UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Zero Trust with Guaranteed Accuracy *Architecture* Implementation for *Intrusion Detection Systems (ZTA-IDS)*

by Abeer Alalmaie

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Dr Priyadarsi Nanda
and the co-supervision of Dr Jia Wenjing

University of Technology Sydney
Faculty of Engineering and IT

November 2023

Certificate of Authorship/Originality

I, Abeer Alalmaie declare that this thesis, is submitted in fulfilment of the requirements for the award of PhD in Computer System, in the school of Electrical and Data Engineering at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note: Signature removed prior to publication.

Date: 26-12-2023

ABSTRACT

As security monitoring advances and cloud computing grows popular, organizations increasingly outsource intrusion detection and monitoring to third-party analysts to save on costs like installation, maintenance, labor, and computational time, thereby enhancing efficiency and focus on services and products. However, due to the data security risks of allowing cloud-based third-party analysts access to network traces, the current ”trust but verify” approach in security monitoring is insufficient. Therefore, new mechanisms such as Zero Trust models, which demand a shift in perspective to “never trust, always verify”, must be built and implemented by network providers. The main challenge, however, is that outsourcing sensitive network traces to untrusted parties is inherently in contradiction with the policy of Zero Trust models. A great deal of effort has been devoted to address such security and privacy issues. Unfortunately, the majority of these sacrifice usability to provide better privacy guarantees, while others sacrifice privacy to maintain usability. A case in point is CryptoPAn, a prefix-preserving anonymization solution that preserves the utility for Internet Protocol (IP)-based intrusion detection analyses but is vulnerable to semantic attacks. Recently, a new notion called the multi-view approach has been proposed to preserve both the privacy and accuracy of the outsourced datasets targeting intrusion detection schemes. In this thesis, we apply multi-view approach, addressing the challenges including use of appropriate partitioning algorithm and interpretation of security rules in each IDS when examining anonymized views. It assesses the model's effectiveness against various intrusions and its resilience to different semantic attacks. Furthermore, we propose a new binary IDS, based on an autoencoder and a convolutional neural network, which outperforms other related works and achieves an accuracy of 92% using small amount of training data. Additionally, we extend binary IDS to a multiclass IDS and we take sequential dependencies into consideration using recurrent neural networks. However, experi-

ments reveal a decline in accuracy on real-world data due to significant domain shift between the training and real-world data domains. This may be due to variety of training data on real-world scenarios and sensitivity to input changes. However, after fine-tuning with a limited set of samples from the real-world domain, our model's accuracy improved significantly, aligning with unique characteristics of the collected data.

Dedication

To my family

Acknowledgements

To mark the completion of this thesis, I first offer my profound thanks to Allah, whose guidance has been a constant source of hope and strength, allowing me to navigate this journey with resilience and ultimately enjoy the fruits of success.

I am deeply thankful to my principal supervisor, Dr Priyadarsi Nanda, for his invaluable mentorship and unwavering support. I also appreciate the insights provided by my co-supervisor, Dr Jia Wenjing. Special recognition goes to my former principal supervisor, Professor Xiangjian (Sean) He, for his early guidance.

My heartfelt appreciation goes to my father, Professor Zaher Alalmaie, my mother, Mrs Halima Alalmaie, and my stepmother, Mrs Fatimah Alalmaie, for their love and wisdom. The encouragement from my siblings, relatives, and friends has been a pillar of support.

To my daughters, Hadeel, Yasmeen, Sawsan, and Savana, your love has been the greatest source of motivation throughout this endeavour.

Lastly, my thanks go to King Khalid University for the opportunity to complete my Ph.D., the Saudi Arabian Cultural Mission (SACM) in Australia for their support, and the University of Technology Sydney (UTS) for fostering a conducive learning environment that has greatly aided my academic pursuits.

Abeer Alalmaie

Sydney, Australia, 2023.

List of Publications

The following is a list of my research papers during my PhD study.

Journal Papers

J-1. **Abeer Alalmaie**, Priyadarsi Nanda and Xiangjian He, Zero Trust Network Intrusion Detection System (ZTA-IDS), (Submitted).

Conference Papers

C-1. **Abeer Alalmaie**, Priyadarsi Nanda, Xiangjian He and Mohrah Saad Alayan: Why Zero Trust Framework Adoption has Emerged During and After Covid-19 Pandemic. 1 Jan 2023 Lecture Notes in Networks and Systems 655 LNNS:181-192.

C-2. **Abeer Alalmaie**, Priyadarsi Nanda and Xiangjian He: Zero Trust Network Intrusion Detection System (NIDS) using Auto Encoder for Attention-based CNN-BiLSTM. 30 Jan 2023, ACM International Conference Proceeding Series1-9.

C-3. **Abeer Alalmaie**, Priyadarsi Nanda and Xiangjian He: ZT-NIDS: Zero Trust-Network Intrusion Detection System. 10 Jul 2023, International Conference on Security and Cryptography1:99-110SciTePress.

C-4. **Abeer Alalmaie**, Priyadarsi Nanda and Xiangjian He: Zero Trust-NIDS: Extended Multi-View Approach for Network Trace Anonymization and Auto-Encoder CNN for Network Intrusion Detection. 28 Oct 2023, IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)IEEE449-456IEEE.

C-5. **Abeer Alalmaie**, Priyadarsi Nanda and Xiangjian He, Zero Trust for Intrusion Detection System: A Systematic Literature Review, ICAART, 2024 (Accepted).

C-6. N Waheed, AU Rehman, A Nehra, M Farooq, N Tariq, MA Jan, F Khan, AZ Alalmaie and P Nanda: FedBlockHealth: A Synergistic Approach to Privacy and Security in IoT-Enabled Healthcare through Federated Learning and Blockchain. (2023).arXiv preprint arXiv: 2304.07668.

C-7. N Waheed, F Khan, S Mastorakis, MA Jan, AZ Alalmaie and P Nanda: 6. Privacy-Enhanced Living: A Local Differential Privacy Approach to Secure Smart Home Data. 2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS) IEEE, pp. 1-6.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Introduction

As monitoring network security becomes more challenging, there is a growing need for outsourcing such tasks to third-party analysts. In this respect, organizations are often unwilling to share their network traces because they do not trust analysts and fear the traces will reveal sensitive information, such as compromising network and system configuration, that could potentially be used for attacks. Networks are vulnerable to data security breaches and unauthorized access, because of their “Implicit Trust” or “Trust but Verify” characteristics. Currently, the cloud is the “New Network Edge” and it cannot be structured under this implicit and old trust model. Implicit trust networks are not capable of working in an atmosphere where network ‘edges’ have disintegrated and broken down. Particularly, such security breaches result in breaching confidential databases, with the number of breaches growing; both in severity and quantity. Hence, the current trend of security monitoring with a “Trust but Verify” method is not adequate in any scenario when third parties are involved.

This thesis explores the impact of Zero Trust models in relation to the accuracy of intrusion detection analysis, merging it with an anonymization method named multi-view and suggests the most appropriate model to preserve both accuracy and security against semantic attacks.

In this thesis, the terms Intrusion Detection System (IDS), Network Intrusion Detection (NID), and Network Intrusion Detection System (NIDS) refer to closely related concepts and are used interchangeably.

This chapter is organized as follows: Section 1.2 details the research background and the motivation for a case where a company outsources its data to a third-party analyst which might be untrusted and by making this decision, it places its trust in an external entity which may cause semantic attacks. To address these security challenges, various solutions have been proposed to tackle these concerns, one of which is Zero Trust. Section 1.3 discusses the need for a Zero Trust approach when outsourcing intrusion detection tasks.

Section 1.4 details the Zero Trust concept. Section 1.5 details the Zero Trust mechanism. Section 1.6 describes the Zero Trust components. Section 1.7 outlines the thesis objectives. Section 1.8 presents the research questions. Section 1.9 gives the scope of this thesis. Section 1.10 presents the research significance. Section 1.11 details the research methodology. Section 1.12 overviews the thesis organization and Section 1.13 concludes the chapter.

1.2 Background and motivation

Organizations are often unwilling to share their network traces because they do not trust analysts and fear the traces will reveal sensitive information which may compromise network and system configurations, and could potentially be used for attacks. Networks are vulnerable to data security breaches and unauthorized access, because of their "implicit trust" or "trust but verify" characteristics. Currently, the cloud is the new network edge [1] and it cannot be structured under this implicit and old trust model. Implicit trust networks are not capable of working in an environment where network edges have disintegrated and broken down. Particularly, such security breaches result in confidential databases being compromised, with the number of breaches growing both in severity and number. Hence, the current trend of security monitoring with a “trust but verify” method is not adequate in any scenario where third parties are involved.

The significant increase in mobile device usage, cloud computing, and the Internet of Things has removed old-fashioned network boundaries. Strengthened network perimeters alone are no longer adequate to ensure security in a world of increasingly sophisticated threats.

As security monitoring becomes more sophisticated, organizations are willing to outsource intrusion detection and monitoring tasks to third-party analysts. In addition to saving the costs associated with installation and maintenance, labor, and overheads, the reasons companies outsource data include improving efficiency, saving time, and focusing more on services and products (e.g., planning, management, etc.). On the other hand, when a company outsources data to a third party, it is relinquishing control and placing trust in third parties, which is inherently in contradiction to the policy of Zero Trust models.

Numerous solutions have been proposed to tackle the security issues which arise when the Zero Trust model is not implemented (e.g., encryption, poisoning, and perturbation). The overwhelming majority sacrifice utility (the accuracy of the outsourced data) to provide better privacy guarantees while others sacrifice privacy to preserve utility. For instance, CryptoPAn [1] is a prefix-preserving anonymization solution that preserves utility for IP-based intrusion detection analyses, but it is vulnerable to semantic attacks (e.g., injection, fingerprinting). Hence, there is a trade-off between the level of attack mitigation a security strategy can achieve and the level of utility which can be preserved on the original data. In particular, a Zero

Trust model is designed to resist different security threats and attacks. There is a trade-off between the level of attack mitigation a security strategy can achieve and the level of utility which can be preserved on the original data [1] . In particular, a Zero Trust model is designed to resist different security threats and attacks.

Currently, the use of Information and Communication Technology (ICT) is exponentially increasing for communication transactions, computation, storage, and exchange of information in government, medicine, e-commerce, agriculture, banking, etc. When different operations are performed on Internet-/network-connected platforms, it is necessary to implement different security standards to keep confidential information safe from unauthorized use, change, loss, or disclosure. Furthermore, if any security incident occurs, the system should take appropriate action to reduce the impact and prevent further damage to the organization. Compromised information has serious consequences for the operations of the organization. Stringent cyber security guidelines, regulations, risk management approaches, and technologies should be implemented to keep networks, infrastructure, programs, and data safe from cyber attacks in the cyber environment [2, 3].

Today, almost all private sectors and government organizations depend on technology more than ever before. They store a great amount of important and confidential data on computers and exchange it across networks while dealing with critical infrastructures. Organization's and user's digital assets (for instance cryptocurrency software and applications) comprise associated computing components, infrastructure, applications, local services, network systems, and broadcasted and/or collected information in the cyber ecosystem. To reduce the risk associated with the rising number of cyber threats and actual attacks, ensuring cyber resilience in crucial infrastructures is essential for national security [1]. Cybersecurity refers to any technology, measure or practice for preventing cyberattacks or mitigating their impact [3]. Thus, cyber security is a crucial element in private industry, banking sectors, and government organizations. Technology and cyber security are continually evolving to ensure a better user experience across the world. However, attackers are also advancing their techniques to exploit critical infrastructures and gain unauthorized access to systems and networks. Thus, it is important to prevent, detect, and respond to cyber attacks quickly so damage to critical infrastructures can be reduced. To mitigate risk, organizations have initiated various measures to improve cyber security awareness [4].

Traditionally, organizations have concentrated on perimeter defence and have given authorized access to network traffic once on the internal network. As a result, unauthorized lateral movement within the environment has been one of the most problematic issues for organizations' networks. Perimeter firewalls are less beneficial in recognizing and blocking attacks from inside the network and are not capable of protecting infrastructure outside the enterprise environment (e.g., cloud-based services, edge devices, remote workers, etc).

Network perimeter security alone can no longer provide enterprise security in a world of increasingly complex threats, so organizations should focus on improving their cyber security management systems. As discussed in [5], there has been a significant growth in the cyber-attacks and the degree of their risks targeting ICT infrastructure with the subsequent major propensities:

- Phishing attacks in the form of fraudulent emails are the most common type of cyber attack.
- Nowadays, crypto miners are used as an imperative monetization approach for cyber-criminals.
- State-supported defences increasingly focus on banks through attack routes.
- Cybersecurity experts are scarce in public organizations due to intense com-

petition, as the industry is actively recruiting these specialists.

- Technical training poses a barrier to fostering awareness regarding cybersecurity within security personnel and executive management ranks.
- Cyber threat intelligence should respond to progressively sophisticated threats using innovative solutions based on computerized implementation and services.
- The advancement in the IoT ecosystem is an issue due to the unavailability of lightweight security solutions for low-cost IoT components and services.
- The lack of cyber threat intelligence mechanisms for low-capacity organisations or bottom-level clients is an issue that should be tackled by suppliers and authorities.

The implementation of Information Systems and Technologies (ISTs) cannot necessarily generate the expected benefits without any revision of the related structures as well as the underpinning planned and administrative methods. Maturity models are used as consultants and reference structures for information system management in various types of companies. A cyber security maturity model is a standard to determine the maturity of a protection system and guidance on how to attain a higher level of security. Thus, several cyber security maturity mechanisms have been suggested to reduce the impact of cyber attacks and increase nation-level protection. Moreover, regional and international organisations have conducted other studies, but they aim to score and rank countries corresponding to their national-level cyber security top systems. The cyber security maturity model aims to find processes, methods, and techniques to enhance an organisation's protection and introduce a level-based advancement approach. Cyber security maturity models can be divided in the following domains [6]:

- Risk management: Risk is the possibility of an undesirable and unanticipated occurrence, such as data loss. Thus, companies should reduce this risk to keep their systems healthy.

- Security policy and plan management: This is the process of evaluating a company's systems to develop safety measures and to develop strategies around standards and guidelines.
- Human Resource Management: This is the process to evaluate the level of understanding, learning, culture, and conduct of employees to enhance security awareness within the company.
- Physical Security Management: This is the process of reviewing the organisation's physical safety by improving the physical security of critical infrastructures.
- IT Security Management: This is the process to manage information technology components to understand the requirements and performance of a company's various operations.
- Communication Security Management: This is the process of checking the organisation's communication security to preserve data security within and outside the organisation when transmitting information.
- Security Technology Management: This is the process to deal with the development, implementation, application, or use of technological devices.
- Security incident and its management: This relates to any unwanted/unexpected event which impacts the system and damages the company's resources and reputation. To overcome this, appropriate incident management policies must be in place to ensure the pre-emptive or responsive control of these events.

- Security Audit and Compliance Management: On a regular basis, the organisation should schedule audit procedures to identify the company's security objectives and policy compliance.

Updated IT security solutions are required to incorporate the following important features:

1. Segregating data, users, services, and devices, within a trusted framework, to make sure each access request is validated and either permitted or rejected.
2. Being resilient and resistant to attack without a large administrative burden.
3. Being capable of simply and quickly (if not automatically) adjusting to an ever-changing service environment even without a large administrative burden.

Organisations are being compelled to think about the traditional network security perimeter again and to consider the implementation of a Zero Trust policy which satisfies the aforementioned features by treating all users, data, services, and device requests the same. Zero Trust represents a shift from old-fashioned security policies where every asset in an organisation is open and accessible to a security policy where the default is that no one from either inside or outside the network is trusted and verification is required from any party attempting to gain access to the network.

Most current corporate networks are flat i.e., there is no or little separation of user and data networks. The weakness of the traditional hub-and-spoke network model lies in its architecture. Crossing the chasm from trust to distrust via a firewall is inherently risky. Instead, Zero Trust no longer distinguishes between "inside" and "outside" the network perimeter.

A Zero Trust Architecture (ZTA) increases security by centralising the defence of resources rather than the network environment, because location is no longer considered the fundamental component of the security posture.

Zero Trust can be defined as a set of cyber security principles applied to generate a strategy which concentrates on transferring network protections from static, wide network perimeters to focus more narrowly on systems, users, and individual or small groups of resources [7]. The concept of trust is removed from within the network and interfaces, users, packets, or applications are no longer trusted. Across industries, security experts are shifting the security diameter to a zero security trust state of mind and quickly implementing and adopting the Zero Trust security network model. Zero Trust is more than a concept, it is a robust security model that follows 7 security principles – data, workload, devices, orchestration and automation, users, network, analytics and visibility.

1.3 The need for Zero Trust for outsourcing an IDS

When outsourcing network traces to a third-party IDS, the accuracy of network monitoring is extremely important for data owners. Hence, traditional approaches may lack the capability of providing a set of benefits when dealing with network trace security monitoring.

The Zero Trust notion requires rigorous identity-based verification for each device and user attempting to access resources on either a private network or the cloud, without considering whether they are inside or outside of the network perimeter. No single method is related to Zero Trust owing to the model's complexity and comprehensiveness; it is a holistic technique to network security which incorporates many regulations and methods and incorporates complementary technologies, protocols, and methodologies to ensure security is not perimeter-based but is ingrained at every level of the network and data access [8].

This comprehensive approach is particularly essential in addressing the multi-faceted nature of security threads, including the often-overlooked insider threat. These threats can range from mishandling sensitive information unintentionally to deliberate actions such as data theft and sabotage. ZTA was introduced to protect digital environments by preventing lateral movement, simplifying granular user-access control, leveraging network segmentation, and providing continuous real-time monitoring to detect and take action against anomalies and potential threats instantly [2].

1.4 Zero Trust concept

Zero Trust is a cyber security paradigm with the aim of resource protection and the premise that trust is not granted implicitly but must be continually assessed. Zero Trust implies never trusting any entity outside/inside of the perimeter of the network. It concentrates on removing trust within an organisation and holds that no implicit assumptions should be made about the credibility of devices, users, applications, or data which has been accessed or is being accessed on an organisation's network [8]. It provides the visibility and IT monitoring required to monitor, secure, and manage every application, user, device, and network related to or being applied by either the organisation or its staff to access data. Zero Trust scrutinises all outgoing or incoming traffic. What distinguishes this model from other existing security methods is the fact that even internal traffic, which is the traffic that does not cross the organisation's perimeter, is also considered as a possible danger.

1.5 Zero Trust mechanism

Zero Trust provides scalable protection schemes throughout numerous associations. In [2], the researchers stated that maintaining trust in the cloud and networks is too difficult and suggested that it is better to eliminate the idea of trust. Further, the authors proposed a Zero Trust mechanism to improve the protection structures and technologies. Traditional security systems assume that individuals inside the organisation's network can be trusted. In contrast, Zero Trust inverts the model, directing IT teams in accordance with the guiding principle of “never trust, always verify” and re-defining the perimeter to consider data and users inside the network. In the trusted model, the assumption is that a user’s identity is not compromised, so each user will act responsibly and can be trusted. The Zero Trust approach considers trust to be a vulnerability because it is possible that network users may be malicious insiders and threat actors and hence, they should not be allowed to move freely inside the network and access or exfiltrate data without the proper verification and authorisation.

In this scenario, every element, ranging from packets to clients, is considered as untrustworthy, regardless of its type. Zero Trust redefines the method of resource separation; a fundamental theory where resources must remain safe, are categorised altogether, and separated carefully or kept unconnected from illegal contact at any type of area [3]. Zero Trust restricts access to sensitive applications, devices, and data on a need-to-know basis. Zero Trust mechanisms also present the occasion for micro-segment systems, allowing groups to change their demands without reforming their whole network. In network micro-segmentation, networks are divided into small granular nodes all the way down to a single application or machine. Security protocols and service delivery models are introduced for each unique segment. A free data flow is considered among the cornerstones of the Internet requirements and is limited to protect customers from privacy violations, networks from penetration, and organisations from attacks on operations and infrastructure.

ZTA is an end-to-end methodology for data and network security that encompasses credentials, operations, identity, access, hosting environments, endpoints, and the interconnecting infrastructure [4]. Zero Trust is seen as an architectural technology that focuses on data protection. The base focus must be on limiting resource access to those with a “need to know.” Previously, agencies, and more generally enterprise networks, have concentrated on perimeter defence, and authorised users are given wide access to all resources. Thus, one of the demanding challenges for organisations has been to unauthorised lateral movement within a network. In this regard, the Trusted Internet Connections (TIC) and agency perimeter firewalls prepare powerful Internet gateways. This helps to block attackers on the Internet; however, the TICs and perimeter firewalls are not as effective in identifying and blocking attacks from inside the network. Thus, the Zero Trust separation access framework can be thought of as the next-generation mechanism of a firewall, expanding the micro subdivision of the networks to achieve flexibility and scalability with the advantage of virtualisation [4].

