A. S. M. Kayes[1], Wenny Rahayu[1], Tharam Dillon[1], Ahmad Salehi S.[1], and Hooman Alavizadeh[1]

[1]Affiliation not available

April 16, 2024

Safeguarding Individuals and Organisations from Privacy Breaches: A Comprehensive Review of Problem Domains, Solution Strategies, and Prospective Research Directions

A. S. M. Kayes, Member, IEEE, Wenny Rahayu, Member, IEEE, Tharam Dillon, Life Fellow, IEEE, Ahmad Salehi S., Member, IEEE, and Hooman Alavizadeh, Member, IEEE

***Abstract*—Privacy breaches have become increasingly prevalent, exposing individuals to significant risks. These breaches can have far-reaching consequences, including identity theft and life-threatening situations. Several studies have analysed data and privacy breaches and presented detection or prevention techniques to combat these breaches. However, because the number and type of breaches have significantly increased, these studies have become less relevant or outdated. Previous research on data/privacy breaches compared the techniques and results of various studies, but none attempted to comprehensively analyse the type of information and the type and level of compromise that occurred after such breaches. In this survey, we examine the fundamental concepts of privacy and security and define the security incidents and data/privacy breaches. We propose a set of criteria to evaluate the published studies on privacy breaches. We thoroughly investigate the problem domains and security-related concerns considering six recent breach cases in Australia, elucidating the critical challenges and issues associated with privacy breaches. We comprehensively review and outline the trends of security incidents and data/privacy breaches from 2020 to 2023. Additionally, we review the current state-of-the-art countermeasures to safeguard against these breaches. Finally, we identify an open research direction to develop an artificial intelligence (AI)-powered security framework that can help analyse cyber threats, characterise attackers' behaviours, distinguish between legitimate and illegitimate privacy policies, and restrict access to individuals' information. Overall, this survey will help organisations to reassess and update their security and privacy measures.**

***Index Terms*—Cybersecurity incidents, data and privacy breaches, access control policies, privacy policies, cyber threat modelling, and artificial intelligence.**

I. Introduction

CYBERSECURITY incidents and data/privacy breaches are increasing in number each year, as reported by Verizon [1]. Recent studies reveal that billions of dollars are lost globally to cybercrime and attacks annually, exemplified by cases like the Westpac cyber attacks [2], Medibank personal information breaches [3], and Optus data breaches [4]. These incidents can lead to fraudulent activities such as pay identity hacks and credit card fraud and can also jeopardise brand integrity and result in life-threatening situations and serious emotional distress for those who are affected [5]. It is necessary to examine various types of compromised data (e.g., personal information, financial data, health records) and the potential consequences for individuals and organisations (e.g., reputational damage, financial loss, legal implications) of data/privacy breaches, elucidating the implications of such incidents in alignment with global privacy laws, e.g., the Australian Privacy Act [6] and the EU General Data Protection Regulation (GDPR) [7].

The privacy landscape is evolving with the proliferation of the Internet of Things (IoT), big data, cloud/fog-based smart technologies, and social media platforms [8]–[10]. Methods of data collection, processing, and promotion have undergone substantial shifts, with a notable emphasis on unstructured text data. While there are benefits to storing text data in modern storage systems, the primary challenge lies in ensuring data privacy and security. Cybercriminals are increasingly targeting these contemporary Internet-driven data platforms [11], rendering them vulnerable to data and privacy breaches.

Traditional solutions such as encryption and authentication techniques [12], [13] and access and privacy control mechanisms [14]–[16] are inadequate against these breaches. Existing surveys of privacy/security issues and data breaches are either insufficient or largely outdated [8]–[11], [17]–[22]. There is an urgent need for a comprehensive survey covering the issues and problem domains of data/privacy breaches. Furthermore, current solutions and approaches, despite their inadequacy for application in modern platforms should be investigated. To comprehend the nature of these breaches, including the type of information compromised (personal, confidential, and/or sensitive information), the type of compromise (e.g., data integrity, confidentiality, and/or availability), the level of compromise (e.g., initial and/or network attacks), and the root cause of the attacks (e.g., human and/or system vulnerability), it is essential to scrutinise the existing privacy and security measures and policies of organisations.

The access control system, which has a long history, is one of the predominant security mechanisms to protect data from unauthorised entities. Various access control systems have been designed and implemented as the cornerstone of modelling privacy and security policies in today's pervasive

Dr. A. S. M. Kayes (a.kayes@latrobe.edu.au), Prof. Wenny Rahayu (w.rahayu@latrobe.edu.au), Prof. Tharam Dillon (t.dillon@latrobe.edu.au), Dr. Ahmad Salehi S. (a.salehishahraki@latrobe.edu.au), and Dr. Hooman Alavizadeh (h.alavizadeh@latrobe.edu.au) are with the Department of Computer Science and Information Technology, La Trobe University, Melbourne, VIC 3086, Australia. (Corresponding author: Dr. A. S. M. Kayes.)

Manuscript received April 09, 2024; revised April 09, 2024.

and Internet-driven environments. These systems aim to preserve personal, confidential, and sensitive information and protect data and resources from unauthorised users. However, one of the most challenging aspects of access control is to specify and model the necessary security and privacy policies. In certain domains, such as distributed cloud environments [23], although there are standards available for role-based access control (RBAC) [14], a noteworthy but often overlooked feature is context-aware access control (CAAC) [15]. CAAC introduces the specification of complex context-specific security and privacy policies. The dynamically changing contextual conditions, referred to as 'contexts' [24], play a vital role in specifying and enforcing these CAAC policies.

In summary, existing security and privacy control policies are predominantly predefined and rule-based, encompassing traditional RBAC and dynamic CAAC policies. These classical access control solutions encounter limitations in today's Internet-driven environments, partly due to their predefined rule-based nature and association with complex contextual constraints. The challenge posed by these complex RBAC and CAAC policies suggests a new direction for automated policies to counteract data and privacy breaches.

In this study, our survey methodology is designed to gather insights into the problem domains related to data and privacy breaches, as well as potential solution strategies and future research directions to address these challenges. The key research questions are as follows:

- What types of data and information are more vulnerable to compromise in cyber incidents/attacks, and what are the threat actors and attack vectors of these compromises?
- What solutions can be applied to detect and prevent breaches, strengthening cybersecurity defenses against evolving threat landscapes?

A. The Contributions

The main contributions of this study are as follows:

- We provide the background knowledge and motivation for this study, such as differentiating between security incidents versus data and privacy breaches, categorising different types of information, and different forms/types of compromise.
- We develop a survey methodology and define a set of evaluation criteria to investigate privacy breaches.
- We explore the current problem domains and privacy breach issues, along with existing approaches to security and privacy measures.
- We survey various security incidents and data/privacy breaches over the period from 2020 to 2023 and propose an innovative AI-powered security framework to protect organisations from these incidents and breaches.
- Finally, we evaluate the strengths and limitations of existing surveys of data/privacy breaches compared with our survey.

B. Outline of the Article

The rest of the article is organised as follows. We present the background knowledge and motivation of this study in Section II. Section III introduces the criteria to evaluate this survey, along with the methodology of this study. Current problem domains and issues are investigated in Section IV, including an overview of security incidents and breaches from 2020 to 2023. Section V discusses relevant approaches to protect organisations from privacy and security breaches. Section VI proposes an AI-powered framework for detecting, preventing, and mitigating data/privacy breaches. Section VII compares existing surveys and our study. Finally, we conclude the paper in Section VIII, along with suggestions for potential future research directions to address the issues of data and privacy breaches.

II. Background Studies and Research Motivation

This section overviews the existing studies and details the motivation for this research.

A. Security Incidents and Data/Privacy Breaches

Cybersecurity incidents and breaches are distinct entities. We distinguish between data and privacy breaches and security incidents, with various examples.

- **Incident Case:** An incident represents a security event, whether a true or false alarm, with the potential to impact the confidentiality, integrity, or availability of data/information. For instance, an organisation may detect a denial-of-service attack causing downtime and disruption to services, leading to inconvenience for clients and public embarrassment.
- **Breach Case:** A breach is the consequential outcome of an incident involving the unauthorised disclosure of data/information where individuals' information has been exposed to unauthorised parties. For example, an organisation experiencing a network intrusion detection attack finds that one of its databases has been compromised, resulting in hackers accessing personal details of staff, including names, dates of birth, addresses, and contact numbers.

Data/privacy breaches occur when individuals' data and information are lost, resulting in disclosure to unauthorised parties. For instance, the personal information of an organisation's staff member may be lost or a database containing personal information could be stolen or hacked by a cybercriminal.

- **Data Breach:** A data breach occurs when an individual's personal, confidential, or sensitive information has been compromised.
- **Privacy Breach:** A data breach is classified as a privacy breach when an individual can be potentially identified through leaked or compromised information.

In the following sections, we first classify different types of information that can contribute to data and/or privacy breaches, then we discuss different forms of compromise, and finally, we present the motivation for this study.

B. Different Types of Information

Different local and global privacy laws, such as the Australian Privacy Act [6] and the EU GDPR [7], delineate how various organisations collect and handle different types of information, encompassing aspects such as use and/or disclosure of information. The 13 Australian Privacy Principles (APP), embedded in the Australian Privacy Act [6], govern the collection, use, and disclosure of individuals' information. These principles underscore organisational governance, integrity, correction, and individual rights. A breach of an APP is deemed an interference with privacy, potentially resulting in regulatory consequences and penalties. The utilisation or disclosure of this information necessitates compliance with applicable privacy laws, such as the APP and GDPR.

