Association for Information Systems
AIS Electronic Library (AISeL)

ACIS 2020 Proceedings | Australasian (ACIS)

2020

A Unified Classification Model of Insider Threats to Information Security

Sunitha Prabhu
Waikato Institute of Technology, sunitha.prabhu@wintec.ac.nz

Nik Thompson
Curtin University of Technology, nik.thompson@curtin.edu.au

Follow this and additional works at: https://aisel.aisnet.org/acis2020

Recommended Citation

Prabhu, Sunitha and Thompson, Nik, "A Unified Classification Model of Insider Threats to Information Security" (2020). *ACIS 2020 Proceedings*. 40.
https://aisel.aisnet.org/acis2020/40

This material is brought to you by the Australasian (ACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ACIS 2020 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

A Unified Classification Model of Insider Threats to Information Security

Completed research paper

Sunitha Prabhu
Centre for Information Technology
Waikato Institute of Technology
Hamilton, New Zealand
Email: sunitha.prabhu@wintec.ac.nz

Nik Thompson
School of Management
Curtin University
Perth, Australia
Email: nik.thompson@curtin.edu.au

Abstract

Prior work on insider threat classification has adopted a range of definitions, constructs, and terminology, making it challenging to compare studies. We address this issue by introducing a unified insider threat classification model built through a comprehensive and systematic review of prior work. An insider threat can be challenging to predict, as insiders may utilise motivation, creativity, and ingenuity. Understanding the different types of threats to information security (and cybersecurity) is crucial as it helps organisations develop the right preventive strategies. This paper presents a thematic analysis of the literature on the types of insider threats to cybersecurity to provide cohesive definitions and consistent terminology of insider threats. We demonstrate that the insider threat exists on a continuum of accidental, negligent, mischievous, and malicious behaviour. The proposed insider threat classification can help organisations to identify, implement, and contribute towards improving their cybersecurity strategies.

Keywords Cybersecurity, Information security, Human Factors, Insider threats.

1 Introduction

The introduction of new technologies, network-enabled devices, and increased connectivity has enabled organisations to be globally connected and improve the way they do business. While the growing integration of outsourcing, offshore work, and remote offices offer many opportunities, it also increases the organisation's information systems exposure to cybersecurity threats and risks. These threats arise due to dilution of the organisational protection barriers and an increase in the number of people who have insider access rights from a remote location (Dupuis & Khadeer, 2016; Sasse et al., 2007). An insider is an individual who is currently or was formerly employed by the organisation, or a collaborator, or a partner, or a contractor, or other associates that have authorized access to the organisation's networks, systems, or data (CERT, 2018). We use this reference definition in our study as it refers explicitly to human actors, their relationship with the organisation, and their authorised access to the systems at some time.

Cybersecurity is socio-technical and involves humans and technical systems. Technical systems have vulnerabilities, and so do humans (Sasse et al., 2007). Understanding the different types of threats humans can pose is essential, as this knowledge can help organisations implement the right strategies to protect their information systems. Several researchers have suggested different methods of classification for insider threats (refer to Table 1). But the classification methods are incompatible as they are based on different factors, making it difficult to integrate existing concepts. In addition, we find inconsistencies in the literature with the classification of non-malicious threats and the terminology used. For the field to advance, we need consistent terminology and a unified understanding of the threats posed. This study makes several significant contributions to research on the role of insiders in cybersecurity. We integrate existing literature to present a unified insider threat classification, cohesive terminology, and effective definition for the different types of insider threats. We propose an insider threat classification exist on a continuum of accidental, negligent, mischievous, and malicious nature.

In this paper, we review the literature to identify the different types of insiders and perform a thematic analysis. We then define and explain the identified types of insiders. This is followed by a discussion on the implications of the proposed classification to theory and practice. The paper concludes by considering the contributions made and recommendations for future research.

2 Related Work

Insiders play a considerable role in the use and misuse of information systems within an organisation. The insider poses a unique security threat, as they may know how to achieve the most significant impact to the organisation while leaving little evidence (NAIC, 2008). CERT (2018) (pg. 17) defines an insider threat as

> "*a current or former employee, contractor, or another business partner who has or had authorised access to an organisation's network, system, or data; and has intentionally exceeded or intentionally used that access in a manner that negatively affected the confidentiality, integrity, availability, or physical well-being of the organisation's information, information systems, or workforce; or who through their action or inaction without malicious intent causes harm or substantially increases the probability of severe future harm to the confidentiality, integrity, or availability of the organisation's information, or information systems*."

The risk to the organisation from insider threats can include tangible losses such as a decrease in service availability and effectiveness; and intangible losses such as loss of intellectual property and public damage to the brand (CERT, 2018). According to a report by Ponemon (2020), the overall number of insider incidents has spiked by 47% globally in the last two years, and the average cost of insider threats to an organisation has risen by 31% in that period to $11.45 million.