1.6 Zero Trust components

The American Council for Technology-Industry Advisory Council (ACT-IAC) describes the following six pillars for the Zero Trust model [8].

1. Users: Continuous authentication is essential for trusted users using one-factor, two-factor, and three-factor to ensure the users' dependability.

2. Devices: If a trusted user tries to gain entry through an unconfirmed method, then it may not be reliable, and it creates a security risk. Thus, all devices should be verified before connecting to the system for any kind of process.

3. Network: The ability to segment, isolate, and control the network is a pivotal point of security and is essential for a Zero Trust Network. Zero Trust Networks move perimeters in from the network edge and segment and isolate critical data from other data to strengthen protections and controls. This is important to control privileged network access, manage internal and external data flows and prevent lateral movement in the network.

4. Applications: The Zero Trust model enables the application layer to be protected. Zero Trust is recognised for its precision in access control, attributed

to its capability to detect and control specific security techniques. Multi-factor authentication is a vital element for granting appropriate access privileges to applications.

5. Automation: Security automation response tools are an important component of the Zero Trust model by reducing manual effort and reducing costs. Automated tools can be used for disparate security system management and end-user oversight and interaction, facilitating efficient responses and fostering enhanced communication with end-users.

6. Analytics: The Zero Trust model implements tools such as security analytics platforms, security user behaviour analytics, and other analytics systems to enable security experts to observe real-time processes, understand the existing security risks and the probability of future incidents and to devise proactive security procedures prior to the occurrence of a security breach.

It is imperative that the Zero Trust model is improved to ensure protection against existing and new security threats while dealing with various operations in the organisation.

This thesis aims to address the wide range of current security challenges to ensure an effective level of security for organisations. Our research objectives and research questions are detailed in the following section.

1.7 Thesis objectives

This thesis aims to preserve the accuracy and the security of network traces when outsourcing intrusion detection tasks to untrusted third-party analysts and proposes a model named the Zero Trust Architecture-Intrusion Detection System (ZTA-IDS).

The benefits of our proposed method include improving indistinguishability and being able to send any data to the analyser to be updated online.

The objectives of this thesis are as follows:

Objective 1: To identify the main challenges and the importance of Zero Trust when outsourcing intrusion detection tasks to an outsider analyst. Outsourcing intrusion detection task can be a threat to the privacy of a user's data as the third party has full access to the data. These threats and challenges need to be identified and analysed in order to arrive at an effective solution. This thesis provides a literature review in Chapter 2 to better understand IDS requirements and limitations and the importance of the Zero Trust model is discussed in Chapter 7.

Objective 2: To obtain a deep understanding of the accuracy-related issues prior to outsourcing intrusion detection tasks. In addition to ensuring security, the accuracy of the IDS should be preserved. To ensure a high level of accuracy, a deep understanding of the requirements of IDS is required, as explained in Chapters 3 to 5.

Objective 3: To obtain a deep understanding of the security threats and the existing Zero Trust technical solutions prior to outsourcing intrusion detection tasks.

The existing cyber security models are not able to offer a satisfactory security level for new threats. The security threats related to outsourcing intrusion detection tasks include semantic attacks, which need to be understood so they can be tackled. This is explained in Chapters 3 to 5.

Objective 4: To define the security criteria to select and evaluate an appropriate Zero Trust solution in our experiments. To achieve the objectives related to the issues of privacy and accuracy, the criteria for selecting an appropriate Zero Trust solution are defined. Chapter 5 describes the experiments that were conducted to evaluate each criterion and its applicability as a Zero Trust solution.

Objective 5: To devise a formal meaning of accuracy when selecting and evaluating an appropriate protocol for the experiments. We achieve this objective by evaluating the accuracy and the security of the proposed scheme as explained in detail in Chapter 5.

1.8 Research questions

Since the intrusion detection task is becoming more sophisticated, our research theme in this thesis is centered around the following:

What is the most effective way for organizations to outsource intrusion detection tasks to third-party analysts without violating the security and privacy of the outsourced data?

The contributions of various private sectors and government institutions are mostly relying on the technological and critical infrastructures in the fast-progressing world. Earlier solutions are based on various factors, such as security instances awareness programs, cyber attack detection, rescue, and protection of infected infrastructures. However, we identify that the existing cyber security models are unable to offer a satisfactory security level for new threats, as technology is regularly updated and upgraded on a global level. Therefore, it is essential for organizations to tackle such questions and provide necessary solutions to threat detection and privacy protection, and to act quickly to reduce damage to the infrastructure and other related systems. To address this issue, we present five main research questions to be answered in this thesis.

Research Question 1 (RQ1): What security issues arise by outsourcing intrusion detection tasks to third-party analysts?

Research Question 2 (RQ2): What level of accuracy can be achieved by outsourcing intrusion detection tasks to third-party analysts.

Research Question 3 (RQ3): What existing Zero Trust solutions achieve satisfactory results (preserving both security and accuracy)?

Research Question 4 (RQ4): How can better security be achieved using our proposed scheme?

Research Question 5 (RQ5): How can better accuracy be achieved using our proposed scheme?

The overall objective of this thesis is to identify the challenges prior to outsourcing intrusion detection tasks to third-party servers, provide effective solutions to mitigate these issues, and finally to evaluate the effectiveness of the solutions in addressing the main challenges. Hence, to answer the aforementioned research questions, this thesis achieves the research objectives to identify the challenges and to address them in context.

1.9 Scope of this thesis

This thesis develops the ZTA-IDS model for outsourcing intrusion detection tasks. The scope of the thesis is as follows:

1. Extended Multi-View: The methodology utilizes a multi-view to hide a prefix-preserving anonymised version of an attribute. The anonymised version of attribute is called real-view which is hidden through some fake views. It is able to be applied to IPs and other important security attributes.
2. IDS: The methodology proposes an IDS model which can accommodate additional fields in the input and their type. It utilizes an encoder to accept any input attribute.
3. Evaluating Accuracy and Security: The methodology needs to be evaluated in terms of accuracy and security. In this phase, the robustness of the method

against semantic attacks is evaluated.

The scope of this thesis also includes combining two proposed methodologies to construct the ZTA-IDS model.

1.10 Research significance

The contributions of this thesis are as follows.

1. We propose a customized accurate and secure Zero Trust implementation for outsourcing network traces for the ZTA-IDS. To the best of our knowledge and in light of the results of the literature review (Chapter 2), we believe this is the first known solution that thoroughly studies the impact of Zero Trust models in relation to the accuracy of intrusion detection analysis, merging it with an anonymization method named multi-view and suggests the most appropriate model to preserve both accuracy and security against semantic attacks (e.g., injection and fingerprinting).

2. We carefully customize the multi-view approach for the IDS scenario in relation to several intrusion detection tasks (such as DoS, analysis, etc.). In particular, the partitioning algorithm and the number of views required are specified according to the intrusion detection task.

3. We analyze the impact of possible solutions and attacks, and then experimentally assess the proposed solution by applying real network traces from a major ISP. The experiments prove the proposed method is robust against semantic attacks and its computational cost is reasonable.

1.11 Research methodology

This section overviews the methodologies used to meet the research objectives outlined in section 1.7. To handle the problem of outsourcing intrusion detection tasks to third-party analysts without violating security and privacy, our proposed ZTA-IDS method consists of three phases:

Phase 1- Extended Multi-View: The methodology utilizes a modified multi-view method which preserves the anonymised attribute as the real-view and hide it through some fake views which are like the real one in nature. It can be applied to IPs and other important security attributes.

Phase 2- IDS: The methodology proposes an IDS model which can accommodate additional fields in the input and their type. It utilizes an autoencoder to enable it to accept any input attribute.

Phase 3- Evaluating Accuracy and Security: The methodology needs to be evaluated in terms of accuracy and security. In this phase, the robustness of the method against semantic attacks is evaluated.

These phases are explained in detail in Chapter 3 and the experiments are reported and discussed in Chapters 4 and 6.

1.12 Thesis organization

This thesis provides a methodology to preserve the accuracy and the security of network traces when outsourcing intrusion detection tasks to untrusted third-party analysts, namely the proposed model named ZTA-IDS. This thesis is divided into seven chapters that are organized as follows:

Chapter 2 presents an overview of the existing literature on using multi-view technology in the field of network tracing, homomorphic encryption, and analysis with aggregated results, which corresponds to RQ1.

Chapter 3 presents the technical aspects of Zero Trust, multi-view, and the IDS model, which corresponds to research questions RQ2 and RQ3.

Chapter 4 details the multi-view methodology and the IDS model, which corresponds to research questions RQ2 and RQ3.

Chapter 5 provides the model validation, which corresponds to RQ4 and RQ5.

Chapter 6 details the experiment procedure, which corresponds to RQ2 and RQ3.

Chapter 7 discusses Zero Trust and its importance, which corresponds to RQ1.

Chapter 8 provides a brief conclusion and offers suggestions for possible future works.

1.13 Conclusion

In this thesis, we introduce a new Zero Trust model utilizing multi-view and IDS named ZTA-IDS. The proposed model effectively protects various systems, crucial infrastructure, networks, data, services, and end-users from critical security risks, meeting various security requirements. The proposed IDS model is able to accept any attribute. It preserves both accuracy and security and doesn't trust anyone in the network. In this chapter, we highlighted the Zero Trust philosophy and its requirements, necessities, components, and significance. Then, we explained the research methodology followed by a research plan for the thesis. Chapter 2 presents the literature review to explore the relevant literature in the field of network security.

CHAPTER 2

Literature Review

2.1 Introduction

This chapter describes a systematic literature review (SLR) which was conducted to comprehensively examine the relevant literature in the field of network security. The purpose of the SLR is to identify and analyze existing studies, articles, and publications related to Zero Trust models. This review aims to identify intelligent solutions presented in the literature which allow internet service providers (ISP) in organizations to outsource intrusion detection system (IDS) network monitoring tasks while reducing or avoiding the risks of semantic attack by untrusted analysts. Several solutions including poisoning, encryption and perturbation, have been proposed. However, these solutions tend to reduce the accuracy of data while attempting to preserve privacy. Zero Trust models employ the philosophy of “never trust, always verify” to reduce the risk to data and the network when IDS network monitoring is outsourced, which is the focus of this review.

This chapter is organized as follows: Section 2.2 covers the literature on the shift in philosophy from traditional security to modern security. Section 2.3 details the SLR conducted to explore the research issues related to Zero Trust. Section 2.4 outlines open research gaps. Section 2.5 highlights the limitation of conducting this SLR. Section 2.6 concludes this chapter.

2.2 The shift from traditional security models to modern security models

2.2.1 Overview

Traditionally, enterprise networks have relied on “implicit trust” or “trust but verify” approaches for traffic control and security which prevent external intrusions using mechanisms such as firewalls, virtual private networks (VPNs), and network access controls however, they regard internal traffic as secure [9]. These approaches are no longer sustainable as organizations are increasingly moving to the cloud, where corporate resources may be accessed by third parties such as customers, analysts, outsourcing partners, and others. With the cloud becoming the new normal, new trust models are needed to ensure a high level of corporate cyber security. Zero Trust is a security framework that does not recognize network edges in a traditional sense and does not consider any network areas as trustworthy [2]. It requires continuous verification of access for all network users at all times. The key advantage of Zero Trust is that it has a highly adaptable infrastructure that can be combined with the cloud for organizational security improvements [3].

Zero Trust is especially effective in cases where an ISP outsources an IDS to third-party analysts. Analysts use network tracing to collect information concerning user behavior, packet flows, and other security aspects. However, organizations are concerned about sharing such data due to privacy worries. A number of anonymization approaches have been proposed in the traditional trust model, but none offer acceptable levels of privacy and data sharing.

To address the privacy concerns associated with sharing sensitive network data in the context of outsourcing IDS to third-party analysts, machine learning and metaheuristic algorithms play a crucial role. These are employed to develop advanced data anonymization techniques that not only protect individual identities but also preserve the integrity and utility of the data. These methods can automatically detect and obfuscate sensitive information while still allowing for the meaningful analysis of network behavior and security patterns [10].

The following sections cover the history of Zero Trust and its emergence as follows: defining Zero Trust, US Government cyber security, standardization challenges in cyber security, how the need for Zero Trust Architecture (ZTA) emerged, yesterday's businesses, today's cyber security infrastructure systems, and future cyber security infrastructure models.

2.2.2 Defining Zero Trust

The Zero Trust approach was introduced by John Kindervag of Forrester Research in 2010 [2]. He realized that making the cloud and networks more reliable was too monumental a task and instead suggested that the best point of action would be eliminating the idea of trust itself. Trust, Kindervag argues, is binary, that is, it is either ON or OFF, therefore it is very different from the real world where interactions are not as binary [2].

While legacy network infrastructure only required the management and visibility of north-south architecture, today's business needs require the management of traffic from east to west, which opens up a number of challenges and risks because there is no visibility present in this area. Moreover, as a result of this, Griffy-Brown et al. posit that threats can remain undetected for as long as 229 days in this east-west mechanism. Plus, they argue, to compound the problem, it is estimated that east-west traffic is only increasing, having risen by 80 percent in 2016 alone [11].

Embracing a Zero Trust model means rethinking our cyber security philosophy. Puthal et al. [3] posit that we currently implement cyber security with a "trust but verify" approach. Embracing Zero Trust means to shift this view to "never trust, always verify". In a Zero Trust approach, none of the pillars, such as users, applications, devices, and packets, are trusted without considering the entity type, even if it is part of the network [3]. Zero Trust, therefore, is an opportunity to provide scalable security infrastructure across a number of different kinds of organizations.

Puthal et al. [3] posit that, unlike other principles such as “deny by default”, “least privileges” or role-based access control, Zero Trust completely redefines the method to resource segmentation, a key element where resources that must be maintained protected, are grouped together and isolated safely or kept separate to keep unauthorized access at bay. According to [4], Zero Trust models also provide the opportunity to micro-segment networks, enabling organizations to adapt their requirements without restructuring their entire network.

Assunção [4] points out that by redefining the network, Zero Trust models create new opportunities for the segmentation gateway. Assuncao’s approach focuses on all the resources of the modern network, from content to filtering, access control, cryptography, firewalls, and package forwarding. The Zero Trust segmentation gateway method is viewed as the next-generation technology of a firewall, enhancing the micro-segmentation of the networks while suggesting scalability, versatility, and providing the benefit of virtualization-friendliness. There is no perfect model available, nevertheless, it is likely that the impacts of attack will be reduced using Zero Trust methods.

Puthal et al. [3] stated that there are three tenants of Zero Trust:

1. Each resource should be accessed regardless of location
2. Access control is on a need-to-know basis and is strongly enforced
3. Organizations must inspect and log all traffic to make sure that users are doing the right thing.

   Furthermore, the authors also list the following three core security technologies bringing Zero Trust into the SME market:

4. Identity and Access Management.
5. Cloud Security Platform (CASB).
6. End-point detection and response.

2.2.3 US Government Cyber Security

In 2016, Obama proposed the US Cyber Security National Action Plan encouraging the private sector to share details of security events with one another and with the US government. Offering several strategies to protect the US against cyber threats, the plan concentrated on various problems such as informing the public about the increased threat of cybercrime, improving cyber security protection, and protecting consumers' personal information. A presidential commission was formed to enhance national cyber security. The IT plans of the government have also changed according to this information. In addition, it helps Americans secure their online accounts, assisting in keeping their credentials and personal information safe. The use of multifactor authentication was encouraged, which is one of the major aspects of the Zero Trust model. This official government policy indicates a move towards less explicit trust for users, highlighting the need for a Zero Trust network infrastructure philosophy. Srinivas et al. [12] proposed a range of security measures that they believe can help mitigate the severity of cyber attacks. By incorporating this design into the security infrastructure, they argue attacks can be prevented. These include the following:

1. Firewalls

   There are three common types: 1) packet filters 2) application-level gateways 3) circuit-level gateways. When designing a firewall, it is essential that all the traffic coming in and out of the network goes through the firewall. Additionally, only authorized traffic, as defined by the local security policy should be able to pass through the firewall. Understanding that firewalls can be penetrated is crucial. Zero Trust, therefore, offers a promising solution as it breaks down the notion of perimeters and microsegments the network.

2. Antivirus Software

   There are four generations of antivirus software intrusion detection and prevention systems. Intruder detection merely 'detects' when there is an attacker, however a prevention system actually protects against the attacker. Hence, it is essential to detect as well as prevent, meaning both systems are very important.

3. Encryption

   Data stored in systems must be encrypted and locked with an encryption key, either using a symmetric key encryption or public key encryption mechanism login credentials. High entropy passwords are crucial in authentication, as guessing attacks are more effective than many believe.

4. Awareness

   Educating users about viruses is important, as well as the need for appropriate authentication and effective virus software.

5. Operating System Update

   Built-in software updates should fix bugs and enhance security. Srinivas et al. [12] discussed the architecture of cyber security incidence frameworks and their objectives which include incident management to avoid incidents before they occur, reducing threats and dangers if a cyber security incident occurs, as well as improving the coordination and management of these incidents. Additionally, impact should be minimized to achieve confidentiality, integrity and the availability of the organization's services, as well as protecting its assets and critical operational data which is reported back to executive management. Another major objective involves keeping cyber security costs to a minimum. The Cyber Security Incident Management Framework sets out the roles and responsibilities of individuals in public and private sectors, enabling any organization to effectively take part in a coordinated national cyber incident response.

Confidentially is designed to keep sensitive information private and secure from unauthorized access, disclosure, alteration, or destruction. The goal of confidentiality is to protect an organization's assets and critical operational data using these techniques: Access Control, Encryption, Data Masking and Redaction, Secure Communication Protocols, Regular Security Audits and Monitoring, and Employee Training and Awareness.

2.2.4 Standardization challenges in cyber security

Srinivas et al. [12] identified a number of problems surrounding the standardization of any cyber security framework. Standardization can protect users in a cyber-environment by setting a defined cyber security technique. Standards may include users, networks, devices, software processes, applications, and information as it connects to networks. The standard minimizes risks, preventing and alleviating cyber-attacks. Some of the major problems that arise in attempting to create standardized procedures are as follows:

- Organizational challenges: Standard Development Organizations mainly initiated by industries, often become over-standardized. Only some standards aim to comply with privacy and data protection.

- Lack of agility: Defining and agreeing on standards can take years. Often these become outdated before they can even be used.
- Competing set of standards: There are multiple groups of standards in IT and cyber security which are currently in competition, and it can be difficult to determine the best standards from so many options.
- Economic considerations: Companies in a superior position using their own proprietary standards often fail to implement and support the standards for their products.
- Lack of awareness: Much of the time, people simply are not aware that standards exist. Therefore, awareness programs should be enforced.

2.2.5 How the Need for a ZTA Emerged

This section highlights how research on security solutions has changed, indicating how the need for ZTA has emerged over time as companies design new processes and layers to protect data. It highlights how traditional security models were designed using network perimeters such as firewalls, IDS, VPNs, web gateways and other network security devices to detect untrusted users.

More recently, attackers have begun to use a different type of method to breach network security known as compromised credentials. Therefore, cyber security strategies are moving from “trust but verify” to “never trust, always verify” (Zero Trust). The future highlights of the thesis are related to how the growing use of cloud-based applications and mobile devices have created more opportunity to compromise systems. Additionally, the move towards cloud computing and the Internet of Things (IoT) means old network perimeters are no longer effective.

2.2.6 Yesterday's Businesses

In this sub-section, a brief history of business network security is given.

In the late 1980s and early 1990s, the Internet became a useful tool for companies to manage and store their information. Before this, the Internet was restricted to governments and academic research. The history of corporate cyber security, therefore, has evolved over a period of less than 40 years. Early on, businesses realized the importance of cyber security very quickly, as the number of attacks grew and threatened networks, servers, and private information. Convicted hacker Kevin Mitnick shocked the world when he hacked the computer systems of various companies in the late eighties and early nineties, causing the loss of millions of dollars' worth of code and intellectual property from numerous companies. This proved once and for all to the world that cyber security was a very worthy investment.

