core details of an attack can be found and used to build up an overview of the attack to an investigator.

9.0.1 Full System

To this point, the system has been considered separate elements, each building on top of the last. However, these elements are not independent, and it is vital to consider the system as a whole and precisely how the system meets each objective and the overall aim of the research. The individual work packages: the attack language, causal, temporal and narrative chains, and finally, the topic analysis can be used to answer critical questions about an insider threat incident when used together. This section will reconsider the aim and objectives of the project, reflecting on the system's technical objectives and the overall aim, demonstrating how the system can be used in an insider threat investigation the implications of the system on the broader insider threat domain. Finally, this section will reflect on and critique the work and suggest improvements that could be made or future work that could be completed.

The attack language was the first phase of the research; this provides the initial segmentation of the reports by topic rather than by document. The output of this stage of the system is to create and place the text in a structure able to organise sentences within documents by document or topic. The attack language uses a similar process to grounded theory in the social sciences. While grounded theory uses humans to discover and assign topics, the attack language uses computational topic modelling to do this by machine. The topic model was trained on a general insider threat incidents corpus gathered from publicly available news articles. Public news articles were chosen because these incidents are diverse in archetype, insider, actor, organisation type, asset targeted and other characteristics but share a similar writing style. A shared writing style ensures that the

normalisation process is effective, as these documents share similar words, phrasing and punctuation. The topics are then labelled based on an existing insider threat model; this is done to ground the automatic model to an existing formal model. The attack language output creates the first stage in an insider threat model, where characteristics of attacks are discovered and noted. However, in both topic modelling and the creation of formal frameworks, topics and characteristics are without further context and atomic. To be effective, these atomic characteristics need to be connected, becoming a custom insider threat framework. This connection is completed by the following technical objective of the system, causality and temporality.

To connect these individual characteristics: casual, temporal and narrative links are found, and the relationships between characteristics are visualised. First, the narrative links provide the base for the more context-rich links. The narrative links use Markov chains by representing each topic as a state and assuming a document can be represented as a series of topics, where the current topic changes as the narrative evolves. There may be many different writing styles, such as the use of shorter sentences, bullet points or paragraphs, but there is some logical narrative that topics form one topic following another. These initial links are then refined by examining the context surrounding these narrative links by identifying causal and temporal patterns and verbs. Each link can then be given additional context, and a timeline of events and casual chains can be built. This process produces several directed graphs where the topics created during the attack language stage can be contextualised with others, creating the custom insider threat framework. Although this framework is dynamic, it is still limited as it provides an overview of an attack, but the detail is lost. For example, this connection process could connect the methodology an insider used and their historical behaviour before the attack, but it could not be used to understand what exact behaviour the insider exhibited or what exact methodology they used; this is important as this information allows organisations to make organisational changes in the future.

The topic analysis technical objective attempts to remedy this limitation, combining

all sentences within a topic and visualising the content. This process also creates a directed graph for each topic, highlighting these critical features for each topic. These graphs vary in size depending on the number and complexity of topics. The core advantage of this visualisation method is the vast array of algorithms that can be applied to a mathematical graph to visualise the key concepts or understand the underlying structure. While the causality and temporality chapters look at the 'big picture', it is vital to see the specifics of an incident. For example, if an investigator wished to know which assets were targetted and how they may wish to examine the Attack or Attack Step characteristics. This feature can be vital for understanding how security controls may have been bypassed and what security controls may be required in the future. Therefore, by combining this approach into the system, not only can a custom, dynamic insider threat framework be created, but this can be used to create actionable intelligence.

9.0.2 Aims and Objectives

During the first chapters, this research's overall aim and objectives were set, developed, and discussed. The overall aim of this project was to create capabilities that could aid in the investigation of insider threat cases. While this does not 'solve' the issues of insider threat, offering detection or prevention, this work attempts to better understand insider threat. The literature has shown that this increased understanding can lead to significant progress in the detection capabilities (Costa et al. 2014; Probst et al. 2010; Sanzgiri and Dasgupta 2016). Using reports, which are often already written after an incident, an incident can be visualised, using details from each report.

First, insider threat characteristics must be extracted from the reports; the Attack Language chapter met this objective. Second, these characteristics should then be contextually understood, creating a custom insider threat framework; this objective was met by the Causality and Temporality chapter. Finally, the specifics of each characteristic can be examined so the fine detail would not be lost by the 'big picture' custom framework; the topic analysis chapter met this. However, the final objective, to use these individual

tools together and show how they may be used to investigate an insider threat attack, has not yet been discussed. Although the technical objectives have been met, it is essential to consider this final objective for the aim to have been met sufficiently.

To meet the final objective and, therefore, the aim of this work, the entire system must be considered. In particular, how this system could be used in an investigation and its place in an investigatory digital forensics framework, and in critical tasks such as creating a timeline, mapping out an incident, finding specific characteristics, analysing these characteristics, and make organisational changes. Showing how the system can provide answers to these tasks and questions demonstrates how the aim and objectives have been met, the implications to insider threat research, and where this system may fit in the domain. The following section will discuss these goals and how these essential tasks can be completed.

9.1 Using the System

9.1.1 Creating Custom Corpora and Models

During the data-gathering stage of this research, two corpora were gathered, a general corpus of insider threat incidents, which were gathered from news articles, and a sample corpus of organic narratives, collected during the perspectives experiment. The first corpus provides a report of many different cases of insider threat written in similar writing styles. Therefore, the documents have characteristics that any attack could have when considered together. This corpus was used to train the topic model during the attack language technical objective. The second corpus provides an example of the input of organic narratives and is used throughout the technical objectives to map our understanding of insider threat to the topic model. In this research, a sample topic model was created using a grid search method to choose the model's parameters, such as the number of stopwords, the number of documents used to train the model, and the number of topics. However, these parameters can be tuned, allowing a model to be adapted to an organisati. This

adaptation may be beneficial for an organisation that faces more niche attacks, or older cases could be removed and new ones added to improve the model enabling a focus on newer methodologies, motivations, and particularly newer technology. In addition, this process allows an organisation to customise the model using a custom corpus and model either by adding additional documents to the insider threat model or by including more labelled documents to score the model. The initial corpus was created from news articles; this is particularly advantageous as these use very similar writing styles. This model is not static, and if an organisation did want to implement this initial research, this corpus could be modified and adapted to their needs. For example, amending or creating a new corpus of more tailored insider threat documents.

It is possible to optimise the topic model further using the techniques discussed in the attack language section; this may be useful for organisations to ensure a model is tailored for their organisation. A corpus of organic narratives must be collected and labelled according to a model. By developing the general methods, this process will allow for the use of other models such as the CERT archetypes (Cappelli, Moore and Trzeciak 2015) if an organisation prefers it. In this research, a simple grid search optimisation method was used to find the most effective model parameters to use. A simple grid search was chosen for this initial proof of concept research as it was simple, and the parameter choices were limited. If this work is expanded, this research could be applied to other models rather than limited to insider threat. Another formal model could be selected even in a different domain. The steps outlined in the previous section could be followed; simply changing the corpora to the new domain would be possible. These play a significant role in how an incident is mapped out and understood, particularly in identifying relevant topics by characteristic.

9.1.2 Map Out an Incident

When using the topic model, the most critical decision to make as an investigator is which topics are the most important. This choice can be made in several ways: choosing a

topic with many sentences or links to other topics, choosing topics with high causality or temporality, finding chains of topics, and by insider threat characteristic. Choosing a topic with many sentences or links to other topics can be a good choice to begin; these topics usually involve many aspects of the insider threat case and closely align with those referring to the attack or attack steps. Identifying those that may have many connections can show those topics that are interconnected to others, showing concepts that appear in almost every document. These can be seen in the narrative links in topics with many connections. These topics are primarily related to attack characteristics; however, this could change depending on the respondents. The following method to identify relevant characteristics is the additional contextual information, using causality and temporality clues, either to find those which have a large number of causal sentences or causal chains. Causal and temporal chains can be used to build a timeline or perform root cause analysis. Some topics can be ignored; these may be repeated or belong to a parent code. However, when viewing the narrative graph, merged by code, emerging topics can be recognised; these may be codes with not enough examples to be labelled; however, others may be emerging or unique characteristics that should be investigated.

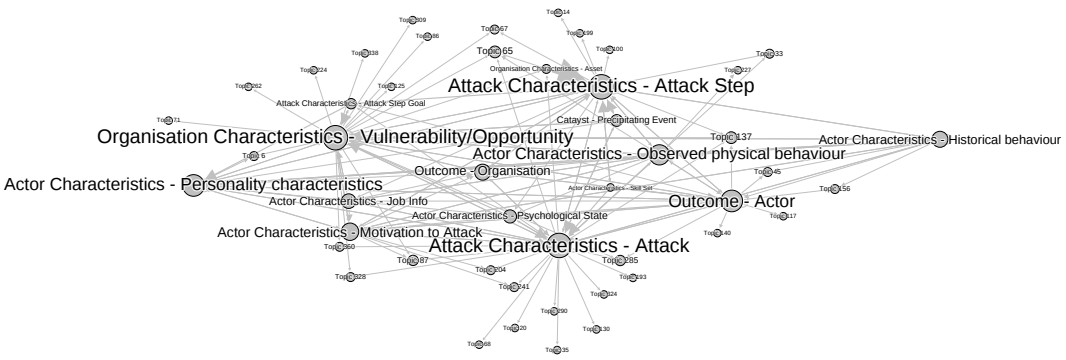


Figure 9.1: Narrative model of the data with the individual topics mapped to codes

When considering the case from the perspectives experiment, the emerging topics are particularly noticeable. Some characteristics are apparent as mislabelled; for example, in Figure 9.1 Topic 20 is close to Attack Characteristics - Attack, which contains one sentence ‘She also had a ring of nine accomplishes who have not yet been charged as

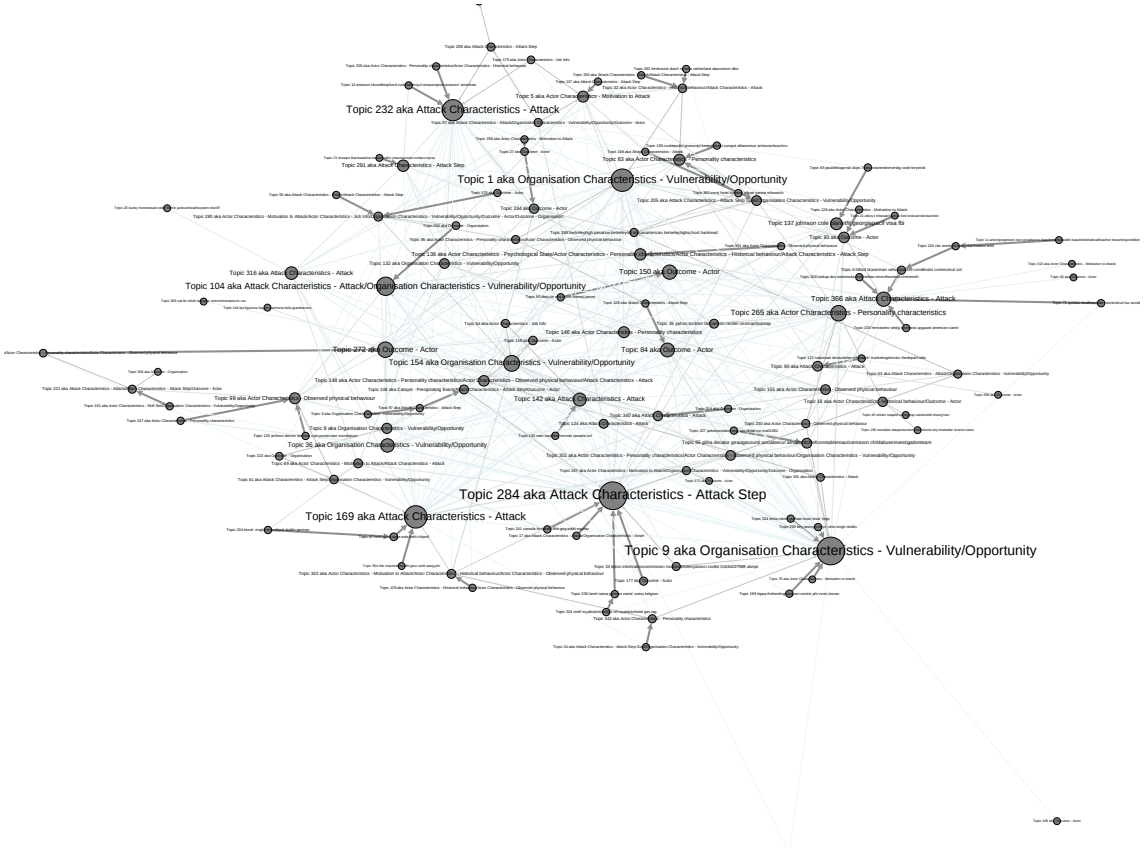


Figure 9.2: Narrative model of the topics using Markov chains

their exact roles in this are still being investigated', which could be labelled as 'Attack' or 'Outcome Actor'. However, others may suggest alternative characteristics which the model has not accounted for; in particular, many of the outlier topics discuss the opinions of those around the insiders, Topic 6 discusses the other employees' suspicions and how it was managed, 'No one of the coworkers believed because she never seemed to be suspicious, only an IT manager did saw something' and 'She was caught and no one could believe it to be true, especially since she was so generous with everyone'. These topics are critical for later decision-making, as recognising how employees react to an insider threat can be used in future training.

Topic 6 also begins a temporal chain shown in Figure 9.3, from Topic 6 with no temporality to Topic 366, to a larger topic, Topic 284. This chain develops from the initial description of how colleagues may have reacted to an incident, to Attack Characteristics - Attack to Attack Characteristics - Attack Step, which can then develop into

Actor Characteristics - Job Info (Topic 150), Organisational Characteristics - Vulnerability/Opportunity (Topic 1) or Actor Characteristics - Motivation (Topic 69). This single chain can describe a timeline of an event. Other chains exist in the causality graph shown in Figure 9.4, Topic 36 begins a chain that is weekly causal to, Topic 125 a highly causal topic, to Topic 99, ending at Topic 3, which has a small amount of causality. When comparing this to the codes, Topic 125 is highly causal but is an emerging topic; in this case, this topic may warrant further investigation. The more prominent topics of Topic 284 (Attack Step), 1 (Vulnerability/Opportunity), 232 (Attack), 169 (Attack), and 9 (Vulnerability/Opportunity), initially Topic 1 and Topic 169 may be disregarded, as they both show the attack or vulnerability and are smaller. This process allows the analysis to refine the number of topics, only analysing those that meet some requirements, such as highly casual those than begin chains. These initial topics have already shown a high-level overview of an attack, however questions such as 'How related are these two topics?', 'What connects these features?' more precisely 'What was the insiders' motivation?', 'And how did they attack?' are then answered by considering the language within the topics.

9.1.3 Idenitfy a Characteristic and Analyse Specifics

Once particular topics have been highlighted in the previous task, each individual topic can be analysed. Providing the missing additional context when compared to the basic labels from the insider threat framework. The investigator can use these insights to begin the process of policy implementation. This step allows an investigator to fully understand the big picture and explore key technical details, such as the exact steps taken to attack an asset or motivation. These details can sometimes be lost when frameworks are applied, with a focus on the overview view of an attack rather than the precise details. However, it is challenging to make exact organisational changes without these precise details. Particularly in ensuring that any organisational changes target the particular motivation or assets that an insider targeted.

From the topics identified previously, the more significant topics - those relating to the

attack, this is a wise starting point as these are likely the topics with the most sentences and should tell the story of the entire attack. Topic 232 in Figure 9.5 is an example of both a large topic and a topic that primarily discusses the attack. Here it is clear that the insider was a well-liked manager with a good track record with management and direct reports. During her attack, she used this reputation to steal $60 million, defrauding the company for over 18 years. The methodology is discussed briefly, with the insider writing fraudulent checks with access to a system. The insider was fined after the investigation and potentially received a prison sentence. Potentially the insider also recruited others with a discussion of accomplices.

Topic 9 in Figure 9.6 is another prominent topic; however, this one concerns the vulnerability and opportunity that the insider abused during their attack. Here, the system discussed previously is discussed in more detail. The insider became exempt from a new system; she put the case forward with support from senior management and discussed some issues with the new system. This exemption allowed the insider to use a paper-based system, where she was able to hide some discrepancies within her department, concealing the fraud. Part of the exemption was likely due to her connections and the fact she was well-liked in her department. A potential control to manage this risk is also visible in this topic, to ensure that everyone is transferred to a new system and no one is exempt, despite the rapport they have with upper management.

Another control could be to understand how the insider's personality played a role in the attack; Topic 265 is another significant topic, closely connected to Topic 366 identified above. When examining this topic, it is clear that part of her ability to manipulate others was her generosity, described as kind and generous by her colleague. The news of her fraud came as a shock. It is clear that an insider can be highly manipulative, but some behaviour such as generosity may still be visible; a potential remedy could be to investigate employees who seem to be spending a lot of money. However, the motivation for this particular insider threat attack is still unclear. So one solution could be to identify topics that have the motivation code, such as Topic 322, 5, 157, 280, 78, 128, 313; of

these, Topic 5 and 322 contain the most sentences.

Although the motivation is unclear, many reports discuss financial gain as a primary motivation, with potentially some financial issues stemming from a gambling addiction. Given the lack of apparent motivation, it seems likely with the insider's generous nature, some references to gambling, and a report which references financial gain. It may be helpful to approach this as the insider's historical behaviour Topic 136, 300, 32, 322, 129, 18, the largest of which are 136 and 18.

Topic 18 give further details on the insider's behaviour. The insider avoided suspicion by claiming that she had received a large inheritance. This creates a clearer picture of the attack; she was not caught sooner as she gained the trust of both her colleagues and upper management. She used this trust to continue using an older system, despite a move to a new system, which would presumably have caught her.

9.1.4 Make Organisational Changes

From this analysis, it is clear that the insider was first primarily motivated by money, potentially due to gambling issues, and she was able to hide the attack using her personality, in particular her social skills, which gave her a good reputation with both her colleague and upper management. In particular, with management, she convinced them to make her team exempt. The insider used the paper-based system to write fraudulent cheques until they were caught and fined for their actions. This case is similar to other insider fraud cases; the insider could continue their fraud over a long period, bypassing restrictions with their employer's trust. With this knowledge, the mitigations suggested by the CERT Insider Fraud model (Cappelli, Moore and Trzeciak 2015) can be examined.

From this model, it is clear that Fraud Prevention Controls were not implemented correctly, therefore contributing to the opportunity to commit fraud. Although it is unclear what the rationalisation was in this particular case, their generous nature could suggest that the insider intended to make things right. This could have contributed to social networking pressures and the insider's financial problems to become an incentive to continue

the fraud. It is possible to use this model to better understand the potential policy decisions that can be made, as the authors suggest mitigation approaches for insider fraud. These are summarised in Figure 9.7. The three most common organisational controls are inadequate auditing, susceptibility to recruitment, and verification of modifications. The inadequate auditing was implemented by the organisation, however the insider's ability to recruit upper management, unknowingly, into the scheme. In addition, the insider's colleagues did notice this behaviour but never reported it. The insider should never have been exempt, and this issue has been remedied. Additional controls could be implemented to train and encourage suspicious contact, specifically including a confidential method to report suspicious behaviour without repercussions to the reporter. In addition, the insider being aware there are security controls can reduce the amount of insider fraud, as they are more aware of the risk of being caught.