While the focus for insider threat is generally on the motive or intentional action of the breach (Hadlington, 2018), harm can also be caused by individuals without any malicious intent (Carroll, Greitzer, & Roberts, 2014). Indeed, while initially omitted as a category, the Computer Emergency and Recovery Team (CERT) have introduced *unintentional insider* as a threat category only since 2016 (CERT, 2016). CERT (2018) and Verizon (2018) classify insider threats as unintentional (those that perform actions without any malicious intent), and malicious (those that intentionally use or exceed

their privileges in a manner that negatively affects the organisation). Their classification is limited to these two categories, and the term unintentional includes all non-malicious acts.

One of the dominant themes used for the threat classification is the intention, which is described using terms such as accidental, intentional, or deliberate. Loch, Carr, and Warkentin (1992) classify threats as accidental and intentional. Im and Baskerville (2005) elaborate on the above classification and include catastrophes and human error as part of accidental threats. Human errors are further classified as skill-based such as data input errors, rule-based such as invalid default values, and knowledge-based such as software update crash (Im & Baskerville, 2005). A similar but brief classification was given by Crossler et al. (2013) for the maladaptive behaviour of insiders as intentionally conducted (deviant behaviour), and unintentionally conducted (misbehaviour). Crossler's definition of unintentional intent includes misbehaviour caused by the action or inaction of an individual and also accidental threats.

Other studies have combined behavioural intent and malicious intent. Wall (2013) describe insiders as a non-malicious negligent insider, non-malicious well-meaning insider, and the malicious insider. Kraemer and Carayon (2007) include deliberate violations of non-malicious nature in their classification and classify insider threats as accidental, deliberate violations of a non-malicious nature, and deliberate violations of a malicious nature. A similar classification is presented by Van Den Bergh and Njenga (2016). They classify insider behaviour as misbehaviour (intentional violation without knowledge of the violation), non-malicious deviant behaviour (intentional non-malicious violation), and malicious deviant behaviour (intentional malicious violations). Though these types are the same as those described by Wall (2013), different terms are used to describe the threats. Carroll et al. (2014) present a conceptual framework with a focus on not only the behavioural intent and malicious intent, but also the individual's action and inaction that could be a threat. They classify insider threats to cybersecurity as non-malicious unintentional information compromised due to lack of action, non-malicious unintentional information compromise due to action, and intentional malicious.

Interestingly, some studies consider the role of the individual's technical skills in their security behaviour. Stanton, Stam, Mastrangelo, and Jolton (2005) focus on the user's technical expertise in addition to their intentions. The technical expertise needed to perform the act is described as high or low; and intentions are described as malicious, neutral, and beneficial. The insider threats are categorised in six categories as (1) intentional destruction (malicious with high technical expertise) with a firm intention to harm IS, (2) detrimental misuse (malicious with low expertise) with an intention to harm through annoyance, harassment, and rule-breaking, (3) Dangerous tinkering (neutral with high expertise) use of skills with no intent to harm, (4) naïve mistakes (neutral with low expertise) no intention to harm, (5) aware assurance (beneficial with high expertise) with an intention to do good, and (6) basic hygiene (beneficial with low expertise) with an intention to do good without technical expertise. In our study, this was among a few classifications that specifically included an individual's good intentions resulting in harm to the organisation's IS.

Some studies outline behaviour and intent as a continuum. Willison and Warkentin (2013) and Aurigemma and Mattson (2014) propose the intent of an insider to be a continuum. Willison and Warkentin (2013) outline intent as a continuum ranging from passive non-volitional non-compliance to volitional non-malicious non-compliance, and intentional malicious acts. Whereas, Aurigemma and Mattson (2014) show non-malicious behaviour as a continuum of volitional to non-volitional. We find this type of classification practical as an individual's behaviour is granular and a gradient rather than being type-cast.

3 Thematic Analysis

Through this review of prior work, we make three observations. Firstly, we observe that researchers have classified insider threats based on many different factors such as intent, motive, technical expertise, or combinations of the above. A unified threat classification method is desirable as it will enable an organisation to identify potential threats and develop appropriate prevention strategies. The mitigation strategy implemented for one type of threat may not be suitable for another. For example, an insider lacking skills to use the system can be provided with user training; however, if a user is deliberately misusing the system, then user training is of no use. Therefore, to improve our understanding of the insider information security threats, we integrate the different approaches taken in prior work and present a new unified method of classifying insider types later in this paper.

The second observation made was that different terms had been used to describe the same threat, for example, as:

- a threat caused by accident is described using terms such as accidental (Im & Baskerville, 2005; Loch et al., 1992), unintentional (CERT, 2018; Elmrabit, Yang, & Yang, 2015; Verizon, 2018), passive non-volitional non-compliance (Willison & Warkentin, 2013), and misbehaviour (Crossler et al., 2013; Van Den Bergh & Njenga, 2016);
- a threat resulting through inaction to follow the prescribed security procedures is described using terms such as negligent (Wall, 2013), naïve mistakes (Stanton et al., 2005), and deliberate non-malicious (Guo, Yuan, Archer, & Connelly, 2011; Kraemer & Carayon, 2007);
- a threat resulting from active-risk behaviour, i.e. through the deliberate action of misuse of privileges is described using terms such as deliberate tinkering (Stanton et al., 2005), non-malicious well-meaning (Wall, 2013), and non-malicious deviant (Van Den Bergh & Njenga, 2016); and
- a threat caused by malicious intent was consistent in all descriptions as malicious and intentional or deliberate.