Yesterday's Internet Architecture and Vulnerabilities

Traditional Internet protocols were designed without any attention to the protection issues and a decade ago, the Internet was far less secure than it is today [24]. Security protocols weren't implemented in older TCP/IP communications stacks, leaving the Internet highly vulnerable to IT attacks. Internet protocol version 4 (IPv4) formed the main architecture of the legacy Internet. Its security mechanisms were naturally flawed. Security systems relied on IPsec as a specific mechanism for securing packet payloads through cryptography. Nonetheless, it was open to skilled hackers capable of breaking through the encryption to obtain access, a fundamental design flaw. Security mechanisms were introduced to each layer of the Internet Protocol Suite, allowing data packets to be protected when being sent across networks. Cyber security philosophies quickly emerged, and a focus on four key attributes became the staple: confidentiality, integrity, privacy, and availability. The various attack methods and solutions in relation to these attributes are shown in Table 2.1.

Table 2.1 : Attack Techniques and Security Technology[13]

Security Attributes	Attack Techniques	Security Technology
Confidentiality	DoS, Phishing, Eavesdropping, IP Spoofing, and Hacking	IDS, Cryptographic Systems, IPSec and SSL, Cryptographic Systems, Firewall
Integrity	Viruses, Worms, Trojans, Eavesdropping, IP Spoofing, and DoS	IDS, Firewall, Anti-Malware Software, IPSec and SSL
Privacy	Email bombing, DoS, Spamming, Cookies, and Hacking	Firewall, IPSec and SSL, IDS, Anti-Malware Software
Availability	Email bombing, DoS, Spamming and Systems Boot Record Infectors	Anti-Malware Software IDS, Firewall

Intranet Architecture

Before many companies migrated their systems to the cloud, corporations tended to avoid using the public internet, favoring the use of private intranets that could control and protect the network traffic. Daya [14], P.4 stated: “Intranets are connected to the Internet, however, protected from it at the same time”. Intranets were set up as private networks which utilized the internet protocols and were therefore only accessible to a small number of users on the network. This differed from extranets which could be accessed by a number of different parties, from suppliers to customers. The bridge between a company’s intranet and the external internet was controlled through a gateway, most likely, one protected by a firewall and various other features like encryption, VPNs and user authentication. Under this legacy model, no access was required from the organization's internal network to the internet itself. VPNs offered intranets existing across multiple network locations with connectivity to the internet, a private network connected to the public network across numerous remote sites.

Daya stated that the intranet architecture model had several benefits; it could be set up very quickly and enabled effective data-sharing. But it was not without flaws. Without ironclad security, it could be compromised. Too much security also meant that data was overly restricted, and members of the company lacked access to important information. As a result, many companies began opting for open networks with a number of safeguards in place like firewalls, encryption, virus checking, and rules for employees to follow regarding email attachments and synchronized password and security certification systems.

Historical Data Centre Traffic Mechanism

Griffy-Brown et al. [11] state that in a legacy environment, systems were far easier to control due to their siloed nature. It was simple to put up a firewall and then open to the internet as the business grew. The IT team was able to see all the incoming traffic flowing from the internet into their domain. As their study argues: "this means north–south traffic was transparent, and the business issue was: how do we move quicker to connect with more strategic opportunities?" This historical data center mechanism is illustrated in Figure 2.1.

Hardware Security Devices

Traditional security models attempted to include a number of hardware devices in the cyber security mix as well as software. Some of these included biometric security

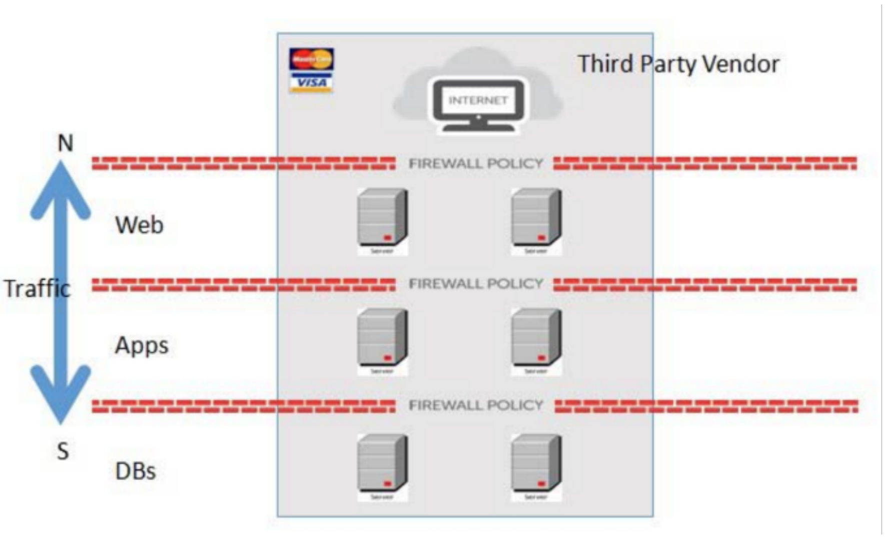


Figure 2.1 : Historical Architecture Data [11]

devices such as finger-print reading mice or keypads and public key infrastructure-based authentication tokens. The keypad or keycard and pin system was also widely implemented.

Legacy Internet Attack Methods

There are seven key attack methods which can be used to compromise legacy security models [14]:

a) Eavesdropping

Communications are intercepted by an attacker. The eavesdropper can covertly 'listen' to the network's messages (passive), or they can distort and compromise the information (active).

b) Viruses

Much like a disease, a virus infiltrates a file and multiplies, spreading until the entire system is infected.

c) Worms

Like a virus, the worm infiltrates a system and multiplies, however it does not require a file. It is commonly spread through email and can be network-aware (selective of targets).

d) Trojans

Trojans are malicious software programs masquerading as harmless programs that infect the user through a virus.

e) Phishing

A phishing attack is one where information is obtained by tricking users into providing personal data and sensitive information.

f) IP Spoofing Attacks

An IP spoofing attack is where the attacker changes the source IP address to impersonate a different computer system, disguise the sender's identity or both. Denial of Service

g) A denial-of-service attack

A denial-of-service attack causes the system to receive too many requests, rendering it unable to return communications. The system struggles to respond to requests, and it is left without service.

Legacy Internet Security Technologies

Legacy internet security systems comprised five core technologies [14]:

a) Cryptology Systems

The use of codes to turn data information into unrecognizable data

b) Firewalls

A border control mechanism, securing the perimeters of a private network from attack. Much like the moat of a castle, it acted as a front line against intruders in both hardware and software.

c) IDSs

An additional layer of protection after the firewall which is able to detect when an attack is occurring, enabling security professionals to understand how they're being attacked and to plan accordingly.

d) Anti-Malware Software and Scanners

Programs able to scan for, detect and cure attempts at attack methods like viruses, trojans and worms.

e) Secure Socket Layer (SSL)

A suite of protocols used to create a secure path between a browser and site, enabling a user's privacy to remain secure as they browse a website. Users are validated through certificates proving their identity.

The Traditional Data Security Model

Previously, cyber security was perimeter-based, in what is recognized as a "trust but verify" or "implicit trust model", where all communication between devices was trusted within a given security group. This, according to [15], assumed that networks were segmented, and that data center architecture could create a boundary, or "demilitarized zone" between trusted and untrusted sections of networks. However, in today's modern cloud-computing world, this static approach has become obsolete: physical or virtual perimeters have been broken down, blurred, and erased, enabling cyber attackers to invade these outmoded perimeters.

2.2.7 Today's Cyber Security Infrastructure Systems

Today, the cloud is the new network edge [15] and it cannot be defended under the old implicit trust model. The old system simply does not work in a climate where network edges have broken down and disintegrated. As a result, we continue to see a growth in attacks around the world, both in number and severity. In 2018, Chinese hackers were able to attack the Australian National University's internal records network, leading to students' university's information and student records being sent overseas (Channel 9 News, 2019).

In today's networks, an estimated 80 percent of data security breaches involve stolen credentials, according to the Verizon Data Breach Investigations Report (2017). Additionally, companies are ten times more likely to experience a breach caused by identity-theft tactics or by compromised credentials than any other vulnerabilities. Many IT departments appear to be stretched to their limits, with the current manually intensive process involved in investigating cyber alerts. According to Eidle et al. [16], as of 2017, this was taking several minutes to hours or longer. Failure to detect attacks, they also argue, is causing immense harm to the productivity of organizations, with companies reporting losses of up to 30 per cent in revenue, and a 4 per cent loss of business opportunities. This is a convincing argument for the implementation of cyber security models like Zero Trust to improve efficiency and response times to immediate threats.

Srinivas et al. [12] discussed the growing need for additional government regulations to help reduce the risk of cyber security threats to several critical government services across the world. They indicate the prevalence of a "growing threat from various cyberattacks with probable implications for consumer confidence, public protection as well as economic growth" (178). The article offers a few suggestions to help improve global cyber security risks and incident management frameworks, while discussing the standardization challenges in cyber security. Ultimately, it highlights that there is a current gap in cyber protection, which may be addressed by Zero Trust.

Agile Data Centre Traffic Mechanism

As companies respond to several technological opportunities that help them to expand, having an agile data center has become crucial. This architecture is increasingly virtualized, encompassing several third parties and business partners. As Griffy-Brown et al. [11],P.90 argue that "the security problem has become the lack of traffic visibility across infrastructure, particularly where "companies have a legacy environment co-mingled with a virtualized environment". The old notions of the perimeter have disappeared, giving rise to a host of problems. Not only does north-south traffic need to be managed, as with the legacy architecture, east-west traffic also must be managed. As discussed in [11], this problem is monumental because companies have no visibility of what is occurring when traffic travels east to west (2016) as shown in Figure 2.2.

Google's Beyond Corp Security Model

Tao et al. [17]highlight how Google is already using ZTA to allow its employees to work remotely from outside of a VPN whilst still guaranteeing the security of privileged information. Beyond Corp shifts access control from the perimeter to individual users and devices with dynamic access control and behavior perception strategies. In doing this, it enables each user access via its core identity service, which provides appropriate permissions for accessing resources only once authentication has been passed and the devices have been proven to have integrity. Therefore, employees can work securely from any location, without the need for traditional VPNs. According to [4], Beyond Corp offers the following benefits:

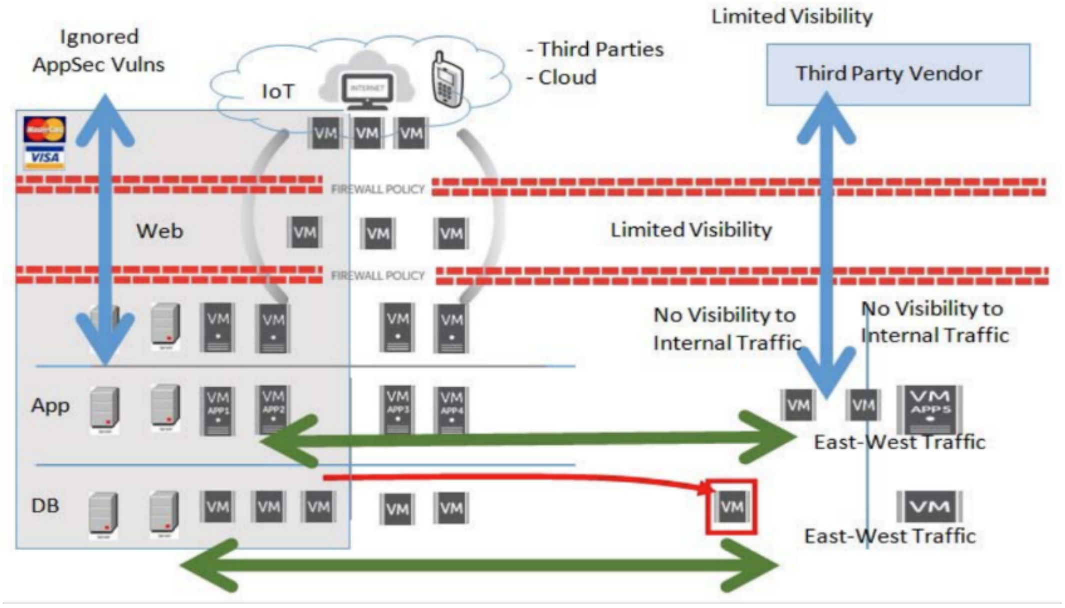


Figure 2.2 : Agile Data Center Mechanism [11]

1. It keeps devices up to date.

2. It maintains an inventory of employee devices.

3. It ensures all endpoints are monitored and all traffic is logged.

4. communicates only over fully encrypted channels.

5. It incorporates multifactor authentication.

6. It eliminates static credentials.

Difficulties Caused by Beyond Corp

Some vendors require network access inside the enterprise to support the installed products and to provide direct services, which makes access somewhat difficult. This opens up further potential to increase the reach to vendors without compromising security. This aligns with the findings of Griffy-Brown et al. [11] where multiple groups are trying to address business requirements through a third-party vendor.

They note that there is a lack of visibility within these vendors, as they show compliance, not controls. This has led to a few breaches across the system (2016).

2.2.8 Future Cyber Security Infrastructure Models

A number of theorists, including Tao et al. [17], predict that cyber security will have to adopt an immune function much like a human being. Autonomic intelligence may be useful in building this capability. Tao et al. state that with the fast development of a number of information techniques, new technologies like cloud computing, big data, the Internet of Things, and artificial intelligence have created new inputs in social development.

Big Data

Tao et al. [17] discuss the importance of combating data security risks for big data. Network applications based on big data are gradually changing the way society uses information. These applications have created a number of new challenges to network security, especially when it comes to data security. The production and flow of data have become far more complicated than in previous network systems. Legacy security technology systems simply cannot keep up with the ever-increasing complexity of new technological operations. In recent years, big data systems have become a major research focus. Traditional network security systems can be used in big data; however, big data has also introduced a number of new uninvolved components such as virtual machines, containers, and data security. Tao et al. [17] argued that a fine-grained big data security method based on a Zero Trust model is the most effective method for solving data security issues in the big data landscape where the main data security risks (Cloud Security Alliance) are:

1. Authorization and authentication Currently, big data systems, as represented by the Hadoop platform, are missing a mandatory system of authentication

and authorization.

2. Access control

   Currently open-source big data platforms struggle to achieve fine-grain data access control compared to legacy relational databases.

3. Auditing

   It is difficult for big data systems to achieve effective auditing, which is an essential part of post-evidence analysis.

4. Sensitive Data Protection

   Big data platforms store increasing amounts of highly sensitive data. With the diverse nature of platform access and the openness of these platforms, big data systems are at risk of leaking this privileged information.

Tao et al. [17] conducted experiments on the fine-grained big data security method based on the Zero Trust model in 2018. They found this method to be successful in ultimately identifying the majority of data security risks. They introduced a novel method for big data security control which comprises three phases: user context recognition based on Zero Trust; fine grained data access authentication control; and a data access audit based on full network traffic to detect and intercept risky data access in ambient big data.

2.2.9 Findings: Why Zero Trust?

The advent of cloud computing has changed the way businesses function. It offers multiple benefits but it has also opened up a number of new security risks. A critical problem has emerged with the new co-mingled network blurring the perimeters of the traditional network, with new devices being connected to networks classically outside of network control. Griffy-Brown et al. [11] interviewed over 200 executive board members of 80 companies between 2014 and 2016 to determine how to secure increasingly dynamic architecture in an environment without a perimeter. The responses to this question revealed that the policy of bringing your own devices presented extremely valuable opportunities but also posed onerous risks. Setting up a centralized and highly scalable mobile device management system through the use of access controls was the most significant challenge. Hence, a more risk-based method to cyber security is required in today's dynamic technological setting.

According to [4], it has never been more significant to structure security models that maintain users' safety. Assunção [4] draws on research by the Computer Security Institute showing that approximately 60 to 80 per cent of network misuse comes from within the network. Zero Trust therefore, offers a solution to both of these issues, with its ability to enhance micro segmentation of a network offering more visibility of overall traffic by inspecting devices and users which connect to the network [4].

DeCusatis et al. [15], P.5 posit that Zero Trust networks incorporate dynamic, automatic security policies that extend past traditional security boundaries, providing fine granularity segmentation and isolation of core critical resources, a technology which is based on a "trust nothing, verify everything" mentality. All traffic must be validated, even between virtual machines which share a common host. They stated "A fundamental aspect of Zero Trust includes guaranteeing secure access to all resources without paying attention to location and assuming all network traffic is a threat until it is inspected, authorized, and secured. security is part of a layered defense in-depth method that avoids kill chains and thus prevents single points of failure from compromising the whole security defense system".

According to [3], Zero Trust treats all network traffic as untrusted, continuously confirming users and endpoints by securing cloud data. The main benefit of Zero

Trust is its highly flexible infrastructure which can be integrated with the cloud to increase organizational security. To ensure the safety of their networks amid new cyber security threats, cyber security experts should embrace new security strategies alongside a Zero Trust mindset.

2.2.10 Core Pillars of Zero Trust

The American Council for Technology-Industry Advisory Council (ACT-IAC) states there are six pillars of the Zero Trust model, as shown in Figure 2.3. These pillars are as follows:

1. Users - People/Identity Security

   Continuous authentication is very important to Zero Trust. The use of multi-factor authentication and a process of constant monitoring and validation of the users' trustworthiness, along with Identity, Credential, and Access Management are effective ways to manage user access. This, combined with technologies for protecting user interactions, like web gateway solutions, are also very effective components of Zero Trust.

2. Devices - Device Security

   Not only is the authentication of users essential to the Zero Trust philosophy, so is the authentication of devices. If a verified user attempts to gain access via an unverified device, the device will not be trusted. Mobile device managers should be used to make device-trust assessments and additional assessments should take place for access requests like software versions, protection status, etc.

3. Network - Network Security

   New technologies have changed how network perimeters are created. The old-fashioned hard perimeter system no longer works. A Zero Trust approach moves “perimeters in from the network edge and segment and isolates critical data from other data” creating a “much more granular” (2019, p6) perimeter, using micro-segmentation to strengthen controls and protection of the network.

4. Applications - Application and Workload Security

   Protecting the application layer is another essential part of Zero Trust. Being able to identify and control virtual applications enables more granular access decision-making. Like users, multi-factor authentication is also very effective at controlling access to applications.

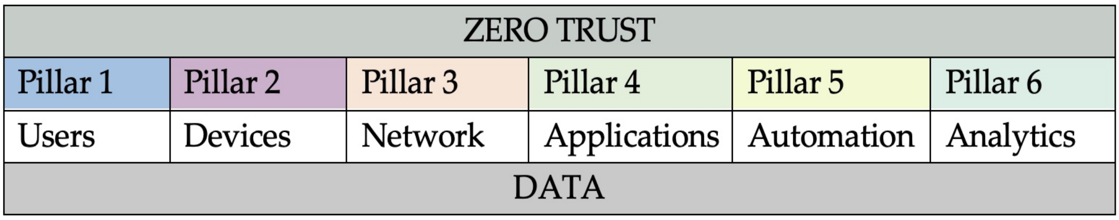
5. Automation - Security Automation and Orchestration

   Zero Trust Security needs to be automated and adaptive. Information pertaining to users, the server, or applications, can be collected and fed into a pool of data for machine learning purposes. The more data that is available, the more the machine will notice any behaviors which are abnormal, creating red flags that can be used to block access. This can greatly reduce manual workload, thereby keeping costs down.

6. Analytics - Security Visibility and Analytics

   Analytics provides a useful method of preventative protection in a Zero Trust model. Using various analytics programs, security professionals can understand when a threat is occurring in real time, leading to a more proactive defense.

The following section covers possible semantic attacks when accessing sensitive assets.



ZERO TRUST					
Pillar 1	Pillar 2	Pillar 3	Pillar 4	Pillar 5	Pillar 6
Users	Devices	Network	Applications	Automation	Analytics
DATA					

Figure 2.3 : The six pillars of the Zero Trust model

2.2.11 Semantic attacks

There is concern about possible semantic attacks by analysts with access to sensitive corporate data, Figure 2.4. When an adversarial third-party analyst gets access to sensitive information in a network for malicious purposes, the system is vulnerable to semantic attacks. Semantic attacks may be either passive or active, with active attacks, such as data injection involving a single class of attack. Fingerprinting and other passive attack types involve multiple classes of attack. In theory, a ZTA could resolve this by requiring continuous access verification. However, additional anonymization mechanisms are required to ensure sufficient levels of privacy (needed by organizations) and data availability (needed by analysts). This chapter reviews the existing approaches to network trace anonymization under ZTA and identifies the key challenges and possible solutions to these challenges.