Individuals' information can be classified into various types, and different laws and regulations are applicable to handle different categories of information. Based on the APP, the collection of sensitive confidential information and highly sensitive health information must align with the primary and/or secondary purposes. We classify individuals' information into the following categories and provide examples of the purpose of collecting this information, in line with the APP.

- **Personal Information:** Information about an individual, such as names and addresses, can be classified as personal information. According to the APP, an entity or organisation can collect an individual's personal information if it is reasonably necessary and done lawfully.
- **Confidential Information:** Information including credentials or identities (biometrics, passport numbers), dates of birth, and financial details (bank accounts) is considered confidential information. The APP stipulates that (i) confidential information, being sensitive, can be collected with the individual's consent and for a specified purpose; (ii) an organisation may use/disclose confidential information for secondary purposes (e.g., direct marketing) if the individual has provided consent.
- **Sensitive Information:** Information such as critical health records and other highly sensitive details (gender orientation, religious beliefs) is categorised as sensitive information. The APP imposes specific protections on the collection and handling of sensitive information, for instance, (i) health information can be collected with consent for primary purposes, and (ii) sensitive information can be used for primary purposes (e.g., care/treatment) but not for direct marketing.

C. Different Types of Compromise

Based on the widely accepted CIA (confidentiality, integrity, and availability) security triad or model [25], there is a possibility that confidentiality, integrity, or availability can be compromised due to data/privacy breaches. Table I discusses the terminologies and definitions of CIA security compromises resulting from these breaches.

TABLE I
COMPROMISE OF CIA.

Terminology	Definition
Confidentiality has been compromised	This means information has been disclosed to an unauthorised user.
Integrity has been compromised	This means information has been modified by an unauthorised user.
Availability has been compromised	This means information is not accessible to an authorised user.

D. The Motivation of This Study

Data and privacy breach incidents are a significant concern, exemplified by notable breaches involving personal and sensitive health information in the networks of prominent entities. Various personal, confidential, and sensitive information breaches have occurred globally during and post-COVID such as Australia's Optus [4], Medibank [3], Latitude Financial [26], and My Health Record [27] breaches over the past few years [28]. These types of breaches cost billions of dollars every year globally [29].

Existing state-of-the-art research on privacy and security breaches and current surveys of these incidents are limited to the specific domains that we study in this research. Thus, there is a need for a literature survey to protect individuals and organisations from privacy breaches. We need to identify the problem domains of these privacy breach cases and potential approaches to identify, prevent, and mitigate these breaches and security incidents. It is necessary to investigate the key issues and concerns, specifically examining recent data and privacy breach incidents, such as exploring the type of information breached (e.g., users' personal or sensitive information), the form of compromise (confidentiality, integrity, and/or availability), the level or severity of the breach (e.g., the attack was initial and/or network-level), the root causes of the incidents (e.g., human or system-level vulnerability), and propose a possible solution direction to combat these privacy breaches.

E. Understanding Privacy Breaches

In a discussion of personal, confidential, and/or sensitive information breaches, it is necessary to carefully analyse the fundamental concepts of privacy and security and illustrate their significance through examples.

- **Privacy:** In the context of privacy breaches, privacy refers to the fundamental principle that individuals or entities have the right to control who accesses their information. It encapsulates the idea that data subjects should retain control over their confidential or sensitive information, ensuring that they have the authority to determine how their data is utilised. Privacy breaches occur when this control is compromised, typically through misuse, or exploitation of personal, confidential, or sensitive information.

  Breach Scenario: Consider a scenario where a social media platform's privacy settings allow users to control who can view their posts or profile information. In this context, a privacy breach occurs when the platform fails to uphold these settings, enabling unauthorised access to users' personal information. This breach could result from loopholes in the platform's privacy policies, which may

TABLE II
KEY FACTORS EXPLORED IN PRIVACY BREACH INVESTIGATIONS.

Type of Information	Cyber Threat Modelling	CIA Security Triad
• Personal information • Confidential information • Sensitive information	• Initial attack • Network attack	• Confidentiality • Integrity • Availability

allow the platform to intentionally misuse or share users' data beyond what the users intended or consented to.

- **Security:** A security breach involves the unauthorised intrusion into information systems. Security encompasses various measures aimed at protecting information from unauthorised access, alteration, or destruction, thereby reinforcing data integrity and confidentiality. In this context, privacy breaches occur when external malicious actors violate security protocols or firewalls and involve unauthorised access to personal, confidential, or sensitive information, potentially compromising its confidentiality or integrity.

  Breach Scenario: Consider a scenario where a healthcare organisation experiences a privacy breach due to the inadequate protection of a health information management system. Hackers gain unauthorised access to the system and extract sensitive medical information, compromising patient privacy. Similarly, a financial institution may suffer a privacy breach resulting from insufficient cybersecurity measures, leading to unauthorised access to customers' sensitive financial data.

III. Survey Methodology and Evaluation Criteria

The survey presented in this article serves dual objectives: firstly, to gain insight into the nature of data and information compromised in data breach incidents, and secondly, to ascertain the extent of compromise experienced. The overarching aim is to facilitate a comprehensive investigation into the root causes of cyber-attacks. Through an understanding of the types of data affected and the level of compromise incurred, the study aims to contribute to the formulation of effective privacy and security measures essential for safeguarding organisations against potential threats. The Australian Privacy Act 1988 [6] and the EU GDPR [7] are used to distinguish different types of personal information. The cyber threat model [30] is also used to inspect the level of compromise encountered. The CIA security model [25] is used to explore the components (confidentiality, integrity, and/or availability) that have been compromised.

Table II details the factors we investigate to analyse various data and privacy breaches. We define a set of evaluation criteria to apply to the survey, which are as follows.

- **Criteria 1: Compromised information**. We investigate the type of data and information that has been compromised after cyber incidents/attacks, such as personal, private/confidential, or sensitive information (i.e., highly sensitive health information or critical business data).

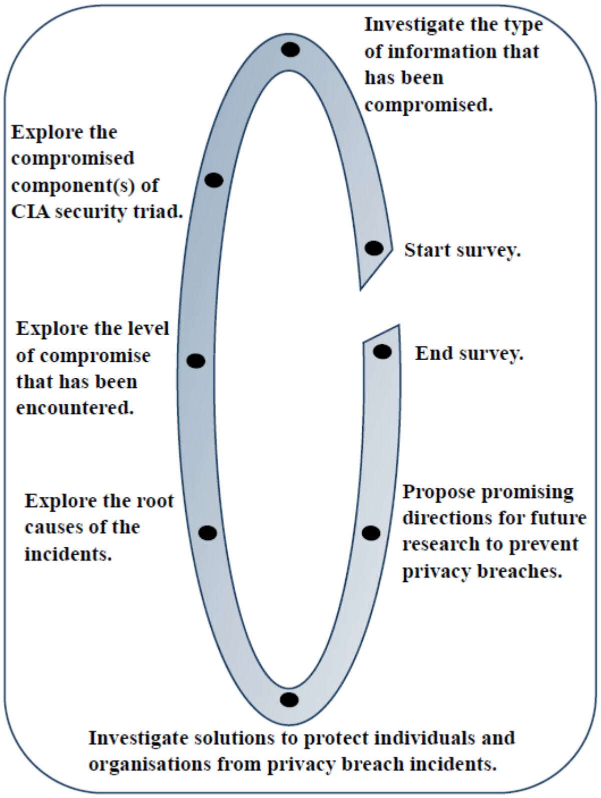
Investigate the type of information that has been compromised.
Explore the compromised component(s) of CIA security triad.
Start survey.
Explore the level of compromise that has been encountered.
End survey.
Explore the root causes of the incidents.
Propose promising directions for future research to prevent privacy breaches.
Investigate solutions to protect individuals and organisations from privacy breach incidents.

Fig. 1. A six-step process employed to examine relevant studies from the current state-of-the-art literature and investigate promising directions for future research.

- **Criteria 2: CIA model**. We explore what has been compromised after the attack using the CIA model, confidentiality, integrity, or availability.
- **Criteria 3: Initial/Network attack**. We explore the level of compromise that has occurred due to cyber incidents.
- **Criteria 4: Human/System vulnerability**. We thoroughly investigate the incidents to explore the root causes or weak links such as vulnerable humans and/or systems.

In addition to these four evaluation criteria, we also **investigate potential solutions** to safeguard individuals and organisations against breach incidents. We finally **propose promising directions for future research** to prevent privacy breaches. Figure 1 illustrates the six-step methodology that was followed to conduct the literature review for this study.

Table III shows the keywords and phrases that we use to identify the relevant articles to be reviewed in this study.

We employed various online repositories, such as IEEE Xplore, ACM Digital Library, and Google Scholar to identify pertinent articles. Figure 2 shows the number of articles containing the keyword "privacy breaches" on Google Scholar over the past decade. Our search yielded 18,510 results within a timeframe spanning from 2014 to 2023. While privacy breaches represent longstanding concerns, they have attracted substantial research attention, as evidenced by a marked increase in the number of articles, particularly in 2023, compared to preceding years (e.g., 2,230, 2,530, and 4,350 articles in 2021, 2022, and 2023, respectively). We meticulously

TABLE III
KEYWORDS AND PHRASES TO IDENTIFY THE KEY ARTICLES.