Our third observation was the use of inconsistent definitions to describe threats, thereby making it impossible to compare studies directly. For example, some researchers describe threats caused by the intentional non-malicious violation (such as negligent and mischievous threats as shown in Table 1) as being intentional (Guo et al., 2011; Kraemer & Carayon, 2007), while others describe them to be unintentional (Carroll et al., 2014; CERT, 2018; Crossler et al., 2013; Verizon, 2018). To advance the field, we need a common vocabulary.

We classify an insider threat to cybersecurity based on intent, motive, and action as (1) accidental: unintentional, non-malicious, (2) negligent: intentional, non-malicious, due to inaction, (3) mischievous: intentional, non-malicious, due to actions through misuse of privileges, and (4) malicious: intentional, malicious. Table 1 summarises the thematic analysis of insider threats.

Author(s)	**Accidental**	**Negligent**	**Mischievous**	**Malicious**
Loch et al. (1992)	Accidental			Intentional
Stanton et al. (2005)	Basic Hygiene & Aware Assurance	Naïve Mistakes	Dangerous Tinkering	Detrimental Misuse & Intentionally Malicious
Kraemer and Carayon (2007)	Accidental	Deliberate Non-malicious	Deliberate Non-malicious	Deliberate Malicious
Guo et al. (2011)		Intentional Non-malicious		Intentional Malicious
Crossler et al. (2013)	Unintentional (Misbehaviour)	Unintentional (Misbehaviour)	Unintentional (Misbehaviour)	Intentional (Deviant)
Wall (2013)		Non-malicious Negligent	Non-malicious Well-meaning	Malicious
Willison and Warkentin (2013)	Passive, Non-volitional Non-compliance	Volitional Non-malicious Non-compliance		Intentional Malicious
Carroll et al. (2014)		Non-malicious Unintentional compromise due to lack of action	Non-malicious Unintentional compromise due to action	Intentional Malicious
Van Den Bergh and Njenga (2016)	Misbehaviour	Non-malicious Deviant		Malicious Deviant
CERT (2018) Verizon (2018)	Unintentional	Unintentional	Unintentional	Malicious

Table 1. Thematic Analysis of Insider Threats

We use the terms accidental, negligent, and malicious as they best describe the behaviour of the individual acting as a threat and being consistent with naming used by most of the researchers. We introduce the term "mischievous" to describe the active-risk behaviour of the individual performing the non-malicious intentional action through the misuse of privileges, and the name describes the act, consistent with the nature of the description of the other three types of threats identified here.

From Table 1, we see that while all researchers identify threats caused by malicious intent, most have variations when describing non-malicious threats such as accidental, negligent, and mischievous. Also, some researchers have not considered accidental threats to be a threat. While it is possible that some of these researchers' focus was exclusively on non-accidental threats, this was not explicitly stated. While all researchers address intentional non-malicious insiders, only some have differentiated between non-malicious threats resulting from inaction to those resulting from actions through the misuse of the privilege of the insider.

4 Typology of Insiders

An insider may act in a manner they are not supposed to – either accidentally or deliberately – and the threats can be challenging to predict. Understanding the types, motivation, and implications of insider threats are essential for an organisation to identify the vulnerabilities, develop prevention strategies, and protect their systems from breaches. The mitigation strategy implemented for one type of threat may not be suitable for another. Based on the extensive review of prior research, we classify, define, and explain the different types of insiders in this section.

4.1 Accidental Insider

Accidental refers to an unexpected act without malicious motive or deliberate intent. We define an accidental insider as:

> *an insider who has no malicious intent associated with their action or inaction; and due to a lapse makes an error that caused harm or increased the probability of future harm to the organisation's information systems.*

Accidental incidents may occur when the individual knows the right procedure, but mistakes happen when performing familiar tasks or due to the person's ignorance of the system (Ahola, 2019). There are many ways in which an insider can accidentally have an impact on security. Examples include:

- Misdelivery: sending something to a wrong recipient (Verizon, 2018); posting sensitive information publicly on a website (CERT, 2018). According to Verizon (2018), misdelivery was the fourth most common cause of all cybersecurity incidents.
- Unintentional disclosure: unwittingly being influenced to divulge confidential or sensitive information to an unauthorised person, which subsequently allows the unauthorised person to access or breach the protected system (Mouton, Leenen, & Venter, 2016).
- Errors and omissions: data entry errors when users create or edit data, errors by system developers and programmers in the form of bugs (Nieles, Dempsey, & Pillitteri, 2017).
- Physical loss: misplacement or loss of a portable device (Li, No, & Boritz, 2020). Personal, confidential, and sensitive information from the devices can be accessed and used for future attacks.

CERT (2018) explain the reasons for unintentional incidents by insiders as simple human error, fatigue or sleepiness, feeling of stress, lack of attention, the effect of drugs, and mood. A study conducted in 2018 by Shred-it (sample size unknown) revealed that over 40% of small business owners and executives reported employee negligence or accidental loss to be the root cause of their data security breach (Shred-it, 2018).