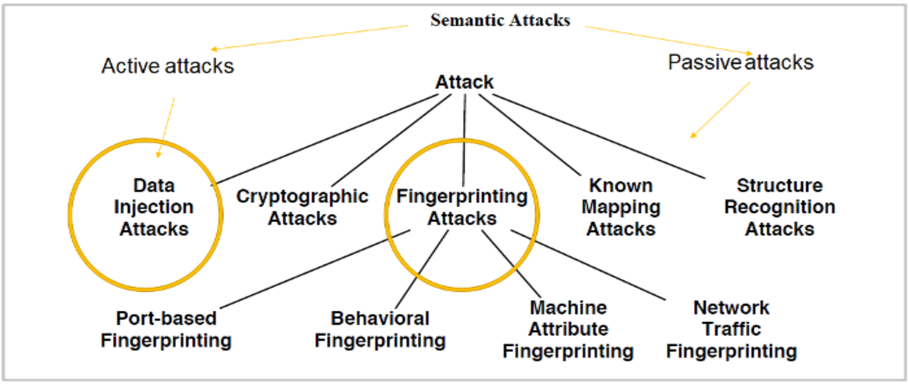

Semantic Attacks
Active attacks
Passive attacks
Attack
Data Injection Attacks
Cryptographic Attacks
Fingerprinting Attacks
Known Mapping Attacks
Structure Recognition Attacks
Port-based Fingerprinting
Behavioral Fingerprinting
Machine Attribute Fingerprinting
Network Traffic Fingerprinting

Figure 2.4 : Taxonomy of attacks against network trace anonymization [18]

A range of anonymization tools are currently employed to increase data security in networking tracing environments. Some examples are as follows:

a) Anontool is an extendable command line tool based on the anonymization API. It can anonymize both live and stored traffic while allowing users to create anonymization applications and policies according to their needs. They can be applied either to selected single fields or groups of fields. Additionally, Anontool offers a sizeable number of anonymization tools, such as block ciphers, constant/random values, basic mapping functions, regular expression matching and replacement, prefix-preserving anonymization, and various hash functions [19].

b) CANINE (Converter and Anonymizer for Investigating NetFlow Events) is a tool that converts and anonymizes NetFlow logs. It exclusively enables one type of NetFlow to operate on data from NetFlows in other formats. Moreover, it supports the anonymization of various fields in multiple ways. CANINE supports Cisco NetFlow V5/V7, which is widely used on the market. CANINE also enables users to select source/destination files, version settings, and fields for anonymization [20].

c) FLAIM (Framework for Log Anonymization and Information Management) is a tool based on the C++ programming language. It is designed to address anonymizing specific logs supporting multi-level anonymization to balance privacy/security concerns and information quality. Flexible in comparison to other applications, FLAIM is designed to share security-related data with other professionals safely [21].

All these tools, however, have weaknesses by being vulnerable to semantic attacks and requiring heavy sanitization leading to useless data for network analysis. Hence, a tool for preserving accuracy and security is required. In light of the aforementioned points, a literature review is essential in the context of privacy concerns. It allows researchers and practitioners to gain an understanding of the existing methods, challenges, and advancements related to data anonymization and privacy preservation in network security.

2.3 Systematic literature review on Zero Trust

2.3.1 Search method

This study applied an SLR approach to data collection and analysis [22]. The following research requirements (R1-R3) were investigated:

R1*: The existing approaches using Zero Trust in outsourcing with IDS monitoring.*

R2*: The existing approaches utilizing anonymization encryption approaches.*

R3*: The existing approaches preserving both security and accuracy.*

2.3.2 Review Protocol

This section overviews the different stages used to identify the relevant studies used in this SLR. This SLR is conducted based on the guidelines presented in [22].

Inclusion and Exclusion Criteria

The search was conducted across four major IT databases: ACM Digital Library `https://dl-acm-org`, IEEE Explore `http://ieeexplore.ieee.org/`, Elsevier Science Direct `https://www-sciencedirect-com`, and Springer Link `https://link-springer-com`. The keywords used for the search are provided in Table 2.2. The timeframe for the study was between 2010 (introduction of ZTA [2]) and 2023. Studies other than those that focus on Zero Trust, anonymization, and IDSs were excluded from the review. Only full papers in English were included. To ensure a high level of rigor, only papers from peer-reviewed journals and conferences were considered. To capture the relevant information from the selected databases, we used search terms based on the formulated search requirements. Table 2.2 shows the search terms used to conduct the search in the selected databases.

Table 2.2 : Search terms/phrases used to find the existing literature

Domain	Keywords
Trust	Zero Trust models
Anonymization	Multi-View (MV), Homomorphic Encryption (HE), Computation on Encrypted Databases (CED)
IDSs	IDS

The Search Process

After the initial search process using the key terms in the selected databases, a total of 731 papers were selected, as shown in Table 2.3. After removing the duplicates and irrelevant studies based on the title and abstract, 46 papers were selected. Then, these were examined more deeply, using the quality assessment criteria outlined in Table 2.4.

Table 2.3 : First Iteration of digital library paper statistics

Digital Library	Keywords: (Zero Trust, MV approach, Homomorphic Encryption, and Computation on Encrypted Databases)
IEEE Xplore	142
ACM Digital Library	96
(Elsevier) Science Direct	103
SpringerLink	390

After applying the quality assessment criteria detailed in Table 2.4, only 15 papers were selected as being relevant to this SLR for further analysis.

Table 2.4 : Quality assessment criteria

	Quality Assessment Criteria
1	The paper must be a research paper.
2	The paper must have a clearly stated goal.
3	The paper must clearly define the study context.
4	The paper must have a proper framework to meet the study goal(s).
5	The paper must have a rigorous, well-defined approach to data analysis.
6	The paper's findings must be evidence-based.
7	The research in the paper must be validated and/or implemented.

2.3.3 Classification of the selected studies

Analysis of the selected papers against the requirements R1 to R3

The analysed studies proposed a number of models for the anonymization of network tracing in ZTA. The models were based on two major approaches: homomorphic encryption (HE) and its modifications and the MV approach.

Homomorphic Encryption (HE)

HE is a technique to process encrypted data to protect original information from the third party that processes it [23]. When using HE protocols for a specific operation, it is necessary to encrypt the data based on the specific scheme, which only supports the specific operation on that data. Therefore, it is necessary to know the operation in advance while the data are being encrypted. Otherwise, each data item should be encrypted through multiple encryption methods to allow multiple operations later, which greatly increases computational resources. Furthermore, the HE system should also know the type of queries for which operations are to be performed so that an appropriate HE mechanism is selected to encrypt the original data [23].

HE and Intel SGX are used to preserve privacy even while performing different operations on encrypted data in untrusted third-party systems. However, there are some limitations, such as dramatic ciphertext expansion with low bandwidth, the need for off-premises support, being strictly bonded to the hosting server, and limited usable memory. Therefore, the authors in [23] proposed the Virtual Secure Enclave (VISE) method that effectively combines HE and Intel SGX techniques to overcome the limitations. VISE performs the execution of sensitive HE data on the cloud, where the SGX enclave is tested remotely. Because all computations of the sensitive HE data are performed externally on the cloud, VISE frees up the local memory resources.

Since private data must be transferred to the cloud server for large-scale operations, there is a clear need to protect the transferred information so that the original data is not exposed. The authors in [24] proposed a single-server HE mechanism (named CMP-SWHE) using the confused modulo projection theorem to process encrypted data without learning from user data. The designed blind computing scheme uses batch processing which improves computational efficiency while in the process. Data owners often use encryption for the original data before transferring it to the cloud. Because analyzing encrypted data is challenging, some HE schemes create multikey environments for privacy-preserving data mining. One such scheme was proposed by Pang and Wang [25] who designed a privacy-preserving association rule mining mechanism to reduce the encrypted data on a cloud server in large-scale shopping malls.

Cheon and Kim [26] designed a hybrid mechanism by combining the ElGamal and Goldwasser-Micali schemes that merge public key encryption and SHE for cloud computing environments to reduce the circuit exponentiation at the cost of extra public keys. The proposed mechanism, according to the authors, offers efficient encrypted data computing with lower requirements for storage and bandwidth. Thus, a good balance between the volume of the transmitted ciphertexts and the costs of their conversion is preserved.

Kim et al. [27] proposed an original algorithm utilizing wildcard pattern matching between encrypted data (i.e., string and keyword) based on leveled FHE and additional encrypted inputs. The researchers proposed original compound query protocols for wildcard search conditions on encrypted databases. The performance evaluation results show that the efficiency of the proposed protocols is comparatively better, but they are not adequate for real-world applications.

Qiu et al. [28] designed a scheme using Paillier HE and the data masking technique to perform privacy-preserving linear regression (PPLR) on the data that were partitioned horizontally in advance. They used two servers operating simultaneously to regress the shared user-submitted data without the need to decipher their contents. The users submit their data in an encrypted form to a server, and then, two servers collaboratively develop a regression prototype on the shared data without the need to understand its contents (Figure 2.5). The results show that data masking techniques offered a higher level of efficiency.

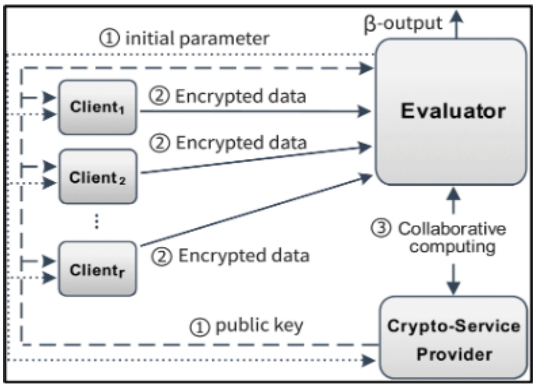


Figure 2.5 : The proposed distributed data analysis system in [28]

Some FHE schemes are based on multi-key identity to overcome non-adaptive chosen ciphertext attacks [29]. The proposed protocols satisfy the homomorphic (i.e., partially or fully) and compactness properties. Since these puncturable encryption instantiations are centered on indistinguishability obfuscation, the scheme is not usable in practical terms.

Alagic et al. [30] designed a protocol for Quantum Fully Homomorphic Encryption with verification (vQFHE) to validate random polynomial-time quantum calculations without involving a client and the server. The proposed scheme also provides correctness, authentication, compactness, and security of verification. It can withstand indistinguishability under a chosen plaintext attack. However, there is an open research question, namely whether any vQFHE schemes exist to confirm quantum computations publicly (while not using the decryption key) or not.

Dyer et al. [31]proposed HE protocols to perform integer arithmetic calculations, which can be used for secure single-party computation in the cloud. In their study, four different schemes were designed to enhance the security level, but the computation overhead is also increased, requiring more computational resources to process the calculation. The proposed scheme relatively outperforms the related protocols in the computation time of arithmetic operations on encrypted data.

Saha et al. [32] developed an efficient scheme for processing conjunctive and disjunctive private queries over an encrypted database. Their approach applies to lower-depth equality circuits based on the packing methods that result in efficient batch computations. Based on both theoretical analysis and practical evaluations, the proposed schemes were found to be more efficient in terms of running time and cipher text size (due to the use of a base-N encoding technique) than the existing protocols.

Xu et al. [33] proposed a new technique called the fully homomorphic encryption based Merkle tree (FHMT) for streaming authenticated data structures (SADS) to verify streaming computation performed on untrusted servers. FHMT transfers most of the computation operations to the server, resulting in no computation overhead. However, the typical FHMT cannot support a dynamic scenario well because its height is fixed. Also, they designed a fully Dynamic FHMT (DFHMT) mechanism to authenticate an infinite number of data elements while incurring lower computational overhead. The results of DFHMT demonstrate that the user is required to perform only simple numerical multiplications and additions instead of hash functions while providing the same security as FHMT.

Catalano and Fiore [34] proposed a method to transform a linear HE into a scheme capable of performing second-degree computations on encrypted data. The benefit of the proposed technique, according to the authors, is its light transformation requirement. Specifically, it only requires the message space to be defined as a public ring for the arbitrary modeling of elements in a uniform manner. As such, full compactness is achieved while performing encrypted data computations on two non-interactive services.

Bellafqira et al. [35] designed a Homomorphic Proxy Re-Encryption protocol to share outsourced cipher text data which were encrypted by applying public keys for remote data operations. The proposed scheme is designed based on encrypted noise (using a secret key) in which the differences between the encrypted noise and data are calculated on the cloud. These differences are encrypted by the cloud using the delegate's public key with the subsequent noise removal to reduce computational workload.

Finally, Ahmed and Ogalo [36]discussed a Zero Trust model to provide sensitive data access in which the request is granted based on the type of access, user, device, application, and data. They showed significant developments in the HRM area. They investigated the reasons behind the growing shift from HRM to E-HRM. Their study also explained the features and prospects of E-HRM and its advantages over traditional HRM works.

MV Approach

Another class of encryption that preserves both the security of the underlying dataset and the accuracy of the analysis is the class of property preserving encryption (PPE). This category of encryption retains certain properties of interest through its scrambling procedure, e.g., distant preserving and prefix preserving. Unfortunately, these encryptions are shown to be vulnerable to a serious class of attacks called semantic attacks, which are designed to collect and extrapolate confidential data on the original datasets through fingerprinting and injection when observing the PPE's ciphertexts.

The MV approach was proposed to mitigate this vulnerability while preserving the appealing accuracy of these solutions. The idea behind this approach is the creation of a number of relatively indistinguishable fake views along with the original view, making it impossible to distinguish the latter. The only paper which used an MV approach was [1]. The authors described a specific algorithm using CryptoPAn encryption with two distinctive keys, one of which is preserved exclusively by the data owner. After this, the original trace is partitioned using the key which is given to the analyst. The analyst generates N views based on partitioning, analyzes them all, and generates the corresponding reports. However, only by using the second key can the original view be identified. The experiment results demonstrate that the proposed scheme is able to expressively decrease the level of information leakage while maintaining comparable utility. Furthermore, since it is only necessary to send one seed view to the analysts, the proposed solution does not incur additional communication overhead. Therefore, the authors concluded that the system offers a better solution for outsourced encrypted data applications to preserve privacy and utility with less computational overhead.

Table 2.5 : Summary of Anonymization Approaches and Results

SLR Papers	Using Zero Trust in outsourcing with IDS monitoring	Anonymization Encryption Approach	Achieves better accuracy	Preserve security	Preserve privacy
[23]	No	HE+Intel SGX	Yes	Yes	Yes
[24]	No	Single-server HE	No	Yes	No
[25]	No	Full HE	No	No	Yes
[26]	No	Somewhat HE	No	Yes	No
[27]	No	Full HE	No	Yes	No
[28]	No	Partial HE	No	Yes	No
[29]	No	Full HE	No	Yes	No
[30]	No	Full HE	No	Yes	Yes
[31]	No	Full HE	No	Yes	No
[32]	No	Full HE	No	Yes	No
[33]	No	Merkle Tree HE	Yes	Yes	No
[34]	No	Partial HE	Yes	Yes	No
[35]	No	HE Proxy	Yes	Yes	No
[36]	Yes	None	Yes	Yes	No
[1]	No	MV	Yes	Yes	No

2.3.4 Discussion

A summary of the anonymization approaches for data tracing identified in the literature as well as the outcomes of their uses is presented in Table 2.5.

The literature review demonstrated that the dominant approach to data anonymization remains homomorphic encryption. Fourteen of the fifteen papers used some kind of HE for network tracing issues. Unfortunately, HE schemes usually suffer from a number of issues, making them impractical architectures for many real-life scenarios. Some common problems with HE models mentioned in the literature are memory consumption and computation time overheads, a lack of expressibility and linkability, as well as integrity issues [23, 33]. A few proposed systems managed to achieve the desired levels of privacy and information accuracy however only [23] offered a model that ensured a relatively high level of data accuracy along with acceptable levels of privacy and security. Unfortunately, the scope of IDS considered in this work is quite limited (to only one intrusion detection task, i.e., code injection).

The MV approach, which conceals an accurate view of the original datasets among seemingly indistinguishable fake versions, could be a better solution in this regard. The real view is possible to detect with a unique key, which is generated for the data owner only [1]. On the one hand, this minimizes the threat of semantic attacks; on the other hand, this ensures that the one analysis based on the true data is highly accurate. Accordingly, this solution does not suffer from the issues inherent to HE as discussed above. However, such systems are still very rare, which means that further research is required to bolster the existing knowledge and ensure the viability of implementing and using MV systems for ZTA-IDS.

2.4 Open research gaps:

The literature review helped identify several significant challenges and limitations in using Zero Trust anonymization to enhance the accuracy and security of organizational networks. The following research gaps are identified:

- Research Gap 1: The security issues of outsourcing IDS tasks to third-party analysts are not identified, i.e., semantic attacks. Currently, little attention has been given to the identification and mitigation of these potential risks involved in such outsourcing arrangements.

- Research Gap 2: Another gap exists in the absence of a well-established, effective solution in the research field of utilizing Zero Trust models for IDS task monitoring by third-party analysts in that there is no effective solution yet established in this research field.

- Research Gap 3: Additionally, the challenges associated with designing a solution that specifically addresses the mentioned security issues in IDS task monitoring by third-party analysts remain largely unknown

2.5 Limitation of this SLR

It is important to state the limitations of this SLR as well. While it has provided valuable insights into the identified research gaps and challenges, it may not include every possible resource or fully capture the dynamic nature of the research field. However, this SLR serves as an initial step in understanding the key research gaps and limitations in the field of utilizing Zero Trust approaches in the monitoring of IDS tasks by third-party analysts.

2.6 Conclusion

The increase in the tendency to apply cloud computing, the Internet of Things, and mobile device use has made traditional network boundaries disappear. Hardened network perimeters alone are no longer effective for providing enterprise security in a world of cloud computing and increasingly sophisticated threats.

The Zero Trust approach combines tight identity-based verification for every person and device attempting to access resources on a private network or the cloud, regardless of whether they are inside or outside the network perimeter. Zero Trust should be considered as a holistic framework rather than be associated with any specific security approach or method. Indeed, it is based on a variety of principles, methods, and ideas of cyber security integrated to ensure digital security. Examples include, among others, the prevention of semantic threats, segmenting networks, granular user-access management, and minimization of lateral movement. ZTA could offer an excellent solution to organizations seeking to outsource IDS services to third-party analysts while reducing semantic attack threats. However, the specific anonymization mechanisms ensuring a good balance between privacy and security of such ZTA-based systems are still being explored.

Based on the reviewed literature, it can be seen that the dominant anonymization approaches are based on HE systems which suffer from a number of setbacks, sacrificing either privacy or analytical data accuracy in the process of encryption. MV-based systems may offer a better solution, but they have not been sufficiently explored yet. This review is part of large-scale research that proposes and evaluates a comprehensive ZTA for the IDS system ZTA-IDS. Such a system could offer a much-needed solution to the edgeless network security where the trade-off between privacy and data utility is minimized. In the following Chapter, we discuss the technical scenarios of the Zero Trust model, MV and IDS.

CHAPTER 3

The Technical Scenarios of the Zero Trust Model, Multi-View and Intrusion Detection

3.1 Introduction

The previous chapter presented on a systematic literature review (SLR) which analyzed the existing studies, articles, and publications to identify the gaps related to Zero Trust Architecture Intrusion Detection System (ZTA-IDS) tasks.

In this chapter, we provide a detailed explanation of the technical scenarios of the Zero Trust model, Multi-View (MV) and IDS. We also explain the technical information of Zero Trust models, which includes IDS and MV integration in the network monitoring service. We also detail the importance of accuracy and security for outsourcing IDS tasks.

This chapter is organized as follows: section 3.2 explores effective Zero Trust models for IDS in outsourcing settings; Section 3.3 explores the defense with regard to the MV approach; Section 3.4 explains the proposed network IDS ; Section 3.5 highlights the accuracy of IDS in third-party settings; section 3.6 explains in detail the integration of Zero Trust in IDS; Section 3.7 details the adversary model. Finally, Section 3.8 concludes the technical discussions.

3.2 Effective Zero Trust Models for IDS in the Outsourcing Setting

An effective zero Trust model to be used when outsourcing IDS tasks is one that retains the accuracy of the outsourced data (in comparison to its original version) as much as possible to enable accurate intrusion detection while providing formal guarantees on accomplishing the objectives of Zero Trust models. To this end, we identified two setups (1) securely outsourcing datasets such as the homomorphic encryption and MV approach), and (2) conducing an analysis with aggregated results like mediated analysis, federated Learning and their corresponding techniques.

3.2.1 Securely Outsourcing Datasets

A common way of enabling secure computation between two parties, i.e., data owners and third-part analysts, is to apply certain encryption techniques which are designed to preserve the most important properties of the original datasets. In the following, we shed light on two main categories of secure computations technologies homomorphic encryption and the MV approach.