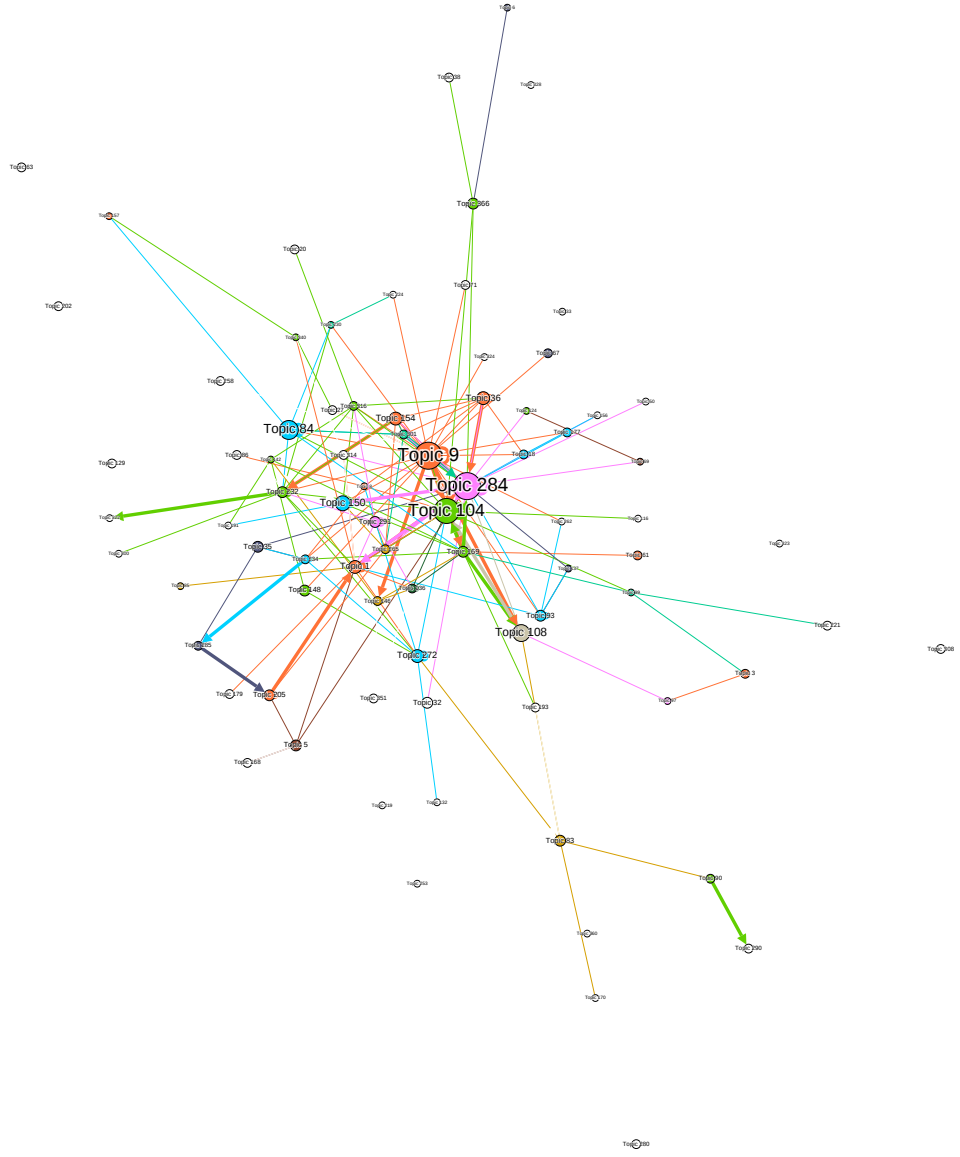


Figure 9.3: The temporal layer coloured by code

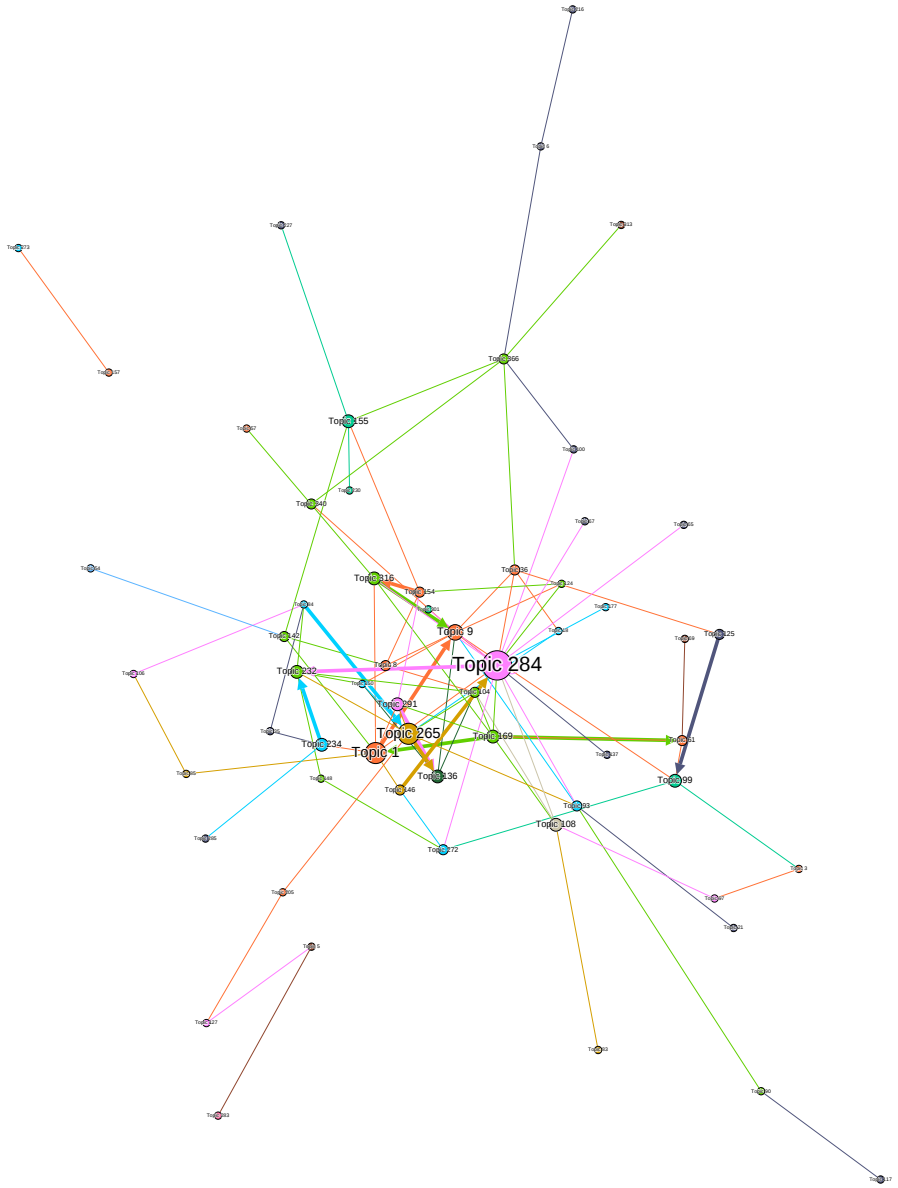


Figure 9.4: The causal layer coloured by code

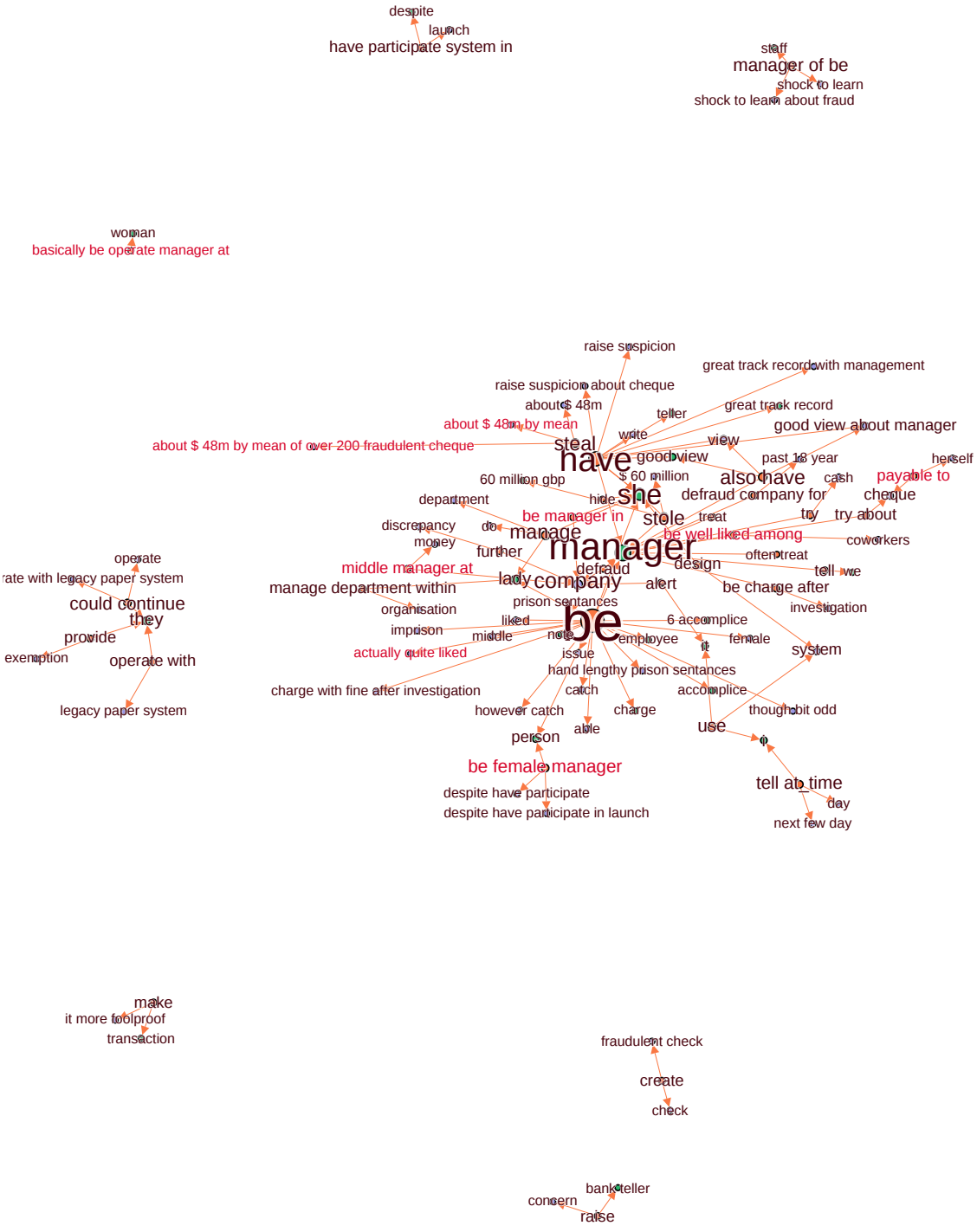


Figure 9.5: Topic 232

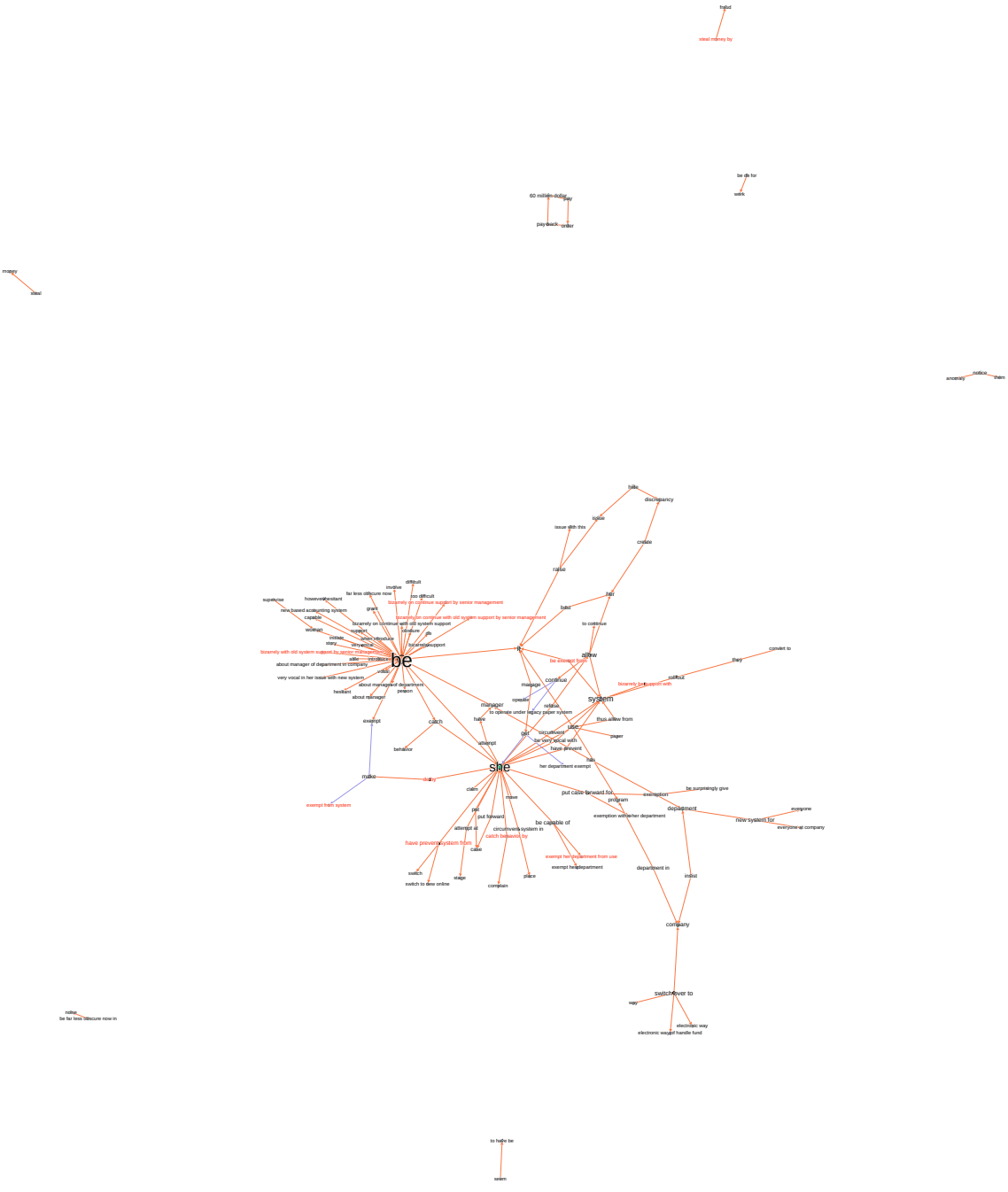


Figure 9.6: Topic 9

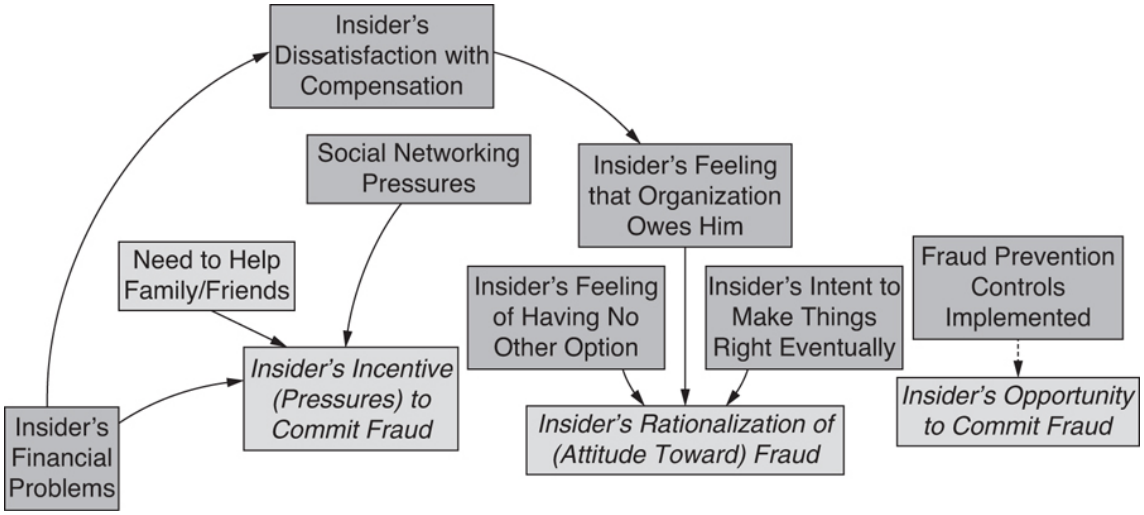


Figure 9.7: CERT Model for Insider Fraud (Cappelli, Moore and Trzeciak 2015)

9.2 Implication for Insider Threat

The research in this thesis has presented a new and novel approach to understanding insider threat, using natural language processing to analyse organic narrative reports within the context of insider threat. Ultimately this research is the initial steps to a tool able to create a custom insider threat framework, which could adapt to changing business requirements built upon an existing framework. In this work, several novel techniques have been developed both in NLP and the insider threat domain. The technical objectives represent these and have been met during the project, supporting the overall aim set out at the start of the thesis. The specific technical objectives identified were; to collect two corpora, collections of documents, a general corpus of insider threat attacks, and a corpus of organic narratives to map these to an existing insider threat framework. This mapping then places machine-generated topics in context, becoming a custom insider threat framework by creating connections between each characteristic using causal and temporal clues, and finally, to analyse each individual topic to allow for the understanding of the details. These technical objectives have also included novel techniques, particularly in the automatic labelling of documents based on minority topics and the mapping of organic narratives to an existing formal model. This work is built upon existing models for a general understanding of insider threat, specifically Nurse et al. (2014b) However, this work addresses a number of the issues identified in the literature gap, existing models aim to provide a method of understanding insider threat, they do have significant disadvantages, during the literature gap these were identified and six-core issues were found.

First, models require a large amount of domain knowledge to be used effectively and are designed for experts in insider threat; during the literature review, this was identified as a major issue preventing the insights of these models from being deployed. Many security professionals are not experts in insider threat or the psychology and behaviour of an insider. The second was the issue of confirmation bias; these models are based on a range of data; in the creation of many models, this is often technical data. Technical data can be easier to generate and process, which leads to its abundance; however, this

can limit a model to just focusing on the technical aspects of an attack. During this work, this focus was shifted to a 360-degree view, using reports from many witnesses rather than just technical members of staff. This 360-degree view allows this work to capture behaviour that may have been limited to an insider's colleagues or line manager, such as a disregard for rules, which has been shown to precede insider threat activity. These two issues both resort from the static nature of models, and particularly the data used to create them. The data ingested into a static model cannot be changed, and therefore the models may not be able to adapt to new attacks, techniques, and insiders. The third issue identified was this static nature; models attempt to combat this by creating general models developed with data over multiple years, allowing an organisation to place new attacks under general characteristics. The use of NLP, by comparison, allows for a dynamic and changing model, particularly adapting to attacks an organisation may face by adding additional documents to the topic model. These additional documents could be newer articles of insider threat attacks, creating a model that can adapt to new methodologies and approaches or reports from an organisation or tailoring the model for the specific attacks an organisation is likely to face; this adaptation is incredibly important as the security landscape changes over time, particularly the shift to working remotely. The fourth issue is the difficulty in creating actionable intelligence, models often focus on understanding an insider threat attack, but due to their high knowledge requirements, this understanding is not necessarily passed to an end-user to take actions, for example, policy changes. The technical solutions, on the other hand, can offer this actionable intelligence, this flags suspicious activity and describes it, e.g. high risk or unusual file access, allowing an end-user to flag an individual. This research attempts to bridge this gap by providing an intuitive model and visualisation. So understanding of an attack can be turned into action, for example, recognising the technical approach used and limiting access.

This then supports the fifth issue, the difficulty in using a model to understand an attack strategically and tactically, or in other words, viewing an attack at a higher level, and viewing the details of an attack. With traditional models, the details are often lost

to more generalised categories; this can make it challenging to act on this information, particularly in creating policy decisions and supporting management chains. The research presented in this thesis shows the overall aspects of an attack and the finer details, with the causal and temporal graphs showing the overall connections of elements within an attack and the topic analysis showing the details of a single topic. While models are limited to supporting strategic, long-term plans, they do not necessarily support tactical decisions, as these require finer levels of detail. Instead, tactical decisions are left to technical solutions such as deployable behavioural monitoring, which may cause ethical concerns, particularly in regards to automated and supportive decision-making.

The literature gap identified was the concept of tool support and solutions; while technical approaches to insider threat promise a solution, which can effectively manage insider threat, models instead approach the issue as supporting existing decision making. While this is early work and not yet a full implementable piece of software, this approach offers an alternative to technical solutions, particularly when considering that insider threat is often managed at various levels rather than purely technical. Although a technical solution is usually preferred, with many organisations deploying technical monitoring, solutions often introduce a range of ethical and moral implications regarding workplace surveillance, the pressure to report behaviour, and the impact on morale in the workplace. In particular, this additional stress may cause more insider threat attacks, increasing the risk further. This work seeks to address this by focusing on supporting existing decision-makers. Particularly in ensuring that insider threat mitigations do not place additional stress, by allowing people to report in as much detail as they would like, without prompting for specific details or constraining their reports, this will enable reports to be collected without introducing additional stress or cognitive load, if an individual is unwilling or unable to remember. For this to be deployable, a range of additional software would have to be created. Although some initial software was written due to time constraints and the novel approach, this work primarily approaches this issue as a proof of concept rather than full software.

The research presented in this thesis has, therefore, met the initial aims and objectives and used these new techniques developed to address some of the current literature gaps within the insider threat domain. By leveraging natural language processing to process a large number of organic narrative reports, a complex insider threat attack can be visualised easily within the context of an existing insider threat framework. In particular, this research has reduced the large amount of domain knowledge required to use insider threat models, the confirmation bias that models may introduce, their static nature, and difficulty in using them to generate actionable intelligence, particularly in supporting both strategic and tactical decisions and the creation of a toolset rather than attempting to solve insider threat, e.g., detecting an insider.

However, this new approach has produced several implications for the domain of insider threat rather than this initial literature gap, the use of a toolset for investigations, and how this can lead to insider threat solutions and particularly how this can be used to generate actionable intelligence, how this work can support the ethics of insider threat. The implications of this work are not limited to the domain of insider threat; the creation of NLP techniques for data gathering and mapping text to an existing formal model presents a range of exciting and useful insights for many domains. This section will discuss these in detail and demonstrate how the initial research gap, technical objectives, and tools created to support a wider impact in the insider threat and other domains.

9.2.1 Investigation Tool

The previous section discussed how the suite of NLP tools could be used during an investigation to answer practical questions regarding an attack; however, the implication of the tools used in investigations has not yet been discussed. Insider threat attacks are complicated to investigate, particularly as many of the tools for insider threat focus on deployable software and struggle with explaining the decisions to potential investigators. This work has sought to address this, giving investigators the tools to understand an attack. However, as noted by Nurse et al. (2014a), understanding insider threat better, particularly as

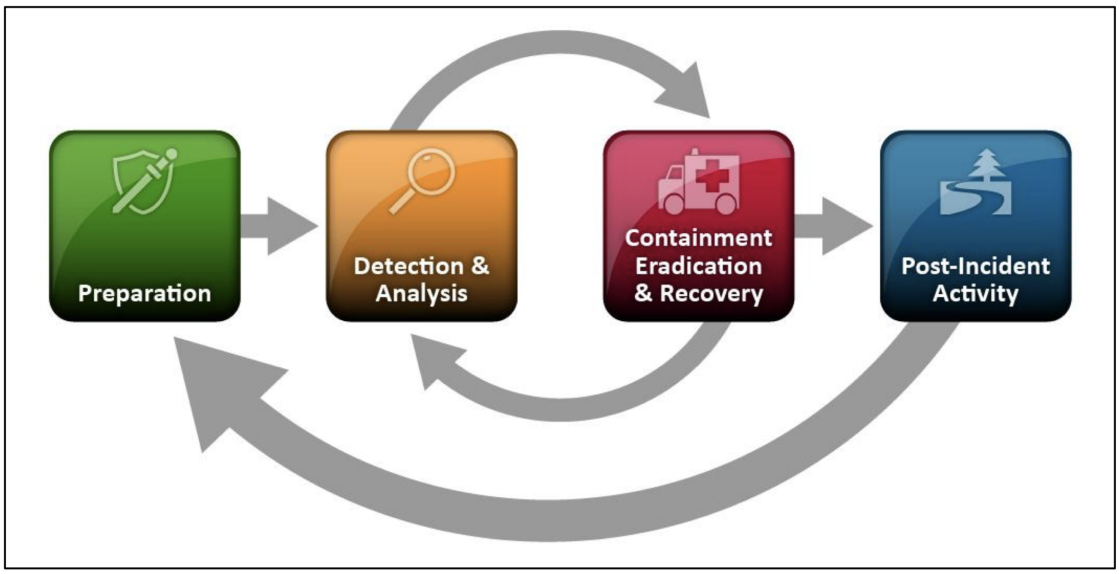


Figure 9.8: NIST Incident Response Framework (Cichonski et al. 2012)

the world of work changes, is the first step in improving detection and management capabilities. One of the key advantages of the approach developed is that this work fits into existing business processes and can also sit alongside existing technical solutions; this allows this research to be effective without interrupting existing processes in an organisation. In addition, by offering a technical solution, more cases can be analysed, understood, and reporting can be increased. In the previous section, the research gap that this work targeted, this section discusses the impact on the broader domain of insider threat and how this research can be used practically to solve some outstanding issues in the field.

This research fits well into existing digital forensic practice, but particularly incident response. For practitioners, the techniques developed during this thesis, implemented into full software, would not greatly impact existing processes. Support existing processes in incident response already being used by investigators. Particularly this work supports the preparation and post-incident activity of the NIST incident response lifecycle (Cichonski et al. 2012) Shown in Figure 9.8. While the tools primarily target the post-incident activity, for example, in gathering and archiving evidence and reports after an incident has taken place, the impact of the research focuses on the preparation stage. During the preparation stage, policies, training, internal processes etc., are developed; although this

is often considered a first stage, the NIST life cycle shows that these are closely linked, and by targeting the post-incident activity, the preparation stage can be directly informed. Many insider threat tools specifically target the detection and analysis stage; when compared with external incidents, insider threat is much more difficult to detect. In some cases, insider threats can go unnoticed for many years, making it challenging to contain, eradicate and recover. However, the individual life cycles of incident response do not occur in a vacuum, and improving capabilities in all areas will impact the ability to improve response capabilities.

The first step in improving detection capabilities is a greater understanding of insider threat. Although detection is the clear end goal of many pieces of insider threat research, an increased understanding is key for ensuring that these detection capabilities can be improved. Even with solutions that can be deployed on a network, detection is an extremely difficult problem. By focusing on increased understanding of the new types of insider threat, such as those identified by Cappelli, Moore and Trzeciak (2015), as workplaces and technology at work change, insider threat will change with it. By offering, in particular, a technical solution, more reports can be read and analysed, and reports from individuals that would not have been considered but may contain potentially valuable information can be quickly analysed. This new approach allows for an increase in data collection, but also of data utilisation. With a focus on custom frameworks and adapting models, the techniques developed during this piece of research can adapt over time with minimal work. Although this is not a technical solution for detection, the technical nature of this approach, particularly with creating custom insider threat models, is key, focusing on understanding many cases. As the types of insider threat change, this work can allow detection capabilities to change with it.

The preference of security teams to focus on detection techniques, especially within the context of incident response, is the difficulty with actionable intelligence. Detection is often the only option that offers actionable, relevant intelligence within an organisational context. However, the research shows that often the risk of insider threat attacks can be

mitigated by introducing organisational measures, for example, offering employee assistance programs, as noted by Cappelli, Moore and Trzeciak (2015). Insider threat frameworks do provide this to a degree, they can still be difficult to approach for non-experts. While technical approaches detect anomalous activity, the lack of explainable decision making creates a challenge for generating actionable intelligence. One of the impacts this work generates is this focus on actionable intelligence to create policy changes. Although there are limitations to this initial work, this focus on policy and actionable intelligence is clearly a much-needed step for insider threat research.

9.2.2 Implications for Applied NLP

In this thesis, the research presented has been focused on specifically insider threat; however, the techniques developed can impact other domains, particularly those with a focus on report writing. By exploring these techniques, any organic narrative report can be processed and understood in more formal structured models. This is particularly advantageous as many formal models are challenging to implement, and report writing is time-consuming. In addition, many of the data gathering approaches have been developed for insider treat; however, as noted in the previous chapters, the techniques could be used in any domain where minority topics should be used to categorise documents. The fields that would benefit from this work is the wider cybersecurity field, where technical reports are widespread, but where often policymakers are not necessarily cybersecurity experts. However, this is not just limited to cybersecurity but also any domain where text can be mapped into formal models. NLP is currently used in many fields to aid in the understanding of reports.

Topic modelling is used for many pure NLP applications, such as Grefenstette and Muchemi (2015), but is also used as a tool for understanding large bodies of text in the social sciences. Using topic modelling for this causes issues with interpretation and the difficulty in understanding how a machine may be making decisions. Some approaches have been developed to combat this, usually by considering additional information, for

example, time, to demonstrate how topics may change over time. Other systems have included labelling topics individually; this is usually performed by a human, with irrelevant topics being discarded. This could be improved by using the techniques developed during this research, specifically, the mapping to an existing structured model. This capability was created during the attack language technical objective, where the topics are labelled based on the code from an existing insider threat model. This interpretation method could be highly impactful to those in the social sciences who would like to use topic modelling but struggle to interpret these results. Although presently, this does take manual human labelled examples, this represents a much smaller load than human-led effort than techniques such as grounded theory; future work could expand on the automated decisions.

The research uses organic narratives rather than formal reports and demonstrates the possibility of event reconstruction despite using these more informal reports. These reports may present fewer ethical concerns than traditional reporting, as frameworks may introduce bias and may pressure respondents to reveal information. The approaches developed for data gathering could be expanded. At the same time, the mapping allows organic narratives and reports to be better understood, the data expansion approach suggested the use of fuzzy topics. Fuzzy topics are topics where a proportionally very small topic can fundamentally change the context of a document. This becomes more challenging when many insider threat cases do not share similar methodologies, motivations, and organisations; rather, the shared aspect is the actor who committed them. For many articles, this represents a much smaller portion of text than more dominating topics such as the motivation or methodology. These combined present useful data gathering opportunities, particularly in these cases where discriminative topic models may not work. The data gathering approaches developed allow more data to be gathered, but importantly, allow for more data to be understood, particularly in existing contexts such as social models. For researchers in these areas, these capabilities represent significant opportunities for data gathering across many fields.

9.2.3 Ethics

The insider threat domain has many ethical concerns due to its nature, where employees are the perpetrator; the insider has many connections across an organisation. Much of the work regarding the ethics of insider threat focus on behavioural monitoring (Palm 2009; Greitzer, Frincke and Zabriskie 2011), and being asked to report others. These issues not only make an incident challenging to investigate but also challenging to manage internally at an organisation. There are also widening concerns about the ethics of machine learning, particularly on automated decision making, and this too extends to natural language processing. Issues such as accuracy, data gathering, bias, and a worry regarding lack of human invention are strong themes in this space (De Vault, Oved and Stone 2006; Bolukbasi et al. 2016). This work has considered these ethical concerns during developed and aimed to reduce ethical concerns and support organisations to ensure policies can be introduced without increasing negative attitudes towards work and straining the relationship between employees and their employer. The most significant ethical concerns in the domain of insider threat are the use of behavioural monitoring tools and the desire to punish potential insiders, potentially before any activity has occurred Greitzer, Frincke and Zabriskie 2011. These are designed to monitor employees and flag behaviour which could be insider threat activity, usually using technical approaches. These encourage employers to become reactive, reacting to these behavioural signs; these reactions may involve disciplinary actions against an employee. However, these behavioural monitoring techniques often strain the relationship between organisations and their employees (Palm 2009; Greitzer, Frincke and Zabriskie 2011), which can increase the risk of both malicious and unintentional insider threat Greitzer et al. 2014; Cappelli, Moore and Trzeciak 2015. This is not the only approach to managing insider threat, and positive management at organisational levels is often used.