Unwitting or passive insiders range from individuals who share information unaware of the security implications to those who are manipulated and coerced into active participation (NAIC, 2008). However, it would not be justified to assume that these human vulnerabilities are the same as undesirable behaviour (Sasse et al., 2007). Security behaviours are highly context-bound, and behaviour that may be highly desirable in one context may potentially cause a security risk in another – for example, trusting or assisting a colleague. In a social context, discussing and sharing information with a colleague would be considered to be team-spirited and desirable. In contrast, in a security context, it may be considered as information leakage and a security breach. It is natural for an individual to trust a colleague, even if it is someone they have not met earlier, simply because they

work for the same organisation. Insiders, therefore, often underestimate the likelihood of falling victim to a cybersecurity breach, and a dishonest colleague could leverage this trust.

Another dimension to consider is the frequency of the incidents. This can give some indication of the intent of an insider. The first time an insider performs one or more of these acts, it may be accidental. However, when such actions are repeated, the classification might more appropriately be considered negligent or mischievous (Giandomenico & Groot, 2018).

4.2 Negligent Insider

Negligence refers to the deliberate omission of information security measures. We define a negligent insider as:

> *an insider who has no malicious intent associated with their intentional inaction; and through their passive-risk behaviour has caused harm or increased the probability of future harm to the organisation's information systems.*

Despite having security measures such as firewalls, security patches, and virus scanners, the threat to the confidentiality, integrity, and availability of information systems still exists through the negligence of the information security management teams or their users (Workman, Bommer, & Straub, 2008). Incidents may be a result of ignoring the security policy as the risk is perceived to be low, having naïve and careless attitude, and taking "short cuts" to save time among others (Gyunka & Christiana, 2017; Parsons et al., 2015). Users often use rationalisation to justify deviant actions (Barlow, Warkentin, Ormond, & Dennis, 2013). For example, an individual may share the network password with colleagues because they rationalise that their actions cause no harm.

Some examples of negligent insider threats to an organisation include:

- Failure to follow password policies: using default or easy to guess passwords, writing down the password, sharing the password with colleagues, or using the same passwords over many systems (Verizon, 2018). As per NCSC (UK) (2019), 123456 is the most popular password, and 45% of people reuse the password of their email account on other services.
- Failure to update patches: Software developers fix vulnerabilities to software and send updates for installation. More often than not, end-users delay installation of updates and with dire results (Ahola, 2019).
- Ignorance: Not having the required level of knowledge or enough information about a specific circumstance (Ahola, 2019) may result in making an incorrect entry or giving out the wrong information. Using unsecured Wi-Fi networks without understanding the risks can result in personal information and credentials being harvested (Ahola, 2019).

Negligent acts have no malicious intent or misuse of privileges but cause harm to the organisation due to inaction such as failing to update the password, leaving systems unattended, or failing to install security patches. The motivation of negligent acts may be due to convenience, high workload, policy complexity, or habitual bypassing of security mechanisms (Sasse et al., 2007). Negligence can also be from the management failing to upgrade the systems or software. Canner (2020) informs that only around 50% of the surveyed IT specialists used software to combat phishing attacks, and less than 50% used email encryption or provided secure collaboration tools. Even with the knowledge of the threats and risks, people do not use these technologies. Often non-malicious violations are not officially reported due to their unidentifiable nature or due to the violations seemingly so minor (AlHogail, 2015).

4.3 Mischievous Insider

We introduce the term mischievous to describe an intentional misuse of privileges without malicious intent. We define a mischievous insider as:

> *an insider who has no malicious intent associated with their intentional action; and through their active-risk behaviour has misused their privileges and has caused harm or increased the probability of future harm to the organisation's information systems.*

A mischievous insider is often aware of the security risks and knows the right course of action, but still chooses to violate policies as a shortcut or because they simply do not think that the rules are crucial (Ahola, 2019; Wall, 2013). Guo et al. (2011) conceptualise intentional non-malicious security violations to have all of the following four characteristics (1) intentional behaviour: non-malicious security violation with a conscious decision to follow a course of action, (2) self-benefiting without malicious

intent: such as to save time and effort, (3) voluntary rule-breaking: choosing to violate security policies at their own will, and (4) possibly causing danger or security risk. These characteristics apply to a mischievous insider. Examples of mischievous behaviour include:

- Accessing unauthorised confidential files for fun or out of curiosity without an intention to cause harm (Hunker & Probst, 2011);
- Downloading unauthorised files or software applications to finish a job quicker (Guo et al., 2011), installing unauthorised external applications on work devices to improve efficiency (Willison & Warkentin, 2013);
- Violating network usage policy (Stanton et al., 2005), using unauthorised removable media, allocating excessive privileges to users, disabling security configurations (Guo et al., 2011);
- Sharing the password with a colleague who has forgotten theirs may result in an unanticipated incident negatively effecting the organisation's information security (Willison & Warkentin, 2013).

Lacombe (2017) refers to a study from CompTIA and reports that human errors account for 52% of the root causes of security breaches. 42% of these cited end-user failures to follow procedures; 42% cited general carelessness; 31% cited the inability to get up to speed on new threats; 29% cited lack of expertise with applications; 26% cited IT staff failure to follow policies and procedure. Mischievous acts are caused by the intentional misuse of privileges without an intention to cause harm, such as using their privileges to access confidential information of a celebrity client out of curiosity. While viewing personal details may seem harmless, and a victimless crime, the client's privacy has been compromised.