Homomorphic Encryption

One interesting direction in secure computation is to implement homomorphic encryption (HE) which is a kind of encryption that enables computation on ciphertexts, and generates an encrypted result which, when decrypted, matches the operations' result as if it had been performed on plaintext. Therefore, ciphertexts generated by HE can potentially be used by third-party analysts (e.g., third-party IDS) whenever necessary. Briefly, HE-like asymmetric schemes require two keys, i.e., a process (public) key, and an encrypt/decrypt (secret) key data, which are used respectively by the third-party analysts to perform operations over this encrypted data, and by the data owners for encrypting (the dataset) and decrypting (the analyses results) [46]. The development of practical HE schemes is still in its primary phase. The proposed schemes include (1) a partially homomorphic encryption (PHE) scheme [37] which covers only few primary operations (addition or multiplication), (2) a fully homomorphic encryption (FHE) scheme whose security power relies on noise perturbation into the ciphertexts which itself may lead to undecryptable ciphertexts, and (3) a somewhat homomorphic encryption (SHE) scheme which is placed between the former two schemes [38]. Unfortunately, HE schemes usually suffer from the following negative properties, making them an impractical architecture for many real-life scenarios.

- Memory consumption and computation time overheads. These are still far from meeting practical requirements. For instance, 1 Mb of data results in more than 10 Gb of encrypted data, and while the computation time of "addition" is appealing (1ms), the multiplication could take over 5 seconds.
- Expressibility. One of the common issues among all encrypted databases is the lack of sufficient expressibility when it comes to decision making, interpretation and applying other types of operations (they need a secret key).
- Linkability. Another issue is the potential security vulnerabilities pertaining to the linkage of one computation to another. This can be especially effective in the case of HE since both encrypted databases are typically generated under the same conditions and with the same key to enable multi-round computation.
- Integrity. Due to its native hidden nature, HE cannot ensure that every computation is carried out as expected.

Because of these limitations, there are very few state-of-the-art studies which use HE schemes in our scenario of interest (secure and accurate IDS). To the best of our knowledge, Coppolino et al. [23] is the only work which tries to marry the two worlds. Unfortunately, the scope of IDS considered in this work is quite limited (to only one intrusion detection task, i.e., code injection).

This is reasonable because of the level of expressibilty and linkability required to conduct different IDS tasks simultaneously. Specifically, in this detection method first a signature for code injection is built and then a single string field is examined to check if it includes special characters, i.e., semicolons, dots, brackets, commas, or any other non-alphanumeric characters in the ASCII table, which could lead to an injection attack. Due to the simplicity of this procedure (matching a single string field), HE can be successfully leveraged.

Kasongo and Sun [39] proposed the use of feature extraction for IDS, where each feature extraction method does not show promising results on its own. However, a combination of several features results in 90.85% accuracy for binary classification [39]. Tama et al. [40] proposed the use of a hybrid feature selection method and a two-stage ensemble classifier based on bagging for IDS, outperforming the other methods for binary classification with a 92.27% accuracy. Jiang et al. [41] proposed the use of deep neural networks for IDS. The evaluated methods are AlexNet, LeNet, CNN, BiLSTM, and CNN-BiLSTM. The results show that CNN-BiLSTM outperforms the others in accuracy for the multiclass mode.

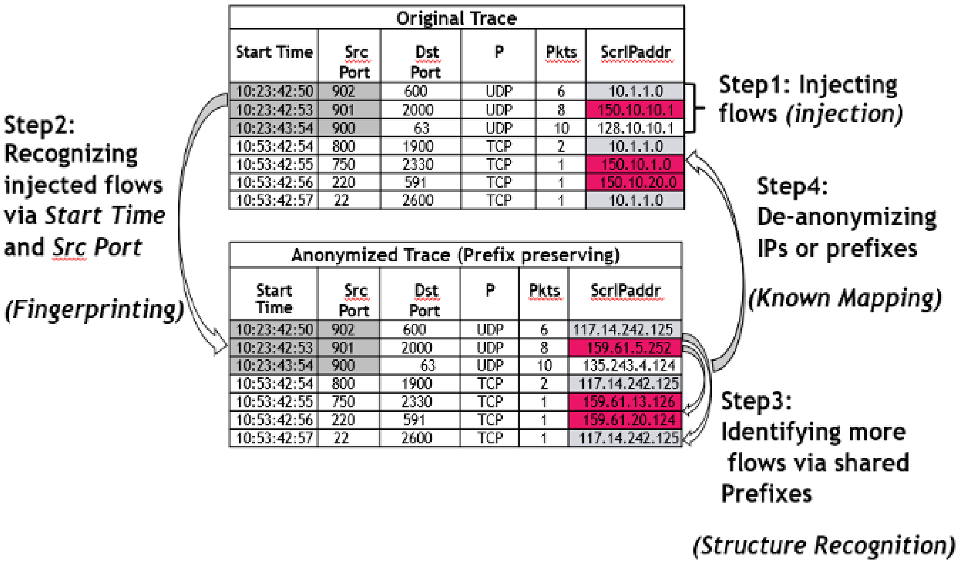
Multi-View Approach and Mitigation Strategy

Another interesting class of encryption which preserves both the security of the underlying dataset and the accuracy of the analysis is property preserving encryption (PPE). This category of encryption retains certain properties of interest using a scrambling procedure, e.g., distant preserving and prefix preserving. Unfortunately, these encryptions are vulnerable to a serious class of attacks called semantic attacks which are designed to collect and extrapolate confidential data on the original datasets through fingerprinting and injection when observing the PPE's ciphertexts.

Figure 3.1 illustrates this vulnerability for prefix preserving encryptions applied to an excerpt of a network trace (obtained from Mohammady et al. [1]). In Figure 3.1, table 1 illustrates the main trace and table 2 presents the trace anonymized applying the prefix preserving encryption. In this example, without loss of generality, we only focus on source IPs. Inside each table, similar prefixes are highlighted through similar shading. In Step 1, an adversary has injected three network flows, shown as the first three records in the original trace (upper table).

In Step 2, the attacker detects the three flows which are injected in the anonymized trace (the second table in Figure 3.1) via unique compositions of the remaining features (Begin Time and Src Port). In Step 3, the adversary is able to extrapolate the knowledge from the flows that were injected to the original flows in this way: for instance, as prefix 159.61 is shared by the 2nd (fake), 5th (original) and 6th (original) flows, the adversary is aware that all three should share the same prefix in the main trace too. These detected connections between flows in the two traces is named matches. In Step 4, the adversary can derive the prefixes / entire IPs of the anonymized flows in the main traces because the adversary is aware of the main IPs of the flows which are injected, such as the 5th and 6th flows should have the prefix: 150.10, and the IPs of the 4th and last flows should be 10.1.1.0. In general, a strong attacker who is capable of probing all the subnets of a network by injection or fingerprinting can possibly de-anonymize the whole anonymized trace.

Recently, a technique called the MV approach was proposed to mitigate this vulnerability while preserving the appealing accurate property of these solutions. Briefly, the core concept is for the data owner to convey sufficient data to the third-parties, therefore, they can produce and investigate various anonymized views of the main data which are introduced to be adequately indistinguishable, even to adversaries armed with previous knowledge and executing semantic attacks, which maintains the security. Simultaneously, one of the anonymized views will yield true analysis outcomes, which will be privately retrieved by the data owner or any other authorized parties which maintains the accuracy. While using HE in practical applications conveys the challenges of Computational Overhead, Increased Data Size, Complexity of Implementation, Limited Operations. Interestingly, this solution does not suffer from the four limitations inherent to HE and therefore is a suitable candidate to be used in our scenario of interest which is Zero Trust anonymization IDS tasks.



Original Trace					
Start Time	Src Port	Dst Port	P	Pkts	ScrIPaddr
10:23:42:50	902	600	UDP	6	10.1.1.0
10:23:42:53	901	2000	UDP	8	150.10.10.1
10:23:43:54	900	63	UDP	10	128.10.10.1
10:53:42:54	800	1900	TCP	2	10.1.1.0
10:53:42:55	750	2330	TCP	1	150.10.1.0
10:53:42:56	220	591	TCP	1	150.10.20.0
10:53:42:57	22	2600	TCP	1	10.1.1.0

Anonymized Trace (Prefix preserving)					
Start Time	Src Port	Dst Port	P	Pkts	ScrIPaddr
10:23:42:50	902	600	UDP	6	117.14.242.125
10:23:42:53	901	2000	UDP	8	159.61.5.252
10:23:43:54	900	63	UDP	10	135.243.4.124
10:53:42:54	800	1900	TCP	2	117.14.242.125
10:53:42:55	750	2330	TCP	1	159.61.13.126
10:53:42:56	220	591	TCP	1	159.61.20.124
10:53:42:57	22	2600	TCP	1	117.14.242.125

Step1: Injecting flows *(injection)*
Step2: Recognizing injected flows via *Start Time* and *Src Port*
(Fingerprinting)
Step4: De-anonymizing IPs or prefixes
(Known Mapping)
Step3: Identifying more flows via shared Prefixes
(Structure Recognition)

Figure 3.1 : An example of injection attack in prefix-preserving anonymization [1]

3.2.2 Analysis with Aggregated Results

An alternative method for enabling a data-driven networking study is to run several analyses on behalf of the scientists by the data owners instead of publishing sanitized data to maintain privacy. Conditions are placed in relation to which investigations are allowed and which output is returned.

Mediated Analysis

One possible way of implementing such a setting was initially designed by Mogul and Arlitt [42] and then Mittal et al. [43] called mediated trace analysis. Figure 3.2 illustrates this model which can be seen as an interaction between a third-party analyst and an encrypted trained model (rather than having access to the raw data).

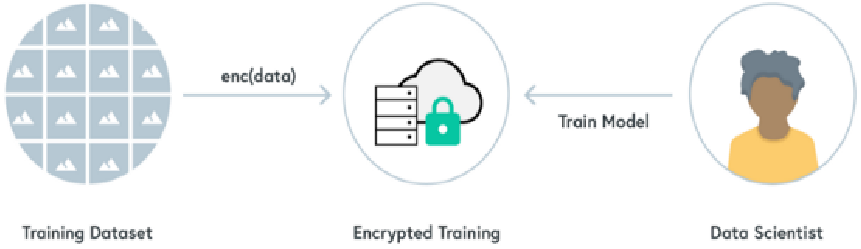


Figure 3.2 : Mediated analysis as a Zero Trust model [42]

Considering the complications of ensuring the security of sensitive data and the previous failures, it is agreed that formal and strong privacy guarantees are necessary in building trust and ensuring that data owners are confident in the security measures applied during the data encryption process for training purposes. For instance, Mittal et al. [43] proposed an investigation where leaks less than a predefined number of bits (within a data theoretic sense) are permitted, even though limiting information leakage and maintaining privacy are not equal. One analysis shows if hosts A and B link leaks which are only 1 bit in size, this might represent an unreasonable privacy loss for the hosts. Another definition of privacy considered in the literature is differential privacy [42]. Informally, the variation of privacy guarantees that the absence or presence of individual records is difficult to derive from the analysis output. However, it is not clear if a variation in privacy is a suitable guarantee for networking analysis, or even if one single definition exists that has been used in all investigations and datasets. McSherry and Mahajan [44] considered it as it gives the most powerful known privacy guarantees. Interestingly, it is resilient to collusion, helps various interactive queries, and is not dependent on any auxiliary information which an attacker may possess (which includes semantic attacks). Similarly, a variation in privacy is capable of providing a powerful basis for mediated data analysis even though powerful guarantees of variation in privacy cannot be freely achieved. Privacy is maintained by aggregating noise in the analysis output, improving its accuracy. The additional noise is scaled to mask the absence or presence of the records' small sets. However, the magnitude of the noise is not usually big, and the analyst is aware of the distribution, so it can render sensitive analyses of no use. Additionally, ensuring the expressiblity for different classes of analysis is challenging when applying data privacy.

Federated Learning

Federated learning (FL) is a recently proposed distributed learning technique that can leverage private training data held on thousands or millions of client devices [45]. As shown in Figure 3.3, the FL algorithm proceeds in a multi-round fashion where each client is responsible for submitting a locally trained model to an aggregation server that produces a joint model. This server and the aggregated client model are referred to the central server and the central model, respectively. Since client models are shared rather than training data, data confidentiality is guaranteed throughout the learning process. This feature of FL is highly desirable for learning tasks on sensitive data (i.e. health-related) but makes it impossible to verify the validity of the data used to train the model. Being unable to ensure full expressibility is another issue when FL is executed.

Each class of the techniques previously mentioned can be efficiently run for a specific type of application. The IDS mission however requires a global view of the

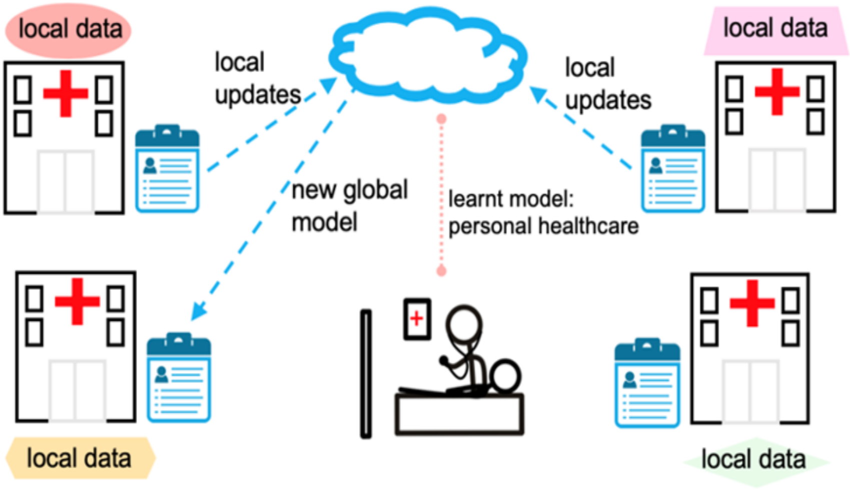


Figure 3.3 : Federated Learning [45]

Table 3.1 : Comparison of different anonymization techniques

Methodology	Semantic Attack	Accuracy	Overhead
Homomorphic Enc	Robust	Limited classes of simple queries	High
MV	Robust	Any query (raw data is being sent)	Low
Analysis with Aggregated Results	Semi-Robust	Many classes of queries but still not comprehensive	Low

dataset to accurately distinguish benign activities from suspicious ones. The only solution that can provide such view while ensuring the security of the underlying dataset is the MV approach. Table 3.1 compares these solutions in detail.

In this thesis, we consider the problem of securely outsourcing network IDSs while trying to preserve the accuracy of detection. For this purpose, we rely on the MV approach.

3.3 Defense: The Multi-View Approach

Before describing the MV approach, we first need to formally introduce the prefix preserving anonymization PP(-,K) which is basically a cryptographic mapping function like CryptoPAn [1] that relies on a secret key K. The core of CryptoPAn is based on the Rijndael cipher (the algorithm behind AES - Advanced Encryption Standard) [46]. However, instead of using Rijndael for encryption in the traditional sense, CryptoPAn uses it to generate a pseudorandom bitstream that is XORed with the original IP address to produce the anonymized IP address. The most important property of this function is that it can preserve the prefixes of numeric-value attributes. In other words, if two real addresses share the first X bits, e.g., 150.10.10.1, 150.10.20.1 share 14 bits in their prefixes, they are mapped to another two anonymized addresses, e.g., 97.61.5.252, 97.61.5.252, which share the first X bits. As detailed in the previous sections, PP is vulnerable to different classes of semantic attack, and the MV approach presented in Figure 3.4 is designed to secure its output [1].

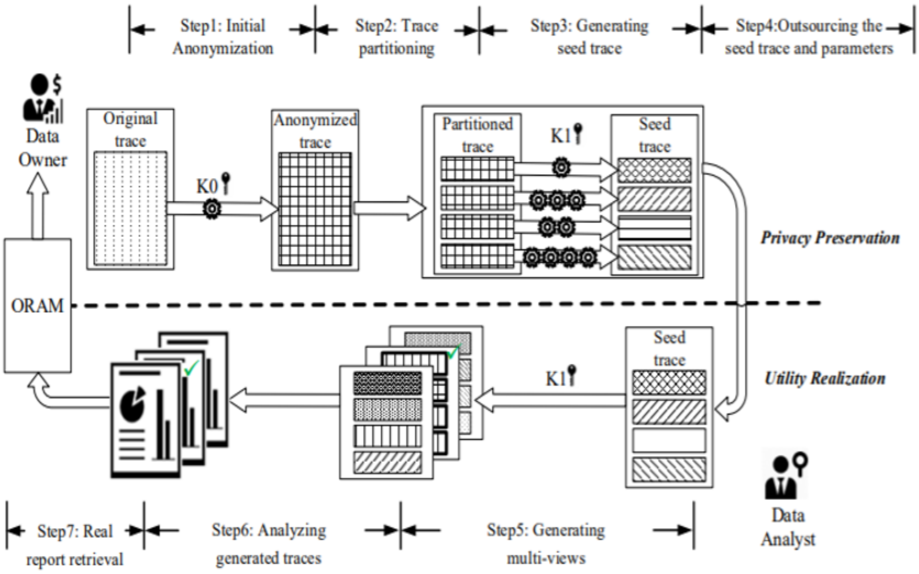

Step1: Initial Anonymization
Step2: Trace partitioning
Step3: Generating seed trace
Step4:Outsourcing the seed trace and parameters
Data Owner
Original trace
K0
Anonymized trace
Partitioned trace
K1
Seed trace
Privacy Preservation
ORAM
Seed trace
K1
Utility Realization
Data Analyst
Step7: Real report retrieval
Step6: Analyzing generated traces
Step5: Generating multi-views

Figure 3.4 : Multi-View approach [1]

There are seven steps in this approach, the first four are initiated on the data owner side and the final three steps are initiated on the data analyst side. We note that this approach assumes that the confidential attribute is the IP address, and as long as IPs are kept secure, the adversary cannot infer any sensitive information.

3.3.1 Implementing effective Zero Trust model at the Data Owner Side

Step 1: Data owner produces two 256 bit cryptographic symmetric keys (K_0 & K_1) and the original data is anonymized using $PP(-, K_0)$. The symmetric keys are critical components since it determines the output of the anonymization process.

Step 2: The anonymized trace is partitioned.

Step 3: Each partition is anonymized but repeated for another number of times at different partitions. Therefore, the seed trace is not prefix-preserved.

Step 4: The seed trace and some supplementary parameters (a pseudo random vector whose elements are the number of times PP(-,K) must be applied to each partition in each view generation) are outsourced to the data analyst. The pseudo vector and the seed view generation are designed such that after r number of times view generation is undertaken, the real view (which is identical to the view in step one) will be retrieved. Here, r is a random number ∈ [1, N] which is used to generate seed traces.

3.3.2 Accuracy Guarantee at the Data Analyst side

Step 5: Analyst produces N views based on the seed view and supplementary parameters.

Step 6: Analyst analyses all N views and generates corresponding reports.

Step 7: Data owner retrieves report corresponding to the real view, using a private information retrieval (PIR) protocol in a way so that the analyst cannot identify which view was retrieved [47].

Clearly, the quality of the views generated in the MV approach is the main factor of the adds-on confidentiality the network trace will receive. Specifically, if all other (fake) generated views are too far or too close from the original trace (prefix-wise and in the presence of some adversary knowledge), the MV approach may end up compromising a high level of privacy. In particular, the adversary can discard many of the (far) fake views by looking at the prefix relations in the IP addresses and compare them to his/her adversary knowledge to identify inconsistencies. Conversely, a design with fake views generated too close to the real view incurs drastic privacy leakage. In the latter case, the fake views are actually not so fake. Therefore, in [1], the authors proposed a metric called indistinguishability to reflect the distance of each view from the real view. Based on this formalization, they suggest two schemes for their partitioning algorithms, i.e., IP-based and distinct IP based partitioning, each with their own partition sizes. They concluded that a distinct IP-based partitioning with a customized pseudo vector can significantly reduce privacy violation 50 fold.

Since IP based prefix length is very useful to partition, particularly in the context of partitioning network datasets for analysis, the choice of prefix length can significantly influence the following parameters [48, 49]:

- Granularity of Data Analysis: A shorter prefix length (indicating a broader network range) can lead to more aggregated data analysis, providing insights at a higher level but potentially missing finer details. Conversely, a longer prefix length allows for more granular analysis, capturing detailed behavior within smaller subnets but may require more computational resources and could raise privacy concerns by making it easier to infer information about individual hosts.