Keywords	Phrases
• Privacy breaches	• Personal and confidential information breaches
• Personal information breaches	• Sensitive and business data breaches
• Sensitive information breaches	• Compromise of patients' health record
• Data breach investigation	• Survey on data breach challenges and issues
• Privacy breach investigation	• Survey on privacy breach challenges and issues
• Access control	• Security and access control policies
• Privacy	• Privacy laws and policies
• Authentication	• One, two, and multi-factor authentication techniques

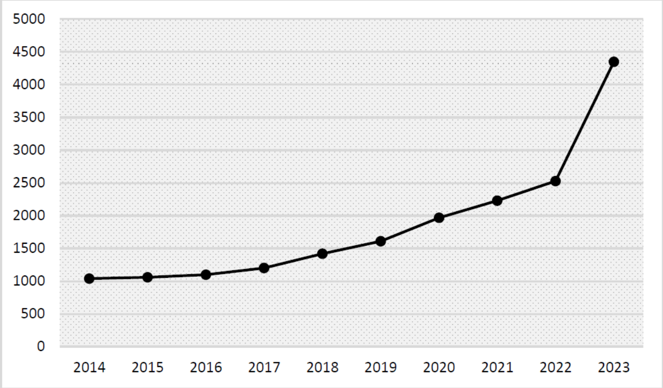

Fig. 2. The number of articles containing the keyword "privacy breaches" by year on Google Scholar.

reviewed the articles retrieved from multiple digital libraries using the keywords and phrases delineated in Table III (both exact and relative matches). Initially, we conducted a cursory examination of the titles and abstracts, excluding irrelevant articles, followed by a thorough evaluation of the selected articles by scrutinising their full content.

IV. Privacy Breaches: A Survey of Current Problem Domains and Issues

This section summarises recent security incidents and data/privacy breach cases and reviews various studies to investigate the type of information and the CIA security pillars that have been compromised, the level of compromise, and the root causes of these incidents.

A. Review of Studies to Investigate the Type of Information Compromised

We analysed some of the recent cybersecurity incidents in various organisations, such as personal and/or health information breaches in Australian hospitals and Medibank [3], Optus data breaches [4], and Westpac's PayID attacks [2] and investigate the type of information that has been compromised, including personal, confidential, or sensitive information such as financial records and medical details.

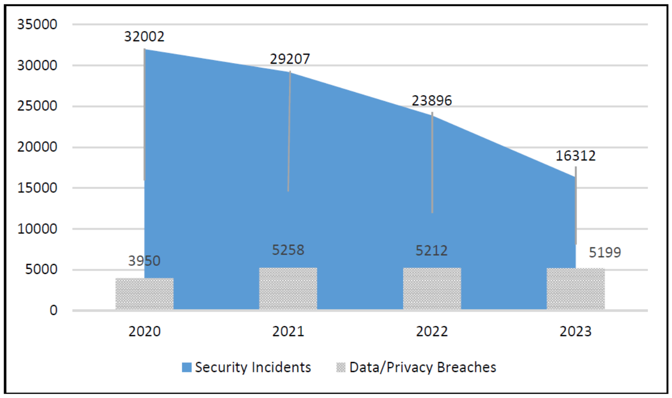

Fig. 3. Total number of security incidents and data/privacy breaches: trends from 2020 to 2023.

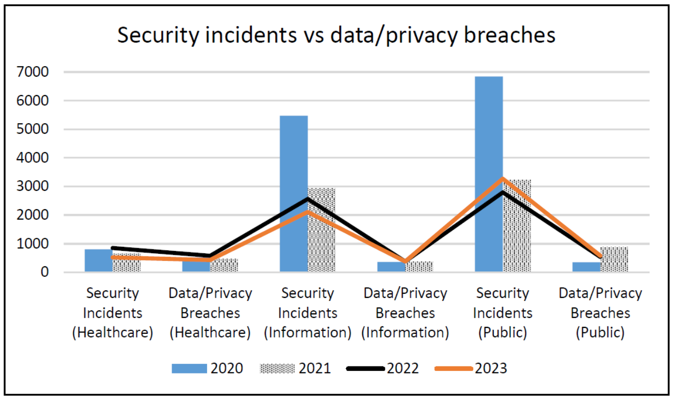

Fig. 4. Exploring security incidents vs data/privacy breaches in different organisations (2020-2023).

Using Verizon's Data Breach Investigation Reports (DBIRs) (e.g., 2023 DBIR [1]), we analysed various breach occurrences during and post-COVID from 2020 to 2023 and summarise our findings in Figure 3. A security incident refers to a security event that compromises one or multiple components of the CIA security triad, and a data breach specifically denotes a security incident resulting in the confirmed disclosure of information to an unauthorised user. The data reveals an increase in the number of breaches in 2023 involving cryptocurrency (refer to Verizon's recent DBIR [1]).

Following Verizon's DBIRs from 2020 to 2023, we analysed the number of incidents and breach occurrences in healthcare, information, and public administration organisations (see Figure 4), such as the recent breach cases which occurred in Australia's Optus [4] and Medibank [3] data and privacy breaches, resulting in the compromise of personal and sensitive information.

The data shows that of 522 security incidents, 433 occurrences resulted in breaches in healthcare organisations, constituting approximately 8% of the total breaches which occurred in 2023. A similar trend is observed in other sectors, with approximately 11% in public organisations in 2023. However, these DBIRs are not adequate for categorising the numerous data and privacy breaches resulting in the compromise of

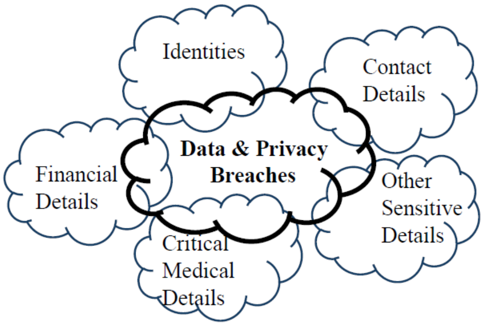

Fig. 5. Types of information which may be compromised in a data/privacy breach.

different information types (e.g., personal, confidential, and sensitive information).

As reported in existing studies, a serious consequence of a privacy breach is the compromise of confidential or sensitive information. Figure 5 illustrates the types of personal, confidential, and sensitive information which may be compromised in a cybersecurity incident. Adhering to the regulations outlined in the EU GDPR [7] and the Australian Privacy Act [6], and based on the information types listed in Table II, several privacy breaches in Australia's healthcare, information, financial, and public systems have been investigated [2]–[4], [26], [27]. These breaches have compromised various types of information, resulting in significant ramifications.

In summary, the types of data/information which may be compromised in a personal information breach, a confidential information breach, and a sensitive information breach, are shown in Figure 5.

- *Personal Information Breach*: This encompasses the compromise of **contact details** such as names, email addresses, and physical addresses.
- *Confidential Information Breach*: This type of breach involves the compromise of credentials or **identities**, including tax file numbers, passport numbers, and driving license numbers, as well as **financial details** such as bank accounts.
- *Sensitive Information Breach*: This type of breach involves the compromise of **critical medical details**, such as medical history (disability, illness), diagnosis reports, and genetic information. Additionally, it includes **other sensitive details** like criminal records, political affiliations, religious beliefs, race information, and sexual orientation.

B. Review of Studies to Explore the Compromised Component(s) of the CIA Security Model

The CIA security model [25] has been devised to analyse data/privacy breaches by identifying the components of the model that have been compromised. We investigate recent cyber incidents in Australia and summarise the findings as follows:

- **Ransomware Attacks on Victorian Hospitals [3]:** Ransomware attacks have disrupted the normal operations of hospitals across Gippsland, Geelong, and Warrnambool in Victoria. Some health services have been forced to shut down, including certain IT systems such as booking and management systems, as well as electronic health record systems. There has been no evidence to suggest that personal, confidential, or sensitive (health) information was accessed by cybercriminals. However, several systems needed to shut down, leading to delays in surgeries, as well as inpatient and outpatient care, due to computer networks being compromised by ransomware attacks.
- **Westpac PayID Attacks [2]:** Westpac experienced a significant breach of personal information affecting one hundred thousand customers, as cybercriminals exploited vulnerabilities in the PayID lookup function. This breach exposed customer data, including mobile numbers and emails, which in turn exposed bank account details. These incidents underscore the severity of the breach, as attackers abused the vulnerable PayID lookup system to gain unauthorised access to personal and confidential (financial) information. These attacks not only compromised the integrity and confidentiality of customer data but also raised concerns about the security of Westpac's online banking features.
- **Optus Cyber Incidents [4]:** The personal and confidential data of millions of current and former customers, such as names, dates of birth, phone numbers, email addresses, driver's licences, and/or passport numbers, were compromised in the Optus cyber incidents. These incidents not only jeopardised the confidentiality of customers' personal and confidential data but also raised concerns regarding the integrity of Optus's systems and the trustworthiness of their security measures. Upon discovery of the data and privacy breaches, Optus acted swiftly to shut down the attacks and initiated collaboration with the Australian Cyber Security Centre and the Information Commissioner to mitigate potential risks to affected customers.
- **My Health Record Incidents [27]:** The My Health Record system, a significant initiative by the Australian government, has faced challenges in managing cybersecurity risks, as revealed by a review conducted by the national audit office. The audit highlighted inadequacies in the management of cybersecurity and privacy risks within the system, raising concerns about the protection of personal and sensitive (health) information. It was reported that not all Australian healthcare providers maintained minimum cybersecurity standards, contributing to the sector reporting the highest number of notifiable data breaches across other sectors. Incidents involving the unauthorised collection, use, or disclosure of health information have led to breaches compromising both the integrity and confidentiality of patient's health data. These incidents and data breaches made the system opt-out.
- **Medibank Cyber Incidents [3]:** Millions of customers were affected by the cyber incidents at Medibank, which are considered to be among the most significant breaches of personal and confidential information in recent history (e.g., names, birth dates, and passport numbers).