This work focuses on supporting this positive effort rather than the use of behavioural monitoring. Behavioural monitoring is often supplied as a software package, usually preferred due to its ability to learn over time with modern machine learning methods

and integrate the process into existing practices. The use of natural language processing allows for similar advantages to manage large amounts of textual data, which can adapt with an organisation. However, this does bring its own drawbacks, primarily in asking people to report their colleagues, Forte (2019) suggests this may be due to lack of training on these behaviours and what to report. To mitigate these ethical concerns, we use organic narratives through which individuals can decide on what and how they choose to share with an organisation. This is in contrast to insider threat models, which can pressure individuals to report information that they may not be comfortable disclosing. Using this approach, a large number of documents can be collected from a range of individuals in the hope that if a single individual chooses not to share a piece of information, another individual may choose to, and this can be captured. Therefore, the mapping of reports is hoped to allow insider threats to be investigated without asking witnesses to report others, which can impact the ability of an organisation to investigate attacks.

This does not imply that natural language processes have no ethical concerns. As discussed in the previous sections, there are several ethical concerns that NLP presents, and mitigations have been put into place to support this. These usually focus on two issues, how data is gathered and processed, but for an issue in insider threat, the main concern is accuracy. Data gathering and specifically how data is collected is a continuous issue in all forms of NLP (De Vault, Oved and Stone 2006; Bolukbasi et al. 2016), this is particularly true for security concerns. For this research, this issue, how data is gathered was considered during the data collection. Due to the sensitive nature of insider threat news articles were chosen for the general corpus of insider threat as these are public and do not require an organisation to collect or share privileged information. By using organic narrative reports, individuals can choose how to retell a story; this allows the individual to choose what to share with their employer; these are also used anonymously, collecting no data about a writer. The second issue regards data processing; this issue was discussed previously in the previous section; in order to ensure this work is grounded in the societal context, the model is mapped to an existing model. By framing the work in the post-

attack and preparedness stage of an incident, rather than detection, the additional ethical detection concerns are limited.

9.2.4 Conclusion and Impact

In conclusion, this work has addressed many issues from the literature and developed novel techniques to better understand insider threat attacks. The capabilities developed have had a range of impacts both for the field of insider threat, digital forensics, and wider computational social sciences research. The research has resulted in a practical toolset that could be used at an investigation phase rather than at a detection phase, allowing the impact to be placed primarily in the post-attack and preparation stages of the NIST Framework (Cichonski et al. 2012). Although this was the primary impact, improving the understanding of insider threat and how it is changing does inevitably lead to increased detection capabilities in the future. Although the tools were made for insider threat, by employing this methodology on training data for another domain, the research can be used for other fields. The research has been created with this consideration and is discussed further in Chapter 10 and how the impact discussed in this chapter can be realised in other fields. This chapter discussed other issues such as the ethics of the work and the mitigations introduced and how this allows the work to have a wide-reaching impact for insider threat; however, there are critiques and limitations. The next chapter discusses the limitations and critiques and discusses in depth how this work has attempted to mitigate some of these and how improvements can be made during future research.

9.3 Critique and Limitations

Although this research has presented several implications to insider threat research and provides a novel approach to understanding insider threat attacks, there are various limitations of this work and improvements that could be made to mitigate the limitations. This section will evaluate each stage of the research, first the attack language, causality and

temporality, and topic analysis, and then the system as a whole, focusing on the aim and objectives. Many of the limitations stem from this research providing a proof of concept and representing the first stages; future research could build upon these initial directions and work, building upon this platform. This section will critically evaluate the research presented in this thesis suggest mitigations and improvements which have or could be made. The next chapter offers future work to build upon this work.

9.3.1 Data Gathering

The first issue is regarding the general insider threat dataset used for training; this has a variety of insider threat attacks, written in very similar styles, downloaded from known insider threat news articles, and automatically categorised articles. During the automatic categorisation, various articles were miscategorised. However, some of these were removed by highlighting articles with a large number of irrelevant topics and manually checking these, and removing irrelevant documents. This was necessary due to the sparsity of the data; for some words, they only appear in a few documents, and although sparse terms are removed, this could impact the topic modelling and if these sparse terms are technical, insider threat terms that are being lost. Although this does impact the final topic model, it should not affect the conclusions, as the topic model will generate an out-of-domain topic; these are common in topic modelling; they can be disregarded when applied in domain.

During the data gathering, there was not an investigation into the insider threat archetypes represented, so it is unclear what the diversity of archetypes among specific cases may be. One limitation is that because this corpus was gathered from public data sources, it may be skewed to certain cases that are more likely to be reported. Equally, it is unlikely that insider threat as a domain has an equal class balance between each archetype, that is to say, that some insider threat archetypes may be more common than others, potentially due to the difficulty in each methodology. However, this is a limitation of any corpus created from public news sources, and there are undoubtedly many attacks that do not get

reported particularly to the press . One approach to resolve this could be to add additional documents to the corpus allowing an organisation to easily tailor the model to attacks they might have faced in the past. Alternatively this could represent an interesting piece of future work, comparing the types of attacks reported to the kinds of attack experienced by an organisation. This is a limitation of any research conducted on insider threat, and it is not clear on the exact rate of any archetype, though there has been some work completed in certain domains (Randazzo et al. 2005).

The next issue is that the general insider threat dataset was not continually updated; this was done so experiments could be repeated during the project. However, this does limit the final topic model, as it is not dynamically created. The mitigation chosen for this issue was the creation of a UI in the form of a website that can be used to train new models with new data. In addition, APIs were created independent of the dataset, allowing for new datasets to be created, trained, and used, adding additional data. Although these were not utilised during this project, it is feasible that further work could build upon this, creating an insider threat dataset to be used for future projects, or future work could use this research in a deployable product. This would allow practitioners to develop new models based on their own reports of insider threat or weight the model for newer insider threat attacks to react to new methodologies or motivations.

9.3.2 Attack Language

During the Attack Language stage of the project, several topic models were trained on the corpus of general insider threat corpus, and the best performing topic model was chosen as the final topic model. This final topic model was applied to the sentences from the perspectives experiment, creating a corpus of sentences segmented by topic. This allows the sentences to be organised by document in order or by topic so that sentences that share a topic can be grouped together. Although this works well, and many sentences are correctly marked as related, various issues could be addressed with further work. One drawback is the difficulty in representing topic meaning. This is an issue that has been

discussed in the topic model literature (Jacobi, Atteveldt and Welbers 2016) and addressed usually using manual labels for each topic.

This issue was addressed by grounding the topic model using an insider threat model, by asking human coders to apply an existing insider threat model, created using grounded theory (a similar technique to topic modelling, but performed by humans assigning codes to text). This created a dataset linking each sentence to a topic and a human-assigned code. The performance of the topic model can then be evaluated by considering how closely it matches the human assigned codes; this does not assume that a topic model will always create the same topics as the codes, but instead that sentences that appear in the same code should appear in the same topic. In order to decide and evaluate these parameters, a grid search optimisation was employed, and the final topic models were ranked, choosing the model with the highest score between the boundaries. This approach, however, was slow, and although a large number of models were scored and created, this was very time-consuming and limited the overall number of potential models that could be scored. There were also issues with the scoring with models with very few topics and models with a large number of topics causing outliers, although this was addressed by including boundaries.

The use of this dataset for scoring was then expanded to labelling the topic based on these codes. However, this can limit the representation of meaning, as the machine-based topic models are more specific. In contrast, the human codes are more general and emerging topics (which may not have a label) can be lost. One approach from the literature to interpret topic modelling has been automatically labelling topics; however, the technical difficulties in applying this, especially for a new domain, made this particular approach unfeasible for this project. Another approach considered was creating another labelled dataset with more sentences, to classify the topics using a large dataset of many sentences. These two approaches were not chosen, as labelling using human codes was agreed upon as a representation of meaning. This had the additional benefit of providing a topic hierarchy. When topics were visualised using the Markov Chain approach, it

became clear that some of the emerging topics were closely related to one or more codes, and this could allow these topics to be labelled. However, this was later realised to be not necessary.

The final issue of the attack language topic model was that some topics were over-represented in the dataset, some topics contain a small number of sentences, yet others have a large number of sentences. These are usually describing the general attack, as an overview, likely due to the scoring approaching discussed above, and the over-representation of the code 'Attack Characteristics - Attack' in the labelled data; this is likely due to the over-representation of this label in the data. Although these sentences are related, these large topics are varied and can produce confusing graphs at a later date. Although no solution was found to mitigate this issue, a solution may be to expand the total number of topics, as topics are often hierarchical, and consider these subtopics of the parent code 'Attack Characteristics - Attack'. This could be addressed by improving the scoring algorithm rewarding more granular topics. However, this was beyond the scope of this project. During the visualisation stage, analysing individual topics with many sentences creates large graphs but where key concepts can still be visualised presently.

9.3.3 Causality

The causality, temporality, and narrative, which builds upon the topic modelling was the next technical objective completed. Using Markov chains, a document can be represented as a state transition where a state is a specific topic. These Markov chains create a directed graph for an individual document or over the entire corpus. This Markov chain approach works well, especially for individual documents, where the topics are clear. This does get difficult to visualise effectively over an entire corpus primarily this is caused by the number of topics (over 300) compared to the number of codes in an insider threat model (21 codes). In particular, the causality and temporality are limited due to the domain and limited data on causality itself. Despite these minor issues, this stage of the project aimed to allow an investigator to highlight topics of interest, providing additional context and

links between topics, and as shown in the previous section, this is possible, and therefore this objective has been completed.

The initial issue with the causality and temporality was the original, narrative, Markov chains. These chains show how a document is written and how one topic leads to the next. This can be done by document or over the entire corpus. Although this works well, it is difficult to interpret, particularly over the entire corpus of organic narratives. This was helped by visualisation tools such as Gephi, allowing the graph to be manipulated to highlight more prominent topics and highlight chains. Another solution presented was aggregating topics with the same code; this reduces the number of topics and can be used to show relationships between codes. Other approaches that were considered include alternative aggregation techniques, highlighting important chains with many topics that follow each other, and alternative visualisation approaches. The solution decided was to highlight the key chains, and this needed additional context, notably the addition of causal and temporal information.

The causal and temporal layers are built on top of the initial narrative graph by identifying causal and temporal words and phrases in the text and describing the relationship (this topic must come after, before, during the next). This uses a collection of patterns and verbs designed for an open domain causal system. The key issue with these, in particular, is that they are not domain-specific. There are words that may be unique to insider threat attacks that may be causal, particularly when describing technical methodologies or assets targeted. The graph may, therefore, be missing some information, depending on the words used. Causality mining is an ongoing research area in NLP, particularly the training of models. Although this project used open domain verbs and patterns, this research is still an active area in the NLP community. The patterns used were limited, and to create new, domain-specific causal and temporal patterns would be out of scope for this project. However, it is important to note that although these patterns are limited, overall, the narratives, it is likely that some concepts regarding causality may be described differently. Although more domain-specific causal words and patterns may improve these

graphs, the aggregate nature of detecting causality across the whole corpus allows many of these relationships to be captured. In addition, this system's aim is to support existing decision-making processes, such as investigations into attacks, and to visualise reports; the accuracy is not as much of a concern as it may be in automated decision-making tools as mistakes can be noticed by an investigator and disregarded.

This stage and the previous allows for a custom insider threat model, with topics identified, labelled, and contextually placed; however, one of the limitations of models is the difficulty in creating actionable insights. In particular, these give an overview of an attack; however, the individual details are extremely important to allow an organisation to change internal policies. Therefore, the final technical objective of this system was to be able to analyse and visualise the individual topics, particularly in understanding these causal and temporal links. This stage allows an investigator to understand the specific topic and understand what is being discussed.

9.3.4 Topic Analysis

The topic analysis is critical as it supports tactical decision-making, with the other parts of the system supporting strategic decision-making. It uses IE triples to create a visualisation of all the sentences within a topic, no matter the overall length of a sentence. OpenIE is used to extract the triples, and an algorithm was developed to merge the triples from different documents, providing a summary of the topic without losing information. This is particularly important if any information is lost during this process, the model becomes difficult to use, and it becomes difficult to make the precise policy changes needed. The core critique of this tool is this merging algorithm; although it is able to merge information, there is still a lot of repeated information, creating challenging to understand visualisations in topics with a large number of sentences. However, as this data is merged in larger topics, some data is lost in smaller topics; in addition, this algorithm could be constantly improved, and therefore there is a continuing issue of diminishing returns. However, the approach presented here demonstrates the concept and is tuned to perceive

here fine details, even in smaller topics.

This part of the project comprises two core objectives, the merging algorithm, and the visualisation. The algorithm, in particular, is the main objective where improvements could be made. The difficulty of this was the merging of sentences from different authors, as each author will use different language, structure, and level of details. Therefore, joining these structures where words may differ is challenging, made more difficult due to the limitation of open-domain off-the-shelf tools. Off-the-shelf and open-domain tools were used because these tasks are challenging, and creating custom models involves a large amount of labelled data, computational power, and time. With the differing language across a topic, the challenge became automatically detecting the same sentence, despite the difference in wording or structure. This was solved by comparing the edit difference between the two sentences in terms of the number of words and keeping the simpler sentence structure. The simplest sentence structure was chosen as this would be easier to visualise, with more complex structures often containing additional words without context. For other sentences which were more distinct, a focus on ensuring the graph linked together with the same words, for example, 'She'- referring to the insider, is discussed often in some topics, to merge more complex discussions of the insider, all instances of 'she' were merged into one vertex on the graph. Although it is possible that 'she' may refer to more than just the insider within a topic, this was unlikely. However, additional information may aid in reducing this ambiguity. Finally, word2vec was considered to find where two words were related but were not the same word; in this model words are represented as vectors, and the distance between vectors can show how related two words are, although this can depend on context. This caused two major issues, first in deciding limits, and second, in information loss, information loss was the primary concern when developing these tools, both in determining how close one word must be to another to be merged and the loss in removing the different words, particularly subtle language use.

There were many additional NLP tasks available in CoreNLP, the off-the-shelf tool that was chosen for the topic analysis objective. These were explored to give further

context and richness to the graphs created; in particular, named entity recognition, co-reference resolution, and sentiment were all considered. These are aimed at improving the visualisation and allowing an investigator to understand a topic more deeply. This additional information was difficult to import due to the cognitive load associated with interpreting each topic and provided little benefit. Co-reference resolution was particularly difficult, and the co-references were often spread across different topics rather than the same topic. This additional information became challenging to visualise simply and effectively, with the more limited information producing simpler graphs that were easier to understand. This presents an opportunity for future work within this project's scope, and as many of the techniques developed are not limited to insider threat, this could be explored as part of future work in another domain.

9.3.5 Using the System

The final objective is how each stage of the project and each technical objective come together to create a system. The use of the system and a demonstration on how it can be used was presented in the previous section, with a focus on practical applications and how this may fit into larger incident response and digital forensics context. The system is made upon a web-based tool that can run additional scripts and Gephi to produce the graph visualisations. This objective demonstrated how each of these individual tools could be used together to address an insider threat attack, from an investigator's point of view, and answering key questions such as the technical methodology used, the insider's motivations, and how these can be used to implement policy changes. The issues regarding the system primarily involve usability, as the system as a whole is not a complete product. In particular, the reliance on Gephi and the UI for the visualisation. In addition, there are other issues such as the labels being based on an existing insider threat model, limiting the ability of the model to deal with emerging topics as the current model is static, and the NLP model can change over time. This system has only been demonstrated with one type of insider threat case.

The current UI was developed as a research tool rather than a product, and therefore it is limited to topic modelling. The graphs were developed using Gephi, with the topic analysis as a separate python script that would output Gephi files. The use of Gephi was chosen because of the flexibility of the graph visualisations, particularly the ability to change the graph layout and easy access to graph layouts such as Hu (2005). Although this does introduce another tool to access the results, the benefits of this are clear for the visualisations, including not just graph layout but the ability to modify the size, colour, and other properties of vertices and edges. These graphs are interactive, and certain attributes can be filtered, the graph expanded, and graph algorithms such as Hu (2005) can be performed. However, it is envisioned that future work could develop these initial tools and provide a complete interface to allow this research to be deployed easily within an organisation exploiting techniques such as user-centered design, to create the software.

As discussed in the previous sections, an issue with the topic modelling is most visible when viewed through the lens of the system as a whole. In particular, the difficulty in providing labels for topics; presently, this is done using an existing insider threat framework. As discussed in the previous sections, this impacts the system's usability in interpretation. Particularly when analysing the causality and temporality. The differences between these topics are subtle but can be extremely useful; therefore, it is vital to understand the potential information loss between codes and topics and the impact of labelling a topic as a particular code. The subtle nature of insider threat may be essential to capture by labelling these differences, with the downside of potentially causing information overload. However, as future work, if this was to be expanded into software, it could go through usability testing with potential users. These issues could be resolved by creating a hierarchical topic model using the codes and allowing the level to be adjusted. Although out of scope for this research, the ability to use the underlying tools and methodology into a piece of deployable software was considered, and many of the tools have been developed to be adaptable for potential future work.

One limitation of this work is the experimental data; this research has only been shown

on a single case of insider fraud. As identified by Cappelli, Moore and Trzeciak (2015), insider threat is complex, and there are many archetypes of insider threat under the two types of malicious and unintentional. This work targets organic narrative reports, which are difficult to gather, as these are often not made public by organisations. To counter this, the perspectives experiment was launched, and data was gathered from a fictional account of an insider threat attack. However, this experiment is a time-consuming process and requires the application of an existing model to aid in the grounding and interpretation of the results. However, the insider threat model and additional tools were not trained on this specific case; therefore, they are likely appropriate for future experiments. To further expand this work, an approach could be to focus specifically on unintentional insider threat or other insider threat archetypes, and a case of unintentional insider threat could be analysed to evaluate the model's efficacy on different insider threat types and archetypes. To complete this a new insider case from the literature would have to be selected, new perspectives written and recorded to allow for a new version of the perspectives experiment. Once collected, this new perspectives experiment could be mapped to the insider threat framework chosen or to a new framework; for example in the case of unintentional insider threat, a different model may be appropriate such as Greitzer et al. (2014). This could allow the existing model to be compared on both new insider threat types or archetypes as well as using a different model for mapping. While this is a large amount of work, it could effectively demonstrate this particular model. However, the novelty presented in this thesis is not this specific model but the process used to create it and showing that process using an example case of insider threat to create capabilities to explore an insider threat attack.

Despite all these critiques of the work, the tools created perform well and create interactive visualisations of an insider threat attack, grounded with a social model to aid in interpretation. These critiques have been considered throughout the project, and several mitigations have been put in place to limit the impact on the results. Many of these result from issues being out of scope for this project; this research represents a first step in using

natural language processing in the insider threat domain, creating initial tools which could later be implemented in a full software package. In addition, this project was time-bound therefore a focus was placed on ensuring the objectives were completed, and there was no delay in delivering the work. Many of the limitations introduced are because of the management of the project's scope and the time-bound nature. During this work, many novel approaches have been developed, particularly in creating tools to understand an insider threat attack. These novel approaches, despite some limitations, do demonstrate how NLP can be used in the insider threat domain and represent a first step to creating a piece of software that can be deployed within an organisation.

Chapter 10

Future Work

As identified in the previous section, there are several areas for follow on research. Although these improvements are out of this project's scope, they could be addressed in future work. The first piece of future work would be to apply these methods to new cases, either by collaborating with other researchers who have used organic narratives to create insider threat models or by running another experiment similar to the perspectives experiment. Second, improving each tool, for example, the merging algorithm could be improved to recognise less salient information better. Finally, producing each tool as a software package with the existing web interface allows an investigator to understand an insider threat attack more easily. Future research could also seek to support existing research by improving existing frameworks and recognising new attack vectors or attack characteristics. The impact of the research presented was fully discussed in the previous section alongside the critiques of the work. This chapter will discuss how this work can continue to impact the domain and the wider field of cybersecurity and human aspects of security.

One key direction would be to confirm these findings on additional cases, particularly in another insider threat archetype. The case chosen was insider fraud, and although this is very common, there are other archetypes, insider IP theft, insider sabotage and the smaller insider archetypes. In addition, the case of insider fraud could be expanded

to cover recruitment by an outsider, though this is most common in insider IP theft. The work could also be expanded to accidental insider threats, and creating a general model for understanding all types of insider threats computationally would be a great advantage for investigators and researchers. Creating these validation experiments would require a new insider threat case, using this case in another perspectives experiment to gather organic narratives and potentially labelling a subset of this to evaluate performance. Although this was beyond the scope of this project, demonstrating the potential for all insider threat cases would be extremely beneficial, demonstrating the effectiveness of organic narratives in many cases.

Another key improvement that would increase the impact of this work and could be done in the future is to expand the information shown to the user. Presently this was limited to using the Gephi software package, being difficult to visualise enriching information such as named entities (people, places) and co-reference resolution. These could simplify the graphs by improving the merging algorithms. The challenge is ensuring that the graph is still interpretable while adding relevant additional information and not overwhelming a user. Developing this additional information would require additional experiments to examine user experience, ensuring that this information is appropriate. This was similarly beyond the scope of this research. However, if this work were to become a deployable piece of software, this would be necessary.

The future research that this initial piece of research demonstrates is not only limited to the project itself but there is also still additional work that can be spawned in both insider threat and more widely in cybersecurity. This work has developed several techniques to download and use corpora, specifically organic narratives. These narratives have many advantages over traditional reports, providing a method of reducing bias while also allowing individuals to report their view of an incident in however much or little detail they would like, and informal, informal or technical language. Additionally, this work has made heavy use of topic models to understand and process these large corpora. These developments can be key to limiting insider threat further and providing additional tools for

individuals to use and, more generally, other types of insider threat facing an organisation.

There is still a lack of data for insider threat, which can be a barrier to researching insider threat domain, as each piece of research often needs to collect data. CERT provides the main dataset for insider threat, but this is the only public dataset. Future work could clean and label the full dataset of insider threat cases. Potentially categorising each by insider threat archetype (fraud, IP theft, sabotage) or by type (malicious or accidental). Many of the capabilties discussed in this work have been in data collection for insider threat. Either allowing alternative organic narratives of insider threat to be easily mapped to existing models or expanded existing corpora. By cleaning and labelling the data sets created, this work released a methodology and a research-ready dataset of insider threat cases, but this has not yet been released publicly. Data gathering remains a challenge for all future development of insider threat tools, especially models.

Building upon the existing research into insider threat was identified as part of the research gap for this work. However, the work presented here can be improved upon further and models developed based on this work. One assumption from the insider threat literature is that the insider threat archetypes are extremely distinct and therefore require distinct models, this has been challenged, and general models of insider threat have been created, such as Nurse et al. (2014b), which work has built upon, specifically. This general framework for insider threat cases, however, the view that a general model can be created is debated in the domain with many preferring to consider the archetypes individually such as Cappelli, Moore and Trzeciak (2015) and earlier work such as Randazzo et al. (2005). Therefore, confirming this approach could be a particularly valuable direction for future work; it is difficult to measure this presently, however with the technical approaches developed, models can be scored and tested with each other, and this is measured. In addition, a future direction could improve the software support for using all models, rather than purely this work, improving the accessibility and deployment of models in an organisation.

In insider threat, in particular, one such advantage of this would be to identify new

cases or new trends in insider attacks before an insider chooses to attack their organisation. Future work could build upon the emerging topics initial work and deployment of models to evaluate changes in patterns of insider threat attacks. By examining topic models and how topics change over time, new techniques, methodologies, motives, and assets can be identified sooner. This could allow attacks to be prevented using traditional security controls. This could be key to exploring new archetypes such as those identified by CERT and ensuring that new archetypes are discovered sooner, potentially by continuous scanning over public data.

In general, the use of organic narratives for not just insider threats but wider cybersecurity is likely to be impactful. Presently, this work has focused on insider threat; however, organic narratives and reports in cybersecurity are very common. Traditionally, the reports are usually very technical. However, witness reports may be useful when attacks, such as social engineering, do not target infrastructure directly but instead target individuals. One such use could be to identify emerging attacks by examining reports of general cybersecurity incidents, using topic models mapping incidents to an existing list of techniques such as the MITRE ATT&CK framework, and then plotting incidents over time. New attacks could be identified by examining where a topic cannot be well mapped to an existing technique or where an older technique suddenly becomes more common. This uses much of the work developed in this thesis; however, by widening this up to cybersecurity in general, rather than insider threat, other attacks can be identified in the same manner.

10.1 Out-of-domain impact

Although the majority of the future work is limited to insider threat and cybersecurity, this work has developed many techniques that can be applied to other fields and used by researchers. Two potential areas for future work is pivoting the work with organic narratives to the technical report, the models have performed well identifying technical topics

this may be an effective method to aid in the understanding of medical or engineering reports, the second is the impact that this work can make to social science, enabling a different approach to grounded theory or allowing interpretable topic models to be created and utilised.

Technical reports are not just limited to the insider threat domain or, more widely, the cybersecurity domain. Understanding technical reports remains an issue in several domains, such as the medical or engineering fields. Although these may not use organic narratives as widely, preferring technical reports written in formal language. Future research could build upon the more general models which have been created and use a similar technique to understand technical reports better. Research in this area is ongoing, and NLP has been a staple in the medical field for many years due to the large amount of reporting and the desire to improve patient outcomes. Topic modelling is not as widely used, and future research could seek to apply topic modelling in this domain, potentially in addition to other ongoing research efforts. However, topic modelling has shown its effectiveness for data gathering when other, more discriminative approaches have failed, and future research in topic modelling could likely lead to advances in technical report understanding.