- Anonymity vs. Utility Balance: In anonymization processes the prefix length can affect the balance between preserving user anonymity and retaining the utility of the data. Longer prefixes may retain more specific routing or geographical information, useful for certain types of analysis but potentially compromising anonymity. Shorter prefixes increase anonymity but may reduce the utility of the data for detailed network performance or security analysis.

- Network Security and Intrusion Detection Sensitivity: The partitioning of network traffic data by prefix length can affect the sensitivity and specificity of IDS. Finer partitioning (longer prefix lengths) might improve the detection of anomalies within small network segments but could also increase false positives. Coarser partitioning might miss specific, localized threats but could be effective for identifying large-scale anomalies.

- Performance and Scalability: The choice of prefix length impacts the performance and scalability of network analysis tools and systems. Longer prefixes can lead to a higher number of partitions, which might increase the computational and storage resources needed for analysis. Shorter prefixes result in fewer, larger partitions, which might be easier to manage but could lead to less efficient data processing due to the increased volume of data in each partition.

- Compliance and Regulatory Requirements: Regulations concerning data privacy and protection might indirectly influence the choice of prefix length by setting bounds on the level of detail that can be retained in anonymized datasets. Ensuring compliance while maintaining the analytical value of the data requires careful consideration of how network data is partitioned and analyzed.

- Risk of Re-identification: Particularly in anonymized datasets, the prefix length can affect the risk of re-identification of individual users or hosts. Longer prefix lengths, which provide more specific location or organizational information, might increase this risk, challenging the goal of anonymization. Adjusting the prefix length to ensure it supports anonymity without unduly sacrificing data utility is a critical consideration.

3.4 Network Intrusion Detection System

IDSs are critical to the success of Zero Trust models. The fundamental philosophical purpose of an IDS is to minimize examples of false positive alarms and to enhance its detection accuracy. IDS methods are divided into two categories: binary and multiclass. In recent years, neural network-based IDSs have emerged as leading systems in the intrusion detection research domain. Neural networks give systems the capability to learn and grow by applying previous data. To be more precise, neural network-based computer programs do not need to be explicitly programmed [39].

Kasongo and Sun [39] proposed the use of feature extraction for IDS, where no feature extraction method achieves promising results on its own. However, a combination of some features results in 90.85% accuracy for binary classification.

Tama et al. [40] proposed a hybrid feature selection method and a two-stage ensemble classifier based on bagging for IDS. It outperformed the other methods for binary classification with a 92.27% accuracy.

Jiang et al. [41] proposed deep neural networks for IDS. The evaluated methods are AlexNet, LeNet, CNN, BiLSTM, and CNN-BiLSTM. The results show that CNN-BiLSTM outperforms the others in relation to accuracy for the multiclass mode. Consequently, we use a customized CNN-BiLSTM model similar to the other blocks such as MV for IDS.

3.5 Accurate IDS in Third-party Settings

Conducting reliable IDS heavily depends on the accuracy of the received dataset compared to its original version. Specifically, for most detection algorithms to be effective, they rely on learning benign (or suspicious) activities from high-dimensional and large-sized datasets. Therefore, only those defense mechanisms with minimal modification to the format/entries in the dataset can lead to trustworthy intrusion detection. In addition to the existing homomorphic encryption and computation over aggregated data techniques, the MV approach [1] has recently been introduced as an effective methodology to maximally benefit the best of both worlds. This approach outputs (1) a prefix preserving version of IP address attributes, and (2) 100% accurate copy of values in other attributes. Such a minimally modified version of the network traces (the real view) can be used to accurately conduct so many different analyses. Confidentiality is ensured by hiding the real view among a set of indistinguishable fake views.

3.6 Integrating Zero Trust in IDS

Another important aspect of a Zero Trust implementation of an IDS in an outsourcing setting is to ensure the accuracy of the required analyses. Ensuring the accuracy of the model must be guaranteed alongside its confidentiality because the first motivation of such outsourcing frameworks is for data owners to benefit from more accurate analysis tasks (as described earlier). For this purpose, as we clarified in the literature review section, we believe the MV approach is an ideal candidate. In contrast to other solutions, MV allows third-party analysts to receive and interact with the raw data. Nevertheless, as MV anonymization entails, the original versions of some of the attributes must be replaced with a cryptographic prefix preserving version. We now shed light on the impact of such a transformation over the outcomes of certain classes of IDS algorithms.

- Since our focus is network traces, and we only apply PP on the IP address attributes, all traffic-level analyses (packet level) are expected to return trustworthy results, e.g., the traffic of the entire network, or the traffic on a certain port.

- With similar reasoning, the flow level analyses also remain intact.

- The main issue relates to the graph-level set of analyses, e.g., subnet-based analyses, throughputs of a subnet, and reachability analyses.

  On the other hand, recently, solutions with deep neural networks have undertaken traditional machine learning approaches in various applications [50] [51] due to their strong learning power and IDS is not an exception.

  In [52], a misuse-based IDS was proposed to detect five categories in a network namely: Probe, Exploit, DOS, Generic and Normal. This system was based on a misuse-based model, which permitted it to act as a firewall with some extra information added to it. Moreover, unlike most related works, in their paper, the UNSW-NB15 dataset was considered as the offline dataset to design their own integrated classification-based model to identify malicious activities in the network. Furthermore, they generated their own real-time data set at NIT Patna CSE lab (RTNITP18) which acted as the working sample of their intrusion detection model.

  In [53], a IDS model was introduced which was a fusion of a CNN and a gated recurrent unit. The main concerns of the researchers were to tackle the issues associated with the low accuracy of existing intrusion detection models for the multiple classification of intrusions and the low accuracy of class imbalance data detection. In their method, a hybrid sampling approach merging adaptive synthetic sampling and repeated edited nearest neighbours was applied for sample processing to solve the positive and negative sample imbalance issue in the original dataset. The feature selection process was carried out by combining the random forest algorithm and Pearson correlation analysis to address the feature redundancy problem. Then, the spatial features were extracted by applying a CNN, and were then extracted by fusing Averagepooling and Maxpooling by exploiting the attention mechanism to assign different weights to the features, thereby decreasing the overhead and increasing the model's performance. Simultaneously, a gated recurrent unit (GRU) was applied to extract the long-distance dependent information characteristics to achieve universal and effective feature learning. Eventually, a softmax function was applied for classification.

We consider employing the NN-IDS in our solution and we elaborate on the accuracy of these special types of IDS.

3.7 Adversary Model

As previously mentioned, our model is proposed against a semi-honest adversary who will be conducting the task of IDS but might have various incentives to disclose the identity or the exact value of some attributes of some of the records. We now present our threat model while making the following assumptions.

- The adversary is semi-honest.
- The adversary can observe the anonymized version of the data. The adversary also knows the underlying anonymization algorithm.
- The adversary has knowledge of the original version of a subset of the records. The adversary has collected this information by running some semantic attacks, e.g., injection attacks. Accordingly, we assume that the adversary knows the original prefix values of the prefixes (subnets in the case of network traces, and the most significant bits of the other numeric-value datasets).
- Finally, the adversary's objective is to find the original version of the highest possible number of anonymized records. From this point forward,

we refer to these estimations of the adversaries of the original version of the anonymized records as matches. The larger the number of true matches, the higher the adversary advantage becomes, and higher privacy leakage incurs.

As an example of adversary models, Byzantine attacks represent a scenario where the adversary possesses the capability to perform arbitrary malicious actions. This includes not only passive attacks, such as eavesdropping, but also active interference, such as sending conflicting information, corrupting data, or impersonating other nodes, to undermine the reliability and consensus of a distributed system.

While Byzantine attacks encompass a broad spectrum of malicious activities, the inclusion of semidishonest behaviors within this categorization warrants careful consideration. On one hand, certain actions by semi-dishonest users—such as providing false information-mirror the deceptive aspects of Byzantine faults, potentially complicating consensus and trust within the system. On the other hand, the less malicious intent and potentially limited impact of semi-dishonest behaviors suggest that they might not always necessitate the stringent countermeasures designed for Byzantine faults.

Notably, the critical distinction lies in the intent and impact of the behaviors. Byzantine fault tolerance mechanisms are designed to counteract the most disruptive and deceptive actions that threaten system consensus and integrity. In contrast, semi-dishonest behaviors, while disruptive, often lack the malicious intent and scale of impact characteristic of Byzantine faults.

Addressing semi-dishonest behaviors requires a nuanced approach that balances the need for system security and fault tolerance with the recognition of the varied intentions behind these actions. Strategies may include:

- Adaptive Trust Models: Implementing trust models that adapt to user behaviors over time, gradually restricting or expanding access based on adherence to protocols.
- Behavioral Analysis and Anomaly Detection: Employing machine learning algorithms to identify patterns indicative of semi-dishonest behaviors, facilitating early intervention.
- Protocol Design and Incentives: Designing system protocols that are resilient to minor deviations and implementing incentive mechanisms that encourage honest participation.

The strategy does not affect our proposed method, since it works at firewall or IDS cache.

More information on our solution to combat adversary attacks is given in Chapter 6.

3.8 Conclusion

In this chapter, we explored the technical information relating to Zero Trust models, which includes an IDS and MV integration in network monitoring services. We also highlighted the importance of accuracy and security for outsourcing IDS tasks. In the following chapter, the research methodology selected for this study is detailed.

CHAPTER 4

Research methodology

4.1 Introduction

Chapter 3 highlighted the technical information and the various scenarios of Zero Trust, Intrusion Detection System (IDS), and Multi-View (MV). This chapters details the research methodology employed in this thesis. The methodology has been designed to address the research gaps identified in Chapter 2.

This chapter is organized as follows: Section 4.2 describes the research methodology; Section 4.3 explains the proposed schema, namely the Zero Trust Architecture Intrusion Detection System (ZTA-IDS) followed by an overview of the ZTA-IDS phases in Section 4.4. Section 4.5 defines the criteria of security and accuracy. Section 4.6 introduces the proposed IDS and describes each module. The implementation of ZTA-IDS is discussed in Section 4.7. The challenges and solutions are explained in Section 4.8. Finally, Section 4.9 concludes the chapter.

4.2 Selected Research Methodology

We employed the design science research methodology (DSRM) in this research to achieve the thesis objectives provided in Chapter 1. The DSRM is appropriate for validating and testing new design systems. By adopting this methodology, we can define the problem, design the solution and then evaluate the proposed solution to advance knowledge on network security [54].

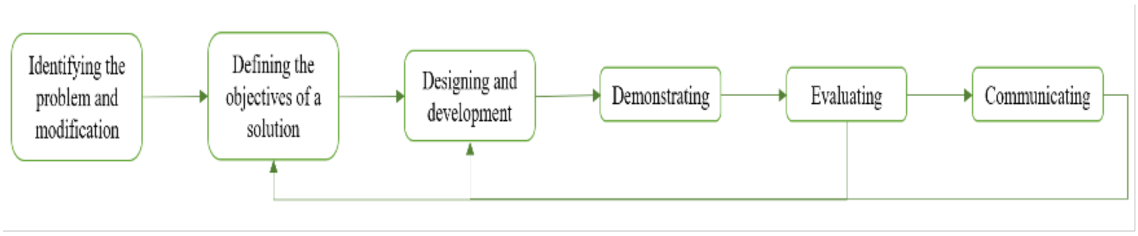

Identifying the problem and modification
Defining the objectives of a solution
Designing and development
Demonstrating
Evaluating
Communicating

Figure 4.1 : Design science research methodology (DSRM) [54]

The DSRM follows a predefined sequence of steps as shown in Figure 4.1 which begins with the identification of the research problem and the motivation as highlighted in Chapter 1, followed by a literature review to identify the research gaps in Chapter 2. Then, the design and development phases are explained in Chapters 3 and 4. An evaluation phase to assess the performance of the proposed model is demonstrated in Chapter 5 followed by the effectiveness and efficiency of it through the development of a proposed system as explained in Chapter 6. The importance of Zero Trust solution are detailed in Chapter 7. The research outcomes are communicated through publications in reputable journals and conferences.

Our contributions in this chapter have been published in the 2023 ACM International Conference and can be accessed using the following link:

1. Zero Trust Network Intrusion Detection System (NIDS) using Auto Encoder for Attention-based CNN-BiLSTM.

`https://dl.acm.org/doi/abs/10.1145/3579375.3579376`

Also, another work has been published in the 2023 International Conference on Security and can be accessed using the following links:

2. ZT-NIDS: Zero Trust-Network Intrusion Detection System, 10 Jul 2023International Conference on Security and Cryptography.

https://opus.lib.uts.edu.au/handle/10453/171227.

In addition, the extended MV approach for NID has been published in the 2023 TrustCom conference and can be accessed using the following links:

3. Zero Trust-NIDS: Extended Multi-View Approach for Network Trace Anonymization and Auto-Encoder CNN for Network Intrusion Detection, IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom).

https://ieeexplore.ieee.org/document/10063568

4.3 The Proposed Schema: ZTA-IDS

Chapter 3 overviewed the technical information on Zero Trust, MV and IDS in different scenarios. In this section, we incorporate these approaches to design our solution methodology. Figure 4.2 overviews the ZTA-IDS approach. First, we review the MV approach as explained in Chapter 3. In MV, there are eight steps, the first five steps are initiated on the data owner side and the final three steps involve the data analyst side which are described in the following sub-sections.

4.3.1 Implementing an effective Zero Trust model at the Data Owner Side

The first five steps of the MV approach which are designed and run on the data owner side are as follows:

Step 1: Data owner generates two 256 bits cryptographic keys (K_0 & K_1).

Step 2: The original data is anonymized using CryptoPAn and key K_0.

Step 3: The anonymized trace is partitioned.

Step 4: Each partition is anonymized but repeated for a different number of times at different partitions. Therefore, the seed trace is not prefix-preserved.

Step 5.1: The seed trace and supplementary parameters are outsourced to the data analyst.

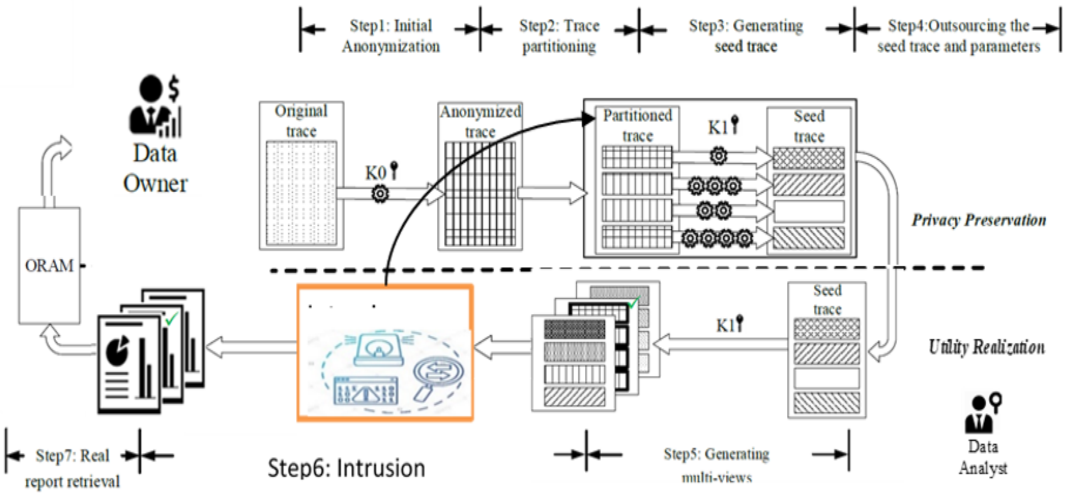

Step1: Initial Anonymization
Step2: Trace partitioning
Step3: Generating seed trace
Step4:Outsourcing the seed trace and parameters
Data Owner
ORAM
Original trace
K0
Anonymized trace
Partitioned trace
K1
Seed trace
Privacy Preservation
Utility Realization
Seed trace
K1
Data Analyst
Step7: Real report retrieval
Step6: Intrusion
Step5: Generating multi-views

Figure 4.2 : An overview of the ZTA-IDS approach

4.3.2 Accuracy Guarantee on the Data Analyst side

The remaining three steps of the MV approach, which are implemented on the data analyst side, are as follows:

Step 5.2: Analyst generates N views based on the seed view and the supplementary parameters.

Step 6: Analyst analyses all N views and generates corresponding reports.

Step 7: Data owner retrieves report corresponding to the real view using a private information retrieval (PIR) protocol so the analyst cannot identify which view was retrieved.

4.3.3 Actions on Data Owner Side (Privacy Preservation)

On the data owner side, ZTA-IDS performs in a similar way to the MV approach. The only modification is that the partitioning algorithm (mainly the parameters of the algorithm) is defined based on both security and accuracy requirements. An appropriate choice of partitioning could vary from distinct IP-based partitioning for the prefix groups of length three octets when conducting anomaly-based IDS, to distinct IP-based partitioning for the prefix groups of length only one or even less when conducting packet/flow level analyses.

4.3.4 Actions on Data Analyst Side (Intrusion Detection)

On the data analyst side, ZTA-IDS first follows the MV approach to generate a different view of the dataset. Next, the analyst preprocesses all the views to a format suitable for the operations which are about to be performed. For instance, when verifying security rules such as subnet *.*. is not allowed to communicate with subnet #.#., the analyst must apply PP the same way as they applied it to the network traces. Moreover, the analyst may need to re-order the entire trace based on time after generating the views. Prefix-to-prefix communication is another set of queries for which the analyst must run some preprocessing to reduce the impact of anonymization. The data analyst will then run the IDS algorithms on each of the views (together with their unique security rules) and return the outcomes to the data owner.

Our proposed method utilises the CryptoPAn model, which gives a baseline for prefix-preserving anonymization. The advantage of CryptoPAn is that it is deterministic and allows consistent prefix-preserving anonymization under the same K (key of the CryptoPAn anonymization). This model gives a baseline for prefix-preserving anonymization. However, as previously discussed, it is vulnerable to semantic attacks. Consequently, a more advanced methodology for anonymizing data is needed.

4.4 ZTA-IDS phases

Our proposed ZTA-IDS method is shown in Figure 4.3 which consists of three phases including Extended MV, IDS and Evaluating Accuracy and Security which are described in detail in the following. In this chapter, the challenges of the first two phases are discussed in Chapter 4 up to Chapter 6 and phase 3 which outlines the validation, is covered in Chapter 5.

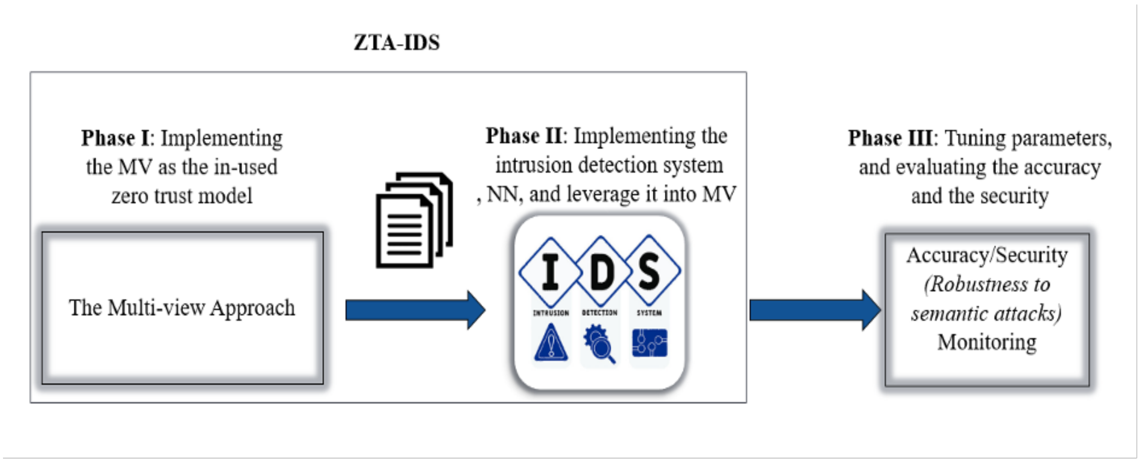

ZTA-IDS
Phase I: Implementing the MV as the in-used zero trust model
The Multi-view Approach
Phase II: Implementing the intrusion detection system , NN, and leverage it into MV
I D S
INTRUSION DETECTION SYSTEM
Phase III: Tuning parameters, and evaluating the accuracy and the security
Accuracy/Security *(Robustness to semantic attacks)* Monitoring

Figure 4.3 : Various phases of ZTA-IDS

4.5 Criteria on Security and Accuracy

Clearly, the characteristics of the framework, e.g., partitioning algorithm, number of views, etc., rely on IDS tasks. If the IDS task is independent of the prefix relation of the IP addresses, e.g., counting the number of packets of size larger than 300KB, then we can strengthen the confidentiality aspect of the solution arbitrarily. On the other hand, if the IDS task depends on the trustworthiness of the prefix relations, an appropriate degree of MV must be applied. Interpreting one IDS task's requirements is one of the main contributions of our ZTA-IDS framework. For an NN-IDS, depending on the type of intrusion, the learning module may require fingerprinting a larger number of attributes to predict malicious activities from benign ones. Therefore, the partitioning algorithm, the number of partitions, and the number of attributes involved in the partitioning must be carefully selected to guarantee a maximal level of accuracy for an NN-IDS. These are the most important parts of our experiments on which we elaborate in the next chapters.