Much of prior work does not differentiate a negligent insider from a mischievous insider. Instead, these are grouped as intentional non-malicious (Dupuis & Khadeer, 2016), non-malicious deviant (Van Den Bergh & Njenga, 2016), or volitional non-malicious non-compliance (Willison & Warkentin, 2013). While both types do not intend to cause harm and are aware of the associated risks, they are differentiated in terms of any misuse of their privileges. A negligent insider does not misuse their privileges but rather fail to comply with the security procedure, whereas a mischievous insider deliberately misuses their privileges. The action and behaviour behind negligent acts and mischievous acts are different. We, therefore, conceptualise these in a new way as separate classes of insider threats based on behaviour.

4.4 Malicious Insider

The definition of a malicious insider and their intentional action to breach security was consistent among prior researchers. We define a malicious insider as:

> *an insider who has a malicious intent associated with their action or inaction; and has caused harm or increased the probability of future harm to the organisation's information systems.*

A malicious insider is an insider who makes a conscious choice to misbehave and cause harm to the organisation (Gyunka & Christiana, 2017; Parsons et al., 2015; Wall, 2013). Some examples of malicious insider threats include:

- Insider social engineering: an insider manipulating a colleague to forcefully, intentionally, or unintentionally release sensitive information (Elmrabit et al., 2015).
- Insider fraud: an insider obtaining and retaining information such as credit card details for fraud and identity theft; Modifying information without authorisation with the intent to self-benefit (CERT, 2018; Nieles et al., 2017).
- Insider IP theft: an insider theft of trade secrets, modifying or stealing confidential or sensitive information; includes industrial espionage and colluding with outsiders (CERT, 2018)
- IT sabotage: an insider's use of their IT experience and knowledge to cause specific harm to an individual or organisation. (Elmrabit et al., 2015)
- Exploit security gaps: An insiders curiosity to discover vulnerabilities in existing or new technologies to launch attacks in future (Gupta, Arachchilage, & Psannis, 2018)

The motivation for an attacker to commit malicious acts could be for self-benefit or personal gains, sense of entitlement, disgruntlement, revenge (Hadlington, 2018)

4.5 Summary of Insider threats

An *accidental* insider has no intent to violate security policies nor a motive to do so. Often, they are fully skilled to do their jobs, and any incident may be a one-off and caused by a lapse in judgement. While both *negligent* and *mischievous* insiders are intentionally non-compliant, these intentions are not malicious in nature and they are differentiated in terms of any misuse of their privileges. Neither *accidental* nor *negligent* insiders deliberately misuse their privileges, whereas a *mischievous* or *malicious* insider may deliberately and skilfully misuse their privileges. Furthermore, a *negligent* insider knows that they are violating the security policies but either underestimate the risks or lack the motivation to comply. In contrast, a *mischievous* insider is aware of the risks and knowingly decides to break the rules without malice, while a *malicious* insider does so to self-benefit or to cause harm.

Through our profiling of accidental insider, negligent insider, mischievous insider, and malicious insider, we found that there was a crossover in the intent, motives, and actions between them. This crossover of characteristics blurs the distinction between conventional accidental, negligent, mischievous, and malicious acts. We attribute this overlap to the fact that intent is granular and a gradient rather than distinct. The threats caused by accident or oversight by careless or unmotivated employees can be the precursor to more extreme incidents (Willison & Warkentin, 2013). Accidental acts can soon turn to negligent acts then to mischievous acts and ultimately to malicious acts.

Figure 1 illustrates the differentiating characteristics of the types of insiders. Note that the types are conceptualised as a continuum rather than being distinct.