The most valuable future research outside of cybersecurity is in social science, grounded theory and the ability to use topic modelling as an automation approach. This work has demonstrated that using an existing grounded theory model and using topic modelling to automate some of the coding process is possible. For this, it would be necessary to investigate potential accuracy measures and compare grounded theory topic models to the more organic models created by analysing the optimal number of topics such as those created by Arun et al. (2010). One of the challenges with this has been the difficulty in interpreting models, and this has limited the use of topic modelling for this use, however by mapping topics to an existing model, some of the interpretation issues can be mitigated. Unfortunately, there is still much work for topic interpretation, and these methods have only been tested in the insider threat domain. However, with future work, topic

interpretability may be improved.

10.2 Summary

This work has many potential directions that can be explored to directly improve the tools created in this thesis, create a software package, or make an impact more widely in insider threat, cybersecurity, and other domains. This work represents the initial proof of concept work that can impact many fields. For insider threat, this work focused on creating a toolset that could be used to investigate an insider threat attack. This work utilised organic narratives mapped to an existing insider threat framework, which are then visualised. The future work from this initial work has many directions, depending on what is addressed specifically. Based on this research, this work could be improved and turned into a full software package. For insider threats, the primary direction is to improve this work and develop new tools to understand attacks, but as an aim, future work could build upon this work to not just understand attacks but to understand and recognise emerging attacks. These insider threat specific techniques could be expanded to other types of attacks for the wider cybersecurity domain. The primary out of domain directions could be to develop topic modelling further with a mind to include it in a grounded theory model development, either to aid in the creation of grounded theory models or to allow grounded theory models to be easily mapped to pieces of text. There are many directions that future work can take in the future, and the impact of this work may not be limited to insider threat, with the general techniques developed able to affect many different fields.

Chapter 11

Conclusion and Contribution

This final chapter discusses the contributions and conclusion of this work, precisely, in how this work meets the aim presented in the Aims and Objectives chapter, to provide a tool able to understand an insider threat incident by leveraging natural language processing. This thesis shows how this aim has been met and provides a new suite of tools to enable this capability, enabling the creation of a full view of an incident without focusing on just one element, such as the technical aspects or artefacts. Specifically presenting; a method to map reports to known insider threat frameworks, build a custom framework for a specific incident, draw connections between events linked causally or temporally, and finally explore an insider threat characteristic in detail and rationalise changes to prevent the next attack. This contribution does not focus on the general nature of insider threat but instead allows for the customisability of an insider threat framework to investigate threats that may be unique to that organisation.

This research has resulted in three capabilities that can be used together as a system to investigate insider threat attacks. This is accomplished using natural language processing techniques, a type of machine learning applied to the human-language text. This work builds upon existing models and focuses on enhancing existing processes rather than building new techniques from scratch. First, two corpora (collections of documents) were created a corpus of general insider threat cases and a corpus of organic narrative reports

of an existing insider threat case. Using topic modelling on the corpus of general insider threat cases, elements of all insider threat cases can be captured in a topic model, which is then applied to the corpus of organic narrative reports. To provide additional context, an insider threat model is mapped to the topics, giving topics more meaning; however, these topics are still isolated. This aims to provide a similar understanding as an insider threat model causal, and temporal links between topics were discovered. This process created an overview, similar to an insider threat model. However, this only allows for a broad view of an incident rather than the granular details, so each topic must be analysed to enable this. Using off-the-shelf tools, the topic analysis objective was completed, mapping a topic and visualising each sentence within the topic. The process of using this system and the tools provided effectively was discussed at length, with an investigator targeting topics that might be of interest to the investigation, either by a casual chain, insider threat characteristic, size or temporal clues. The aim to improve the human understanding of insider threat and the work presented completed demonstrates that using these techniques, reports can be processed, visualised, identified and analysed for essential details about an attack. Although this does not offer detection capabilities, this greater understanding is key to improving detection efforts in the future, particularly as insider threat detection systems become more sophisticated, exploiting recent advantages such as machine learning.

This work has shown a wide array of novel approaches, techniques and outcomes. These include automatically categorising text based on topic proportion for some corpora, mapping the result of a topic model to an existing grounded theory model, visualising causality and temporality and finally joining all these together to produce a system capable of exploring an attack. This work has produced numerous novel techniques based on the individual capabilities discussed. However, the largest amount of novel work is the interaction between the different techniques, which provide his holistic view of an insider threat incident. The system was developed with these techniques, but the novelty presented in the thesis is the system itself, the ability to invest in a large number of organic narratives and produce a visualisation of an attack, exploring the features and characteristics

better to understand the attack at a much lower skill level, allowing policymakers to act on this information and make changes to prevent the next. Although this does not produce a detection system, the literature shows that a better understanding of attacks is vital to improving detection efforts. Furthermore, this system can be used within an existing digital forensics or incident response framework to reflect on an attack using witness reports of the incident. A key advantage of this work is that the initial model is built entirely using public data, using news articles published across the web rather than data which may be private or confidential to an organisation specifically, but allows an organisation to build custom models with this custom data to create a more tailored model.

There are, however, improvements that could be made, and many of the novel techniques created represent initial approaches for using NLP in the insider threat domain. These were discussed at length in the Discussion chapter, which critiqued and presented various solutions to improve this work. These critiques primarily could be resolved with additional time, and if the research was presented in this thesis was continued and expanded upon, potentially becoming a piece of deployable software. Many of these issues would aid, primarily, in creating models which are more easily understood, for example, topic labelling adding further understanding using named entity recognition, this is an ongoing issue in the adoption of machine learning in the insider threat domain. Other improvements could be made by improving the initial topic model, labelling more data and adjusting the scoring mechanism to improve the overall model performance. These improvements do not counteract the impact that this work creates, and future work could make these improvements specifically to improve the performance of this research as a software package

The techniques are not limited to insider threat, and the work present has the opportunity to impact many more fields, in particular, wider in cybersecurity. One piece of future work which has been discussed was expanding the techniques used to map the organic narratives to the existing insider threat model to general cybersecurity incidents. This could be done by mapping techniques to news articles of general incidents and plotting

these over time as a correlated topic model, showing the prevalence of different techniques. Other future work could build upon the insider threat work, as this work has demonstrated the effectiveness of organic narratives; this could be a potentially useful direction. Organic narratives allow this work to use many different witnesses to build an overall picture of an attack. The use of organic narratives can also be used for future work in other fields, specifically social science. Grounded theory models are used regularly in social science and are often compared to topic models. However, topic models do not offer some advantages that grounded theory models do. The future work from these initial steps discussed in this thesis can continue to impact many fields.

In conclusion, this thesis presents new techniques for understanding insider threat attacks using natural language processing to visualise and analyse not just technical reports but any report, allowing for more witnesses to become part of the investigation process. This novel approach considers all aspects of an attack, in contrast to the traditional approaches, which consider only technical or psychological. This work builds upon the existing body of research on insider threat and offers a new approach to using insider threat models. The aims and objectives of this research were decided after an extensive literature review, and these have been completed and met. This thesis has presented three key technical objectives and demonstrated how these could be used to investigate an insider attack. The future work in this field that this initial work demonstrates can be used for more than just insider threat, particularly for using organic narrative reports rather than purely technical and mapping these to existing models and frameworks. This work represents the first step in improved detection capabilities by fully understanding insider threat attacks.

Appendix A

Full List of Topics

Topic 1 - denton losangel michaelpoter req thoma victoriapolic

Closely related to: Organisation Characteristics - Vulnerability/Opportunity

- She was caught when one of the tellers had questions about suspicious check for over $400000. *(Document: 5d07dfe6a9344)*
- Coworkers thought she was weird, but she might have a gambling problem. *(Document: 5d03fbe74cb6c)*
- She helped create the new It system, but she refused to use it. *(Document: 5d03fbe74cb6c)*
- They didn’t ask too many questions. *(Document: 5d03fbe74cb6c)*
- She bought lunch and drinks for people, so they just went along with her. *(Document: 5d03fbe74cb6c)*
- good verdit for all. *(Document: 5d03dc42937ba)*
- She was well-respected by her supervisors, and seemed to do good work. *(Document: 5d03cbf7e42dc)*
- She must have known she would have been caught earlier. *(Document: 5d03c72fc0576)*
- from using it. *(Document: 5d03baf2e2694)*
- She had helped the IT Dept. *(Document: 5d03baf2e2694)*
- She was able to do so by using a paper accounting system (non-computerized monitoring). *(Document: 5d03baf2e2694)*
- She was condemn to jail and pay a fine. *(Document: 5d039b5fd1d3a)*
- It seemed like she was doing it to use the money to live comfortably and be well-liked by giving things away and spending on her friends and coworkers. *(Document: 5d03998389dc8)*
- She had to pay restititution, fines & taxes. *(Document: 5d03984dc2628)*
- A lot of people were suspicious of her earnings, but never questioned it too much. *(Document: 5d02a27f6c1ec)*
- The lady has since been jailed. *(Document: 5d024dc0f1798)*
- No one questioned it, even though they should have. *(Document: 5d024811e412e)*
- However, such a rumour would have gathered attention sooner. *(Document: 5c87a852e0b46)*
- She did it because she could. *(Document: 5c879871c869a)*
- She was using her involvement with the new IT system as a cover. *(Document: 5c701353965dc)*
- She helped create the new system and though the higher Up's was able to get a exemption from the new system. *(Document: 5c6fe9ae6757a)*
- 9 accomplices have been identified but not charged. *(Document: 5c6f111d877e4)*
- When she was caught, she was fined $60mil. *(Document: 5c6ecc9f2513e)*
- She is required to pay back 48million with 28 in tax and 3.2 in state taxes. *(Document: 5c6e9bcfbe44c)*
- This activity was often overlooked however and she was not caught out on this regard. *(Document: 5c6e8d80126d8)*
- This was possible because a paper-based system was in place. *(Document: 5c6d857487398)*
- Other staff thought of her as flaky and weird. *(Document: 5c6d530129acc)*
- her exemption had backing from others to do this. *(Document: 5c6d530129acc)*
- She had done so over 18 years. *(Document: 5c6d17f2c581e)*

Topic 3 - hoffman ico undertak veritytrusteesltd virginia dataprotectionact

Closely related to: Organisation Characteristics - Vulnerability/Opportunity

- There was also a failure of senior management to question why she pushed back on the implementation a system she helped to design. *(Document: 5c6d37122eeb4)*

Topic 5 - amazon mrglanvill ream american pleb createspac

Closely related to: Actor Characteristics - Motivation to Attack

- Prior, rumours regarding her wealth existed, though the nature of these rumours is unknown. *(Document: 5c87a852e0b46)*
- Her motives for committing fraud are unknown other than obvious financial gain. *(Document: 5c73c9a927a4c)*
- It is unknown why she did it. *(Document: 5c6fc9bcf2c2e)*
- It's unknown and not reported why she did this. *(Document: 5c6d530129acc)*
- She had an unknown financial status but seen to buy things for others within the office. *(Document: 5c6d530129acc)*
- Unknown what will happen to the accomplices or if they will be prosecuted. *(Document: 5c6d530129acc)*

Topic 6 - hillsid braverman wilsonsonsini condénast connecticut oct

Closely related to:

- No one of the coworkers believed because she never seemed to be suspicious, only a IT manager did saw something. *(Document: 5d039b5fd1d3a)*
- She said it was from an inheritance, and people believed her, others joked that she gambled and won the money. *(Document: 5d03983ac17c8)*
- She was caught and no one could believe it to be true, especially since she was so generous with everyone. *(Document: 5d024811e412e)*

Topic 8 - bocaraton bofa mattei nahra pixili tambasco

Closely related to: Organisation Characteristics - Vulnerability/Opportunity

- She wanted to keep the old paper system, in fear the new system would out her. *(Document: 5d079eb114fb4)*
- It was unclear as to why she did it but it can be assumed that she did it to keep up with her lavish lifestyle. *(Document: 5d03ea6d73c8a)*

- It could be said she did it to keep up a life style, possibly gambling or drugs however that was pure speculation and was not mentioned in the news report and was rumours from a colleague. *(Document: 5d03d0e7b9686)*
- Upper management allowed her to keep using the old paper based accounting system in her department. *(Document: 5d03984dc2628)*
- It's likely the managers pushing for the exemptions were the 9 accomplices, since they helped her keep the operation running longer. *(Document: 5c879871c869a)*

Topic 9 - kessing statecolleg veteransaffair dann parksmil royalcanadianlegion

Closely related to: Organisation Characteristics - Vulnerability/Opportunity

- During the 18 years of producing this crime she claimed she was unable to operate with the system that was in place for all departments. *(Document: 5d0800d5b2674)*
- The manager responsible for the fraud had been involved in designing the replacement system, but had also lobbied successfully to exempt her own department from using the new system. *(Document: 5d079fc2cf7ce)*
- The person the story is about is a manager of a department in the company. *(Document: 5d079eb114fb4)*
- Her department was surprisingly given exemption from adopting this new computer system. *(Document: 5d0481de7bb2e)*
- When they converted to a computer system, she insisted her department not be a part of it, which enabled her to continue her scheme. *(Document: 5d04581fba50e)*
- At some stage she attempted to have her department made exempt from the system which initially worked. *(Document: 5d044fa461008)*
- A new program was developed and her department was exempt from it, stating they couldn't run it. *(Document: 5d03de5cd6ad8)*
- She claims her department cannot use the new system. *(Document: 5d03d83dbd8f8)*
- Days later the IT department received notice that the manager's department could indeed work outside from the program and continue to use the old paper system. *(Document: 5d03d561c08b4)*
- This manager stated her department could not work within this program and the company insisted every department had to comply. *(Document: 5d03d561c08b4)*
- Every department in the company had to use this program. *(Document: 5d03d561c08b4)*
- Since the discovery, the company has made it mandatory that every department has to be on the Fiscal IT program, no exceptions. *(Document: 5d03d561c08b4)*
- The IT department recently implemented a new system for everyone at the company to use, however this manager was hesitant. *(Document: 5d03cffd39760)*
- When the IT department implemented a new financial system she continued to use paper. *(Document: 5d03cc65150e0)*
- At some point, she help design a new audit system for the work her company was doing, yet when it was initiated, she pushed back and refused to allow her department to use the new program. *(Document: 5d03cbf7e42dc)*
- This woman worked in an office and supervised her department. *(Document: 5d03cbf7e42dc)*
- She was caught when they rolled out a new system that she had helped design and she had refused to use it while the other departments had. *(Document: 5d03c707b6dfa)*
- She circumvented a new IT system in place to catch such behaviors by complaining until her department was made exempt. *(Document: 5d039b3974a04)*
- She avoided detection by getting her department exempt from the new computer program IT had installed. *(Document: 5d039950a0ffa)*
- IT said she refused to put her department on a new system. *(Document: 5d03984dc2628)*
- When the company switched over to an electronic way of handling funds, she said that her department could not switch over. *(Document: 5d02a27f6c1ec)*
- She kept it secret and hidden using the old paper filing where as they had tried to implement a new procedure doing it electronically but she had managed to get her department bypassed. *(Document: 5d024dc0f1798)*
- She was capable of exempting her department from using the new system, and the media reports that there was accomplices. *(Document: 5c87a852e0b46)*
- She had managed to get her department exempt from using the system that would have better audited what was going on, but this would have exposed her activities. *(Document: 5c7f9270c1494)*
- As the woman was a manager she was able to move her department outside of a new IT system that would audit all transactions. *(Document: 5c7f9270c1494)*
- IT staff busted her after rolling out new fiscal system *(Document: 5c7c178e1c62e)*
- Her request was granted and her department was exempt from the new requirement and continued to operate under the legacy paper system which enabled her to create discrepancies and hide issues. *(Document: 5c73c9a927a4c)*
- Bizarrely, her insistence on continuing with the old system was supported by senior management despite the IT department raising an issue with this. *(Document: 5c6fee6c3665a)*
- A new IT system was rolled out that she was involved in designing so would have understood how it operated and due to this she was very vocal in her issues with the new system and managed to get her department exempt from using the new system thus allowing her to continue. *(Document: 5c6fce7530a52)*
- This did mean however, that her fraud was far less obscured in noise now it was only her department using it, leading them to notice anomalies. *(Document: 5c6ea42c1647c)*
- When a new computer-based accounting system was introduced, she stated that her department would not use it. *(Document: 5c6e74be61b98)*
- She had prevented her department from switching to a new online, auitable system. *(Document: 5c6d7e0262872)*
- When she was found out, the IT department were unable to find any evidence as it had all been paper based, not electronic. *(Document: 5c6d591491dac)*
- The IT department did not ask too many questions regarding why this department did not have to use the new system, they just took what they had been told. *(Document: 5c6d591491dac)*
- A Manager had been caught stealing money by fraud, she stole over 60 million over 18 years, she did this by making sure that when her company changed to an electronic based system, her department did not use this. *(Document: 5c6d591491dac)*
- other departments were using this new system. *(Document: 5c6d530129acc)*

- She was involved in the design of the new IT system and put case forward for exemption within her department. *(Document: 5c6d530129acc)*
- Her department was exempt from the new IT system. *(Document: 5c6d530129acc)*
- Her reason seemed to have been that it was too difficult to use for people in her department, which they agreed with, but this was likely a cover story to enable her to carry on with it. *(Document: 5c6d37122eeb4)*
- She had helped in the design of a new accounting/auditing software system, but had managed to get her department exempt from using. *(Document: 5c6d37122eeb4)*
- A department manager had been stealing from a company for around 18 years. *(Document: 5c6d37122eeb4)*
- The manager was found guilty and ordered to pay back 60 million dollars, and as a result all of the departments within the company now use the computerised system without exception. *(Document: 5c6d2abb29bf8)*
- Her department were made except from the new IT system that did auditing and accounting. *(Document: 5c6d17f2c581e)*

Topic 11 - arlenejorgensen ceo socialservic beaufortdeltahealth beaufortdeltahealthauthor breachingconfidenti

Closely related to:

- A manger or CEO of a company who is a woman, was taking money from work, and making it look like it was going towards other things. *(Document: 5d03983ac17c8)*

Topic 13 - showpo blackswallow hanumanthu msaroutunian mrbaro reyna

Closely related to:

- There was this shocking news about a middle manager,a tax office employee who got arrested for her involvement in issuing over 200 fraudulent cheques over a period of 18 years.she *(Document: 5d03dc42937ba)*

Topic 14 - amazon closedloopfund curbsiderecycl amazonprim amazon’ american

Closely related to:

- co-workers told that she was very popular and was considered to be a kind and generous lady with a lavish life style and a lot of influence around but as she was very good to all none actually complained cared or noticed anything wrong with her lifestyle or spending habits.although *(Document: 5d03dc42937ba)*

Topic 17 - martin martin’ nsa prosecutor boozallenhamilton cia

Closely related to: Attack Characteristics - Attack/Organisation Characteristics - Asset

- The fraud was only initially detected when a bank teller queried a fraudulent cheque. *(Document: 5c6e74be61b98)*

Topic 18 - shell theparadisepap thesüddeutschezeitung applebi neworient paradisepap

Closely related to: Actor Characteristics - Historical behaviour/Outcome - Actor

- Her generosity and lifestyle did not raise any alarms as she had previously told colleagues that she had received a substantial inheritance. *(Document: 5d03d0e7b9686)*
- To evade suspicion, he had claimed to have a remarkable family legacy. *(Document: 5d037c3adef6c)*
- She avoided suspicion by claiming to be in receipt of a substantial inheritance. *(Document: 5c73c9a927a4c)*
- The woman involved was ultimately charged and was made to pay back $63.2 million in various fines and taxes. *(Document: 5c6fee6c3665a)*

Topic 20 - burley homestead collier herrin jacksonhealthsystem sheriff

Closely related to:

- She also had a ring of nine accomplishes who have no yet been charged as their exact roles in this are still being investigated. *(Document: 5d03cafcd1c68)*

Topic 21 - alteryx infowatch ymca ford knesset deniszenkin

Closely related to:

- At a tax office in London, a manager has been accused of issuing 200 fraudulent checks. *(Document: 5d0800d5b2674)*

Topic 24 - mamba prudenti internet ang belarus commission

Closely related to: Attack Characteristics - Attack Step Goal/Organisation Characteristics - Vulnerability/Opportunity

- She exploited weaknesses in paper-based systems - where it was easier to hide discrepancies - to write fraudulent cheques that she paid in herself. *(Document: 5c6ea42c1647c)*

Topic 27 - starwood hilton klein vodafon chadam' denizen

Closely related to: Outcome - Actor

- When she was caught there were some people that were shocked because of her otherwise overall nice demeanor. *(Document: 5d03c707b6dfa)*
- She now has to pay back 60 mil overall. *(Document: 5c6fe9ae6757a)*

Topic 32 - sec iowa bachmann sec' heki octob

Closely related to: Actor Characteristics - Historical behaviour/Attack Characteristics - Attack

- As a result of her crimes she was sued for 6 million dollars after a suspicious check of $400,000 was uncovered by a fellow employee. *(Document: 5d0800d5b2674)*
- Put in context with other accounts, an immediate jump would be to suggest the rumours were of the crime, and thus the accomplices may be fellow office workers. *(Document: 5c87a852e0b46)*
- The subsequent investigation found that the individual been committing this fraud for 18 years. *(Document: 5c6d2b8c9ad3a)*

Topic 33 - tipton informationcommission multistatelotteryassoci rootkit 01634227989 akmpt

Closely related to:

- She sometimes offered to buy a round at the bar or pay for lunches, but no one ever questioned it, they assumed she had won the lottery or won at gambling. *(Document: 5d03d561c08b4)*
- She was writing fradulent checks, and essentially making herself rich that way. *(Document: 5d0398cc6fa54)*

Topic 35 - yahoo lockser lal carlson cecilel ceomarissamay

Closely related to:

- She was caught when a till employee came across a check of $400,000 and reported it as suspicious. *(Document: 5d03d59ae268a)*
- An office employee was fined several million: a bank teller reported a suspicious check for over $ 400,000. *(Document: 5d037c3adef6c)*
- Over the last 18 years, the middle-manager alluded detection until a bank teller reported a suspicious cheque for more than $400,000. *(Document: 5d036a7f85570)*
- The manager in question attempted to cash a large check for $400,00 which caught the eye of the banker, who then reported it. *(Document: 5d02dc8e97cb6)*

Topic 36 - delet skimmer ame keepitsaf bensonclermont cbs

Closely related to: Organisation Characteristics - Vulnerability/Opportunity

- The company rolled out a new electronic system, of which she was part of the design, that everyone was suppose to join and log their information into. *(Document: 5d079eb114fb4)*
- They had managed to avoid detection by easily manipulating the paper based records despite having a computer system designed to avoid the possibility of fraud. *(Document: 5d076e64a7364)*
- A departmental manager spent 18 years cashing fraudulent cheques to the tune of $60m with no assured motive A new IT system designed to add an audit level to the process was designed, and the manager involved in the design phase. *(Document: 5d044fa461008)*
- By the information provided in the story, this manager was involved in the design and implementation of the IT program. *(Document: 5d03d561c08b4)*
- I was help design it. *(Document: 5d03c9147ba66)*
- This manager had also been exempted by others from using the new system which would've undoubtedly caught her - even if she had a say in its design. *(Document: 5c879871c869a)*
- Her fraud went undiscovered for so long as no one in the office questioned certain abnormalities in her circumstances, for example there were rumours and it was noticed she never counted down to pay day like the rest of the employees, and objected to the new computer system despite being involved in designing it. *(Document: 5c73c9a927a4c)*
- The manager was involved in the design of a new IT system that would have detected her fraud. *(Document: 5c6f111d877e4)*
- The IT system designers found it unusual she complained that her department couldn't use the new system since she had been involved in its design. *(Document: 5c6ecc9f2513e)*
- The lady was involved with the design of the system but refused to adopt the system when it came into service. *(Document: 5c6e9bcfbe44c)*

- When it was decided to introduce a computer based system to reduce the risk of fraudulent behaviour she was involved in its design and introduction. *(Document: 5c6e877f639e8)*
- This was the case even though she had helped design it, and must have known that her frauds would be detected by its use. *(Document: 5c6e867e1b300)*
- She had managed to hide her activities within the paper-based system in use at the bank and had fought off the introduction of a new computer-based system designed to improve auditing within the company. *(Document: 5c6e80c32e77a)*

Topic 38 - github mcintosh orellana augustin avalo businessinsid

Closely related to: Outcome - Actor

- Afterwards she was required to pay back the money (over $60 million), another 9 conspirators may also be charged. *(Document: 5c6d567ba889a)*

Topic 45 - refrain naaptol semgroup catsimatidi zhang han

Closely related to:

- The court issued a severe sentence to prevent others from behaving in the same way. *(Document: 5d037c3adef6c)*

Topic 50 - infowatch deniszenkin marketingdirector medasset socialsecur firstniagara

Closely related to: Attack Characteristics - Attack/Attack Characteristics - Attack Step

- This might suggest instead that her co-workers were not accomplices, and their inspection of the system would reveal the crime. *(Document: 5c87a852e0b46)*

Topic 54 - infowatch mull deniszenkin hawaii kodak socialsecur

Closely related to: Actor Characteristics - Job Info

- I think at 42, 13, and 4 million respectively. *(Document: 5d03c707b6dfa)*
- She was well respected and seen as kind but odd. *(Document: 5d035a10e5b0a)*
- This was a lady in middle-management, apparently well established an respected in her role. *(Document: 5c6d6e434c6a8)*

Topic 57 - lexington gumtre krekelberg’ thompson usd рв,¬

Closely related to: Attack Characteristics - Attack/Organisation Characteristics - Vulnerability/Opportunity/Outcome - Actor

- The paper filing system that was in place, and potentially the involvement of 'accomplices' (although their involvement had yet to be identified) had allowed her to siphon away sums of money through the creation of fraudulent cheques over a long period of time (200 instances and $60m I think it was). *(Document: 5c6d6e434c6a8)*
- Fined a lot of money and the discovery of 9 potential accomplices and a lot of taxes to be paid back. *(Document: 5c6d530129acc)*

Topic 61 - chien infowatch brigham chinohil detect htc

Closely related to: Attack Characteristics - Attack Step/Organisation Characteristics - Vulnerability/Opportunity