Cybercriminals initially gained access to internal credentials belonging to an individual and then leveraged these stolen credentials to gain unauthorised access to Medibank's systems, ultimately identifying the location of a customer database. It was during the reconnaissance phase of the cyber kill chain that Medibank's security team allegedly detected suspicious activity. Despite their efforts to intervene by shutting down backdoor access, approximately several hundred gigabytes of customers' personal and confidential information had already been exfiltrated by the cybercriminals.

- **Latitude Financial Cyber Incidents [26]:** A significant breach of customer data occurred across Australia and New Zealand in the cyber attacks on Latitude Financial, another alarming data breach in the region. Millions of present and past customers fell victim to this breach, with cybercriminals gaining access to passport numbers, driver's licence numbers, and other personal information. While many details about the breach are still emerging, it has undoubtedly compromised the confidentiality of customer data, raising serious concerns about privacy and security.

Table IV shows the components of the CIA security triad that have been compromised in recent cyber incidents in Australia's healthcare, information, public, and financial sectors.

TABLE IV
CYBER ATTACKS VERSUS COMPROMISED CIA SECURITY COMPONENTS.

Cyber Attacks	**Compromised CIA Components**
Ransomware Attacks on Victorian Hospitals [3]	Availability has been compromised.
Westpac PayID Attacks [2]	Confidentiality and integrity have been compromised.
Optus Cyber Incidents [4]	Confidentiality and integrity have been compromised.
My Health Record Cyber Incidents [27]	Confidentiality, integrity, and availability have been compromised.
Medibank Cyber Incidents [3]	Confidentiality and integrity have been compromised.
Latitude Financial Cyber Incidents [26]	Confidentiality and integrity have been compromised.

C. Review of Studies to Explore the Level of Compromise Encountered

Cybersecurity incidents can compromise individuals and/or networks, resulting in data/privacy breaches. To understand the level of compromise encountered, we analyse recent incidents and summarise the findings across various types of organisations. According to Verizon's 2023 DBIR [1], healthcare, information, financial, and public systems are more susceptible to being compromised by cybercriminals.

Following the concept of cyber threat modelling, this section demonstrates the level of compromise after any cyber incident. Particularly, we utilise the cyber kill chain model [30] to delve into breach cases and identify the initial/individual versus group/network-level compromises. We also examine the various threat actors and attack vectors that have been utilised by cybercriminals to breach privacy.

- **Healthcare Systems:** Healthcare data is a growing target for cybercriminals. Ransomware attacks on hospitals are growing in number [31]. Cyber attacks on several Victorian hospitals across some regional areas have been discovered [3] and cybersecurity experts suggest that ransomware is likely the cause of the disruption. In response to the attacks, hospitals isolated affected systems and networks and disconnected them from the internet to prevent further escalation.

  Personal and confidential information breaches as a result of the Medibank cyber security incident [3] were also discovered recently, which enabled the theft of internal credentials due to privileged system access. This exposed a vast array of customer data (e.g., names, dates of birth, and passport numbers). The threat actor crafted convincing phishing emails purporting to be from legitimate sources, enticing unsuspecting recipients to click on malicious links leading to credential-stealing websites. Once the attackers obtained the credentials, they exploited their access to infiltrate Medibank's network and locate the customer database, leading to a massive privacy breach.
- **Information Systems:** In the Optus cyber incidents [4], cybercriminals exploited a vulnerable application programming interface to infiltrate the company's systems and access millions of customers' personal and confidential information. These breaches highlight the threat posed by external actors seeking to exploit weaknesses in Optus' software systems. Furthermore, the compromise of email addresses and phone numbers exposed customers to additional risks, as cybercriminals leveraged this information to launch phishing attacks. In these Optus impersonation scams, fake messages informed recipients that their personal information had been compromised in the data breach, potentially leading to further exploitation of unsuspecting individuals.

  In Westpac's PayID system attacks [2], the threat actors utilised both internal and external resources to orchestrate these incidents, taking advantage of the vulnerable PayID lookup system. Attackers repeatedly pinged the PayID name lookup service thousands of times and successfully compromised account holders' names and other bank details (personal and confidential information) associated with phone numbers and emails.
- **Financial Systems:** The cyber incidents at Latitude Financial [26] marked one of the largest-known data breaches of an Australian financial institution, following massive cyber attacks at Medibank and Optus that had already compromised millions of customers' personal information. In the case of Medibank, hackers leaked the stolen data onto the dark web after Medibank refused to meet their ransom demands. Similarly, in both the Optus and Latitude cases, cybercriminals demanded ransoms, but the organisations did not comply with their demands. However, specific details regarding the cybercriminals, threat actors, and attack vectors involved in the Latitude case remain undisclosed.
- **Public Systems:** Similar to the data and privacy breaches

encountered by other sectors such as healthcare, information, and financial, one of Australia's large public systems My Health Record [27] was targeted by both internal and external threat actors, resulting in a range of cybersecurity incidents. Human errors, such as the entry of incorrect patient details, and unauthorised Medicare claims are among the examples of incidents that have occurred within the system.

Table V shows the level of compromise(s), threat actor(s), and attack vector(s) encountered in the cyber incidents in Australia.

TABLE V
THREAT ACTORS, ATTACK VECTORS, AND LEVELS OF COMPROMISES ENCOUNTERED AFTER CYBER INCIDENTS.

Cyber Incident	Threat Actor	Attack Vector	Level of Compromise
Ransomware Attacks on Victorian Hospitals [3]	Unknown	Unknown	Network
Medibank Cyber Incidents [3]	Internal & External	Email	Network
Westpac PayID Attacks [2]	Internal & External	Vulnerable System	Network
Optus Cyber Incidents [4]	Internal & External	Vulnerable System, Email, & Phone	Network
Latitude Financial Cyber Incidents [26]	External	Unknown	Network
My Health Record Cyber Incidents [27]	Internal & External	Vulnerable People	Initial

These incidents and data/privacy breaches underscore the escalating threat landscape faced by organisations across various sectors, emphasising the critical need for robust cybersecurity measures. Such measures are necessary to safeguard personal and confidential information, including sensitive information such as medical records and customer data, and to mitigate the risks posed by cyber threats.

D. Review of Studies to Explore the Root Causes of the Incidents

To understand the underlying reasons behind breaches involving various types of information, we explore the contributing factors. These incidents relate to breaches of personal, confidential, and/or sensitive data in Australia's hospitals and healthcare organisations, such as those in regional hospitals in Victoria and Medibank [3], to those impacting major financial organisations like Westpac banking systems [2] and Latitude Financial Services [26], and major telecom companies like Optus [4]. They highlight the complex and varied nature of the cybersecurity challenges we face.

- **Weak/Inadequate Security Measures:** The lack of encryption for sensitive health or business information, can leave it vulnerable to interception. Similarly, insufficient access controls and/or failure to implement multi-factor authentication mechanisms can allow unauthorised users to gain access to confidential or sensitive information.
- **Human Vulnerability in Cybersecurity:** Human error can lead to various cybersecurity vulnerabilities, including the accidental disclosure of sensitive information through email or other communication channels. Misconfiguration of security settings on systems or applications is another common consequence of human error. Additionally, failure to follow established security procedures, such as using weak passwords or sharing credentials, can significantly compromise cybersecurity defenses.
- **Phishing and Social Engineering:** Email and web attacks are common cyber attack vectors, where deceptive emails or web messages can trick employees into revealing sensitive information or downloading malware. Impersonation of trusted entities can lead to unauthorised access to crown jewels or sensitive data, while manipulation of individuals through psychological tactics can bypass security controls.
- **Lack of Regular Security Audits:** Frequent and adequate security assessments of systems and networks are essential for identifying vulnerabilities and weaknesses. Failure to regularly apply security patches or updates leaves systems exposed to known vulnerabilities that malicious actors can exploit. Additionally, adequate monitoring and logging mechanisms can detect and respond to security incidents effectively.
- **Insufficient Data Protection Policies:** The absence of clear guidelines for handling and storing personal, confidential, and sensitive information securely can lead to data/privacy breaches. The lack of data classification policies to prioritise protection efforts based on data sensitivity and insufficient controls for data access and sharing throughout its lifecycle are also additional vulnerabilities.
- **Poor Incident Response:** The lack of a documented incident response plan detailing roles, responsibilities, technical measures, and business processes for responding to security incidents, such as data and privacy breaches, can lead to a further impact on individuals, systems, or networks. Inadequate training and rehearsal of incident response procedures can result in delays or errors during investigations, affecting effective resolution and further escalation. Failure to establish communication channels and contacts for reporting and escalating security incidents effectively may impede a timely response and mitigation efforts.
- **Vulnerabilities in Applications and System Software:** Weaknesses in software applications or operating systems allow unauthorised access or data leakage and leave applications or systems exposed to hackers. The exploitation of vulnerabilities in third-party applications used by the organisation, coupled with insufficient oversight of third-party vendors' security practices, can lead to the compromise of customers' information.
- **Insider Threats and Lack of Cybersecurity Awareness:** Vulnerable employees who may engage in malicious actions can harm the organisation. Unauthorised access to sensitive information by employees for personal gain, as well as accidental data exposure due to a lack of

TABLE VI
A COMPREHENSIVE ASSESSMENT OF ROOT CAUSES IN AUSTRALIA'S RECENT DATA AND PRIVACY BREACHES.