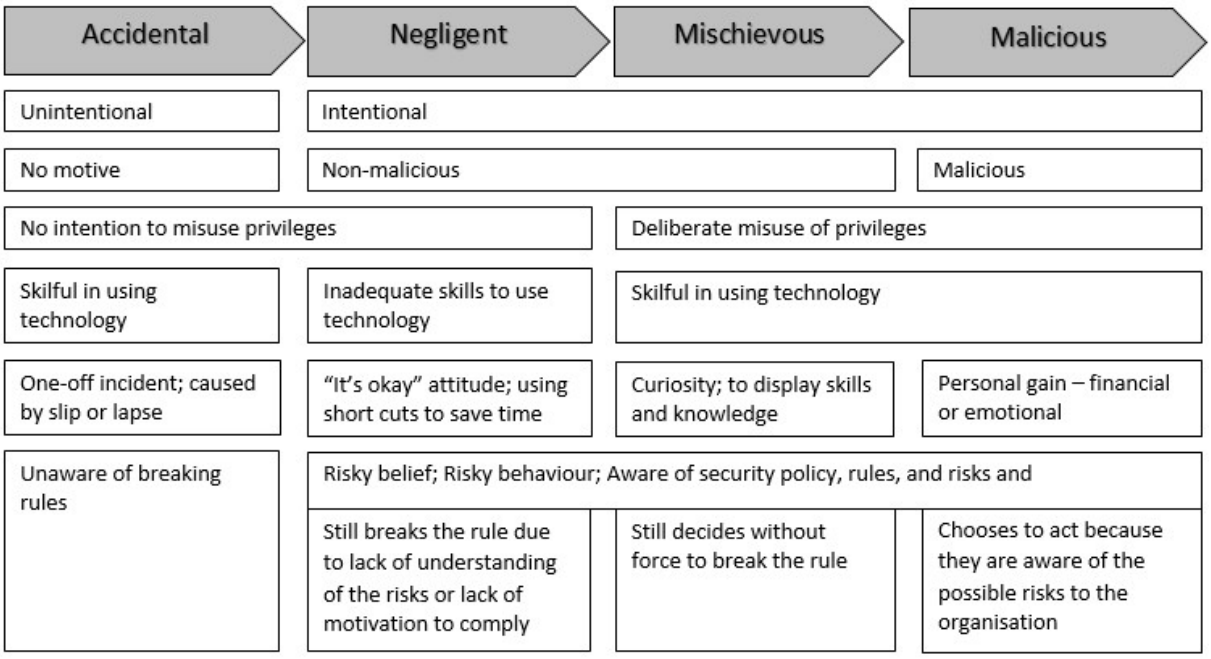

Figure 1: Summary of Types of Insiders

The motivation that differentiates a malicious insider from a non-malicious insider may not always be evident, and it may not be possible to differentiate these groups based on their actions. A survey conducted by Dupuis and Khadeer (2016) confirmed that the characteristics of a malicious and a non-malicious insider were similar. It is therefore hard in practice to make a clear distinction between the execution of malicious acts from naïve or accidental acts (Hunker & Probst, 2011). Hence, depicting each insider type as mutually exclusive may be simplistic or unrealistic.

5 Discussion

Organisations need to be able to differentiate the different types of threats to develop appropriate mitigation and limit the impacts on their cybersecurity. While a technical security vulnerability can be well-defined, the nature of insider threats may not be well understood. To address this issue, we present a classification of insider threats, including accidental, negligent, mischievous, and malicious insiders. The characteristics of an insider threat are differentiated based on an individual's intent, motive, and actions to inform organisations on the types of threat. Understanding the characteristics

of the different threat types can help organisations to identify potential threats and enhance their cybersecurity strategies.

This classification is key to improve the effectiveness of detection and prevention strategies. Organisations need to understand that it may no longer be sufficient to have one standard approach to address threats to cybersecurity. It is important to address each type of threat, identify suitable preventative strategies, and apply the right type of corrective actions towards accidental, negligent, mischievous, and malicious insiders. The proposed insider threat classification extends prior work by integrating the different classes of insider to present a consistent terminology and definition of insider threat types. Having a common vocabulary facilitates a unified view of insider threats and aids with the progression of future research in this direction.

The profiles of an accidental, negligent, mischievous, and a malicious insider, show a crossover in the intent, behaviour, and actions between the different types. This nexus suggests that individuals engaging in behaviour without any malicious intent could transition to engaging in behaviour to cause harm to the organisation. Therefore, in our insider threat classification model, we describe intent as a continuum of accidental, negligent, mischievous, and malicious nature. Figure 2 illustrates where this insider threat classification would fit in a broader security threat model. The classification of external threat sources, non-human internal threat agents, and the impacts of the threats are not in the scope of the current study and are promising candidates for further research.

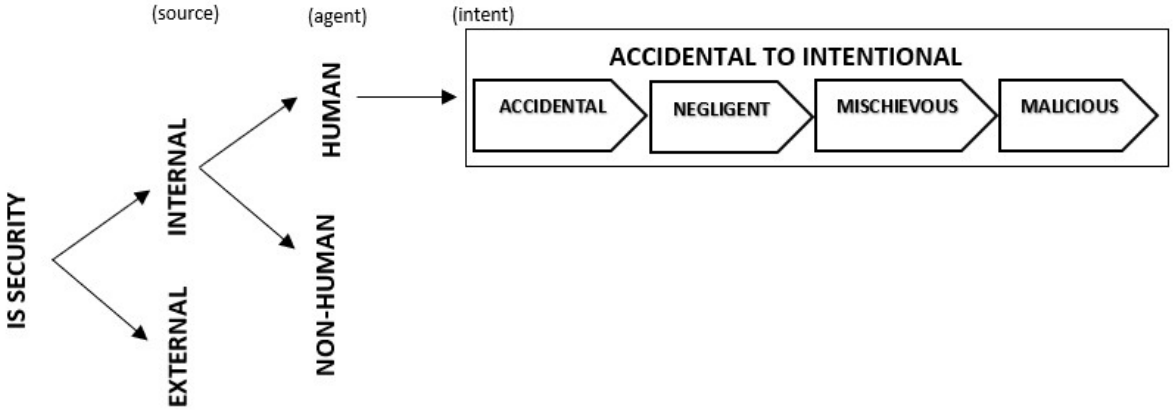

Figure 2: Insider Threat Classification (adapted from Willison and Warkentin (2013))

Our proposed classification extends Willison and Warkentin's (2013) model, which shows intent as a continuum of passive non-volitional non-compliance to volitional non-malicious non-compliance to intentional malicious computer abuse. We show intent as a continuum of accidental to negligent to mischievous and ultimately to an intentional act of cybercrime by an insider. The nature of the intent conceptualised as a continuum, as opposed to distinct, justifies the inconsistencies in prior literature on the classification of non-malicious threats.