- She avoided detection due to there not being strong audit systems in place. *(Document: 5d074d3715040)*
- She evaded detection by abusing the flaws in a paper-based accounting system, which is much harder to audit. *(Document: 5c91e851a6fd6)*
- She was detected eventually through a suspicious $400,000 cheque. *(Document: 5c6e867e1b300)*

Topic 63 - aug mrajmal cedarroad netcar theft acleveland

Closely related to: Attack Characteristics - Attack/Organisation Characteristics - Vulnerability/Opportunity

- The police investigation revealed that this was a planned theft with a network of people involved including 9 additional people arrested. *(Document: 5d0399cfc2b32)*
- Even though the bank had a computer system in place to prevent theft she managed to get her department to avoid using the computer system, and one of her fraudulent cheques was caught by the bank teller. *(Document: 5c7939498b4b6)*

Topic 65 - gliha decatur geaugacounti socialsecur assistantchiefconstablemauricemason childabuseinvestigationteam

Closely related to:

- It turns out she was using the old paper based system to write fraudulent checks for many years before she finally got caught. *(Document: 5d0410658893c)*
- It turned out she had been stealing millions of dollars from the company for 18 years. *(Document: 5d03fbe74cb6c)*
- It turns out it was stolen money from the company. *(Document: 5d03cffd39760)*
- Turned out, she was stealing money from the company for 18 years and had been doing it by hiding the information in the paper systems. *(Document: 5d024811e412e)*

Topic 67 - gregori morrow peterick bellevill skimmer uddin

Closely related to:

- Later it was figured out because the new program would audit and find out she had been stealing from the company. *(Document: 5d03de5cd6ad8)*
- the only way they caught her was when they installed a new program and it caught on to her. *(Document: 5d0398cc6fa54)*

Topic 68 - pauldebogorski dcps 7new brianentininvestig code keystrok

Closely related to:

- Colleagues had notice her more lavish spending and generosity with money, and rumours circulated about the source of her wealth, but she played this off as being a substantial inheritence. *(Document: 5d044fa461008)*

Topic 69 - rosenfeld correct economicpolicyinstitut employ harri jackson

Closely related to: Actor Characteristics - Motivation to Attack/Attack Characteristics - Attack

- A person who had worked at the same company for more than eighteen years was embezzling by way of check fraud for over eighteen years. *(Document: 5d074d3715040)*
- The manager wrote tons of fraudulent cheques, who according to her staff, was not her personality, or at least that's what they thought. *(Document: 5d03c72fc0576)*
- A woman had been embezzling money for 18 years at least but was recently caught over a $400,000 check that a teller thought was suspicious. *(Document: 5d024904967da)*
- It's hard to tell what her motives were for the embezzlement. *(Document: 5c6d6e434c6a8)*
- A longstanding staff member at middle-management level was found to be stealing money from her employer (the Tax Office). *(Document: 5c6d2b8c9ad3a)*

Topic 71 - quinlan healthnet lad connecticut luu socialsecur

Closely related to:

- She wanted to continue using the paper trail, but once the IT company caught on, they realized what she had done. *(Document: 5d03cffd39760)*

Topic 78 - devumi financ pearson octob publicprosecutor rtl

Closely related to: Actor Characteristics - Motivation to Attack

- The main person involved was a female tax employee, she was using the system to finance her lavish lifestyle. *(Document: 5c7f9270c1494)*

Topic 83 - privacypolici townsquaremedia inc googl googleanalyt internet

Closely related to: Actor Characteristics - Personality characteristics

- She opted instead to stay on a paper system. *(Document: 5d079e20e985e)*
- In her branch she opted out of using a new system that would help find fraud within the company. *(Document: 5d061d1464eba)*
- She was using an old accounting system to hide the discrepancies and enabling her to take out money for 18 years. *(Document: 5d039b5fd1d3a)*
- The manager was also the only one who successfully opted out of the new software - she was involved in helping create it, because she might have been aware her fraudulent activity would be detected. *(Document: 5d0399cfc2b32)*
- The manager was arrested alongside a fraud network and an additional 9 accomplices. *(Document: 5d035b56d3d10)*
- The investigation also revealed that the individual was popular with both subordinate and superior staff members and used her popularity to enable the fraud, including obtaining an exemption from using a new IT-based accounting system which was more robust than the paper-based system which she exploited. *(Document: 5c6d2b8c9ad3a)*

Topic 84 - järvet appleinc' appstor beij biodiversityparti calif

Closely related to: Outcome - Actor

- Eventually they were caught and found guilt of the fraud. *(Document: 5d07a0d1acdfa)*
- After being found guilty, the manager was required to pay back $60 million of funds, in addition to $3.2 milllion in state taxes. *(Document: 5d079fc2cf7ce)*
- --> Her office was surprised when they found out this news. *(Document: 5d0481de7bb2e)*
- She got caught when a bank teller found a suspicious check for 400,000 GBP. *(Document: 5d04581fba50e)*
- She was caught when a teller reported a suspicious large check, and once found guilty, made to pay restitution to the bank as well as taxes on the ill-gotten gains. *(Document: 5d03998389dc8)*
- Law enforcement officials found that some scammers who were easily manipulating documents without anyone noticing. *(Document: 5d037c3adef6c)*
- a lady has been found to be stealing from her company for over 18 years at a total of 60 million dollars. *(Document: 5d024dc0f1798)*
- As consequence, she was found guilty in a court and fined in the order of millions, purportedly to set an example. *(Document: 5c87a852e0b46)*
- She was found to be manipulating paper based records, and now everyone is mandated to use the new IT system. *(Document: 5c77c731bba1c)*
- The cheque that she was found out from was a 400,000 dollars cheque. Investigations found that 9 other people were involved but their charges were not yet determined. *(Document: 5c6e9bcfbe44c)*
- Despite the co-worker being surprised at the fraud it is indicated that a further 9 co-fraudsters and also been found. *(Document: 5c6e877f639e8)*
- Once found out the manager was brought before the courts, found guilty, ordered to pay back the money with taxes. *(Document: 5c6e877f639e8)*
- The manager was found guilty and suffered sever penalties, including fines of over $60M. *(Document: 5c6e867e1b300)*
- The computer system was difficult to use and tax office staff found it an extra burden. *(Document: 5c6e867e1b300)*

Topic 85 - nha tuyen vietnam hanoi hcmc tuoitr

Closely related to: Actor Characteristics - Personality characteristics/Actor Characteristics - Observed physical behaviour

- People had their suspicions about her in the office but they didn't know it was stealing. *(Document: 5d024dc0f1798)*
- In contrast, the impression given is that the coworkers were not close enough to know the lifestyle in question. *(Document: 5c87a852e0b46)*

Topic 86 - dwp cis moj taylor themoj james

Closely related to:

- Eventually, her actions caught up with her when a bank teller questioned a check she was trying to cash/deposit. *(Document: 5d079eb114fb4)*

Topic 87 - ford octob sage asia betti chipotl

Closely related to:

- She has to pay back millions in taxes according to the courts. *(Document: 5d03c72fc0576)*
- She was a bit flaky according to coworkers but was nice and would often pay for lunch or rounds at a bar but no one thought too much of it. *(Document: 5d024904967da)*

Topic 90 - armacost homan hough oct socialsecur acrookedbeverlyhillsstylist

Closely related to: Attack Characteristics - Attack

- Also, a bank teller grew suspicious of a check she brought to cash and reported her. *(Document: 5d079e20e985e)*
- She seemed to do by using some type of old system that allow he to cash in 400,000 worth of check through the past 18 years. *(Document: 5d0408ac26ed0)*
- A tax office manager committed fraud for many years until she was caught cashing a suspicious $400,000 cheque. *(Document: 5c730ce477c72)*
- according to the news report and the IT personnel she was cashing fraudulent cheques under an old paper system that should have been phased out, she was pushing for the old system to stay. the person of interest was cashing these fraudulent cheques to maintain a lavish lifestyle. *(Document: 5c6e8d80126d8)*
- A manager of a team in the tax office of a bank was finally caught after 18 years of cashing over 200 fraudulent cheques. *(Document: 5c6e80c32e77a)*

Topic 93 - ech mod fbi batra comcast karn

Closely related to: Outcome - Actor

- Though, described by others as generous and supportive she was unsuspected for nearly 18 years of doing such a crime. *(Document: 5d0800d5b2674)*
- It was then that her crime was uncovered. *(Document: 5d0481de7bb2e)*
- --> The manager was fined more than 60 million dollars for her crime. *(Document: 5d0481de7bb2e)*
- She got caught and fined up to millions of dollars that she got to eventually pay back she was harshly penalized for he crime. *(Document: 5d0408ac26ed0)*

- She helped cover up her crime by claiming she had an inheritance. *(Document: 5d03984dc2628)*
- She was charged for her crimes and had to pay some X millions in restitution *(Document: 5c79375ae3ab2)*

Topic 97 - nav trilliumhealthpartn pringl creditvalley ipc privacycommission

Closely related to: Attack Characteristics - Attack Step

- She was only caught because a ban teller question a large check she attempted to cash. *(Document: 5c6d37122eeb4)*

Topic 99 - gdxdata infowatch deniszenkin doctoralthesi florida hawk

Closely related to: Actor Characteristics - Observed physical behaviour

- She had avoided suspicion on claims that she had received a substantial family inheritance.Today however, she has been fined with more than $60 million, a sever sentence to serve as a warning to others who would engage in fraudulent acts. *(Document: 5d076e64a7364)*
- She had her co-workers fooled as well, but there was some suspicion. *(Document: 5d074d3715040)*
- No suspicion had arisen a that point. *(Document: 5c91e851a6fd6)*
- This activity would often arise suspicion as while she had money to throw around her peers quite often didn't. *(Document: 5c6e8d80126d8)*
- She had claimed that her wealth came from a large family inheritance and whilst rumours and jokes had circulated, she had remained above suspicion. *(Document: 5c6e80c32e77a)*
- She was thought of as a bit odd, eccentric, but generally well liked and considered good at her job so there was little suspicion she was a thief. *(Document: 5c6d37122eeb4)*

Topic 100 - hernandez sledg tasmania upguard american camh

Closely related to:

- and researched it by alerting the company. *(Document: 5d03d561c08b4)*

Topic 104 - knight sydney condénast consumeraffair detectiveactingsuperintendentwatson navi

Closely related to: Attack Characteristics - Attack/Organisation Characteristics - Vulnerability/Opportunity

- They were able to continue their fraud for many years, possibly as long as 18 years, and were able to hide their dealings in an antiquated paper-based documentation system. *(Document: 5d07a0d1acdfa)*
- The person was a middle manager at a firm who committed a large scale cheque fraud. *(Document: 5d07a0d1acdfa)*
- In examing the situation after the fraud was discovered, law enforecement identified up to 9 other individuals who may have participated in the fraudulent scheme. *(Document: 5d04316c13164)*
- will also be a warning for other people who practice fraud to support their lavish lifestyles. *(Document: 5d03dc42937ba)*
- The employee, who has been convicted of fraud, was imposed a large fine ($14,000,000), ordered to make restitution, pay state and local taxes . *(Document: 5d03d561c08b4)*
- She was charged with committing fraud and order to pay over $60 million dollars. *(Document: 5d03cafcd1c68)*
- A tax office employee (middle manager) committed fraud. *(Document: 5d03baf2e2694)*
- Manager at a banking institution was caught creating fraud cheques to the effect of 60 million dollars. *(Document: 5d0399cfc2b32)*
- A company worker committed fraud using an old, paper based system. *(Document: 5d0398b780778)*
- After the fraud was discovered no one is exempt from using the system. *(Document: 5d03984dc2628)*
- The court handed down a severe sentence that served as a warning to others who may be tempted to commit fraud to support their lavish lifestyles. *(Document: 5d036a7f85570)*
- Her, along with some other colleagues are being charged with fraud for a very large sum of money. *(Document: 5d02a27f6c1ec)*
- A tax office manager commited fraud by avoiding the new system implemented to avoid fraud, she nos has to pay a fine. *(Document: 5d027997859c6)*
- The news report stated that the manager had committed a fraud of money and there were no proper security measures in place. *(Document: 5d0256cdddc48)*
- A news report a coworker and a member of the IT staff have expressed their views about a fraud committed by a manager. *(Document: 5d0256cdddc48)*
- The news report stated that the manager had committed a fraud of money and there were no proper security measures in place. *(Document: 5d02545126e8a)*
- A news report a colleague and a member of the IT staff have expressed their views about a fraud committed by a manager. *(Document: 5d02545126e8a)*
- The fraud was discovered when a bank teller questioned a check. *(Document: 5c91e851a6fd6)*
- A tax office employee who worked as a middle-manager for more than 18 years committed fraud to the tune of more than $60 million by writing fraudulent checks. *(Document: 5c91e851a6fd6)*
- She was also responsible for the creation of a new system intended to avoid such fraud. *(Document: 5c87a852e0b46)*
- She also managed to get an exemption from a new computerised system, which would have helped spot the fraud earlier. *(Document: 5c6ff1ec5696e)*
- She was able to commit the fraud due to the lack of auditing and accounting in the paper-based system then in place. *(Document: 5c6f111d877e4)*
- The frauds were committed using the paper based systems that the tax office used. *(Document: 5c6e867e1b300)*

- This incident covers fraud by a middle manager, who was a tax office employee. *(Document: 5c6e867e1b300)*
- This is how she continued to commit fraud. *(Document: 5c6e74be61b98)*
- The manager had been committing fraud using cheques when the office used a paper-based accounting system. *(Document: 5c6e74be61b98)*
- The fraud include a further 9 accomplices, whose roles are still being investigated. *(Document: 5c6d857487398)*
- The fraud was discovered when a bank teller questioned a dubious transaction. *(Document: 5c6d857487398)*
- The fraud involved a potential network of fraudsters, manipulating a paper based system. *(Document: 5c6d530129acc)*
- They carried out the fraud by abusing the lack of checks in the paper system and successfully arguing against the use of a new IT system in her area, that would have imposed more rigorous checking. *(Document: 5c6d3e7557904)*
- They appeared to be living a lavish lifestyle, so that may have been the motive for fraud, although there is no hard evidence about her motivation. *(Document: 5c6d3e7557904)*
- The person who committed the fraud was a manager in a tax office. *(Document: 5c6d3e7557904)*
- Nobody suspected that such a well liked member of staff could be guilty of such a huge amount of fraud. *(Document: 5c6d2abb29bf8)*

Topic 106 - william googl cardsystem educ mastercard ead

Closely related to: Attack Characteristics - Attack Step

- The means was setting up cheques, though the direct recipient is in question. *(Document: 5c87a852e0b46)*

Topic 108 - cia lee cypresssurgerycent nbcnew russian aprilgalvan

Closely related to: Catayst - Precipitating Event/Attack Characteristics - Attack Step/Outcome - Actor

- She got caught eventually because things started looking odd and was told to then pay back the 48 million dollars she had been ordered to pay by a judge. *(Document: 5d061d1464eba)*
- Eventually, all the lies and deception caught up with her once they eventually changed the system in her department. *(Document: 5d02a27f6c1ec)*
- the bank manager was steal money for her own gain she was eventually caught and got jail time *(Document: 5d0288cb57c2e)*
- She was eventually discovered by a bank-teller questioning one of the cheques. *(Document: 5c6ff1ec5696e)*
- Eventually, a new computer system with stricter auditing facilities was rolled out as mandatory to all staff. *(Document: 5c6ea42c1647c)*
- She was eventually take to caught and ordered to payby the money as well as federal and state taxes. *(Document: 5c6d37122eeb4)*

Topic 109 - cvs mitsubishiufjsecur kubo pharmaci walgreen customerspip,в„ў

Closely related to: Outcome - Actor

- She now has to repay all of the embezzled money, plus unpaid taxes and settlement costs. *(Document: 5c6ea42c1647c)*

Topic 116 - socialsecur myfitnessp underarmour hanif albert baylor

Closely related to: Outcome - Actor

- An outside bank employee questioned a single check which lead back to the Tax Office manager. *(Document: 5d079fc2cf7ce)*
- She was middle manager who eluded detection by fraudulently altering paper based records to hide the paper trail that would lead to her being discovered. *(Document: 5d03ba168e184)*
- She was successfully prosecuted along with 9 other individuals. *(Document: 5c6ff1ec5696e)*

Topic 117 - nationwid deniszenkin infowatch’ marketingdirector thedepart ndtv

Closely related to:

- There was no apparent reason why she did it except for her being a theft. *(Document: 5d0408ac26ed0)*

Topic 124 - sawer dataprotectionact procuratorfiscalservic compasshealth dhanju infowatch

Closely related to: Attack Characteristics - Attack

- An employee, mid-manager level, had stolen 60 million over 18 years using over 200 faudulent checks. *(Document: 5d03984dc2628)*
- Maybe, having successfully stolen for a short period without discovery, it became ingrained and she began to simply enjoy the financial spoils of amore affluent lifestyle, as suggested by her colleague. *(Document: 5c6d6e434c6a8)*
- 18 years of money stolen, due to use of a paper system. *(Document: 5c6d530129acc)*
- The total amount stolen was about $60m. *(Document: 5c6d37122eeb4)*

Topic 125 - anthoni denver lawrenc coin prosecutor socialsecur

Closely related to:

- It came as a shock to the company and her co-workers because they thought that she was just being generous and kind when she offered help to those in need. *(Document: 5d076e64a7364)*

Topic 127 - businessinsid norwalk cablevis webb associ atoothbrush

Closely related to: Attack Characteristics - Attack Step

- She was caught as a teller saw an excessively large cheque and has since been charged, whereas her associates have yet to be. *(Document: 5c6fc9bcf2c2e)*

Topic 128 - deutschepost santand mancuso dataprotectionact germany' cambridgeanalytica

Closely related to: Actor Characteristics - Motivation to Attack

- She probably did it for personal gain as I don't recall the story mentioned any clear reason why did she do it. *(Document: 5c79375ae3ab2)*

Topic 129 - danskebank estonian borgen estonia ceo dansk

Closely related to: Actor Characteristics - Historical behaviour/Actor Characteristics - Observed physical behaviour

- Prior to this there had been rumours about the financial situation of the manager after she claimed that she had received a family inheritance. *(Document: 5c77c731bba1c)*

Topic 130 - man iraq liber toronto canada oct

Closely related to:

- The system was designed to combat against fraud and illegal actions. *(Document: 5d0410658893c)*

Topic 132 - finra gao cfi lfs fisma lfa

Closely related to: Organisation Characteristics - Vulnerability/Opportunity

- In the world new system to manage financial transaction, but we have worked old auditing and accounting paper systems, So did not provide & operating new system. *(Document: 5d03c9147ba66)*
- This old system did not have any controls for auditing or inspection. *(Document: 5c77c731bba1c)*
- However it was not straightforward to discover the extent of her fraudulent activities as there was no clear audit trail for her as she had been operating outside the computer system. *(Document: 5c73c9a927a4c)*
- Person was a female middle-manager at a company, operating with the assistance of nine others, managing to abuse the paper auditing (?) system to steal large amounts of money. *(Document: 5c6fc9bcf2c2e)*
- From a work perspective she was deemed experienced, important and knowledgeable enough to be involved in the creation of the new auditing system, but when it was implemented and she protested that it was not workable within her group, and despite the IT Group's insistence, did not have to work to it. *(Document: 5c6d6e434c6a8)*

Topic 136 - googl bochum german hipaa lgttreuhand liechtenstein

Closely related to: Actor Characteristics - Psychological State/Actor Characteristics - Personality characteristics/Actor Characteristics - Historical behaviour/Attack Characteristics - Attack Step

- She claimed she got a money as a result of receiving a family inheritance. *(Document: 5d07afae00e24)*
- She claimed to get her money from a family inheritance *(Document: 5d07ad29c4338)*
- She said she received a large family inheritance, so that was why she could spend so lavishly. *(Document: 5d03fbe74cb6c)*
- She claimed it was from a family inheritance. *(Document: 5d03baf2e2694)*
- She had told everyone that she had a large family inheritance. *(Document: 5d039950a0ffa)*
- She had avoided suspicion by claiming to be in receipt of a substantial family inheritance. *(Document: 5d036a7f85570)*
- People questioned her wealth through roumers about having a gambling addiction however, she claimed it was through family inheritance. *(Document: 5c6e9bcfbe44c)*
- She covered up her access to large sums of money by referring to a family inheritance. *(Document: 5c6e867e1b300)*
- She claimed the money came from family. *(Document: 5c6d530129acc)*

Topic 137 - johnson cole planetfit georgiapacif visa fbi

Closely related to:

- Her employees like her even though they thought she was a bit weird, but no one suspected her in stealing money. *(Document: 5d07dfe6a9344)*
- --> It is suspected that she also had accomplices who helped her in covering up the paper based records. *(Document: 5d0481de7bb2e)*
- --> Nobody suspected her and her fraudulent activities went undetected until one day a bank teller reported a suspicious cheque. *(Document: 5d0481de7bb2e)*
- Nobody suspected her of theft until it was uncovered. *(Document: 5d044fa461008)*
- No one suspected her. *(Document: 5d03fbe74cb6c)*
- Her co-workers never suspected it. *(Document: 5d03e10296384)*
- Her coworkers joked about it, but never really suspected ti was as bad as it was. *(Document: 5d039c516ae96)*
- The manager is a nice person that no one suspected would be involved in theft for the last 18 years. *(Document: 5d035b56d3d10)*

Topic 140 - lee figueroa hagen wachovia bofa grandlarceni

Closely related to:

- There are 9 possible accomplices but no specific information on involvement has been released. *(Document: 5d03baf2e2694)*

Topic 142 - showpo blackswallow internet web abovethelaw dd4bc

Closely related to: Attack Characteristics - Attack

- She has been writing herself checks for various amounts over the last 18 years. *(Document: 5d079eb114fb4)*
- The manager was well known as being affluent by her coworkers, there were rumors that she inherited a fortune or had a gambling addiction. *(Document: 5d0410658893c)*
- A middle level manager at the tax office was writing fraudulent cheques for over 18 years. *(Document: 5d03ea6d73c8a)*
- One IT person comment that she possibly realize using the new system will get her caught red handed so she refuse to use the new system. *(Document: 5d03d83dbd8f8)*
- She was a well regarded employee by her coworkers and management. *(Document: 5d03d561c08b4)*
- A well like and popular middle manager at a bank was caught and arrested for stealing from the company. *(Document: 5d03cafcd1c68)*
- A manager at a tax office was stealing money by manipulating paperwork and writing bad checks. *(Document: 5d039c516ae96)*
- A middle manager was writing fraudulent checks for 18 years. *(Document: 5d039b3974a04)*
- A manager was able to write fraudulent checks from the company for over 18 years. *(Document: 5d039950a0ffa)*
- A tax office bank manager wrote fraudulent cheques over 18 years to steal money from her employer. *(Document: 5c6ecc9f2513e)*
- A manager at a Tax Office was involved in writing / authorising fraudulent cheques. *(Document: 5c6e687b04128)*
- She was well liked by her staff, is sometimes though to be slightly 'odd', but certainly generous and protective of those she managed (on the face of it). *(Document: 5c6d6e434c6a8)*
- Tax office manager was writing fraudulent cheque's, 200 of them to be exact. *(Document: 5c6d17f2c581e)*

Topic 146 - darren fiatfinancialservic cdrom countydurham montgolfi norway

Closely related to: Actor Characteristics - Personality characteristics

- She probably wanted the money, but she was a nice and generous person with her employees, like buy food and drinks and help those who needed help. *(Document: 5d079eb114fb4)*
- All of her coworkers thought she was a nice person. *(Document: 5d056c1ca9952)*
- Everybody thought she was weird, but was nice. *(Document: 5d03fbe74cb6c)*
- When uncovered, everyone was surprised because the employee was always nice and generous towards everyone and could never imagined her stealing money for so long. *(Document: 5d03d561c08b4)*
- Staff thought she was a little quirky and sometimes joked about where she got her money, but they never did anything about it because she bought the occasional round at the bar and did other nice things for staff. *(Document: 5d03cbf7e42dc)*
- Co-workers were shocked as she seemed really nice. *(Document: 5d035b56d3d10)*
- Her colleagues were somewhat surprised by this as she always came across as a nice person to be around, although some did point at her being a bit flaky. *(Document: 5c6fee6c3665a)*
- Her colleagues had viewed her as being nice, caring and happy. *(Document: 5c6e9bcfbe44c)*
- The manager was seen to be a nice person, always willing to help and be there for her staff - even though some of them thought she was a 'bit flaky'. *(Document: 5c6d591491dac)*

Topic 148 - ladbrok themail daniel mrsubasingh usd commiss

Closely related to: Actor Characteristics - Personality characteristics/Actor Characteristics - Observed physical behaviour/Attack Characteristics - Attack

- Got caught stealing by one of the bank tellers noticing something wrong with a check. *(Document: 5d083763eb280)*
- The tax lady was stealing money to pay for her lavish lifestyle.The tax lady was real sneaky for 18 years *(Document: 5d07ee6fa058c)*
- They thought she was odd and were annoyed that she was richer than they were but never thought she was stealing. *(Document: 5d03e10296384)*

- Later on it was discovered that she had been stealing money for over 18 years. *(Document: 5d03d83dbd8f8)*
- She was stealing money from the company though she was so kind and sweet and the workers never though that she could steal from the company. *(Document: 5d03ba168e184)*
- She had been stealing money in this way for up to 18 years. *(Document: 5d035a10e5b0a)*
- Her workmates always thought she was a bit weird but still loved her, and it came as a surprise that she was stealing from the company. *(Document: 5c88f8eb2d622)*
- \A manager was using the cover of trying to stop corruption and stealing from the company, while stealing from the same company for 18+ years. *(Document: 5c701353965dc)*