Root Causes	Data/Privacy Breaches	Incidents Description
Weak or inadequate security measures	• Medibank cyber incidents [3]	• In Medibank's cyber incidents, hackers stole millions of customers' data and demanded ransoms.
Human vulnerability in cybersecurity	• Optus cyber attacks [4]	• In the recent Optus attacks, human error was identified as the primary issue, stemming from a lack of understanding regarding cybersecurity practices.
Phishing and social engineering	• Medibank cyber incidents [3] • Westpac PayID attacks [2] • Optus cyber attacks [4]	• In Australia's Medibank cyber incidents, stolen credentials with privileged access to internal systems were sold on the dark web, allowing unauthorised access to millions of customers' data. • Almost a hundred thousand Australians' confidential information, such as mobile numbers and email addresses, have been compromised through phishing attacks in Westpac's cyber incidents. • A similar mechanism was employed in the Optus cyber attacks, where stolen credentials facilitated unauthorised access and data theft.
Lack of regular security audits	• Ransomware attacks on Victorian hospitals [3]	• Following ransomware attacks on Victorian hospitals across Gippsland, Geelong, and Warrnambool, the state's auditor general emphasised that Victorian patients' health data was very vulnerable to attack, underscoring the urgent need for improved cybersecurity measures and regular security audits.
Insufficient data protection policies	• Cyber attacks on Latitude Financial [26]	• In Australia's recent Latitude cyber incidents, millions of Australians and New Zealanders' personal and sensitive information, such as driving licence numbers, passport numbers, bank account numbers, and credit card numbers, have been stolen by hackers.
Poor incident response	• Cyber attacks on Latitude Financial [26]	• In Latitude's cyber incidents, the lack of appropriate incident response preparedness further exacerbated the consequences, potentially prolonging the exposure of personal and sensitive records to unauthorised access by hackers.
Vulnerabilities in applications and system software	• Optus cyber attacks [4]	• In the Optus attacks, personal information about millions of Australians, including names, email and postal addresses, phone numbers, and dates of birth, was stolen by exploiting unauthorised access to current and former customers' information systems.
Insider threats and lack of cybersecurity awareness	• My Health Record cyber incidents [27]	• Incorrect patient details were mistakenly entered into Australia's My Health Record system due to human negligence. This oversight resulted in additional personal information breaches, underscoring the significant repercussions of such errors in data management and security.

awareness about cybersecurity best practices, can lead to serious consequences.

Table VI outlines some notable root causes of Australia's recent cybersecurity incidents, such as data and privacy breaches. In summary, addressing the root causes of data/privacy breaches necessitates a multifaceted approach encompassing technical safeguards (privacy and security measures that we discuss in Section V), adequate security audits, appropriate data protection, proper incident response, and cybersecurity awareness. By addressing these underlying factors, organisations can fortify their cybersecurity posture and mitigate the risk of privacy breaches in an increasingly interconnected digital landscape.

V. PRIVACY BREACHES: A SURVEY OF EXISTING SOLUTION APPROACHES

The previous sections investigated the types of compromised information, CIA security attacks, and the root causes of cyber incidents leading to data and privacy breaches. This section details the existing solution approaches (such as security and privacy measures) to protect individuals and organisations from possible privacy breaches.

Table VII summarises the potential solution approaches, including privacy laws/regulations, authentication techniques, access controls, and dynamic decision making.

TABLE VII
DATA AND PRIVACY BREACHES VS POTENTIAL SOLUTION APPROACHES.

Data/Privacy Breaches	Potential Solutions
• Medibank cyber incidents [3] • Optus cyber attacks [4] • Westpac PayID attacks [2] • Ransomware attacks on Victorian hospitals [3] • Latitude cyber incidents [26] • My Health Record cyber incidents [27]	• Compliance with relevant privacy standards, laws, and regulations. • Identity verification and management with appropriate authentication techniques to ensure user privacy. • Proper access controls to protect data from unauthorised entities. • AI involvement to automate privacy and access control decision making, supporting various types of information and distinguishing between legitimate and illegitimate privacy policies. • Robust cybersecurity measures, proactive threat detection, and timely response strategies to safeguard against external threats and protect sensitive customer data and/or health information from malicious exploitation.

A. Review of Privacy Models and Approaches to Detect and Prevent Breaches

1) Privacy Laws and Regulations: The Australian Privacy Act 1988 [6] and its 13 APP have been introduced to promote and protect the privacy of individuals and to regulate how entities, such as Australian agencies and organisations, collect, use, store, and distribute information about individuals. Specifically, APP 3 outlines various requirements for the collection of personal and sensitive information, as well as obligations regarding professional confidentiality and the type of consent needed to collect different types of information about individuals. For instance, APP 3 stipulates that an entity can only collect personal information about individuals directly from the individuals themselves, and only if it is directly required for or directly related to the entity's functions or activities. Furthermore, an APP entity is only permitted to collect sensitive information about an individual with their explicit consent.

Similar to the APP, the EU GDPR [7] imposes additional requirements for collecting, processing, and distributing special categories of personal data, which are defined as sensitive information by the Australian Privacy Act 1988 [6]. These special categories include the genetic, biometric, and health data of individuals, as well as personal data revealing racial or ethnic origins, political opinions, religious or philosophical beliefs, trade union memberships, and data concerning sexual orientations. Under the GDPR, the processing of such sensitive personal data is prohibited unless it is expressly allowed by law or the data subject, who is the owner of the data, has provided explicit consent for specified purposes.

Some studies have adopted the Australian and EU privacy laws as a basis for introducing privacy models and frameworks. For example, Notario et al. [32] introduced the concept of "accountability" to ensure user privacy in accordance with the Australian and EU privacy laws. In line with the principles of the EU GDPR, Alshamsan and Chaudhry [16] proposed a privacy framework for data protection. This framework includes a web-based application capable of visually displaying potential risks associated with online privacy policies.

Appropriate privacy policies, following legal and regulatory strategies, can detect and prevent privacy breaches. Non-compliance with regulations such as Australia's Privacy Act [6] and/or the EU GDPR [7] can lead to breaches of personal, confidential, and sensitive information. Failure to implement the controls required by regulatory standards, such as encryption, authentication, or data access controls, and inadequate monitoring and reporting mechanisms to demonstrate compliance with regulatory requirements can exacerbate these risks.

2) Verifying Entities and Protecting Information Using Authentication and Encryption Techniques: Given the prevalence of data and privacy breaches and their potentially severe impact on individuals and organisations, robust encryption and authentication techniques are indispensable for mitigating such risks. These techniques verify the identities of users or entities and protect information control, thereby preventing unauthorised access to personal, confidential, and sensitive information.

Notable authentication methods such as one-factor authentication (e.g., passwords or fingerprints) and two or multi-factor authentication (MFA) [13] are used to protect users' credentials from being compromised. Additionally, different cryptographic techniques and algorithms such as identity-based encryption (IBE) and attribute-based encryption (ABE) [12] are used to protect data from being compromised, stolen, or unauthorised modification and sharing.

Adherence to best practices is essential for effective privacy protection. This entails employing strong encryption techniques, implementing one, two, or multi-factor authentication mechanisms, regularly updating security protocols related to encryption and authentication, and complying with pertinent privacy regulations and standards such as the EU GDPR or the Australian Privacy Act [32].

3) Machine Learning (ML)-Based Models and Approaches to Ensure User Privacy: Several studies have proposed ML-based models and approaches to ensure user privacy. Zimmeck and Bellovin [33] advocated a classification-based automatic solution employing ML models. Their work focused on scrutinising diverse privacy policies on the web to enhance transparency regarding online privacy. Das et al. [34] introduced an ML-based methodology titled "privacy assistants for IoT" to empower users in managing their data, considering individual privacy expectations and personalised preferences.

Liu et al. [22] conducted a survey of existing ML approaches and models to discern and mitigate the privacy and security risks posed by malicious IoT devices. Additionally, Zaeem et al. [35] proposed an ML-driven tool tailored for web users, with a focus on assessing online privacy policies.

In summary, ML techniques exhibit considerable potential for augmenting various facets of user privacy protection.

4) Semantic-Based Approaches for Enhancing Data Privacy and Security: Semantic-based approaches, such as various ontology models, have been proposed to enhance data privacy and security. Hecker et al. [36] introduced a privacy ontology that encompasses different security principles and mechanisms as conceptual entities, along with the associations between them. This ontology aids in achieving interoperability and ensuring compliance with data subject rights as mandated by privacy regulations.

Alkhariji et al. [37] presented a semantics-based privacy approach tailored for IoT applications. Aligned with the Australian Privacy Act and EU GDPR, they advocated for the integration of privacy-by-design practices during the system design phase, incorporating privacy patterns and strategies.

Kayes et al. [15] proposed a policy ontology for modeling security policies. Their policy enforcement architecture facilitates the management of authorised and unauthorised access to individuals' information. The efficacy of the ontology concepts has been validated through applications in healthcare, demonstrating their completeness, correctness, and consistency.

In summary, incorporating security principles and mechanisms into a privacy ontology reinforces methodologies aimed at empowering data subjects to control their information. When coupled with authentication and other security measures, semantic or ontology-based approaches offer comprehensive privacy protection.

B. Review of Security Models and Approaches to Protect Data from Unauthorised Entities

We discuss classical and ML-based access control models aimed at safeguarding data from unauthorised entities and restricting access rights and privileges associated with personal, confidential, and sensitive data.