4.6 The Proposed Intrusion Detection Method

As network security monitoring gradually becomes increasingly challenging, there is a growing demand to outsource these kinds of tasks to third-party analysts. Organizations are often reluctant to share their network traces due to a lack of trust in the analysts due to concerns over sensitive information appearing in the traces, e.g., compromising network and system configuration, which may potentially be used for malicious attacks. Networks are vulnerable to data security breaches and unauthorized access, due to their “Implicit Trust” or “Trust but Verify” characteristics. These days, the cloud is the new network edge, and it cannot be structured under this traditional implicit trust model. Implicit trust networks do not work in a climate where network ‘edges’ have broken down and disintegrated. Such security breaches result in breaching confidential databases, with the number of breaches growing both in number and severity. Therefore, the current trend of security monitoring with a “Trust but Verify” method is not sufficient in any approach when third parties are involved. In this thesis, as previously discussed, we address the challenges when a data owner outsources their private network traces to a third-party IDS to monitor and detect any suspicious incident in real-time or retroactively. This method assumes that networks are segmented, and that data center architecture is able to create a boundary or “demilitarized zone” between trusted and untrusted portions of networks [55].

The architecture of the proposed method is shown in Figure 4.4. It comprises four modules namely auto-encoder (AE) feature extractor, convolution blocks, attention mechanism and long short-term memory (LSTM) layers. The proposed approach is inspired by the auto-encoder convolutional neural network (AE-CNN) for binary IDS. The effectiveness of AE-CNN for binary intrusion detection has been detailed in the related works [53]. Thus, we use the extracted features as the input to our proposed neural network for categorical intrusion detection. We also utilize the attention module to focus on more important features, as well as LSTM layers to handle the temporal dynamics.

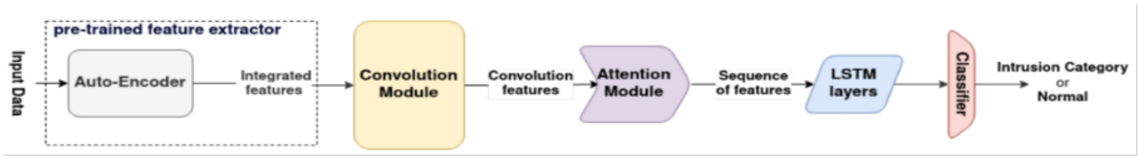


Figure 4.4 : An overview of the proposed method for an IDS

4.6.1 Auto-Encoder Feature Extraction

Network trace-related bottleneck features are extracted via a deep AE. The bottleneck layer in AEs includes the latent variables which forces a compressed knowledge representation of the original input. An AE consists of two components: the encoder, and the decoder as well as a loss function to compare the output with the ground truth (target). The encoder's output part is named bottleneck and can be used as compressed features, and it represents a compressed view of the knowledge of the input. Therefore, a deep auto-encoder is used to extract the combined and compressed features from network trace attributes. The compressed bottleneck features of network traces are extracted using a pre-trained deep AE. The bottleneck layer in the AE maps the original input into a compressed representation where the input features are much more correlated.

In AEs, both the encoder and decoder contain fully connected layers, where the activation for each layer can be different. The encoder function ψ maps the original data X to a latent space z, which is called the bottleneck. The decoder function ϕ maps the latent space z to the output $\widehat{X}$, where the output is expected to be the same as the input. An AE includes 2 components: the encoder which compress input features, and the decoder which is discarded after pre-training. As a result, a deep AE can be used to extract a merged and compressed feature from network trace attributes. In the AE, the bottleneck feature z is extracted using the encoder function from the original data X. The decoder function maps bottleneck z to output $\widehat{X}$ the decoder reconstructs the input as shown in Equation 4.1.

$$\begin{gathered}\psi = X \Rightarrow z \\ \phi = z \Longrightarrow \widehat{X} \\ \psi, \phi = argmin\|\widehat{X} - (\phi(\psi(X)))\|^2\end{gathered} \tag{4.1}$$

In other words, the mean squared error (MSE) loss function of the AE is calculated using Equation 4.2.

$$L(X, \grave{X}) = \|X - \widehat{X}\|^2 \tag{4.2}$$

where X-$\widehat{X}$ is usually averaged over a mini-batch input training set. W, W_0 are weight matrices and b, b_0 are bias vectors for the encoder and decoder, respectively. Bias is not used for the encoder part to aggregate the input feature only. The encoder and decoder parameters can be trained independently, or the decoder parameters can be considered as encoder transposed parameters. In this thesis, these parts are examined independently and the whole network is trained at once. Since we aim at aggregating the input features into the bottleneck layer, our encoder part does not have any bias parameter.

The structure of the AE is shown in Figure 4.5, where the dotted lines are discarded after training the AE. The bottleneck features of the trained AE, which are more spatially related than raw features, are used as input to CNN-LSTM. We use the pre-trained AE as the input features for both methods.

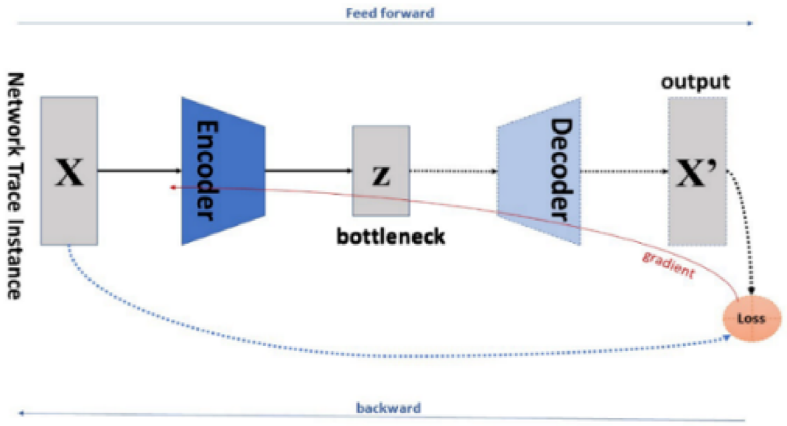


Figure 4.5 : AE feature extractor for network traces

Since AE with a bottleneck layer accepts any numerical value and compresses the information available in the input numerical values, pre-processing and feature selection is not needed. The bottleneck features of the trained auto-encoder, which are spatially related, are exploited as input to the CNN classifier. The structure of the designed auto-encoder is shown in Table 4.1.

Table 4.1 : Auto-encoder structure for network trace feature extraction

Part	Layer	Size
Input	-	196
Encoder	Linear + Sigmoid	128
	Linear + Sigmoid (Bottleneck)	64
Decoder	Linear + Sigmoid	128
	Linear + Sigmoid	196

4.6.2 Convolutional Neural Network (CNN)

We introduce a model to use a CNN to consider the spatially related features extracted using the AE. A CNN classifier has been applied on a bottleneck feature extracted from a trained AE for NID because CNNs work well with data that have a spatial relationship. The CNNs are also known to be good feature extractors because of local convolution filters, repetitive filters among the whole input data, and pooling layers which makes it robust. In this thesis, a 1D CNN was tuned to handle spatial dependencies within traces of data. Our CNN structure is presented in Figure 5.6 and LeakyReLU with a 0.2 negative slope is considered to be an activation function for hidden layers. In the convolution layers, the first number is the number of the filters and the number in parentheses is the convolution filter size, e.g., the first layer has 128 filters, where 11 is the convolution filter size. A pooling with a size of 2 is only used in the first convolution layer.

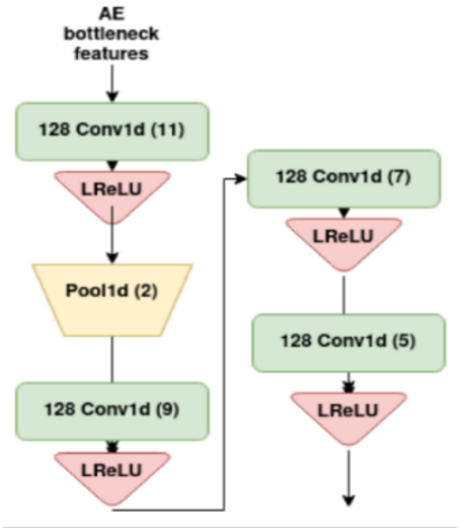


Figure 4.6 : The proposed CNN to handle the spatial dependencies of network traces

In the output of the CNN layer, we have 128*5 features, which its knowledge needs to be combined, since it has a high dimension of features to feed into any regular (non-convolutional) layer.

Our CNN structure is illustrated in Table 4.2 and LeakyReLU with 0.2 negative slope is considered an activation function for hidden layers. Obviously, softmax is applied in the output layer which outputs the classifier results.

Table 4.2 : CNN structure for network trace feature extraction

Layer	#Filters	Filter Size	Output Shape
Input	1	-	64
Convolution	128	11	54
Pooling	-	2	27
Convolution	128	9	19
Convolution	128	7	13
Convolution	128	5	9
Linear	512	-	512
Linear	128	-	128
Linear	64	-	64
Linear (Classifier)	2	-	2

4.6.3 Attention Module

We used a multi-head self-attention module to merge the information available in the extracted features and handle the relation between the CNN features and LSTM components (subsequent layers). The attention module dimension is the number of channels from the last CNN layer (256) and uses 8 heads. The output is mapped into 64 dimensions to limit the features. The attention learns to focus on intrusion-related features.

The alternative structure for multi-head self-attention on top of CNN would be a linear flatten layer, which maps the CNN multi-dimensional features into one large dimension. The total number of features (neurons) in this layer is the same as the total number of CNN features in all dimensions. We report the results of the proposed approach with a linear flatten layer instead of the attention mechanism.

Finally, we also propose to use BiLSTM layers after the attention mechanism to handle the temporal dynamics between the sequences of network traces.

4.6.4 BiLSTM Classifier

Since LSTMs can hold/forget information for a long time, we use LSTMs to handle the temporal dynamics. Also, BiLSTM is able to take forward and backward sequences into consideration, which can be important in handling temporal dynamics. We use two BiLSTM layers with 128-dimensional representations. A dropout with a probability of 0.2 is applied between the two layers of BiLSTM. Finally, a linear layer with the same number of neurons (from 2 to 10) as the target groups is applied.

4.7 ZTA-IDS implementation

We selected a deep learning-based approach for our IDS. We used the existing implementation from [37] and customized it for deployment in the MV framework. This implementation is specifically designed for analyzing the well-known KDDcup99 dataset which consists of 500K “good” and “bad” network connections simulated in a military network environment. The datasets contain a total of 24 training attack types (bad connections), with an additional 14 types in the test data only (refer to Table 4.3 for the attacks). This implementation employs a relatively simple neural network with two hidden layers. The objective is to predict what type of attack is underway. The outcome column specifies the attack type. A value of “normal” shows that no attacks are underway. The outcome is quite accurate (around 99%). The types of attacks in UNSW-NB15 dataset are listed in Table 4.4.

Table 4.3 : Progress made on the ZTA-IDS implementation on KDDcup99

IDS		MV	
Environment	Python 3.8	Environment	Python 3.8
Methodology	Training Neural Network	Operations	Encryption & Partitioning
Types of Intrusions	24 training attack types including: · DoS: Denial-of-Service. · R2L: unauthorized access from a remote machine, e.g., guessing password. · U2R: unauthorized access to local superuser (root) privileges, e.g., various "buffer overflow" attacks. · Probing: surveillance and other probing, e.g., port scanning.	Tasks & Dataset	1. Implementing PP (-, K) 2. Partitioning 3. Random Vector Generation 4. View Generation 5. Pre-process for IDS 6. Dataset: KDDcup99

Table 4.4 : Progress made on the ZTA-IDS implementation on UNSW-NB15

IDS		MV	
Environment	Python 3.8	Environment	Python 3.8
Methodology	Training Neural Network	Operations	Encryption & Partitioning
Types of Intrusions	6 training attack types including: · DoS · Reconnaissance · Exploits, · Fuzzers · Generic · Normal	Tasks & Dataset	1. Implementing PP (-, K) for IP and none-IP values 2. Partitioning: compare number of instances for different partitioning methods 3. Random Vector Generation 4. View Generation 5. Pre-process for IDS 6. Customizing MV to deal with dataset: UNSW-NB15
Tasks	1. Analyse data structures 2. Pre-process data types3. 3. Train auto-encoder for feature aggregation. 4. Train CNN-Attention classifier.		

We implemented MV using the Cryptopan package. A Coefficient of Variation (CV) analysis of the data revealed that partitioning method 1 which divides all IP addresses based on a number (N) is not balanced, so we propose a new partitioning method based on percentiles which is more balanced. Additionally, we extended Cryptopan to consider other attributes (e.g., ports) in the construction of MVs. So, the partitioning method should be extended to include other anonymized attributes.

We proposed a new deep neural network-based IDS which is independent from the number of attributes due to the auto-encoder feature aggregator. We implemented this method on the UNSWNB15 dataset. In its clean format, the UNSW-NB15 includes 42 features. Of these 42 features, three instances are non-numeric (categorical) attributes and 39 are numeric in nature. The UNSW-NB15 is subdivided into these primary datasets: UNSW-NB15 TRAIN, which is applied for training different models and the UNSW-NB15-TEST (100%) which is employed for testing the trained models. The UNSW-NB15 includes instances with these network attack categories: Backdoor, Reconnaissance, Shellcode, Fuzzers, DoS, Generic, Worms, Analysis, Shellcode and Exploits [41].

Our proposed method includes a deep AE-CNN for NID. Although CNNs and recurrent neural networks (RNNs) have been proposed for NID [42, 44, 50], CNNs work well with data that has a spatial relationship (e.g. images) [51], and RNNs work well when predicting sequences of the same source (e.g. speech) [52, 53]. However, network trace attributes have different meanings (e.g., port, IP, etc.) and structures (Boolean, string, numerical) because they are not spatially related. On the other hand, since they may not come sequentially from the same source, they may not be related sequentially. Consequently, using a method that can combine and compress attributes in a record of data can be beneficial.

We exploit a deep AE to extract a merged and compressed feature from the network trace attributes. Then, we use these integrated features for classification, using CNNs. Since these features are higher-level features and continue, we can consider them to be spatially related.

4.8 Challenges and Solutions

This section reviews the challenges in implementing MV and IDS which our solutions address.

4.8.1 Challenges of phase 1

Table 4.5 outlines the challenges encountered in the implementation of MV and our solutions to tackle them.

Table 4.5 : MV challenges in UNSW-NB15 dataset

	MV
Environment	Python 3.8
Operations	Encryption & Partitioning
Tasks & Dataset	1. Implementing PP(:,K) for IP and non-IP values 2. Partitioning a: compare number of instances for different partitioning methods 3. Random Vector Generation 4. View Generation 5. Preprocessing for IDS 6. Customizing MV to deal with Dataset: UNSW-NB15

Multi-View

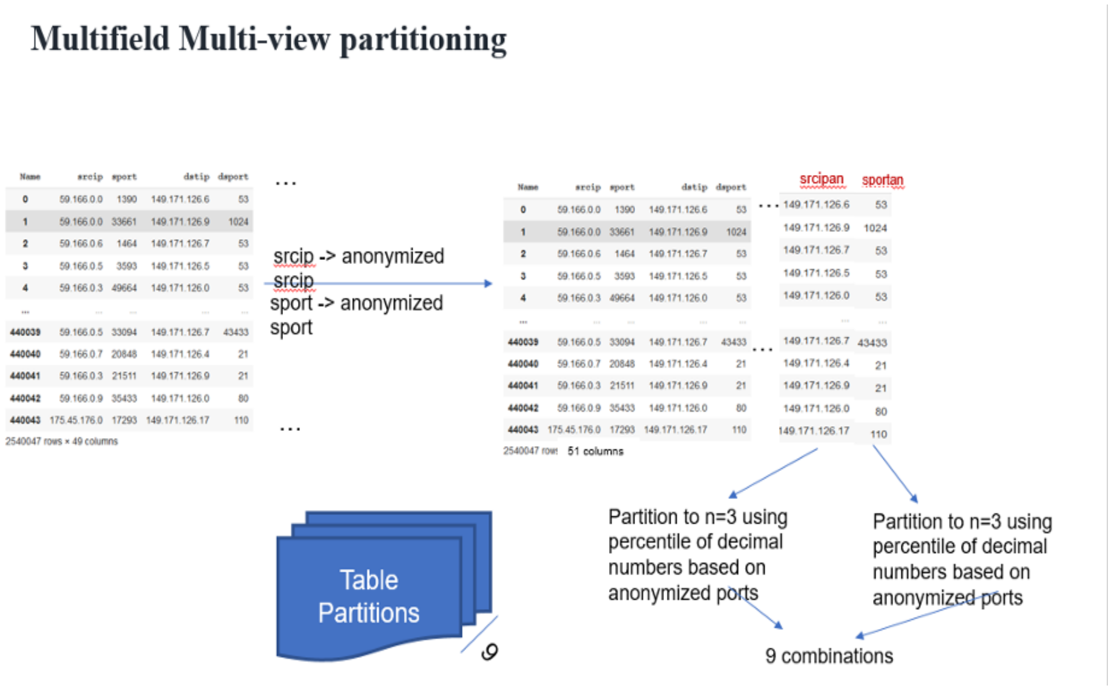
To implement the MV approach to outsource the intrusion detection task, we encountered the following challenges, and we propose solutions for each one.

1. Missing data: Datasets may contain some missing values which causes problems during the implementation. We removed the data records where there are missing values in any important field.

2. Non-IP attributes anonymization: The MV uses prefix-preserving IP anonymization which is not implemented for non-IP values. Therefore, we implemented prefix-preserving anonymization for non-IP values.

3. Unbalanced data partitioning: The number of data records in the partitioning method that partitions whole the range of IPs based on a number N are not balanced. We used N-quantiles instead of only splitting whole valid IP addresses.

4. IDS input: MV generated data needs pre-processing for feeding to IDS, as shown in Figure 4.7.



Multifield Multi-view partitioning

Name	srcip	sport	dstip	dsport
0	59.166.0.0	1390	149.171.126.6	53
1	59.166.0.0	33661	149.171.126.9	1024
2	59.166.0.6	1464	149.171.126.7	53
3	59.166.0.5	3593	149.171.126.5	53
4	59.166.0.3	49664	149.171.126.0	53
...	...	...	...	...
440039	59.166.0.5	33094	149.171.126.7	43433
440040	59.166.0.7	20848	149.171.126.4	21
440041	59.166.0.3	21511	149.171.126.9	21
440042	59.166.0.9	35433	149.171.126.0	80
440043	175.45.176.0	17293	149.171.126.17	110

2540047 rows × 49 columns

srcip -> anonymized srcip
sport -> anonymized sport

Name	srcip	sport	dstip	dsport	srcipan	sportan
0	59.166.0.0	1390	149.171.126.6	53	149.171.126.6	53
1	59.166.0.0	33661	149.171.126.9	1024	149.171.126.9	1024
2	59.166.0.6	1464	149.171.126.7	53	149.171.126.7	53
3	59.166.0.5	3593	149.171.126.5	53	149.171.126.5	53
4	59.166.0.3	49664	149.171.126.0	53	149.171.126.0	53
...	...	...	...	...	...	...
440039	59.166.0.5	33094	149.171.126.7	43433	149.171.126.7	43433
440040	59.166.0.7	20848	149.171.126.4	21	149.171.126.4	21
440041	59.166.0.3	21511	149.171.126.9	21	149.171.126.9	21
440042	59.166.0.9	35433	149.171.126.0	80	149.171.126.0	80
440043	175.45.176.0	17293	149.171.126.17	110	149.171.126.17	110

2540047 row 51 columns

Partition to n=3 using percentile of decimal numbers based on anonymized ports

Partition to n=3 using percentile of decimal numbers based on anonymized ports

9 combinations

Figure 4.7 : Multi-fields partitioning for Modified MV

4.8.2 Phase 2 challenges and solutions

The first challenge we encountered in implementing phase 2 is non-equivalent input data types. As a result, they should be converted to numerical values (using one-hot vector, IP to integer conversion, etc.) in the pre-processing step.