Topic 150 - walker armenianpow cleveland kellywalk losangel ohio

Closely related to: Outcome - Actor

- A tax officer employee was fined 16 million dollars for issuing over 200 fraudulent cheques. *(Document: 5d07afae00e24)*
- A tax officer employee was fined 16 million dollars for issuing over 200 fraudulent cheques. *(Document: 5d07ad29c4338)*
- She then went to court and was find 60 million dollars, of which was broken up into different groups of where the money had to go. *(Document: 5d079eb114fb4)*
- She was ordered to pay 60 million dollars. *(Document: 5d056c1ca9952)*
- The Tax manager was stealing million of dollars from the company. *(Document: 5d0408ac26ed0)*
- She was fined 18 million dollars and caught after years of doing this. *(Document: 5d03e10296384)*
- She was fined 60 million dollars. *(Document: 5d03baf2e2694)*
- The tax office employee was fined more than 60 million dollars. *(Document: 5d03ba168e184)*
- The story is about a tax office employee who fraudulently issued over 200 million dollars worth of fraudent cheques. *(Document: 5d03ba168e184)*
- She is being charged several million dollars in fines. *(Document: 5d039c516ae96)*
- She was fined 60 million dollars. *(Document: 5d039b3974a04)*
- She was caught when a teller questioned the nature of a 400,000 dollar check. *(Document: 5d039b3974a04)*
- A middle manager (and a network of fraudsters) was stealing several million dollars from a banking company. *(Document: 5d03998389dc8)*
- After being caught the manager was asked to pay over $60 million dollars in reprimands and taxes, both federal and local. *(Document: 5d02dc8e97cb6)*
- A tax office middle manager was fined 60 million dollars. *(Document: 5d0255245f406)*
- A manager working at a tax office committed fraud over 18 years, stealing tens of millions of dollars. *(Document: 5c6f111d877e4)*
- After events, the person involved was fined and had to repay an amount that was in millions of dollars. *(Document: 5c6e74be61b98)*
- She received a prison sentance and had to pay back millions of dollars. *(Document: 5c6d7e0262872)*

Topic 154 - accuweath newindianexpress revealmobil claxtonhepburn claxtonhepburnmedicalcent hipaa

Closely related to: Organisation Characteristics - Vulnerability/Opportunity

- The paper based system was a legacy system which had been in the process of getting replaced. *(Document: 5d079fc2cf7ce)*
- She manipulated paper based system in order to write fraudulent cheques, that is how she was able to avoid detection. *(Document: 5d03ea6d73c8a)*
- The company had a paper based system which was easy to manipulate and she refused to use a new IT system for transactions when it was implemented. *(Document: 5d035a10e5b0a)*
- She was manipulating paper based transactions and issuing fake cheques for over 18 years. *(Document: 5d0255245f406)*
- As she was still using a paper based system she was able to better hide her activities. *(Document: 5c7f9270c1494)*
- A new computer system had been introduced which the manager had been quite vocal against, despite being involved with its development, and fought to be kept on the own paper based system. *(Document: 5c77c731bba1c)*
- A manager at a bank was stealing money by manipulating a old paper based system. *(Document: 5c6fe9ae6757a)*
- This had been going on for a while as the current system was a paper based system which didn't leave a paper trail. *(Document: 5c6fce7530a52)*
- She had obviously done it to pocket additional money on top of her salary. *(Document: 5c6ecc9f2513e)*
- She had been stealing money for 18 years through the abuse of the paper based cheque system. *(Document: 5c6e9bcfbe44c)*
- Even the more senior management relented and allowed her group to continue with the paper based system. *(Document: 5c6e877f639e8)*
- The individual concerned used lack of accountability within the paper based tax system to hide fraudulent activity over a period of 18 years. *(Document: 5c6e877f639e8)*
- No traceability due to paper based system. *(Document: 5c6d530129acc)*

Topic 155 - swiss german lapour germani umfogl germany’

Closely related to: Actor Characteristics - Observed physical behaviour

- This came as a shock to many because she was always kind and generous buying rounds at the pub for everyone. *(Document: 5d03cffd39760)*
- Some people were more than happy about her splashing around her money, buying drinks and food for them, she'd explained it had come from inheritance though some of them speculated it may have come from a gambling addiction and were more suspicious. *(Document: 5c6fee6c3665a)*
- She had avoided suspicion by being kind and generous to her colleagues (buying lunch for example), but she was described as odd/weird at times. *(Document: 5c6ecc9f2513e)*
- She would buy the team drinks and food and be quite gregarious at times. *(Document: 5c6d6e434c6a8)*

Topic 156 - mrwallac dataprotectioncommission doj mrshatter revenu swiss

Closely related to:

- She was given a lengthy sentence by the judge. *(Document: 5d03d0e7b9686)*

Topic 157 - usd boe newburgh staton ftc clear

Closely related to: Actor Characteristics - Motivation to Attack/Organisation Characteristics - Vulnerability/Opportunity/Outcome - Organisation

- It's not clear why she did it. *(Document: 5c7939498b4b6)*
- What is not clear is if they were from the same group nor how large that group was. *(Document: 5c6e877f639e8)*

Topic 161 - gunn colonialcommunitycorrect jamescitycounti guardian disney fsa

Closely related to: Actor Characteristics - Skill Set/Organisation Characteristics - Vulnerability/Opportunity

- The manager was involved in setting up the new IT system so that she could determine whether or not she could still continue her current fraudulent activities without being caught and when she realised this wasn't going to be the case she used her status and known personality to get around the new system and have her whole team exempt to avoid detection / questioning as to why only her and her accomplices required the exemption. *(Document: 5c6e687b04128)*

Topic 168 - houston southkorean ncsoft spec beij globaltim

Closely related to: Attack Characteristics - Attack

- She submitted over 200 fraudulent cheques over a time period of approximately 18 years. *(Document: 5c73c9a927a4c)*

Topic 169 - ddos gammel august minnesota usdepart bhhc

Closely related to: Attack Characteristics - Attack

- The scheme centered on a Tax Office department manager and several other associates, who utilized the Office's paper based system to get the checks processed. *(Document: 5d079fc2cf7ce)*
- A tax office worker embezzled money for over 18 years. *(Document: 5d079e20e985e)*
- The story is about a tax office employee. *(Document: 5d076e64a7364)*
- She helped develop a new accounting system on computer, but insisted her office continent to use paper, so that she could hide her activities. *(Document: 5d056c1ca9952)*
- A tax office manager was found guilty for embezzling funds from the office. *(Document: 5d056c1ca9952)*
- This helped her to continue her fraudulent activities. *(Document: 5d0481de7bb2e)*
- When caught, the employee was fined over $60 million and ordered to pay restitution to the company, as well as back state and federal taxes. *(Document: 5d04316c13164)*
- A middle management employee in a tax office was discovered to have been embezzling from the company for over 18 years. *(Document: 5d04316c13164)*
- A manager in an office refused to use a new payment system claiming that her office couldn't get the hang of it. *(Document: 5d0410658893c)*
- She was going to have to pay back over $50 million in restitution. *(Document: 5d03fbe74cb6c)*
- A tax office employees was stealing money by issuing checks to herself. *(Document: 5d03e10296384)*
- It is a shock to everyone in the office because she is consider kind, generous and helpful to people. *(Document: 5d03d83dbd8f8)*
- This way, the manager and some accomplices, who have yet to be charged, were able to continue stealing money from the company and hiding the transactions on their paper system. *(Document: 5d03d561c08b4)*
- 48 million in restitution. *(Document: 5d03baf2e2694)*
- There were oddball rumors but they were dismissed because the office always has rumors. *(Document: 5d03984dc2628)*
- A tax office employee who issued over 200 fraudulent cheques was fined more than $60 million today. *(Document: 5d036a7f85570)*
- They thought she had, perhaps, a gamboling addiction. *(Document: 5d02a27f6c1ec)*
- She was handed down a pretty severe sentence having to pay back 70 million in restitutions including taxes out of that. *(Document: 5d024904967da)*
- She was fined and is required to pay restitutions. *(Document: 5c91e851a6fd6)*
- She was reluctant to agree to switching her sector of the office to the new, electronic system, claiming the office couldn't operate with it even though she helped design it. *(Document: 5c91e851a6fd6)*
- She was seen as an easygoing person not bothered by the pressures of an office environment. *(Document: 5c91e851a6fd6)*
- This person was a manager at a tax office. *(Document: 5c879871c869a)*
- Tax Office employee committing fraud and embezzling money. *(Document: 5c7c178e1c62e)*
- An updated computer system that she helped implement to prevent exactly the kind of fraud she was committing failed to stop her because she was granted an exception from using it and was instead allowed to continue using the older paper-based system. *(Document: 5c730ce477c72)*

- This happened due to her continuing to use an old system that was inefficient for verifying questionable transactions. *(Document: 5c701387c463a)*
- A tax office middle manager with the collusion and possible participation of her team has been issuing fraudulent cheques for 18 years. *(Document: 5c6ff1ec5696e)*
- A middle manger in a tax office was charged for issuing over 2,000 fraudulent cheques. *(Document: 5c6fce7530a52)*
- It was only office rumours and so nobody took much notice, taking her generosity at face value. *(Document: 5c6ea42c1647c)*
- The person was a tax office employee in middle management who embezzled $14 million over the course of 18 years. *(Document: 5c6ea42c1647c)*
- People at the tax office reported that she was considered kind and generous, although somewhat weird. *(Document: 5c6e867e1b300)*
- Subsequently, management at the tax office have made use of the computer system mandatory for all staff. *(Document: 5c6e867e1b300)*
- A manager at a Tax Office had been stealing (via fraudulent cheques) for over 18 years. *(Document: 5c6d857487398)*
- Manager in the tax office, and 9 other accomplices. *(Document: 5c6d7e0262872)*
- She worked for the tax office. *(Document: 5c6d530129acc)*

Topic 170 - universityhealthcar publichealth sanfranciscogeneralhospit juror kroll martinsburg

Closely related to: Outcome - Actor

- Criminal charges were bought and the individual was fined $60 million. *(Document: 5c6d2b8c9ad3a)*

Topic 177 - chen williamsstevenson california knowl mrjone sherifftierney

Closely related to: Outcome - Actor

- She was prosecuted and given a sentence but also order to pay back te money and taxes and costs *(Document: 5d035a10e5b0a)*
- After she was caught she was charged a server sentence and fined for millions. *(Document: 5c6fce7530a52)*
- The news reports describe her being fined about $63M, although there is no mention of a custodial sentence. *(Document: 5c6d3e7557904)*

Topic 179 - infowatch deniszenkin palomarhealth rosbank russian moscow’

Closely related to: Actor Characteristics - Job Info

- The person was a female middle manager in the tax office, although it is thought she was operating as part of a criminal network as a further 9 individuals have been linked to this fraud case but their roles are not yet fully understood. *(Document: 5c73c9a927a4c)*
- The person after the events had to pay restitution and the company mandated no more exemptions from using the new IT system. *(Document: 5c701353965dc)*
- The person of interest in this case an employee in a company, more specifically she is part of middle management. *(Document: 5c6e8d80126d8)*

Topic 191 - lavarnway yto investig wsoc americanqualityexterior bethani

Closely related to: Attack Characteristics - Attack

- She was a flakey but good boss and co-worker. *(Document: 5d039b3974a04)*
- A women who works for a company was stealing from them for over 18 years, she was well-loved and appreciated by her colleagues to the point where her bosses allowed her and her team to be excepted from using the new system. *(Document: 5c88f8eb2d622)*

Topic 193 - cookiepolici javascript keepabreast comput allianzinsur anheuserbuschco

Closely related to:

- She was finally caught when a new computer system was installed and she refused to uses. *(Document: 5d079e20e985e)*

Topic 198 - queensland brisban sember sember’ wentworth awoman

Closely related to: Actor Characteristics - Motivation to Attack

- She did it to keep funding a certain lifestyle she was accustomed too. *(Document: 5c6fe9ae6757a)*

Topic 199 - hipaa thelendingcompani centric phi vvmc brown

Closely related to:

- Other employees wondered where she was getting her money from but never questioned it a great deal. *(Document: 5d03de5cd6ad8)*

Topic 202 - cfaa unitedst bolen clayton thefifthcircuit thoma

Closely related to: Outcome - Actor

- An employee stole from a bank by cashing fake checks she eventually got caught and was charged and sentenced *(Document: 5d079f96dae62)*
- After these events the employee was given a 'severe sentence' and was forced to pay back the monies, but I can't quite remember the exact break down - something like $48M, $12M and $3.2M costs. *(Document: 5c6d6e434c6a8)*

Topic 204 - brexit ★★★★☆ england scotland dublin german

Closely related to:

- She was easily able to hide her trail in an outdated paper system while everyone else had upgraded to a digital system. *(Document: 5d024904967da)*

Topic 205 - turkish css dropbox turkey banccentr googledr

Closely related to: Attack Characteristics - Attack Step Goal/Organisation Characteristics - Vulnerability/Opportunity

- come up with a new computerized system but she insisted she could not work with it and was allowed to exempt her Dept. *(Document: 5d03baf2e2694)*
- Despite a recently implemented computer system, designed to avoid the possibility of fraud, the manager was allowed to operate outside the system allowing her scheme to continue for so long. *(Document: 5d036a7f85570)*
- She was able to commit the fraud by manipulating paper based records without detection as, at her insistence, higher management allowed an exception to be made: her department were allowed to operate outside the recently implemented computer system which added a layer of auditing and accounting. *(Document: 5c73c9a927a4c)*
- She began to help IT install a new, computer-based system, and upon realising that this would expose her she used her nine associates and managed to get her team to stay on paper, thus allowing her to continue her theft. *(Document: 5c6fc9bcf2c2e)*

Topic 216 - cbc anintrud arab eilat huawei leed

Closely related to:

- she was essentially stealing money from the company and nobody knew about it. *(Document: 5d03983ac17c8)*

Topic 219 - tulli inpex keddi cbcnew colleg druginformationsystem

Closely related to: Actor Characteristics - Motivation to Attack/Actor Characteristics - Personality characteristics/Actor Characteristics - Observed physical behaviour

- She stole the money she did to maintain her extravagant lifestyle. *(Document: 5c730ce477c72)*
- She maintained a friendly and helpful persona to try and 'blend in' with the team and her demeanor hid her true intentions. *(Document: 5c6e687b04128)*

Topic 221 - rbi ernst hitachi yesbank atm ceo

Closely related to: Attack Characteristics - Attack/Attack Characteristics - Attack Step/Outcome - Actor

- Following an extensive police investigation, she will have to repay over $60m, including federal and state taxes of over $15m. *(Document: 5c6e80c32e77a)*
- She managed to steal $60M and was caught when a $400,00 cheque was questioned by a bank teller. *(Document: 5c6e687b04128)*

Topic 224 - shell royaldutchshellplc oil royaldutchshel gas tag

Closely related to:

- Interestingly, she helped design the new computer system. *(Document: 5d04581fba50e)*

Topic 227 - policescotland aug css detail inc ma01801

Closely related to:

- Some people questioned where she received so much money, but figured it was from her addiction to gambling. *(Document: 5d03cffd39760)*

Topic 230 - noozhawk fleur delanoi fresno healthcarecent investig

Closely related to: Actor Characteristics - Observed physical behaviour

- Her scheme was connected to a paper record system that the bank used. *(Document: 5d04581fba50e)*
- She explained her wealth through getting inheritance money. *(Document: 5c6ecc9f2513e)*

Topic 232 - oath trustedcloud coredesktop likeoath manag mip

Closely related to: Attack Characteristics - Attack

- The bank manager. *(Document: 5d083763eb280)*
- The middle manager was stealing money from financial institution. *(Document: 5d07dfe6a9344)*
- She is the middle manager who over the past 18 years had issued over 200 fraudulent cheques. *(Document: 5d076e64a7364)*
- A woman basically was operating as a manager at her business. *(Document: 5d061d1464eba)*
- A Manager at a bank stole 60 million GBP over the course of 18 years. *(Document: 5d04581fba50e)*
- The manager and 6 accomplices were handed lengthy prison sentances. *(Document: 5d044fa461008)*
- The person who was guilty was a manager. *(Document: 5d03fbe74cb6c)*
- there was a new system in place to manage financial transactions and make it more foolproof the middle manager was able to bypass it using her influence and was able to continue with her fraud practices until the bank teller notice suspicion.the *(Document: 5d03dc42937ba)*
- The person was a female manager who despite having participated in launching a new system, refused to use it. *(Document: 5d03d0e7b9686)*
- The manager had been stealing money for 18 years. *(Document: 5d03cc65150e0)*
- we told the manager, i how was used the new system, and next few days told by our management, they could continue to operate with the legacy paper systems. *(Document: 5d03c9147ba66)*
- However she was caught because a bank teller reported a suspicious cheque the manager had issued. *(Document: 5d03ba168e184)*
- She was a manager in a company and stole more than $60 million. *(Document: 5d039b5fd1d3a)*
- Though noted as being a bit odd, the manager was well-liked among coworkers and subordinates (possibly because she frequently bought them drinks and so forth). *(Document: 5d03998389dc8)*
- The manager was involved in the creation of the new program - but she must have been aware she would be caught and preffered the old style of doing things. *(Document: 5d035b56d3d10)*
- a female manager was making cheques payable to herself as a way of stealing money from her company. *(Document: 5d035a10e5b0a)*
- The manager was charged with a fine after the investigation. *(Document: 5d0256cdddc48)*
- The manager of the IT staff also had good views about the manager and was shocked to learn about the fraud. *(Document: 5d0256cdddc48)*
- The co worker stated that the manager looked very king and generous and no one expected such a behavior from her. *(Document: 5d0256cdddc48)*
- The manager of the IT staff also had good views about the manager and was shocked to learn about the fraud. *(Document: 5d02545126e8a)*
- The co worker stated that the manager looked very king and generous and no one expected such a behavior from her. *(Document: 5d02545126e8a)*
- A manager at a firm had been developing a new software system with the IT team. *(Document: 5d024811e412e)*
- The woman managed to fool not only her staff and friends but also those in a position to find out what she had been doing. *(Document: 5c7f9270c1494)*
- With the help of a group of other people she managed to hide her activities for 18 years and managed to siphon off the money by issuing over 200 cheques. *(Document: 5c7f9270c1494)*
- Employee was a manager and was issuing bogus checks. *(Document: 5c7c178e1c62e)*
- The person was a female manager. *(Document: 5c79375ae3ab2)*
- The manager had been caught creating fraudulent checks after a teller had raised suspicions about a cheque she had written. *(Document: 5c77c731bba1c)*
- A middle manager at a company was stealing money over a period of 18 years, she managed to do this by using a paper based accounting system rather than the new IT based one, even though she helped design the new system. *(Document: 5c6fee6c3665a)*
- The manager was given a multi-million pound fine. *(Document: 5c6f111d877e4)*
- She had a great track record with management, so they provided an exemption. *(Document: 5c6ea42c1647c)*
- The lady in question was a middle manager and managed a department within the organisation. *(Document: 5c6e9bcfbe44c)*
- This manager was actually quite liked and would often treat her colleagues to lunch or drinks. *(Document: 5c6e8d80126d8)*
- A recent move to a computer based financial transaction system with built in audit and accounting measures was by-passed by the manager, by saying that she couldn't use it and gaining an exemption from using it from higher management. *(Document: 5c6e867e1b300)*
- She managed to steal about $48M, by means of over 200 fraudulent cheques. *(Document: 5c6e867e1b300)*
- The manager was imprisoned and had to pay restitution (and taxes). *(Document: 5c6d857487398)*
- When a bank teller raised a concern about a cheque the manager tried to cash, it alerted the company to further discrepancies, and an internal investigation discovered that the manager in question had been defrauding the company for the past 18 years, to the tune of millions of dollars. *(Document: 5c6d2abb29bf8)*
- A manager in a company was seen as generous and helpful, however she insisted that her department should not use a new computerised system for managing financial transactions because it was deemed as too complicated. *(Document: 5c6d2abb29bf8)*

Topic 234 - snowden carr nsa hawaii crummel islandhealth

Closely related to: Outcome - Actor

- She had to pay back millions in restitution, state, and federal taxes. *(Document: 5d03c707b6dfa)*
- She was doing it by manipulating paper records to hide large sums being moved around, and she was seemingly having her accomplices do the same. *(Document: 5d03998389dc8)*
- Law enforcement discovered a network of fraudsters who were easily manipulating paper-based records without detection. *(Document: 5d036a7f85570)*
- She had to pay back several millions in restitutions, federal taxes, and state taxes. *(Document: 5c879871c869a)*

Topic 241 - canada fernando dhb gog addit equifax

Closely related to:

- tons of people were arrested additional 9 people were taken into custody. *(Document: 5d03c72fc0576)*

Topic 247 - shapeshift autopac cree hellsangel introductori luke

Closely related to: Actor Characteristics - Personality characteristics

- A colleague involved in the development of the new system described this as "suspicious in hindsight" but stated that it was simply "hard to see her as a thief". *(Document: 5c6e80c32e77a)*

Topic 253 - suntrust kraus lincolnshir pera eddiemerlot' hyde

Closely related to: Attack Characteristics - Attack/Attack Characteristics - Attack Step

- The scheme came to light when a bank teller queried a suspicious cheque for $400,000. *(Document: 5c6d2b8c9ad3a)*

Topic 258 - atm nashvill passcod fattah folad svt

Closely related to: Attack Characteristics - Attack Step

- When a bank teller queried a $400k cheque, an investigation was launched and the fraud uncovered. *(Document: 5d044fa461008)*
- I think I'm guessing with the cheques bit just because a bank teller had finally picked up on a $400k fraudulent one? *(Document: 5c6d6e434c6a8)*

Topic 262 - henkovink dutch verizon netherland dawnzimm dike

Closely related to:

- She was able to convince upper management to exempt her department from utilizing the IT system which allowed her to hide her crimes through her paper system. *(Document: 5d0800d5b2674)*

Topic 265 - deloitt nasd infowatch morganstanley touch armstrong

Closely related to: Actor Characteristics - Personality characteristics

- The employee, a middle level manager, was said to be kind and generous. *(Document: 5d07afae00e24)*
- The news came as a surprise to many who say that she was kind and generous. *(Document: 5d07ad29c4338)*
- The manager is described as kind and generous by her colleagues. *(Document: 5d0481de7bb2e)*
- The news was a shock to the office because the manager had always behaved in a kind and generous way with everyone. *(Document: 5d037c3adef6c)*
- The news came as a shock to the office who had found the manager to be a kind and generous co-worker who frequently supported those in need. *(Document: 5d036a7f85570)*
- Reports say that the manager was a kind and friendly person and no one suspected her of being a thief, despite her oddities. *(Document: 5d02dc8e97cb6)*
- The woman spun a web of lies to hide where the money was coming from and was kind to her friends and staff, enabling her to keep her secret. *(Document: 5c7f9270c1494)*
- But she was kind and generous to her colleagues and was known to frequently support those in need. *(Document: 5c73c9a927a4c)*
- Therefore due to her kindness and generosity, the fraud seemed out of character and came as a shock to her colleagues. *(Document: 5c73c9a927a4c)*
- A new IT system was brought in which would have made it harder to manipulate the records, therefore making it harder to commit fraud, but the manager used her influence to give her department exemption from using the new system. *(Document: 5c6ecc9f2513e)*
- The manager appeared kind, generous and supportive of people in her group. *(Document: 5c6e877f639e8)*
- The actions of the manager, who was described as kind and generous by one of her colleagues, was discovered after a teller at the bank questioned a cheque she had written for over $400,000. *(Document: 5c6e80c32e77a)*
- Colleagues perceived the manager as kind and generous, but there were some jokes and rumours about where they got their money from. *(Document: 5c6d857487398)*
- A middle manager, female, known as kind and generous/nice and understanding committed financial fraud. *(Document: 5c6d530129acc)*

Topic 272 - nhs airforc allen benion million mrgraham

Closely related to: Outcome - Actor

- Now she has to pay $60 million in fines. *(Document: 5d07dfe6a9344)*
- She is to pay back $48 million in restitution, $12 million in federal taxes and $3.2 million in state taxes. *(Document: 5d076e64a7364)*
- She was ordered in the end to pay 48 million in restitution, 12 million in back taxes, and another quantity of money as well *(Document: 5d04581fba50e)*
- She owes 14 million in restitutions and 10 million in taxes. *(Document: 5d0410658893c)*
- She was arrested and has to pay 48 million dollars in restitution and some millions in federal and state taxes. *(Document: 5d03ea6d73c8a)*

- court had severed her with severe sentence to pay $48 million in restitution,$12 million in federal taxes and $3.2 million in state taxes.this *(Document: 5d03dc42937ba)*
- The manager has to pay back 48 million and 3.2 million in taxes. *(Document: 5d03cffd39760)*
- 12 million in federal taxes and 20 million in state taxes. *(Document: 5d03baf2e2694)*
- She had to make a payback of 48 million dollars in restitution, 12 million dollars in federal taxes and 3.2 million dollars in state taxes. *(Document: 5d03ba168e184)*
- In the end, she has to pay back millions in taxes. *(Document: 5d0399cfc2b32)*
- In the end she was charged and was to paid upwards of $60 million. *(Document: 5d039950a0ffa)*
- The manager will have to pay back $48 million in restitution, $12 million in federal taxes and $3.2 million in state taxes." *(Document: 5d036a7f85570)*
- Court directed her to pay back 48 million, 12 million tax, and 3.2 state tax. *(Document: 5d035b56d3d10)*
- She was caught out by a bank teller who noticed a discrepancy in one of her cheques, after she was caught out, everybody at the company was required to use the new system - zero exceptions - and she was fined $3.8 million, even though it was estimated that she stole as much as $60 million. *(Document: 5c88f8eb2d622)*
- After the events, she was fined over 60 million. *(Document: 5c879871c869a)*
- Nevertheless it was discovered and she incurred a severe sentence and was fined more than $60 million which comprises: $48 million in restitutions, $12 million in federal taxes, and $3.2 million in state taxes. *(Document: 5c73c9a927a4c)*
- After being found out, she was fined $60 million including $48 million in restitution. *(Document: 5c730ce477c72)*
- She ended up serving a severe sentence and ordered to pay back 48 million dollars, 12 million in tax and 3.2 million in state tax. *(Document: 5c6d591491dac)*
- She was ordered to pay back $48 million in restitution, $12 million in federal taxes and $3.2 million in state taxes *(Document: 5c6d17f2c581e)*