1) Classical Access Control: Discretionary access control (DAC) and mandatory access control (MAC) [14] are two fundamental access control models used to enforce security policies and regulate access to data and information resources in computing systems. MAC is commonly used in high-security environments such as government and military systems, where strict control over information access is necessary to prevent unauthorised access or disclosure. Unlike MAC, where access control decisions are centrally managed, DAC allows data subjects/owners to grant or deny access rights independently in decentralised but less restrictive environments such as personal computers and small-scale networks.

Traditional RBAC [14] and attribute-based access control (ABAC) [38] play crucial roles in ensuring authorised versus unauthorised access to data and information resources in different computing environments. These RBAC and ABAC models are two widely used access control models that provide a flexible and efficient means of managing access to data. In RBAC, access control decisions are based on user's roles, and access permissions are granted to roles rather than directly to individual users. In ABAC, attributes include various characteristics such as user roles/identities and departments, and access control policies define rules that evaluate these attributes to make access control decisions.

Both users' roles and dynamic attributes have been incorporated into the CAAC models [24], where access control decisions are based on contextual attributes associated with users, resources, and the environment. Kayes et al. introduced a family of CAAC models in pervasive or dynamic environments to access data from single and/or multiple sources [15], [23], [24], [39]. The CAAC policies are defined using relevant rules formed using dynamically changing contextual conditions (e.g., user profile and spatial/temporal information). Different types of contexts such as general context, relationship context, situational context, and fuzzy context information are considered in different application settings.

In summary, RBAC simplifies access control by organising permissions around user roles, while ABAC provides more granular control by considering a broader range of attributes. On the other hand, the CAAC models offer efficient ways to manage access to data, enhance decision-making capabilities using dynamic contexts, and ensure security in diverse computing environments. However, these classical access control solutions are typically complex and largely inflexible. They are usually static rule-based and unable to automatically adapt to the constantly changing access settings of different users and/or environments.

2) ML-Driven Access Control: ML-driven access control models have been introduced to automate decision-making capabilities [40]–[42].

Mayhew and Atighetchi [40] proposed a behaviour-based ML-driven access control for anomaly detection. Statistical ML models and behavioural patterns are used to predict the number of HTTP requests and establish TCP connections. The proposed solution is an automated ML-based system capable of decision making without manual intervention. Outchakoucht et al. [41] proposed a reinforcement learning-based access control approach for distributed IoT environments. Recently, Argento et al. [42] proposed an ML-based access control mechanism as the first line of defense, based on users' behavioural patterns. The authors considered the amount of data and the frequency of access as behavioural patterns.

Unlike classical access control models such as ABAC, RBAC, and CAAC, these ML-driven access control models employ various ML techniques and algorithms to enhance decision making.

C. Discussion

In summary, we have discussed the current state-of-the-art privacy, security, and ML-based techniques and approaches to detect and prevent data and privacy breaches and protect relevant information from unauthorised entities.

- **Privacy Policies for Handling Diverse Types of Data and Prevent Data Breaches:** It is imperative to establish robust privacy policies tailored to address various categories of information, with a focus on preventing data breaches and safeguarding information from unauthorised access.
- **Identity Management:** Individual and group-based identities need to be managed with proper encryption and authentication techniques to verify authorised versus unauthorised entities and subsequently protect data/information from being compromised.
- **Security Policies for Protecting Data from Unauthorised Entities:** It is necessary to formulate comprehensive access control policies to govern and control unauthorised access, enhancing the resilience of organisational assets against potential threats and privacy breaches.
- **ML-Based Approaches for Data and Privacy Breach Detection:** Implementing advanced ML techniques to identify and mitigate potential personal, confidential, and sensitive information breaches can ensure a proactive and adaptive security posture.

VI. An AI-Powered Security Framework for Detecting, Preventing, and Mitigating Data and Privacy Breaches

This section introduces an AI-powered security framework against data and privacy breaches.

Cyber threats that exploit vulnerabilities in security and privacy policies are unavoidable. This issue can make organisations' security and privacy protocols vulnerable due to the poor setting of policies (e.g., access control policies, privacy policies). Thus, it is necessary to improve security and privacy measures against these vulnerabilities. Currently, AI-powered learning approaches are increasingly being used to develop and automate new technologies for different purposes, such as providing real-time responses and alarms as well as personalised task reminders and modelling dynamic policies. AI-powered solutions can safeguard organisations from cyber threats and attacks that can potentially exploit vulnerabilities in security and privacy policies.

We propose an innovative AI-powered framework as a countermeasure against cybersecurity incidents and data/privacy breaches, encompassing the situation-awareness, privacy, and security layers (see Table VIII). These three layers will assist in detecting, preventing, and mitigating breaches.

TABLE VIII
AN AI-POWERED SECURITY FRAMEWORK AGAINST DATA/PRIVACY BREACHES.

Layer	Feature	Description
Situation-awareness layer	Detection	Innovative AI technologies can be applied to analyse cybersecurity threats, hackers' mindsets, and suspicious behaviours. This layer can also help identify malicious IoT devices, gather background information about attackers, and detect potential breaches.
Privacy layer	Prevention	An ML-based privacy model can discern and distinguish legitimate and illegitimate policies, facilitating the identification of organisations' improper collection, processing, and storage of personal, confidential, and sensitive information. Leveraging AI technologies such as large language models (LLMs), this ML-based approach can dynamically audit and examine organisations' privacy and service-level agreements, thereby contributing to the prevention of data and privacy breaches.
Security layer	Mitigation	An adaptable, dynamic access control model can enhance decision-making capabilities by leveraging AI technologies. It enables the specification of new security policies in real time to combat both known and unknown threats and attacks, thereby mitigating data and privacy breaches.

A. Detecting Breaches

Maintaining situational awareness is paramount for fortifying defenses against the evolving landscape of cyber threats. Leveraging AI, as discussed by Alavizadeh et al. [43], proves pivotal in this endeavour, particularly in detecting data and privacy breaches. AI technologies enhance cyber threat intelligence by integrating behavioural analysis, enabling the identification of malicious patterns and anomalies, and monitoring suspicious activities encompassing network traffic and system logs. Using sophisticated models and algorithms, AI can forecast potential threats by analysing historical data, predict future attack vectors and vulnerabilities, and assist in attributing attacks to specific threat actors or groups by scrutinising various indicators of compromise, tactics, and techniques.

However, the utilisation of AI engines, particularly those employing LLMs and proprietary data, poses inherent privacy risks. Chen et al. [44] emphasised the need to implement privacy-preserving techniques throughout the AI lifecycle, encompassing pre-training, in-training, and post-training phases. While AI-powered tools require access to private organisational or individual information to comprehend behaviours and situations and deliver timely responses, safeguarding this confidential or sensitive data remains paramount.

Proactively detecting deviations from normal behaviours and promptly flagging potential threats in real time can enable organisations to maintain a proactive stance against emerging cybersecurity and privacy risks, such as data and privacy breaches.

B. Preventing Breaches

Utilising AI for the classification of legitimate versus illegitimate privacy policies can serve as a crucial component in preventing breaches. By leveraging advanced ML models and algorithms, organisations can analyse vast amounts of privacy policies, identifying discrepancies and red flags that may indicate unauthorised or unethical data collection, processing, or storage practices.

AI technologies such as language models can be used to analyse the linguistics of privacy policies and discern illegitimate actions or practices within them, such as collecting or disseminating sensitive information without consent from data subjects/owners. LLMs can be adopted to thoroughly distinguish between legitimate and illegitimate privacy policies, enabling comprehensive audits of organisations' privacy and service agreements to ensure compliance with regulatory requirements and ethical standards.

Integrating an ML-based privacy model into the cybersecurity framework represents a proactive approach to safeguarding individuals' privacy rights and preventing breaches of personal, confidential, and sensitive information.

C. Mitigating Breaches

Dynamic and adaptable security policies can restrict users from accessing personal or sensitive information, thereby mitigating data and privacy breaches. By leveraging AI-powered security policies, such as a deep neural network-based access control model, organisations can dynamically introduce new access control policies and enhance decision-making capabilities to effectively counter both known and unknown cybersecurity threats and attacks. It can effectively manage access control permissions, determining who can access what information under what conditions, thus providing robust protection against privacy breaches.

Unlike traditional CAAC models, which require the intricate and static specification of access control policies based on contextual conditions, the DNN-based approach alleviates the burden on system and security administrators. This approach enables proactive measures to be taken to mitigate potential data and privacy breaches before they occur.

In essence, the proposed AI-powered security framework can serve as a force multiplier in the realm of cyber threat intelligence, aiding in differentiating legitimate versus illegitimate privacy policies, as well as automating and enhancing access control decision making. This empowers organisations to proactively identify, analyse, detect, prevent, and mitigate cyber threats with unparalleled speed, accuracy, and efficacy. By harnessing the power of AI, organisations can stay one step ahead of adversaries and safeguard their digital assets against evolving threats such as data and privacy breaches.

VII. EXISTING SURVEYS VS OUR SURVEY

Following our proposed survey methodology, evaluation criteria, and survey of problem domains, issues, and solution approaches to detect and prevent data and privacy breaches,

TABLE IX
A COMPARATIVE ASSESSMENT BETWEEN EXISTING SURVEYS AND OUR STUDY IN THIS RESEARCH (‘$\surd$’ MEANS AVAILABLE, ‘X’ MEANS NOT AVAILABLE, AND ‘Δ’ MEANS PARTIALLY AVAILABLE).