A further implication of this model is to illustrate that individuals may transition from one insider type to another. An insider's accidental system misuse can become an opportunity for a future malicious act to breach the system or harm the organisation. This is of relevance to organisational stakeholders, as the transition may be identified only through keen observation. For example, an individual may accidentally discover a system loophole and may abuse the loophole with a perception that there is no negative consequence to their action. Eventually, they may be motivated to deliberately act either to self-benefit or to sabotage the organisation's information systems.

While an individual's intent of a breach may not be apparent when the breach is detected, the frequency of the breach and the insider's knowledge of the vulnerability can help determine the intent of the breach. If vulnerability to the system has been detected, the sequence of actions taken by the individual provides crucial insights into their intent. When a vulnerability is detected, the user must report it to the authority. If this vulnerability was discovered or exploited by accident, the report should still be made so that user training can be provided to overcome future incidents. However, if an individual is found to have repeatedly exploited the vulnerability without reporting it, it may be considered a deliberate malicious act and appropriate actions need to be taken by the organisation. Organisations will need to analyse these scenarios and implement the right policies and prevention strategies to protect their information systems.

6 Conclusion

Threat classification is essential for organisations to implement appropriate and effective cybersecurity strategies. In this paper, we have presented a unified typology and classification of insider threats building on prior research in this domain. The thematic analysis enabled us to first identify the problem areas in the information systems literature. We observed inconsistency in the insider type classification as it was based on different factors, and different terminology and definitions were applied for insider types. We extended prior work by providing a standard classification method, consistent terminology, and definition for insider types.

The proposed insider classification is based on an individual's intent, motive, and actions, and described as accidental, negligent, mischievous, and malicious. This classification can provide the organisation with an insight into the most relevant threats and areas that are vulnerable to these threats. We introduce the term mischievous insider to describe an intentional non-malicious misuse of privileges. We also give new definitions for each type of insider threat which will provide organisations with an insider type profile. Having a common vocabulary and a unified understanding of the threats will be useful for future research and will help to advance the field.

We contribute new knowledge to the field of cybersecurity with the classification of the four types of insiders. The knowledge of the distinction yet similarity between each of the types will allow organisations to have an increased awareness that individuals can move across four types over time. We also make an incremental contribution to the existing security threat model by outlining intent as a continuum of the four types of insiders – because the nature of the intent is granular as opposed to distinct. This justifies the inconsistencies in the literature on the classification of non-malicious threats. These contributions are summarised above in Figures 1 and 2.

Our classification of insider threats will be a necessary component of a comprehensive framework for best practices in cybersecurity. This study provides the foundations for future work to investigate prevention strategies and corrective actions to manage the impact of changing roles of an insider within a continuum model. Additional research will help us better ascertain the motivation, mode, and impact of the different insider type behaviours.

7 References

Ahola, M. 2019. "The Role of Human Error in Successful Cyber Security Breaches." *usecure* from https://blog.getusecure.com/post/the-role-of-human-error-in-successful-cyber-security-breaches

AlHogail, A. (2015). Design and validation of information security culture framework. *Computers in Human Behavior, 49*, 567-575.

Aurigemma, S., & Mattson, T. (2014). Do it OR ELSE! Exploring the effectiveness of deterrence on employee compliance with information security policies.

Barlow, J. B., Warkentin, M., Ormond, D., & Dennis, A. R. (2013). Don't make excuses! Discouraging neutralization to reduce IT policy violation. *Computers & security, 39*, 145-159.

Canner, B. 2020. "Egress Study Reveals Perils of Insider Data Breaches." *Information Solutions Review*, 2020 from https://solutionsreview.com/security-information-event-management/egress-study-reveals-perils-of-insider-data-breaches/

Carroll, T. E., Greitzer, F. L., & Roberts, A. D. (2014). Security informatics research challenges for mitigating cyber friendly fire. *Security Informatics, 3*(1), 13.

CERT. (2016). *Common Sence Guide to Mitigating Insider Threat, Fifth Edition*. Retrieved from https://resources.sei.cmu.edu/asset_files/TechnicalReport/2016_005_001_484758.pdf

CERT. (2018). *Common Sense Guide to Mitigating Insider Threats, Sixth Edition*. Retrieved from https://resources.sei.cmu.edu/asset_files/TechnicalReport/2019_005_001_540647.pdf

Crossler, R. E., Johnston, A. C., Lowry, P. B., Hu, Q., Warkentin, M., & Baskerville, R. (2013). Future directions for behavioral information security research. *Computers & security, 32*, 90-101.

Dupuis, M., & Khadeer, S. (2016). *Curiosity killed the organization: A psychological comparison between malicious and non-malicious insiders and the insider threat*. Proceedings of the 5th Annual Conference on Research in Information Technology.