Topic 273 - mondaq valleychildren' contributor universitypediatricspecialist britishcolumbia cysticfibrosi

Closely related to: Outcome - Actor

- She was fined heavily by the court for a total of about $30,000,000. *(Document: 5c7939498b4b6)*

Topic 280 - deflorin pera reserv dunmil ristad assistantramseycountyattorneyjohnristad

Closely related to: Actor Characteristics - Motivation to Attack/Actor Characteristics - Job Info/Organisation Characteristics - Vulnerability/Opportunity/Outcome - Actor/Outcome - Organisation

- Motive is not given, only rumours of gambling habits, or the news report suggesting an extravagant lifestyle; which the rumours again seem to suggest, but do not confirm. *(Document: 5c87a852e0b46)*
- The higher management gave her department exemption to not use to system and enforced usage on the rest of the organisation. *(Document: 5c6e9bcfbe44c)*
- The person involved was a middle manager who had been with the organisation for some time (given that the fraud have been happening over an 18 year period). *(Document: 5c6e74be61b98)*
- The investigation also cast suspicion on 9 other members of staff, leading to allegations of a network of insiders and spawning further investigations *(Document: 5c6d2b8c9ad3a)*

Topic 283 - swiss cex franc hsbc newyork switzerland

Closely related to: Outcome - Organisation

- The news report stated that there was likely a network of people involved, but this was not mentioned by either of the organisation's employees. *(Document: 5c6e74be61b98)*
- No exemptions from using the system from now on within the organisation. *(Document: 5c6d530129acc)*

Topic 284 - timewarn infowatch repres tjxcompani administr affair

Closely related to: Attack Characteristics - Attack Step

- The company got a new electronic tracking system, but that middle manager refused to use it and company allowed that. *(Document: 5d07dfe6a9344)*
- A tax office employee was fined more than 60 million dollars for issuing fraudulent checks for more than 18 years.Eventually one day a bank teller got suspicious of one the checks and that's how she got caught.After a long investigation 9 more people were found guilty and it is still unclear what charges will be put against everyone since their complete role in this fraud is yet to be completely understood.She used buy a round for her employees at the pub and her co-workers are shocked to hear about this.Everybody loved her but nobody thought of her as a thief so they were all shocked. *(Document: 5d076a47dab6a)*
- They noted that she seemed to have more money than they did and was rather generous, but they presumed she had a gambling problem, not theft. *(Document: 5d04581fba50e)*
- The employee was able to get away with the fraud for so long because she resisted using a new IT accounting system that would have exposed her actions so her managers exempted her department from utilizing the system. *(Document: 5d04316c13164)*

- Fellow employees were shocked to find out about the theft because the employee had always seemed so kind and generous, despite being a bit odd. *(Document: 5d04316c13164)*
- It was only because a bank teller questioned a check she had written for $400,000 that she got caught. *(Document: 5d03fbe74cb6c)*
- That was why she got away with it for so long. *(Document: 5d03fbe74cb6c)*
- over an 18 year period she was taking the money in that way. *(Document: 5d03e10296384)*
- an employee had been stealing from the company for 18 years. *(Document: 5d03de5cd6ad8)*
- She was a kind character and liked by most in the office, she had told people that she had inherited a large sum of money to explain her lavish lifestyle and people liked to be treated by her to free lunches etc.. no one suspected that she was stealing from the company even though there were signs and rumers. *(Document: 5d03d59ae268a)*
- The company had introduced earlier a new computerised system to replace the use of papers for transactions, some of the employees were reluctant to use the new system but went along with it except for that same manager as she refused to use it and demanded that her department be excluded from using that system as she wouldn't be able to continue her scam and would be caught. *(Document: 5d03d59ae268a)*
- The employee wrote 200 fraudulent checks before a bank teller questioned a check for $400,000.00 *(Document: 5d03d561c08b4)*
- It was said that she was generous and liked to help others, it could be suggested that she had low self esteem and wanted to be liked and needed. *(Document: 5d03d0e7b9686)*
- She got caught by writing a check that the bank teller grew suspicious of. *(Document: 5d03cc65150e0)*
- Ultimately, a bank worker questioned an unusual check that eventually led to the discovery that she has been frauding the company for 18 years. *(Document: 5d03cbf7e42dc)*
- I must have realised that the new system would have caught me. *(Document: 5d03c9147ba66)*
- The only location that was not on the new software that the manager opted out of, even though she was a part of the development. *(Document: 5d03c72fc0576)*
- But rumors were that maybe she gets all that money from her previous gambling habits, but no one said a thing because they never wanted the lunches to stop. *(Document: 5d03ba168e184)*
- She was weird though, due to the lunches and party she would pay for the employees. *(Document: 5d03ba168e184)*
- She was finally caught when a bank employee reported a check for $400,000 that came through, and police investigated. *(Document: 5d039c516ae96)*
- It came as a shock to everyone that worked with her that she was stealing money. *(Document: 5d0398cc6fa54)*
- she said that she was rich because of an inheritance, other people joked that she had money because she gambled. *(Document: 5d0398cc6fa54)*
- A tax office employee was stealing money from the company that she worked for. *(Document: 5d0398cc6fa54)*
- SHe was only caught because a bank employee questioned a 400,000 check. *(Document: 5d03984dc2628)*
- She often treated her employees to lunch which was thought as of kind but also suspicious as she never spoke of money. *(Document: 5d035b56d3d10)*
- When it got implemented, she said that her division couldn't use it and would be using the paper system that they had before. *(Document: 5d024811e412e)*
- Also of interest is the necessity of the paper system, suggesting that she was incapable of abusing the new system she was developing. *(Document: 5c87a852e0b46)*
- It might suggest that a higher-up was one such accomplice, but this cannot be confirmed. *(Document: 5c87a852e0b46)*
- The woman was given a massive fine, and this incident ensured that the company tightened up its procedures forcing everyone to use the new IT System so that all transactions could be audited. *(Document: 5c7f9270c1494)*
- It was only the fact that a bank clerk questioned a large transaction that the woman was discovered and this was purely by chance. *(Document: 5c7f9270c1494)*
- She objected to this system, arguing that it was a pain to follow and couldn't be operated. *(Document: 5c73c9a927a4c)*
- I forget the amount, but she is due to repay an amount that she stole. *(Document: 5c701387c463a)*
- It is clear that her insistence on using the old system rather than the new one was that she had found a way to exploit the old system that wasn't going to be possible with the new system in place. *(Document: 5c6fee6c3665a)*
- She then made sure that her department was the only exception to the policy that everyone should use the new system. *(Document: 5c6f111d877e4)*
- It amounted to around 200 cheques over that time. *(Document: 5c6ecc9f2513e)*
- She fought the changes, under the premise that it was making it too difficult for her team to work. *(Document: 5c6ea42c1647c)*
- Once implemented she avoided using that system within her group even though it was mandated that it should be and was required through out the rest of the organisation. *(Document: 5c6e877f639e8)*
- There was no indication among the group that, despite the rumours, they were suspicious that 10 people in a group were engaging in fraud of this extent ($60m) . *(Document: 5c6e877f639e8)*
- It is not entirely clear why she did it, but the employee who worked with her said that she was generous with money - one example given was in terms of paying for team meals. *(Document: 5c6e74be61b98)*
- It would appear that she may have had other members of staff write the cheques and she could authorise them however this is not confirmed. *(Document: 5c6e687b04128)*
- When an IT system, that included fraud detection measures, was introduced, the manager successfully argued that their department did not need to use it. *(Document: 5c6d857487398)*
- Not sure why, but it was mentioned that she had a lavish lifestyle and always bought drinks and meals for others. *(Document: 5c6d7e0262872)*
- This happened over a long period of time, 18 years, so she did not seem to become overcome with guilt in that time. *(Document: 5c6d6e434c6a8)*
- She was found out by a bank teller that questioned a cheque that she was trying to cash. *(Document: 5c6d591491dac)*
- It's therefore unclear whether this had a significant impact upon them, as it is possible that far more than that was stolen. *(Document: 5c6d3e7557904)*
- The IT department had not noticed the suspicious activity due to the fact that the manager had managed to persuade her superiors that the computerised system was too complicated for her team to use (even though she helped to design it), and as such there was only a paper based record that could easily be manipulated. *(Document: 5c6d2abb29bf8)*

Topic 285 - berkeleyhigh pasarow berkeleysid africanamerican berkeleyhighschool hackread

Closely related to:

- The manager although nice always bought her employees' lunch, therefore it was a surprise to the staff when they found out the news. *(Document: 5d0399cfc2b32)*
- After a lengthy investigation, a further nine accomplices were found though it is still unclear if charges will be brought against them until their roles are more fully understood. *(Document: 5d036a7f85570)*

Topic 290 - key anonym keys’ ohio singh dublin

Closely related to:

- After an investigation, the tax office worker has to pay back 60 million dollars in fines. *(Document: 5d079e20e985e)*
- A middle manager in an office had stolen money using the original paper system without using the new computer system which she helped to design. *(Document: 5d03d83dbd8f8)*

Topic 291 - appl pierr citibankkorea zhejiang mrshumphrey albanycounti

Closely related to: Attack Characteristics - Attack Step

- She was discovered after a bank teller reported a suspicious cheque of more than $400,000. *(Document: 5d076e64a7364)*
- was able to run her network along with nine other accomplice ,until a bank teller reported the incident to the law enforcement.her *(Document: 5d03dc42937ba)*
- She was discovered because of a check a bank teller questioned and reported for $400,000. *(Document: 5d03baf2e2694)*
- The bank rolled out a new system to manage financial reports, her unit, was exempt from using the reports as per her request. *(Document: 5d035b56d3d10)*
- She was investigated and caught after a bank teller reported a suspicious cheque for more than $400,000 which she had submitted. *(Document: 5c73c9a927a4c)*
- She was caught as a bank clerk noticed a suspicious cheque and reported it. *(Document: 5c6fce7530a52)*

Topic 298 - infowatch centralbank deniszenkin jimnicholson octob russian

Closely related to: Outcome - Actor

- From the result of her criminal activity she is being fined a large sum of money. *(Document: 5c6e8d80126d8)*

Topic 300 - kinnear delunamartinez wellsfargo charl districtattorney’ miller’

Closely related to: Actor Characteristics - Personality characteristics/Actor Characteristics - Historical behaviour

- She was well liked and generous with her staff, although some of them thought she was a little bit odd and had some history of alcohol abuse. *(Document: 5c6d2abb29bf8)*

Topic 301 - deflorin socialsecur jane labor wsu andrealeighgunderson

Closely related to: Actor Characteristics - Personality characteristics/Actor Characteristics - Observed physical behaviour/Organisation Characteristics - Vulnerability/Opportunity

- This was odd, but her supervisors supported her. *(Document: 5d03cbf7e42dc)*
- Co-workers thought she was odd but overlooked it because she bought drinks for everyone at the pub or treated co-workers to lunch. *(Document: 5d03984dc2628)*
- Although deemed a bit odd and maybe with some unexplicable behaviours the group were generally accepting of her and had only minor suspicions about her behaviours, background and generosity. *(Document: 5c6e877f639e8)*
- Whilst it was noted that this was odd nothing was investigated as to why she refused. *(Document: 5c6e877f639e8)*

Topic 308 - dallashomehealthcar mathew parkland dalla informationcommission kinnucan

Closely related to: Outcome - Organisation

- Investigations continue into the network around her activities, with nine accomplices arrested so far. *(Document: 5c6e80c32e77a)*

Topic 309 - oat ife elliott singapor antoniohortaosorio css

Closely related to:

- A whole network of fraudsters was found when investigated. *(Document: 5d0255245f406)*

Topic 313 - fbi octob socialsecur jeanci standardinsurancecompani name

Closely related to: Actor Characteristics - Motivation to Attack

- Her name was never stated, nor was her motives tho it could be said that she just wanted to make more money. *(Document: 5c701bae6117a)*

Topic 314 - cohen manafort weisselberg coleman trumporgan mueller

Closely related to: Outcome - Organisation

- 9 accomplices were discovered through a lengthy investigation. *(Document: 5c6ecc9f2513e)*

Topic 316 - patel allegro allegro' bean dekalb mcgee

Closely related to: Attack Characteristics - Attack

- A network of other fraudulent staff were discovered after investigations were done. *(Document: 5d076e64a7364)*
- She had written over 200 fraudulent checks and was only caught because a bank teller finally questioned one of them. *(Document: 5d04316c13164)*
- Her scheme last 18 years and involved at least 200 fraudulent checks. *(Document: 5d03cafcd1c68)*
- She wrote 200 fraudulent checks over a period of 18 years. *(Document: 5d03baf2e2694)*
- A woman had been writing fraudulent checks at her job for 18 years. *(Document: 5d02a27f6c1ec)*
- By creating discrepancies in the paper system (easier to justify), she was able to sneak mistakes past everyone until a bank teller picked up on a fraudulent check. *(Document: 5c879871c869a)*
- A lady working at a bank was writing fraudulent checks to herself, taking money over the course of 18 years. *(Document: 5c701387c463a)*
- Done by fraudulently writing checks and making use of the old-fashioned un-auditable paper-based system for accounting. *(Document: 5c6d7e0262872)*
- Caught by someone processing a fraudulent cheque. *(Document: 5c6d530129acc)*
- The scheme involved fraudulent cheques which were disguised within the paper-based accounting system. *(Document: 5c6d2b8c9ad3a)*

Topic 322 - allegion parkland nettalon jone schachter wooldridg

Closely related to: Actor Characteristics - Motivation to Attack/Actor Characteristics - Historical behaviour/Actor Characteristics - Observed physical behaviour

- Some jokingly put this down to gambling or alcohol addiction. *(Document: 5c77c731bba1c)*
- They put this down to a gambling addiction, or past drug abuse history. *(Document: 5c6ea42c1647c)*
- Staff did notice that the manager always seemed to have money, but generally put it down to her having an inheritance of some kind. *(Document: 5c6d2abb29bf8)*

Topic 323 - thehomedepot marquardt admiralchucki investig prosecutor wellsfargo

Closely related to: Outcome - Organisation

- She was able to commit the crime by insisting she use the old system and was able to steal money undetected for 18 years along with accomplices who are still being investigated She was caught when a clerk was suspicious of a check she had authorised. *(Document: 5d03d0e7b9686)*
- It is possible that there were a network of people involved, with 9 possible accomplices still under investigation. *(Document: 5c6e867e1b300)*

Topic 324 - tesla robert graham lowri tesla' tripp

Closely related to:

- She stole 48 million and knew how to rig the system. *(Document: 5d03cc65150e0)*

Topic 328 - lookup dvs minnesota criminallaw seventhamend cheremeh

Closely related to:

- Her friends and colleagues were very surprised to find out about it, although some had suspicions. *(Document: 5d0398b780778)*

Topic 338 - farrel nama german nama' swiss belgium

Closely related to:

- Her colleagues were surprised to hear about this. *(Document: 5d04581fba50e)*

Topic 340 - swisscom bank garnett enervest garnett' mitchel

Closely related to: Attack Characteristics - Attack

- She was only caught because a bank teller questioned a cheque she wrote. *(Document: 5d07bded75b16)*
- There was a person who was a manager at the bank that some would describe as a nice woman who had been stealing millions from the bank overtime. *(Document: 5d03c707b6dfa)*
- She ended up getting caught when a bank teller marked a check as suspicious. *(Document: 5d039950a0ffa)*
- the bank manager was stealing checks she got away with it by telling everyone it was a family inheritance she was able to hide the transactions becasue the bank used a paper based system she was caught by a bank teller who thought a check looked suspicious the bank manager was caught and fined 60 million dollars *(Document: 5d028986a93c6)*
- A female bank manager was caught stealing money from the bank. *(Document: 5c7939498b4b6)*
- generally well-liked and popular female tax office manager had been stealing from taxes over 18 years by exploiting loopholes in paperwork systems and was very against an electronic system which she probably knew would make her theft harder to carry out and easier to detect, and was caught by a bank teller who noticed/questioned a suspicious cheque for $400,000; there were at least nine other accomplices; required to pay at least $45 million in restitution/taxes and other costs *(Document: 5c6fc0970bcac)*
- A bank teller helped catch her when they spotted a suspicious cheque for $400'000. *(Document: 5c6ecc9f2513e)*

Topic 342 - wikileak elmer switzerland ameripris juliusba zurich

Closely related to: Actor Characteristics - Personality characteristics

- Appeared to be an extremely generous person from which corroborates w/coworkers. *(Document: 5c7c178e1c62e)*
- She appeared to her colleagues as generous and personable, albeit somewhat eccentric. *(Document: 5c6ea42c1647c)*

Topic 351 - mrlin dunsworth wellsfargobank espanola healthpei ico

Closely related to: Actor Characteristics - Observed physical behaviour

- During work outings, the lady always bought every one drinks and never had to 'count down until pay day'. *(Document: 5c6e9bcfbe44c)*

Topic 354 - hhc manitobahealth jpmc swift wang phi

Closely related to:

- The story involved a long-term scheme from within the Tax Office to defraud the tax revenue system and collect money by issuing over/ approximately 200 checks during an 18 year period. *(Document: 5d079fc2cf7ce)*

Topic 360 - ernst hotel bienemi gilpatr gretna infowatch

Closely related to:

- Some co-workers considered who odd, or a little off. *(Document: 5d0399cfc2b32)*

Topic 366 - att kelley aventura hca medicalcent target

Closely related to: Attack Characteristics - Attack

- Now every department in the company has to use the new electronic system to avoid such incidents. *(Document: 5d07dfe6a9344)*
- One lady stole allot of money from a company over a long period. *(Document: 5d07bded75b16)*
- --> When the company introduced a new computer system to replace the paper based system so as to avoid any frauds, she denied to adopt the computer system. *(Document: 5d0481de7bb2e)*
- --> Taking a lesson from this incident, the company no longer exempts any department from adopting the computer based system. *(Document: 5d0481de7bb2e)*
- --> A manager at a tax office was able to commit fraud for 18 years and steal from the company because her company allowed her to use paper system for her department. *(Document: 5d0481de7bb2e)*
- A tax company manager was caught cheating the system by using fake checks and was stealing from the company for over 18 years with an amount of over $40 million. *(Document: 5d03d59ae268a)*
- Apparently, the company had established an IT program to have additional check and balances on the company's financial transactions. *(Document: 5d03d561c08b4)*
- This is a fraud study , where a middle management employee of a company was stealing money from the company for 18 years. *(Document: 5d03d561c08b4)*
- Now everyone at the company has to use the new system which tracks money. *(Document: 5d03cffd39760)*

- A very friendly manager at a company was stealing tons of money from the company. *(Document: 5d03cffd39760)*
- Therefore, the company updated their system to avoid any attempt of fraud in the future. *(Document: 5d0398b780778)*
- eventually a new program was installed in the company and caught on to her stealing the money from the company. *(Document: 5d03983ac17c8)*
- A Manager at a company had been secretly stealing funds for 18 years due to discrepancies with the company's paper cheque system. *(Document: 5d02dc8e97cb6)*
- A tax income worker was found to be committing fraud and taking company money not owed. *(Document: 5c87a852e0b46)*
- Lady stole money from her company, was caught by a new system that she helps make. *(Document: 5c701bae6117a)*
- Some people thought she was weird and had a past alcohol abuse issue There was a new computer based system introduced at the company which would help to prevent fraud. *(Document: 5c6e9bcfbe44c)*
- A middle manager at a company committed fraud by using an old paper based system to report false numbers instead of a new electronic system, which would have caught her out. *(Document: 5c6d567ba889a)*

Bibliography

Agrafiotis, I. et al. (2015). 'Identifying Attack Patterns for Insider Threat Detection'. In: *Computer Fraud and Security* 2015, pp. 9–17. ISSN: 13613723. DOI: 10.1016/S1361-3723(15)30066-X.

Aho, A. V. (1991). *Algorithms for finding patterns in strings, Handbook of theoretical computer science (vol. A): algorithms and complexity*.

Althebyan, Q. and B. Panda (2007). 'A Knowledge-Base Model for Insider Threat Prediction'. In: *Proceedings of the 2007 IEEE Workshop on Information Assurance, IAW*. IEEE, pp. 239–246. ISBN: 1-4244-1304-4. DOI: 10.1109/IAW.2007.381939.

Angeli, G., M. J. Johnson Premkumar and C. D. Manning (2015). 'Leveraging Linguistic Structure for Open Domain Information Extraction'. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Beijing, China: Association for Computational Linguistics, pp. 344–354. DOI: 10/gf9f44.

Arora, S. et al. (2012). 'A Practical Algorithm for Topic Modeling with Provable Guarantees'. In: *arXiv:1212.4777 [cs, stat]*. arXiv: 1212.4777.

Arun, R. et al. (2010). 'On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations'. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, pp. 391–402. DOI: 10/fndkt7.

autotldr (2017). *autotldr_ Is a Reddit Bot That Automatically Summarizes Posts. Find It at https://www.reddit.com/r/autotldr/ or Learn More http://autotldr.io Pic.Twitter.Com/twY861wCI1*. en. Publication Title: autotldr_ Type: Tweet.

B., Andrew P. (1997). 'The use of the area under the ROC curve in the evaluation of machine learning algorithms'. In: *Pattern Recognition* 30.7, pp. 1145–1159. ISSN: 0031-3203. DOI: https://doi.org/10.1016/S0031-3203(96)00142-2. URL: https://www.sciencedirect.com/science/article/pii/S0031320396001422.

Ball, K. (2010). 'Workplace surveillance: An overview'. In: *Labor History* 51.1. Publisher: Taylor & Francis, pp. 87–106.

Bastian, M., S. Heymann and M. Jacomy (2009). 'Gephi: An Open Source Software for Exploring and Manipulating Networks'. In: *International AAAI Conference on Weblogs and Social Media*.

Baumer, E. P. S. et al. (2017). 'Comparing Grounded Theory and Topic Modeling: Extreme Divergence or Unlikely Convergence?' In: *Journal of the Association for Information Science and Technology* 68.6. Publisher: Wiley Online Library, pp. 1397–1410. DOI: 10/gbgw78.

Bird, Steven, Ewan Klein and Edward Loper (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc.

Bishop, M. et al. (2014). 'Insider Threat Identification by Process Analysis'. In: *2014 IEEE Security and Privacy Workshops*, pp. 251–264. ISSN: 10816011. DOI: 10/gdrb3t.

Blei, D. M. and J. D. Lafferty (2006). 'Dynamic Topic Models'. In: *Proceedings of the 23rd International Conference on Machine Learning*. ACM, pp. 113–120.

Blei, D. M. and J. D. McAuliffe (2010). 'Supervised Topic Models'. In: *arXiv:1003.0783 [stat]*. arXiv: 1003.0783.

Blei, D. M., A. Y. Ng and M. I. Jordan (2003). 'Latent Dirichlet Allocation'. In: *Journal of Machine Learning Research* 3.Jan, pp. 993–1022. ISSN: ISSN 1533-7928.

Bolukbasi, T. et al. (2016). 'Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings'. In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*. NIPS'16. Red Hook, NY, USA: Curran Associates Inc., pp. 4356–4364. ISBN: 978-1-5108-3881-9.

Bowen, B. M. et al. (2008). *Baiting Inside Attackers Using Decoy Documents*. Technical Report. U.S. Army Research Office.

Brdiczka, O. et al. (2012). 'Proactive Insider Threat Detection through Graph Learning and Psychological Context'. In: *Proceedings - IEEE CS Security and Privacy Workshops, SPW 2012*, pp. 142–149. DOI: 10/gdrb38.

Breiman, L. (2001). 'Random Forests'. In: *Machine Learning* 45.1, pp. 5–32. ISSN: 1573-0565. DOI: 10.1023/A:1010933404324. URL: https://doi.org/10.1023/A:1010933404324 (visited on 12/01/2022).

Brown, C. R., A. Watkins and F. L. Greitzer (2013). 'Predicting Insider Threat Risks through Linguistic Analysis of Electronic Communication'. In: *2013 46th Hawaii International Conference on System Sciences*, pp. 1849–1858. ISSN: 15301605. DOI: 10/gdrb3z.

Bruijn, B. de and J. Martin (2002). 'Getting to the (c)Ore of Knowledge: Mining Biomedical Literature'. en. In: *International Journal of Medical Informatics* 67.1-3, pp. 7–18. ISSN: 13865056. DOI: 10.1016/S1386-5056(02)00050-3.

Cambria, E. and B. White (2014). 'Jumping NLP Curves: A Review of Natural Language Processing Research'. In: *IEEE Computational intelligence magazine* 9.2, pp. 48–57. DOI: 10/gdvjzs.

Cao, J. et al. (2009). 'A Density-Based Method for Adaptive LDA Model Selection'. In: *Neurocomputing* 72.7-9, pp. 1775–1781. DOI: 10/dvh9dh.

Cappelli, D. M., A. P. Moore and R. Trzeciak (2015). *The Cert Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes (Theft, Sabotage, Fraud)*. Vol. 1. Addison-Wesley. ISBN: 978-85-7811-079-6.

Cendrowski, H. et al. (2007). *The Handbook of Fraud Deterrence*. John Wiley & Sons.

Chinchani, R. et al. (2005). 'Towards a Theory of Insider Threat Assessment'. In: *International Conference In Dependable Systems and Networks*, pp. 108–117. ISSN: 1530-0889. DOI: 10/czrkz5.