Survey	**Problem Domains and Issues**				**Solution Approaches**	**AI-Powered Framework**
	Compromised Information	**Compromised CIA Component**	**Initial/Network Attack**	**Human/System Vulnerability**		
Data breaches [9] [17] [45]	Δ	X	X	X	Δ	X
Privacy and security issues in IoTs and healthcare [10] [18]	Δ	X	X	X	Δ	X
Privacy issues and security breaches [11]	Δ	X	X	Δ	Δ	X
Privacy issues and ML-driven innovations [22]	Δ	X	X	Δ	Δ	X
Privacy concerns and related issues [20] [21]	Δ	X	X	X	Δ	X
Privacy breaches [8] [19]	Δ	X	X	X	Δ	X
Our Survey	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$

we assess the strengths and limitations of existing relevant surveys.

Several recent studies [9], [17], [45] examined the challenges associated with data breaches. Barona and Anita [9] conducted a survey addressing data breach challenges, providing a comprehensive summary of various issues and cyber threats in the context of the cloud computing landscape. Neto et al. [45] surveyed data breach challenges and developed a global database of these challenges, including details about the affected clients and the amount of data involved in these breach incidents. Hassanzadeh et al. [17] investigated human-centric issues such as users' mental models related to data breaches, especially users' misconceptions about breach cases.

Other studies investigated privacy and security issues in various computing environments [10], [11], [18], [22]. Yang et al. [18] conducted a survey on privacy and security issues in IoT-driven environments, while Hathaliya and Tanwar [10] conducted a comprehensive survey on privacy and security issues in healthcare. Liu et al. [22] investigated different ML approaches that can support modern industries such as surveillance systems, financial organisations, and smart healthcare, where privacy issues and subsequent security attacks are significant concerns. They conducted a comprehensive study on privacy preservation problems and ML approaches, such as private ML, ML-aided privacy protection, and ML-based privacy attacks, along with corresponding protection schemes as solution directions. Recently, Wang et al. [11] presented different security and privacy concerns in the metaverse. They specifically investigated next-generation Internet-related emerging technologies such as AI and blockchain, where serious privacy concerns and security breaches can be encountered.

Several studies explored the impact and consequences of privacy breaches [8], [19]–[21]. Liginlal et al. [19] conducted an empirical study on human error-specific privacy breaches and proposed a framework for error management. Mamonov and Benbunan-Fich [8] presented their empirical investigation report on privacy breach perceptions among smartphone application users, specifically including human activities while collecting and misusing personal information that potentially leads to privacy concerns. Recently, Abawajy et al. [20] studied privacy concerns on social networks and categorised different adversarial background knowledge used by adversaries to mount privacy breach attacks on these networks, while Kokolakis [21] reviewed the current research on users' privacy attitudes and behaviours.

These existing studies and surveys on privacy and security issues and breaches are not adequate to identify and categorise the type of information (personal, confidential, and sensitive) that has been compromised, as well as the level and form (e.g., data confidentiality and integrity compromised due to a network-level attack), and the root causes of compromise (e.g., vulnerable human or software system). Our survey includes a comprehensive evaluation of recent cyber incidents in Australia, precisely identifying what information was compromised, and introduces potential solutions in contrast to the existing approaches.

Table IX outlines our comparative assessment of existing studies and surveys on privacy and security issues and breaches versus our survey. This comparative analysis provides insights into the strengths and limitations of existing research and underscores the contributions and novel insights offered by our study.

VIII. Conclusion and Future Research

Cybersecurity incidents, resulting in data and privacy breaches, persist as a significant challenge, especially in today's pervasive, dynamic, and technology-driven smart environments, where the privacy landscape is constantly evolving. Consequently, many existing studies and surveys on privacy and security breaches may be outdated or not directly applicable to systematically analyse and identify the aftermath of such cybersecurity incidents. Recent cybersecurity incidents in Australia, including the Medibank data breaches and cyber attacks on Optus and Latitude Financial networks, highlight the pressing need to address this issue effectively.

In this study, we conducted a review of six recent cybersecurity incidents in Australia, which resulted in significant data breaches. Our analysis identifies the type of compromised information, the root causes, and the level and form of compromise associated with these incidents. Additionally, we explored potential solution approaches and proposed an AI-powered security framework as a countermeasure for detecting, preventing, and mitigating data and privacy breaches. Furthermore, we examined existing surveys on security and privacy breaches and presented a comparative assessment table, contrasting them with our survey findings.

A. Promising Future Research Directions

- **Integrating Security and Access Control for Holistic Privacy Protection and to Mitigate Privacy Breaches:** A critical area for future research lies in addressing privacy breaches resulting from inadequacies in privacy, security, and access control mechanisms. By integrating these three facets into a cohesive framework, researchers can develop holistic solutions to mitigate privacy breaches effectively. This entails exploring innovative approaches to harmonise privacy policies with robust security measures, ensuring compliance with regulations and protecting sensitive information from unauthorised access. Additionally, enhancing access control mechanisms to provide granular control over personal, confidential, and sensitive data access is essential.
- **Investigating the Risk and Impact of Privacy Breaches:** Analysing the security principles of confidentiality, integrity, and availability (CIA security triad) to assess the risk of breaches is vital. This involves evaluating whether unauthorised entities gained access to personal, confidential, and sensitive information, whether the integrity of the information was compromised, and whether the availability of the information was affected, such as through denial-of-service attacks or data manipulation. Developing privacy-preserving technologies and automated compliance monitoring systems can significantly contribute to safeguarding individuals' privacy rights and mitigating the impact of privacy breaches. Focusing on these interconnected issues, future researchers can innovate more resilient and secure digital ecosystems, where data protection will be prioritised and privacy breaches will be minimised.
- **AI-Powered Threat Intelligence:** Leveraging AI-powered threat intelligence and predictive analytics can enhance proactive detection and mitigation of cyber incidents leading to data and privacy breaches. By analysing patterns and trends in cybersecurity threats, organisations can anticipate and prevent potential breaches, thereby strengthening their security posture and protecting personal, confidential, and sensitive information.
- **Behavioural Analysis for Insider Threat Detection:** Given evidence from recent cyber incidents in Australia, insider threats pose a significant risk to data and privacy breaches. Future research should explore innovative techniques for detecting anomalous behaviours within organisations. Leveraging ML and behavioural analytics, researchers can develop proactive strategies to identify suspicious activities and mitigate the risk of insider attacks.
- **Self-Supervised Approach to Detect Data and Privacy Breaches:** Exploring self-supervised learning approaches for detecting data breaches can enhance the ability to identify and respond to cybersecurity incidents effectively. By training models on unlabeled data, self-supervised learning algorithms can learn meaningful representations of normal behaviours, enabling them to detect deviations indicative of potential breaches without relying on labeled training data.
- **ML-Driven Privacy Assistant for Automated Compliance Monitoring:** To ensure adherence to privacy regulations and standards, such as the Australian Privacy Act and the EU GDPR, future research should explore the development of an automated compliance monitoring system. By leveraging advanced AI techniques and algorithms, an ML-driven privacy assistant can be developed which can continuously monitor data collection, processing, storing, and distribution activities and detect potential violations in real time, thereby enhancing regulatory compliance and minimising the risk of privacy breaches.

By addressing these key research areas, future studies can contribute to the development of robust frameworks and technologies that effectively mitigate data and privacy breaches and safeguard individuals' rights to privacy in an increasingly interconnected digital landscape.

References

[1] Verizon, "2023 data breach investigations report," *Verizon Team, Available: https://www.verizon.com/business/resources/Tdbc/reports/2023-data-breach-investigations-report-dbir.pdf*, pp. 1–89, 2023.

[2] J. Seaman and J. Seaman, "Compliance - a team effort," *PCI DSS: An Integrated Data Security Standard Guide*, pp. 323–358, 2020.

[3] M. Dart and M. Ahmed, "CYBER-AIDD: A novel approach to implementing improved cyber security resilience for large Australian healthcare providers using a unified modelling language ontology," *Digital Health*, vol. 9, pp. 1–15, 2023.

[4] D. Kolevski, K. Michael, R. Abas, and M. Freeman, "Cloud computing data breaches in news media: Disclosure of personal and sensitive data," in *IEEE International Symposium on Technology and Society*. IEEE, 2022, pp. 1–11.

[5] P. Mayer, Y. Zou, F. Schaub, and A. J. Aviv, "Now I'm a bit angry: Individuals' awareness, perception, and responses to data breaches that affected them," in *USENIX Security*, 2021, pp. 393–410.

[6] M. Paltiel, "Recent amendments to the Australian Privacy Act," *Journal of Bioethical Inquiry*, vol. 20, no. 2, pp. 161–167, 2023.

[7] M. Goddard, "The EU General Data Protection Regulation (GDPR): European regulation that has a global impact," *International Journal of Market Research*, vol. 59, no. 6, pp. 703–705, 2017.

[8] S. Mamonov and R. Benbunan-Fich, "An empirical investigation of privacy breach perceptions among smartphone application users," *Computers in Human Behavior*, vol. 49, pp. 427–436, 2015.

[9] R. Barona and E. M. Anita, "A survey on data breach challenges in cloud computing security: Issues and threats," in *2017 International Conference on Circuit, Power and Computing Technologies*. IEEE, 2017, pp. 1–8.

[10] J. J. Hathaliya and S. Tanwar, "An exhaustive survey on security and privacy issues in Healthcare 4.0," *Computer Communications*, vol. 153, pp. 311–335, 2020.