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

Giandomenico, N., & Groot, J. d. 2018. "Insider vs. Outsider Data Security Threats: What's the Greater Risk?", 2020 from https://digitalguardian.com/blog/insider-outsider-data-security-threats

Guo, K. H., Yuan, Y., Archer, N. P., & Connelly, C. E. (2011). Understanding non-malicious security violations in the workplace: A composite behavior model. *Journal of management information systems, 28*(2), 203-236.

Gupta, B. B., Arachchilage, N. A., & Psannis, K. E. (2018). Defending against phishing attacks: taxonomy of methods, current issues and future directions. *Telecommunication Systems, 67*(2), 247-267.

Gyunka, B. A., & Christiana, A. O. (2017). Analysis of Human Factors in Cyber Security: A Case Study of Anonymous Attack on Hbgary. *Computing & Information Systems, 21*(2).

Hadlington, L. (2018). The "human factor" in cybersecurity: Exploring the accidental insider. In *Psychological and behavioral examinations in cyber security* (pp. 46-63): IGI Global.

Hunker, J., & Probst, C. W. (2011). Insiders and Insider Threats-An Overview of Definitions and Mitigation Techniques. *JoWUA, 2*(1), 4-27.

Im, G. P., & Baskerville, R. L. (2005). A longitudinal study of information system threat categories: the enduring problem of human error. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 36*(4), 68-79.

Kraemer, S., & Carayon, P. (2007). Human errors and violations in computer and information security: The viewpoint of network administrators and security specialists. *Applied ergonomics, 38*(2), 143-154.

Lacombe, Y. 2017. "What is the Greatest Vulnerability in Cyber Security Today?" *Vircom* from https://www.vircom.com/blog/greatest-vulnerability-cyber-security-today/

Li, H., No, W., & Boritz, J. (2020). Are External Auditors Concerned about Cyber Incidents? Evidence from Audit Fees. *Auditing, 39*(1), 151. doi:10.2308/ajpt-52593

Loch, K. D., Carr, H. H., & Warkentin, M. E. (1992). Threats to information systems: today's reality, yesterday's understanding. *Mis Quarterly*, 173-186.

Mouton, F., Leenen, L., & Venter, H. S. (2016). Social engineering attack examples, templates and scenarios. *Computers & Security, 59*, 186-209.

NAIC, N. I. A. C. (2008). *The Insider Threat to Critical Infrastructures*. Retrieved from https://www.cisa.gov/sites/default/files/publications/niac-insider-threat-final-report-04-08-08-508.pdf

NCSC (UK). 2019. "Most hacked passwords revealed as UK cyber security exposes gaps in online security." *National Cyber Security Centre* Retrieved 14-4-2020, 2020 from https://www.ncsc.gov.uk/news/most-hacked-passwords-revealed-as-uk-cyber-survey-exposes-gaps-in-online-security

Nieles, M., Dempsey, K., & Pillitteri, V. Y. (2017). NIST Special Publication: An Introduction to Information Security. *NIST Special Publication 800-12,* (Revision 1). Retrieved from doi:https://doi.org/10.6028/NIST.SP.800-12r1

Parsons, K. M., Young, E., Butavicius, M. A., McCormac, A., Pattinson, M. R., & Jerram, C. (2015). The influence of organizational information security culture on information security decision making. *Journal of Cognitive Engineering and Decision Making, 9*(2), 117-129.

Sasse, M. A., Ashenden, D., Lawrence, D., Coles-Kemp, L., Fléchais, I., & Kearney, P. (2007). Human vulnerabilities in security systems. *Human Factors Working Group, Cyber Security KTN Human Factors White Paper*.

Shred-it. (2018). *State of Industry Information Security 2018 North America*. Retrieved from https://www.shredit.com/getmedia/b5de58fd-7e17-4d18-b718-9eca8d0665a6/Shred-it-2018-North-America-State-of-the-Industry.aspx

Stanton, J. M., Stam, K. R., Mastrangelo, P., & Jolton, J. (2005). Analysis of end user security behaviors. *Computers & security, 24*(2), 124-133.

Van Den Bergh, M., & Njenga, K. (2016). *Information security policy violation: The triad of internal threat agent behaviors*. Proceedings of the 1st International Conference on the Internet, Cyber Security, and Information Systems (ICICIS).

Verizon. (2018). *2018 Data Breach Investigations Report*. Retrieved from https://enterprise.verizon.com/resources/reports/DBIR_2018_Report.pdf

Wall, D. (2013). Enemies within: Redefining the insider threat in organizational security policy. *Security journal, 26*(2), 107-124.

Willison, R., & Warkentin, M. (2013). Beyond deterrence: An expanded view of employee computer abuse. *MIS quarterly*, 1-20.

Workman, M., Bommer, W. H., & Straub, D. (2008). Security lapses and the omission of information security measures: A threat control model and empirical test. *Computers in Human Behavior, 24*(6), 2799-2816.

Acknowledgements

The authors would like to acknowledge Mason Torres, Curtin University, for his valuable comments during the preparation of this manuscript. We also want to acknowledge the three reviewers for their valuable suggestions and insightful comments, which helped to improve this manuscript.

Copyright

Copyright © 2020 authors. This is an open-access article licensed under a Creative Commons Attribution-NonCommercial 3.0 New Zealand, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and ACIS are credited.