Ching, W. K. et al. (2013). *Markov Chains: Models, Algorithms and Applications*. en. Second. International Series in Operations Research & Management Science. Springer US. ISBN: 978-1-4614-6311-5. DOI: 10.1007/978-1-4614-6312-2.

Choi, D. et al. (2014). 'Text Analysis for Detecting Terrorism-Related Articles on the Web'. In: *Journal of Network and Computer Applications* 38, pp. 16–21. ISSN: 1084-8045. DOI: https://doi.org/10.1016/j.jnca.2013.05.007. URL: http://www.sciencedirect.com/science/article/pii/S1084804513001215.

Cichonski, P. et al. (2012). 'Computer security incident handling guide'. In: *NIST Special Publication* 800.61, pp. 1–147.

Coles-Kemp, L. and M. Theoharidou (2010). 'Insider Threat and Information Security Management'. In: *Insider Threats in Cyber Security*. Ed. by C. W. Probst et al. Boston, MA: Springer US, pp. 45–71. ISBN: 978-1-4419-7133-3. DOI: 10.1007/978-1-4419-7133-3_3.

Computer Misuse Act (1990). eng.

Costa, D L. et al. (2014). 'An Ontology for Insider Threat Indicators Development and Applications'. In: *CEUR Workshop Proceedings*. Vol. 1304.

Data Protection Act (2018). eng.

De Vault, D., I. Oved and M. Stone (2006). 'Societal Grounding Is Essential to Meaningful Language Use'. In: *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1*. AAAI'06. Boston, Massachusetts: AAAI Press, pp. 747–754. ISBN: 978-1-57735-281-5.

Dounis, N. P. (2017). 'GDPR Regulatory Compliance and the Role of Internal Audit: Theoretical and Practical Approach'. In: *Int'l. In-House Counsel J.* 11. Publisher: HeinOnline, p. 1.

Eberle, W., J. Graves and L. Holder (2010). 'Insider Threat Detection Using a Graph-Based Approach'. In: *Journal of Applied Security Research* 6, pp. 32–81. ISSN: 1936-1610. DOI: 10.1080/19361610.2011.529413.

Elmrabit, N., S. H. Yang and L. Yang (2015). 'Insider threats in information security categories and approaches'. In: *2015 21st International Conference on Automation and Computing (ICAC)*. IEEE, pp. 1–6.

Elmrabit, N. et al. (2020). 'Insider Threat Risk Prediction based on Bayesian Network'. en. In: *Computers & Security* 96, p. 101908. ISSN: 0167-4048. DOI: 10/ghbwng. URL: http://www.sciencedirect.com/science/article/pii/S016740482030184X (visited on 17/09/2020).

Equality Act (2010). eng.

Fafinski, S. (2013). *Computer Misuse: Response, Regulation and the Law*. Willan.

Feinerer, I. (2013). 'Introduction to the tm Package Text Mining in R'. In: *Accessible en ligne: http://cran. r-project. org/web/packages/tm/vignettes/tm. pdf*.

Flesch, R. (1948). 'A New Readability Yardstick'. In: *Journal of applied psychology* 32.3. Publisher: American Psychological Association, p. 221.

Forcepoint (2015). *Insider Threat*. en. Publication Title: Forcepoint. URL: https://www.forcepoint.com/product/insider-threat (visited on 17/09/2020).

Forte, L. (2019). *Insider Threat Report 2019*. Tech. rep. Red Goat Cyber Security, p. 30.

Gavai, G. et al. (2015). 'Supervised and Unsupervised Methods to Detect Insider Threat from Enterprise Social and Online Activity Data'. In: *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)* 6. ISSN: 20935382. DOI: 10.1145/2808783.2808784.

Gender Recognition Act (2004). eng.

Ghawi, R. and J. Pfeffer (2019). 'Efficient hyperparameter tuning with Grid Search for text categorization using kNN approach with BM25 similarity'. In: *Open Computer Science* 9.1. Publisher: De Gruyter Open, pp. 160–180.

Girju, R., D. I. Moldovan et al. (2002). 'Text Mining for Causal Relations'. In: *FLAIRS Conference*, pp. 360–364.

Goldberg, L. R. (1993). 'The Structure of Phenotypic Personality Traits'. In: *American psychologist* 48.1. Publisher: American Psychological Association, p. 26.

Gordon, M. and M. Kochen (1989). 'Recall-precision trade-off: A derivation'. In: *Journal of the American Society for Information Science* 40.3, pp. 145–151.

Grefenstette, G. and L. Muchemi (2015). 'Extracting Hierarchical Topic Models from the Web for Improving Digital Archive Access'. In: *Expert Workshop on Topic Models and Corpus Analysis*.

Greitzer, F. L., D. Frincke and M. Zabriskie (2011). *Social/Ethical Issues in Predictive Insider Threat Monitoring*. en. ISBN: 9781616922450 Pages: 132–161 Publication Title: Information Assurance and Security Ethics in Complex Systems: Interdisciplinary Perspectives Type: Chapter Published: www.igi-global.com/chapter/social-ethical-issues-predictive-insider/46344. DOI: 10.4018/978-1-61692-245-0.ch007.

Greitzer, F. L. and R. E. Hohimer (2011). 'Modeling Human Behavior to Anticipate Insider Attacks'. In: *Journal of Strategic Security* 4, pp. 25–48. ISSN: 1944-0464. DOI: 10/bqq7mr.

Greitzer, F. L. et al. (2012). 'Identifying At-Risk Employees: Modeling Psychosocial Precursors of Potential Insider Threats'. In: *Proceedings of the Annual Hawaii International Conference on System Sciences*, pp. 2392–2401. ISSN: 15301605. DOI: 10.1109/HICSS.2012.309.

Greitzer, F. L. et al. (2013). 'Psychosocial Modeling of Insider Threat Risk Based on Behavioral and Word Use Analysis'. In: *e-Service Journal* 9.1. Publisher: Indiana University Press, pp. 106–138. ISSN: 15288226, 15288234. DOI: 10.2979/eservicej.9.1.106. URL: http://www.jstor.org/stable/10.2979/eservicej.9.1.106 (visited on 18/02/2022).

Greitzer, F. L. et al. (2014). 'Unintentional Insider Threat: Contributing Factors, Observables, and Mitigation Strategies'. In: *2014 47th Hawaii International Conference on System Sciences*, pp. 2025–2034.

Griffiths, T. L. and M. Steyvers (2004). 'Finding Scientific Topics'. In: *Proceedings of the National academy of Sciences* 101.suppl 1, pp. 5228–5235.

Ha, D. et al. (2007). 'Insider Threat Analysis Using Information-Centric Modeling'. In: *Advances in Digital Forensics III*. Ed. by Philip Craiger and Sujeet Shenoi. Springer New York, pp. 55–73. ISBN: 978-0-387-73742-3.

Hanauer, D. A. et al. (2015). 'Supporting Information Retrieval from Electronic Health Records: A Report of University of Michigan's Nine-Year Experience in Developing and Using the Electronic Medical Record Search Engine (EMERSE)'. In: *Journal of biomedical informatics* 55, pp. 290–300. DOI: 10/f7hdct.

Hanson, J., T. Thorsen and N. Hunstad (2021). 'Insider threat programmes: Time to hit restart'. In: *Cyber Security: A Peer-Reviewed Journal* 4.3. Publisher: Henry Stewart Publications, pp. 213–222.

Ho, M. S. et al. (2016). 'Demystifying Insider Threat: Language-Action Cues in Group Dynamics'. In: *Proceedings of the Annual Hawaii International Conference on System Sciences*. Vol. 2016-March, pp. 2729–2738. DOI: 10.1109/HICSS.2016.343.

Homoliak, I. et al. (2019). 'Insight into Insiders and IT: A Survey of Insider Threat Taxonomies, Analysis, Modeling, and Countermeasures'. In: *ACM Computing Surveys* 52.2. arXiv: 1805.01612, pp. 1–40. ISSN: 0360-0300, 1557-7341. DOI: 10/gg95s3.

Hu, T. et al. (2019). 'An Insider Threat Detection Approach Based on Mouse Dynamics and Deep Learning'. en. In: *Security and Communication Networks* 2019, pp. 1–12. ISSN: 1939-0114, 1939-0122. DOI: 10/ghbwnb. URL: https://www.hindawi.com/journals/scn/2019/3898951/ (visited on 17/09/2020).

Hu, Y. (2005). 'Efficient, High-Quality Force-Directed Graph Drawing'. In: *Mathematica Journal* 10.1. Publisher: Redwood City, Ca.: Advanced Book Program, Addison-Wesley Pub. Co., c1990-, pp. 37–71.

Huang, R. and E. Riloff (2010). 'Inducing Domain-Specific Semantic Class Taggers from (Almost) Nothing'. In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 275–285.

Human Rights Act (1990). eng.

Hunker, J. and C. Probst (2011). 'Insiders and Insider Threats an Overview of Definitions and Mitigation Techniques'. In: *Journal of Wireless Mobile Networks, Ubiquitous \textbackslash ldots*, pp. 4–27. ISSN: 20935374.

Hutchins, J. W. (2004). 'The Georgetown-IBM Experiment Demonstrated in January 1954'. en. In: *Machine Translation: From Real Users to Research*. Ed. by D. Hutchison et al. Vol. 3265. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 102–114. DOI: 10.1007/978-3-540-30194-3_12.

Iliinsky, N. P. N. (2010). *Beautiful Visualization : [Looking at Data through the Eyes of Experts]*. First edition. Theory in Practice. Sebastopol, CA: O'Reilly Media, Inc. ISBN: 978-1-4493-9133-1 1-4493-9133-8 978-1-4493-7988-9 1-4493-7988-5 978-1-4493-9068-6 1-4493-9068-4. URL: http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=536623.

Information Commissioner's Office (2018). *Personal Data Breaches*. URL: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/personal-data-breaches/.

InfoWatch (2019). *Data Leakage News — InfoWatch*. URL: https://infowatch.com/analytics/leaks_monitoring (visited on 25/01/2021).

Jacobi, C., W. van Atteveldt and K. Welbers (2016). 'Quantitative Analysis of Large Amounts of Journalistic Texts Using Topic Modelling'. In: *Digital Journalism* 4.1, pp. 89–106. ISSN: 2167-0811. DOI: 10/f3s2sg.

Jagarlamudi, J., H. Daumé and R. Udupa (2012). 'Incorporating Lexical Priors into Topic Models'. In: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*. EACL '12. Stroudsburg, PA, USA: Association for Computational Linguistics, pp. 204–213. ISBN: 978-1-937284-19-0.

Johnstone, L. (2022). *DataBreaches.Net: The Office of Inadequate Security*. Publisher: DataBreaches.net. URL: https://www.databreaches.net/ (visited on 07/01/2022).

Kandias, M. et al. (2010). 'An Insider Threat Prediction Model'. In: *Trust, Privacy and Security in Digital Business* 6264, pp. 26–37. ISSN: 0302-9743. DOI: 10/dqzj7z.

Keeney, M. et al. (2005). *Insider threat study: Computer system sabotage in critical infrastructure sectors*. Tech. rep. National Threat Assessment Ctr Washington Dc.

Kirk, A. K. and D. F. Brown (2003). 'Employee assistance programs: A review of the management of stress and wellbeing through workplace counselling and consulting'. In: *Australian psychologist* 38.2, pp. 138–143.

Lafferty, J. D. and D. M. Blei (2006). 'Correlated Topic Models'. In: *Advances in Neural Information Processing Systems*, pp. 147–154.

Lau, J. H. et al. (2010). 'Best Topic Word Selection for Topic Labelling'. en. In: p. 9.

Lau, J. H. et al. (2011). 'Automatic Labelling of Topic Models'. en. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, p. 10.

Le, D. C., A. N. Zincir-Heywood and M. I. Heywood (2020). 'Analyzing Data Granularity Levels for Insider Threat Detection Using Machine Learning'. In: *IEEE Transactions on Network and Service Management* 17, pp. 30–44. ISSN: 1932-4537. DOI: 10/ghbwnf.

Legg, P. A. et al. (2015). 'Caught in the act of an insider attack: detection and assessment of insider threat'. In: *2015 IEEE International Symposium on Technologies for Homeland Security (HST)*, pp. 1–6. DOI: 10.1109/THS.2015.7446229.

Legg, P. et al. (2013). 'Towards a Conceptual Model and Reasoning Structure for Insider Threat Detection'. In: *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* 4.4, pp. 20–37. ISSN: 20935374.

Liang, N., D. P. Biros and A. Luse (2016). 'An Empirical Validation of Malicious Insider Characteristics'. In: *Journal of Management Information Systems* 33.2, pp. 361–392. ISSN: 1557928X. DOI: 10/gdrb39.

Liaw, A. and M. Wiener (2018). *Breiman and Cutler's Random Forests for Classification and Regression*. Series: Package 'randomForest'.

Loftus, J. (2009). *IBM Prepping 'Watson' Computer to Compete on Jeopardy!* en-US. Publication Title: Gizmodo. URL: https://gizmodo.com/5228887/ibm-prepping-watson-computer-to-compete-on-jeopardy.

Lu, J. and R. K. Wong (2019). 'Insider Threat Detection with Long Short-Term Memory'. In: *Proceedings of the Australasian Computer Science Week Multiconference*. ACSW 2019. Sydney, NSW, Australia: Association for Computing Machinery, pp. 1–10. ISBN: 978-1-4503-6603-8. DOI: 10/ghbwm9. URL: https://doi.org/10.1145/3290688.3290692 (visited on 17/09/2020).

Lund, J. et al. (2019). 'Automatic Evaluation of Local Topic Quality'. In: *arXiv:1905.13126 [cs, stat]*. arXiv: 1905.13126.

Maasberg, M. and N. L. Beebe (2014). 'The Enemy Within the Insider: Detecting the Insider Threat Through Addiction Theory'. In: *Journal of Information Privacy and Security* 10.2, pp. 59–70. ISSN: 1553-6548. DOI: 10/gdrb4b.

Magklaras, G. B., S. M. Furnell and P. J. Brooke (2006). 'Towards an Insider Threat Prediction Specification Language'. In: *Information Management & Computer Security* 14, pp. 361–381. ISSN: 0968-5227. DOI: 10/czhvkd.

Manning, C. D., P. Ragahvan and H. Schutze (2009). *An Introduction to Information Retrieval*.

Manning, C. D. and H. Schütze (1999). *Foundations of Statistical Natural Language Processing*. MIT press.

Manning, C. D. et al. (2014). 'The Stanford CoreNLP Natural Language Processing Toolkit'. In: *Association for Computational Linguistics (ACL) System Demonstrations*, pp. 55–60.

Martin, K. and E. R. Freeman (2003). 'Some problems with employee monitoring'. In: *Journal of Business Ethics* 43.4, pp. 353–361.

Mathew, S. et al. (2008). 'Insider Abuse Comprehension through Capability Acquisition Graphs'. In: *Proceedings of the 11th International Conference on Information Fusion, FUSION 2008*. DOI: 10.1109/ICIF.2008.4632279.

Michael, M. G., S. J. Fusco and K. Michael (2008). 'A Research Note on Ethics in the Emerging Age of Überveillance'. In: *Computer communications* 31.6. Publisher: Elsevier, pp. 1192–1199. DOI: 10/d76h56.

Mikolov, T. et al. (2013). 'Efficient Estimation of Word Representations in Vector Space'. In: *arXiv preprint arXiv:1301.3781*. arXiv: 1301.3781.

Mirza, P. (2016). 'Extracting Temporal and Causal Relations between Events'. In: *arXiv:1604.08120 [cs]*. arXiv: 1604.08120.

Mohamed, N. and B. Belaton (2021). 'SBI model for the detection of advanced persistent threat based on strange behavior of using credential dumping technique'. In: *IEEE Access* 9. Publisher: IEEE, pp. 42919–42932.

Naveed, N. et al. (2011). 'Searching microblogs: coping with sparsity and document quality'. In: *Proceedings of the 20th ACM international conference on Information and knowledge management*, pp. 183–188.

Nichols, T., A. Danford and A. C. Tasiran (2009). 'Trust, employer exposure and the employment relation'. In: *Economic and Industrial Democracy* 30.2, pp. 241–265.

Noble, C. C. and D. J. Cook (2003). 'Graph-based anomaly detection'. In: *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 631–636.

Noever, D. (2019). 'Classifier Suites for Insider Threat Detection'. In: *arXiv:1901.10948 [cs, stat]*. URL: http://arxiv.org/abs/1901.10948 (visited on 17/09/2020).

Nurse, J. R. C. et al. (2014a). 'A Critical Reflection on the Threat from Human Insiders - Its Nature, Industry Perceptions, and Detection Approaches'. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Ed. by T. Tryfonas and I. Askoxylakis. Vol. 8533

LNCS, pp. 270–281. ISBN: 978-3-319-07619-5. DOI: 10.1007/978-3-319-07620-1_24.

Nurse, J. R.C. et al. (2014b). 'Understanding Insider Threat: A Framework for Characterising Attacks'. In: *2014 IEEE Security and Privacy Workshops*, pp. 214–228. DOI: 10.1109/SPW.2014.38.

Palm, E. (2009). 'Securing Privacy at Work: The Importance of Contextualized Consent'. In: *Ethics and Information Technology* 11, pp. 233–241. DOI: 10/b9x2sb.

Palomino-Garibay, A. et al. (2015). 'A Random Forest Approach for Authorship Profiling'. In: *Proceedings of CLEF*.

Pardau, S. L. (2018). 'The California consumer privacy act: towards a European-style privacy regime in the United States'. In: *J. Tech. L. & Pol'y* 23. Publisher: HeinOnline, p. 68.

Paxton-Fear, K., D. Hodges and O. Buckley (2018a). 'Connected events and malicious insiders: Investigating patterns of insider threat using natural language processing'. In: *Behavioural and Social Sciences in Security*.

— (2018b). 'Increasing the accessibility of NLP techniques for Defence and Security using a web-based tool'. In: DOI: 10.17862/cranfield.rd.10066229.v1. URL: https://cord.cranfield.ac.uk/articles/poster/Increasing_the_accessibility_of_NLP_techniques_for_Defence_and_Security_using_a_web-based_tool/10066229/1.

— (2019). 'Using Topic Distribution to Classify Fuzzy Topics'. In: *In Review Human-centric Computing and Information Sciences*.

— (2020). 'Understanding Insider Threat Attacks Using Natural Language Processing: Automatically Mapping Organic Narrative Reports to Existing Insider Threat Frameworks'. In: *HCI for Cybersecurity, Privacy, and Trust*. Springer International Publishing.

Paxton-Fear, K, D. Hodges and O. Buckley (2021). 'Visualizing an insider threat incident from witness reports using natural language processing'. In: *Conference on Applied Machine Learning for Information Security 2021*.

Probst, C. W. et al. (2010). 'Aspects of Insider Threats'. In: *Advances in Information Security* 49.April. ISSN: 15682633. DOI: 10/b6dx2z.

Radinsky, K., S. Davidovich and S. Markovitch (2012). *Learning Causality for News Events Prediction*.

Randazzo, M. R. et al. (2005). 'Insider Threat Study: Illicit Cyber Activity in the Banking and Finance Sector'. In: *Finance* 38.August, pp. 3–14. ISSN: 07321872. DOI: 10/djxd5b. URL: http://stinet.dtic.mil/oai/oai?&verb=getRecord&metadataPrefix=html&identifier=ADA441249%5Cnhttp://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA441249.

Redscan (2020). *Insider Threat Detection in Cyber Security*. en-GB. Publication Title: Redscan. URL: https://www.redscan.com/solutions/insider-threats-cyber-security/ (visited on 17/09/2020).

Reeves, A., K. Parsons and D. Calic (2020). 'Whose Risk Is It Anyway: How Do Risk Perception and Organisational Commitment Affect Employee Information Security Awareness?' In: *HCI for Cybersecurity, Privacy and Trust*. Ed. by A. Moallem. Lecture Notes in Computer Science. Springer International Publishing, pp. 232–249. DOI: 10/ghnjz7.

Regulation of Investigatory Powers Act (2000). eng.

Rehabilitation of Offenders Act (1974). eng. URL: https://www.legislation.gov.uk/ukpga/1974/53.

Ribeiro, M. T. et al. (2020). 'Beyond Accuracy: Behavioral Testing of NLP Models with CheckList'. In: *arXiv:2005.04118 [cs]*. arXiv: 2005.04118.

Richards, J., D. Tudhope and A. Vlachidis (2015). 'Text Mining in Archaeology: Extracting Information from Archaeological Reports'. en. In: *Mathematics and Archae-*

ology. Ed. by J. Barcelo and I. Bogdanovic. CRC Press, pp. 240–254. DOI: 10.1201/b18530-15.

Riedl, M. and C. Biemann (2012). 'Text Segmentation with Topic Models'. In: *Journal for Language Technology and Computational Linguistics* 27.1, pp. 47–69.

Runeson, P., M. Alexandersson and O. Nyholm (2007). 'Detection of Duplicate Defect Reports Using Natural Language Processing'. In: *Proceedings of the 29th International Conference on Software Engineering*. ICSE '07. Washington, DC, USA: IEEE Computer Society, pp. 499–510. ISBN: 978-0-7695-2828-1. DOI: 10.1109/ICSE.2007.32.

Safa, N. S. et al. (2019). 'Deterrence and Prevention-Based Model to Mitigate Information Security Insider Threats in Organisations'. In: *arXiv:1903.12079 [cs]*. _eprint: 1903.12079.

Sanzgiri, A. and D Dasgupta (2016). 'Classification of Insider Threat Detection Techniques'. In: *CISRC '16 Proceedings of the 11th Annual Cyber and Information Security Research Conference*, pp. 5–8. DOI: 10.1145/2897795.2897799.

Securonix (2020). *Insider Threat Detection - Types of Insider Threats*. en-US. Publication Title: Securonix. URL: https://www.securonix.com/solutions/insider-threat/ (visited on 17/09/2020).

Senator, T. E. et al. (2013). 'Detecting Insider Threats in a Real Corporate Database of Computer Usage Activity'. In: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '13*, p. 1393. ISSN: 9781450321747. DOI: 10/gdrb34.

Shaw, E. and H. Stock (2011). 'Behavioral Risk Indicators of Malicious Insider Theft of Intellectual Property: Misreading the Writing on the Wall'. In: *White Paper*.

Singh, V. (2020). *Vikasing/News-Stopwords*.

Smmry (2018). *SMMRY - Summarize Articles, Text, Websites, Essays and Documents*. Publication Title: Smmry. URL: https://smmry.com/.

Sokolowski, J. A. and C. M. Banks (2016). 'Agent Implementation for Modeling Insider Threat'. In: *Proceedings - Winter Simulation Conference* 2016-Febru, pp. 266–275. ISSN: 08917736. DOI: 10/gdrb36.

Spitzner, L. (2003). 'Honeypots: catching the insider threat'. In: *19th Annual Computer Security Applications Conference, 2003. Proceedings.* pp. 170–179.

Splunk (2020). *Insider Threat Detection Tools & Software*. en-US. Publication Title: Splunk. URL: https://www.splunk.com/en_us/cyber-security/insider-threat.html (visited on 17/09/2020).

Stanford NLP Group (2020). *Pipeline - CoreNLP*. en-US. Publication Title: CoreNLP. URL: https://stanfordnlp.github.io/CoreNLP/pipeline.html.

Sullivan, C. A. (2016). 'Tending the Garden: Restricting Competition via" Garden Leave"'. In: *Berkeley Journal of Employment and Labor Law*, pp. 293–325.

Taylor, M. et al. (2011). 'Forensic investigation of cloud computing systems'. In: *Network Security* 2011.3. Publisher: Elsevier, pp. 4–10.

Thompson, H. H., J. A. Whittaker and M. Andrews (2004). 'Intrusion Detection: Perspectives on the Insider Threat'. In: *Computer Fraud & Security* 2004.1, pp. 13–15. ISSN: 1361-3723. DOI: 10/b62rtp.

Tidy, J. and D. Molloy (2020). 'Twitter Hack: 130 Accounts Targeted in Attack'. en-GB. In: *BBC News*. Section: Technology.

Trzeciak, R. (2011). *The CERT Insider Threat Database*. Publication Title: Carnegie Mellon University: Software Engineering Institute. URL: https://insights.sei.cmu.edu/insider-threat/2011/08/the-cert-insider-threat-database.html.

Verizon (2019). *2019 Data Breach Investigations Report*. 2019 Data Breach Investigations Report DBIR2019. Verizon.

— (2020). *2020 Data Breach Investigations Report*. en. 2020 Data Breach Investigations Report DBIR2020. Verizon.

Wu, Y. et al. (2016). 'Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation'. In: *arXiv:1609.08144 [cs]*. arXiv: 1609.08144.

Yadav, T. and A. M. Rao (2015). 'Technical aspects of cyber kill chain'. In: *International Symposium on Security in Computing and Communication*. Springer, pp. 438–452.

Yaseen, Q. and B. Panda (2012). 'Insider Threat Mitigation: Preventing Unauthorized Knowledge Acquisition'. In: *Journal of Information Security* 11, pp. 269–280. DOI: 10.1007/s10207-012-0165-6.

Yerby, J. (2013). 'Legal and ethical issues of employee monitoring'. In: *Online Journal of Applied Knowledge Management (OJAKM)* 1.2, pp. 44–55.

Young, W. T. et al. (2013). 'Use of Domain Knowledge to Detect Insider Threats in Computer Activities'. In: *Proceedings - IEEE CS Security and Privacy Workshops, SPW 2013*. ISBN: 978-0-7695-5017-6. DOI: 10.1109/SPW.2013.32.

Yuan, S. and X. Wu (2020). 'Deep Learning for Insider Threat Detection: Review, Challenges and Opportunities'. In: *arXiv:2005.12433 [cs]*. arXiv: 2005.12433.

Zhao, H., L. Du and W. Buntine (2017). 'A Word Embeddings Informed Focused Topic Model'. In: *Asian Conference on Machine Learning*, pp. 423–438.