[11] Y. Wang, Z. Su, N. Zhang, R. Xing, D. Liu, T. H. Luan, and X. Shen, "A survey on metaverse: Fundamentals, security, and privacy," *IEEE Communications Surveys & Tutorials*, 2022.

[12] M. Yang, H. Wang, and D. He, "Puncturable attribute-based encryption from lattices for classified document sharing," *IEEE Transactions on Information Forensics and Security*, 2024.

[13] W. Li, H. Cheng, P. Wang, and K. Liang, "Practical threshold multi-factor authentication," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3573–3588, 2021.

[14] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models," *Computer*, vol. 29, no. 2, pp. 38–47, 1996.

[15] A. Kayes, J. Han, W. Rahayu, T. Dillon, M. S. Islam, and A. Colman, "A policy model and framework for context-aware access control to information resources," *The Computer Journal*, vol. 62, no. 5, pp. 670–705, 2019.

[16] A. R. Alshamsan and S. A. Chaudhry, "A GDPR compliant approach to assign risk levels to privacy policies." *Computers, Materials & Continua*, vol. 74, no. 3, 2023.

[17] Z. Hassanzadeh, R. Biddle, and S. Marsen, "User perception of data breaches," *IEEE Transactions on Professional Communication*, vol. 64, no. 4, pp. 374–389, 2021.

[18] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security and privacy issues in Internet-of-Things," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1250–1258, 2017.

[19] D. Liginlal, I. Sim, and L. Khansa, "How significant is human error as a cause of privacy breaches? An empirical study and a framework for error management," *Computers & Security*, vol. 28, no. 3-4, pp. 215–228, 2009.

[20] J. H. Abawajy, M. I. H. Ninggal, and T. Herawan, "Privacy preserving social network data publication," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1974–1997, 2016.

[21] S. Kokolakis, "Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon," *Computers & Security*, vol. 64, pp. 122–134, 2017.

[22] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin, "When machine learning meets privacy: A survey and outlook," *ACM Computing Surveys*, vol. 54, no. 2, pp. 1–36, 2021.

[23] A. Kayes, W. Rahayu, P. Watters, M. Alazab, T. Dillon, and E. Chang, "Achieving security scalability and flexibility using fog-based context-aware access control," *Future Generation Computer Systems*, vol. 107, pp. 307–323, 2020.

[24] A. Kayes, W. Rahayu, T. Dillon, E. Chang, and J. Han, "Context-aware access control with imprecise context characterization for cloud-based data resources," *Future Generation Computer Systems*, vol. 93, pp. 237–255, 2019.

[25] S. Samonas and D. Coss, "The CIA strikes back: Redefining confidentiality, integrity and availability in security." *Journal of Information System Security*, vol. 10, no. 3, 2014.

[26] S. Mckeith, "Cybersecurity: The new frontier," *LSJ: Law Society Journal*, no. 5, pp. 32–42, 2023.

[27] A. Medhekar, "My Health Record and emerging cybersecurity challenges in the australian digital environment," in *Research Anthology on Securing Medical Systems and Records*. IGI Global, 2022, pp. 428–447.

[28] L. Kyi, S. Ammanaghatta Shivakumar, C. T. Santos, F. Roesner, F. Zufall, and A. J. Biega, "Investigating deceptive design in GDPR's legitimate interest," in *ACM CHI*, 2023, pp. 1–16.

[29] H. Tao, M. Z. A. Bhuiyan, M. A. Rahman, G. Wang, T. Wang, M. M. Ahmed, and J. Li, "Economic perspective analysis of protecting big data security and privacy," *Future Generation Computer Systems*, vol. 98, pp. 660–671, 2019.

[30] T. Yadav and A. M. Rao, "Technical aspects of cyber kill chain," in *International Symposium on Security in Computing and Communication*. Springer, 2015, pp. 438–452.

[31] G. Mott, S. Turner, J. R. Nurse, J. MacColl, J. Sullivan, A. Cartwright, and E. Cartwright, "Between a rock and a hard(ening) place: Cyber insurance in the ransomware era," *Computers & Security*, vol. 128, p. 103162, 2023.

[32] N. Notario, A. Crespo, Y.-S. Martín, J. M. Del Alamo, D. Le Métayer, T. Antignac, A. Kung, I. Kroener, and D. Wright, "PRIPARE: Integrating privacy best practices into a privacy engineering methodology," in *IEEE Security and Privacy Workshops*. IEEE, 2015, pp. 151–158.

[33] S. Zimmeck and S. M. Bellovin, "Privee: An architecture for automatically analyzing web privacy policies," in *23rd USENIX Security Symposium (USENIX Security 14)*, 2014, pp. 1–16.

[34] A. Das, M. Degeling, D. Smullen, and N. Sadeh, "Personalized privacy assistants for the Internet of Things: Providing users with notice and choice," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 35–46, 2018.

[35] R. N. Zaeem, S. Anya, A. Issa, J. Nimergood, I. Rogers, V. Shah, A. Srivastava, and K. S. Barber, "Privacycheck's machine learning to digest privacy policies: Competitor analysis and usage patterns," in *IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*. IEEE, 2020, pp. 291–298.

[36] M. Hecker, T. S. Dillon, and E. Chang, "Privacy ontology support for e-commerce," *IEEE Internet Computing*, vol. 12, no. 2, pp. 54–61, 2008.

[37] L. Alkhariji, S. De, O. Rana, and C. Perera, "Semantics-based privacy by design for Internet of Things applications," *Future Generation Computer Systems*, vol. 138, pp. 280–295, 2023.

[38] A. S. Shahraki, C. Rudolph, and M. Grobler, "Attribute-based data access control for multi-authority system," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 2020, pp. 1834–1841.

[39] A. Kayes, W. Rahayu, and T. Dillon, "Critical situation management utilizing IoT-based data resources through dynamic contextual role modeling and activation," *Computing*, vol. 101, no. 7, pp. 743–772, 2019.

[40] A. Adler, M. J. Mayhew, J. Cleveland, M. Atighetchi, and R. Greenstadt, "Using machine learning for behavior-based access control: Scalable anomaly detection on TCP connections and HTTP requests," in *IEEE Military Communications Conference*. IEEE, 2013, pp. 1880–1887.

[41] A. Outchakoucht, E. Hamza, and J. P. Leroy, "Dynamic access control policy based on blockchain and machine learning for the Internet of Things," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 7, pp. 417–424, 2017.

[42] L. Argento, A. Margheri, F. Paci, V. Sassone, and N. Zannone, "Towards adaptive access control," in *32nd Annual IFIP Conference on Data and Applications Security and Privacy*, F. Kerschbaum and S. Paraboschi, Eds. Springer, 2018, pp. 99–109.

[43] H. Alavizadeh, J. Jang-Jaccard, S. Y. Enoch, H. Al-Sahaf, I. Welch, S. A. Camtepe, and D. D. Kim, "A survey on cyber situation-awareness systems: Framework, techniques, and insights," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1–37, 2022.

[44] D. Chen, T. Orekondy, and M. Fritz, "GS-WGAN: A gradient-sanitized approach for learning differentially private generators," *Advances in Neural Information Processing Systems*, vol. 33, pp. 12 673–12 684, 2020.

[45] N. N. Neto, S. Madnick, A. M. G. D. Paula, and N. M. Borges, "Developing a global data breach database and the challenges encountered," *ACM Journal of Data and Information Quality*, vol. 13, no. 1, pp. 1–33, 2021.

IX. Biographies

Dr. A. S. M. Kayes is a Senior Lecturer in Cybersecurity at La Trobe University, Australia. His research interests encompass various areas within cybersecurity, such as data security, access control, fog and cloud security, cyber incidents, and data/privacy breaches. Over the past decade, he has made significant contributions to the field, with more than 75 research articles published in international journals and conference proceedings. He has an h-index of 25 and over 2,400 citations (Google Scholar). He is a member of the Australian Computer Society and IEEE.

Prof. Wenny Rahayu is a Professor and Dean of the School of Computing, Engineering, and Mathematical Sciences at La Trobe University, Australia. Her research interests include data privacy, big data integration and management, and access control. Over the past 15 years, she has published 2 books and more than 260 research articles in international journals and conference proceedings. She has an h-index of 40 and over 8,000 citations (Google Scholar). She is a member of the Australian Computer Society and IEEE.

Prof. Tharam Dillon is an adjunct Professor at La Trobe University, Australia. He has published 8 authored books and more than 500 research papers in international journals and conference proceedings. His research works have been widely cited and therefore have considerable impact. He has an h-index of 67 and over 20,500 citations (Google Scholar), which puts him in the top percentile of researchers globally. He is currently a Life Fellow of the IEEE and a Fellow of the Australian Computer Society and the Institution of Engineers Australia.

Dr. Ahmad Salehi S. is currently a Lecturer in Cybersecurity at La Trobe University, Australia. Before joining La Trobe, he worked as a Research Fellow at RMIT. His research interests include access control, blockchain, cryptography, cybersecurity, and digital health. He has an h-index of 12 and over 500 citations (Google Scholar). He is a member of the Australian Computer Society and IEEE.

Dr. Hooman Alavizadeh is currently a Lecturer in Cybersecurity at La Trobe University, Australia. Before joining La Trobe, he was a Lecturer at the University of Sydney. His research interests include cloud/network security, security modelling, moving target defence, cryptography, and cyber situation-awareness. He has an h-index of 12 and over 600 citations (Google Scholar). He is a member of the Australian Computer Society and IEEE.