The next challenge is the different types and concepts of input data attributes, which makes applying CNN meaningless. Consequently, we implemented an auto-encoder for feature aggregation. Furthermore, we implemented and trained the CNN-classifier for IDS using auto-encoder bottleneck features.

To deploy the ZTA-IDS, we need to leverage this IDS into the MV framework and analyze the impact of different views in terms of both security (how susceptible are the generated views to de-anonymization), and the accuracy of the prediction for the real view. To this end, we need to implement (1) the MV approach, and (2) state-of-the-art semantic attacks (fingerprinting and injection). At this stage, we have successfully implemented the MV approach and also extended it to take other attributes into consideration. We feed them to NN-IDS. Also, we need to simulate the semantic attack to customize the MV parameters and evaluate both accuracy and security, as shown in Table 4.6.

Table 4.6 : IDS part of ZTA-IDS implementation on UNSW-NB15

Methodology	Training Neural Network (Binary)	Training Neural Network (Multi-Class)
Tasks	1. Analyze data structures 2. Pre-process data types 3. Train auto-encoder for features aggregation 4. Train CNN binary classifier using AE features	1. Analyze data structures 2. Pre-process data types 3.Train auto-encoder for feature aggregation 4. Analyze multi-class labels 5. Pre-process multi-class labels 6. Train CNN multi-class classifier using AE features 7. Analyze the sequence relation of the sequences 8. Train CNN-LSTM binary/multi-class classifier using AE features

To implement the multiclass IDS, as detailed in Table 5.6, we first analyze the data according to new targets, because they are imbalanced, and it is better to ensure they are balanced. So, a different type of pre-processing is needed for this purpose. We train the CNNs as in the binary scenario but for multi-classes and we analyze the relationship between sequences and preprocess it to use a proper RNN (LSTM) for multiclass IDS. Finally, we merge these networks to achieve the state-of-the-art method using an attention mechanism.

4.8.3 Phase 3 challenges and solutions

Phase 3 is the attack simulation part of the ZTA-IDS implementation on UNSW-NB15 where we monitor the accuracy and security of the proposed solution and check the robustness of the model against semantic attacks. Chapter 6 provides further details. An example of the main operations of this phase is shown in Table 4.7.

Table 4.7 : Attack simulation part of ZTA-IDS implementation on UNSW-NB15

Environment	Python 3.8
Operations	Evaluate phase 1 and 2 integration using real-world tests
Attack types	DoS, Reconnaissance, Normal, Exploits, Fuzzers, Generic
Tasks	1. A simple website implementation on a server 2. Semantic attacks simulation 3. Capture any request 4. Apply MV on results 5. Apply IDS on requests and report the results 6. Improve IDS accuracy

4.9 Conclusion

This chapter highlights the research methodology employed in our research. The methodology is designed to address the research gaps identified in Chapter 2. We first selected our research methodology. Then, we explained the proposed ZTA-IDS schema followed by the phases of ZTA-IDS. We also defined the criteria on security and accuracy. We introduced our proposed IDS and described each module in it. Furthermore, we explained the implementation of ZTA-IDS in detail. Finally, the challenges and solutions are listed and explained. The following chapter highlights the experimental performance and data analyses.

CHAPTER 5

ZTA-IDS Model Validation

5.1 Introduction

The previous chapter covers the methodology and challenges encountered in phase I and phase II in the implementation of our solution, Zero Trust Architecture Intrusion Detection System (ZTA-IDS).

In this chapter, we focus our attention on adjusting the parameters, assessing the accuracy, and evaluating the security of ZTA-IDS, particularly its resilience to semantic attacks in monitoring scenarios. Phase III involves a detailed assessment of the security and precision of our proposed ZTA-IDS system, with a focus on its defence against semantic attacks. This phase includes creating simulations that closely mimic these attacks, capturing each attack instance and accurately classifying them to establish a reliable ground truth. We conducted these simulations at different intervals, carefully documenting each request and its corresponding true classification. While the UNSW-NB15 dataset provides ten categories of attacks, our replication of certain attack types faced some challenges. Table 5.1 provides a detailed overview of the methodology used. The initial step involves setting up a basic yet secure website on a server. After simulating semantic attacks, we collect all requests on a storage device. These are then processed before being analysed with multi-view (MV) and IDS techniques to improve IDS accuracy. A comprehensive

Table 5.1 : Zero-Trust security and accuracy

Zero-Trust solution	
Environment	Python 3.8
Operations	Evaluate phase I and II Integration by real-world tests
Attack types	• Denial of Service (DoS) • Reconnaissance • Normal • Exploits • Fuzzers • Generic
Tasks	1. A sample website implementation on a server 2. Semantic attacks simulation 3. Capture any request 4. Apply MV on requests 5. Apply IDS on requests and report the results 6. Improve IDS accuracy
Status	Complete

discussion of the results and the evaluation are presented in detail in this chapter.

The dynamic nature of cybersecurity calls for the development of intricate and versatile solutions to combat the diverse array of threats that modern network environments encounter. This chapter contributes to this ongoing endeavour by introducing a novel method that integrates MV anonymization techniques with advanced neural network models, all within the framework of Zero Trust.

Our proposed method offers a comprehensive approach to network security. By combining the privacy-preserving capabilities of MV with the intrusion detection prowess of advanced neural network models, we have crafted a solution that not only aligns with the Zero Trust model but also addresses the dual concerns of privacy and accuracy in intrusion detection.

The experiments and results presented in the subsequent sections provide a rigorous evaluation of our method (implementation available on GitHub: `https://github.com/AbZ1221/ZTA-IDS`). While the initial results using the UNSW-NB15 dataset were promising, the real test of our system's capabilities came from its performance on real-world data. The significant degradation in accuracy when transitioning from the dataset to real-world data highlighted the challenges of generalization in machine learning models, especially in the context of cybersecurity. However, the adaptation of our model using real-world data underscores the potential of our approach. With further refinement and the incorporation of more real-world-like data, we believe our method can achieve even greater efficacy.

Several key takeaways emerge from the points mentioned in this chapter:

- Overall ZTA-IDS Approach: The thesis emphasizes the implementation and effectiveness of a ZTA combined with an IDS, highlighting its critical role in network security.

- Integration of MV and IDS: A significant focus is on the integration of MV analysis with IDS. This approach is shown to provide a robust solution for network security, balancing the need for privacy preservation with the necessity of accurate intrusion detection.

- Importance of Real-world Data Performance: The thesis acknowledges the value of datasets like UNSW-NB15 for training purposes. However, it underscores that the true measure of an IDS's effectiveness is its performance on real-world data, beyond controlled or simulated environments.

- Challenges of Generalization in Machine Learning Models: A major challenge addressed in the thesis is the generalization issue in machine learning models. The research demonstrates how adapting these models using real-world data can lead to improved accuracy, thus enhancing the IDS's effectiveness.

- Need for Continuous Monitoring in Cybersecurity: The dynamic and evolving nature of cybersecurity threats is highlighted. The thesis stresses the importance of continuous monitoring, adaptation, and refinement of IDS to effectively counter these threats and maintain robust network security.

5.2 The overall Zero Trust Architecture Intrusion Detection System (ZTA-IDS)

The comprehensive ZTA-IDS framework mandates that each server request must traverse a screening gateway designed to eliminate malicious inquiries. This gateway needs regular updates from the IDS, which can operate on a distinct server or be managed externally. Before requests are processed by the IDS, they undergo a MV analysis. For a more in-depth understanding of the ZTA-IDS, refer to Figure 5.1.

5.2.1 Proposed Zero-Trust (ZTA-IDS) model

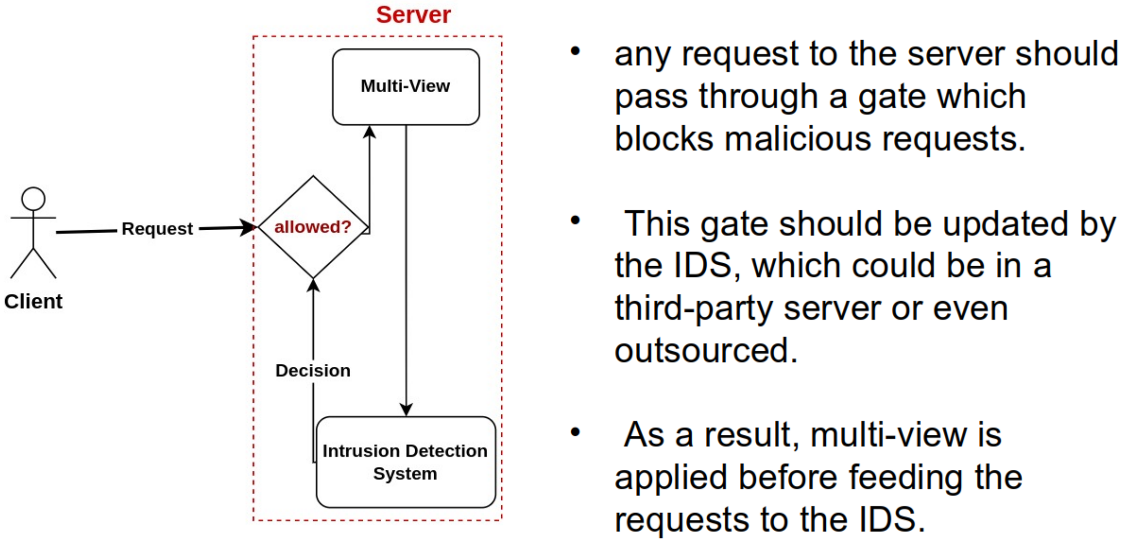
- MV and IDS combination

- Integration with firewall or IPS

- Integrated system validation

Integrating IP anonymization and IDS

IP anonymization is a method used to hide or mask real IP addresses, which can help protect the identities of users or devices on a network. The anonymized IPs can be sent to IDS for the analyses without much concerns.



- any request to the server should pass through a gate which blocks malicious requests.
- This gate should be updated by the IDS, which could be in a third-party server or even outsourced.
- As a result, multi-view is applied before feeding the requests to the IDS.

Figure 5.1 : The overall ZTA-IDS

To implement an IDS using neural networks, an IDS is used to detect malicious activities or policy violations in a network. In a Zero Trust model, this could serve as a continuous monitoring and evaluation mechanism to ensure that every access request is legitimate and safe. By integrating these two methods, it is possible to create a more robust and secure system that aligns with the principles of Zero Trust - namely, that every access request is treated as potentially risky until it has been thoroughly verified. A Zero Trust model could work as follows: every time an access request is made, the IP address is anonymized to protect the user/device identity. Then, the IDS evaluates the request to detect any signs of intrusion or malicious activity. Only if the request is deemed safe is access granted. Data received from the IDS tool should be analysed. If it determines that the traffic is likely an intrusion, it sends a signal to block the traffic.

To block traffic, the system needs to be integrated with a firewall or IPS. When the neural network model sends a signal to block traffic, the firewall or IPS should respond by blocking the IP address or otherwise, preventing the traffic from accessing the system.

As previously mentioned, MV is a method used to hide or mask real IP addresses, which can help protect the identities of users or devices on a network. An IDS is used to detect malicious activities or policy violations in a network. In a Zero Trust model, this could serve as a continuous monitoring and evaluation mechanism to ensure that every access request is legitimate and safe. By integrating these two methods, it is possible to create a more robust and secure system that aligns with the principles of Zero Trust - namely, that every access request is treated as potentially risky until it has been thoroughly verified. A Zero Trust model could work like this:

Encoding and hiding IP addresses: Every time an access request is made, the IP address is encoded and hidden using the proposed MV method to protect the user's/device's identity. This ensures that even if an attacker gains access to network traffic, the real IP addresses remain concealed, preserving privacy.

Intrusion detection evaluation: The IDS evaluates the request using a combination of autoencoder for attention-based CNN-BiLSTM. This advanced detection mechanism analyses the patterns and behaviours within the traffic to detect any signs of intrusion or malicious activity.

Access decision: Access is granted only if the request is deemed safe. The decision is made based on a comprehensive analysis that includes not only the encoded IP but also other factors like user credentials, request patterns, and historical data.

Intrusion response: If the IDS determines that the traffic is likely an intrusion, it sends a signal to block the traffic. This immediate response ensures that potential threats are neutralized before they can cause harm.

Continuous monitoring and adaptation: The system continuously monitors network traffic and adapts to new threats and patterns. This includes updating the models used for intrusion detection and adjusting the MV encoding techniques as needed.

Real-world simulation: To validate the effectiveness of the proposed method, attacks were simulated using a server as a client and multiple servers as targets. This real-world scenario provided valuable insights into how the system would perform under actual attack conditions.

Zero Trust compliance: The integration of MV and IDS within the Zero Trust framework ensures that the system adheres to the core principles of never trusting and always verifying. This alignment with Zero Trust principles not only enhances security but also ensures that both privacy and accuracy are maintained.

By implementing this proposed method, organizations can build a more resilient network security system that aligns with the Zero Trust model. The combination of MV for IP address encoding and advanced intrusion detection techniques provides a comprehensive solution that addresses both privacy concerns and the need for accurate intrusion detection. The real-world simulation further validates the effectiveness of the method, demonstrating its potential for practical application in today's complex and dynamic network environments.

Figure 5.2 illustrates the proposed Zero Trust model with MV and IDS integration.

Integration with Firewall or IPS

In the preceding sections, we explored the synergistic potential of combining the MV approaches with IDS to enhance network security. This combination is pivotal in safeguarding user and device identities within a network and establishing a robust mechanism to detect and respond to malicious activities and policy violations. Building upon this foundation, it becomes imperative to consider how our proposed method could integrate effectively with firewalls and intrusion prevention systems (IPS) to further bolster network defenses.

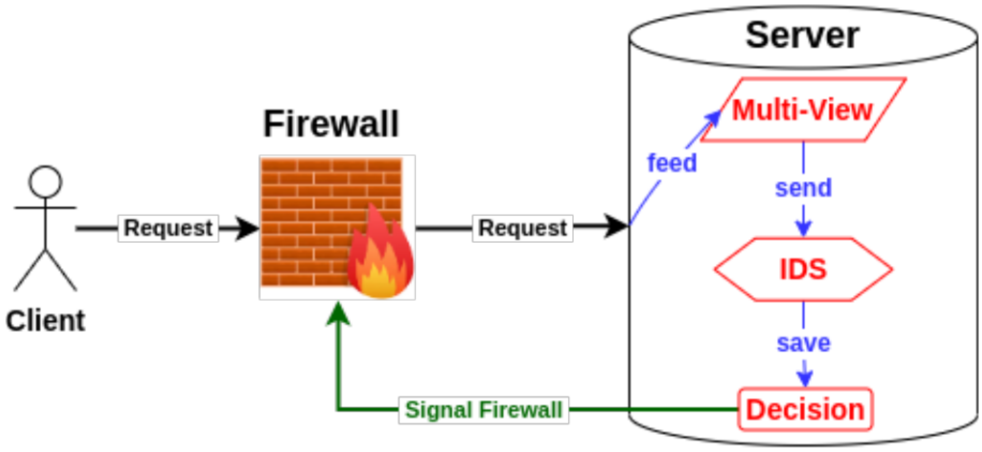


Figure 5.2 : Overview of the proposed Zero Trust model with MV and IDS integration

Firewalls, serving as a crucial barrier between trusted and untrusted networks, regulate traffic based on predefined security rules. They act as the first line of defense in filtering incoming and outgoing network traffic, ensuring that only legitimate data flows are permitted. The integration of our proposed method with firewall technologies enables a more dynamic and intelligent traffic filtering process. This not only enhances the capability to block malicious traffic but to also ensure a seamless flow of legitimate data, thereby maintaining network integrity and performance.

Similarly, the role of IPS extends beyond mere detection, as seen in traditional IDS. IPS are designed to actively prevent and mitigate identified threats, offering a proactive stance in network security. The amalgamation of our proposed method with IPS facilitates a more responsive and adaptive security posture. It empowers the system to not only identify potential threats through advanced detection mechanisms but also to take immediate and effective actions to thwart these threats, thus providing an additional layer of security.

While the integration with firewalls and IPS is critical to achieving a comprehensive security solution, it is essential to delineate the focus of our research. Our study primarily emphasizes the novel methodologies and unique contributions of the proposed method in enhancing network security through the innovative use of MV techniques and advanced IDS capabilities. The specific implementation details pertaining to the integration with firewall or IPS technologies, though important, are not the central focus of this research. Our aim is to lay down the theoretical and conceptual framework that demonstrates the effectiveness and applicability of our method in the realm of network security.

It is also crucial to acknowledge the scope of this study. The practical aspects of implementing the proposed method in conjunction with firewall or IPS technologies, while indispensable for a holistic security approach, fall outside the immediate purview of this research. This study is designed to provide a theoretical foundation and a proof of concept for the proposed methodologies. Future research could build upon this groundwork, exploring the practicalities of integrating these advanced security mechanisms into a cohesive and robust network defense system.

In conclusion, our research underscores the importance of integrating our proposed method with firewall and IPS technologies as part of a comprehensive security solution. However, the primary focus remains on the innovative aspects and unique contributions of the method itself. The practical implementation of such integrations, crucial as they are, is an area ripe for future exploration and falls beyond the scope of this current study.

Integrated System Validation

To rigorously assess the effectiveness of the proposed IDS, it is essential to validate its performance in real-world conditions. This involves a two-step process: collecting information for normal requests and simulated attacks, and then evaluating the accuracy of the IDS on these real-world records. All procedures including data collection, feature extraction and validation are shown in Figure 5.3.

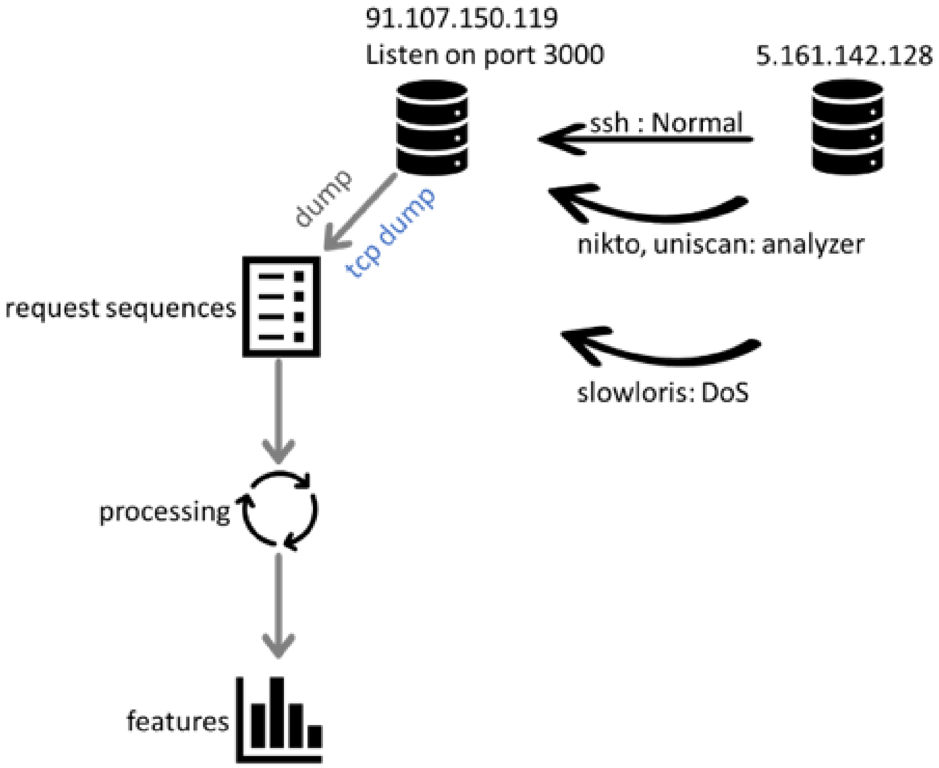


Figure 5.3 : Real-world data validation procedure

In the following, we describe this process in detail.

5.3 Data Collection

Data collection includes two main parts:

1. Normal Requests: Gathering information for normal requests is crucial to understand how the system behaves under regular operating conditions. This includes monitoring and recording legitimate user activities, network traffic, and system interactions.

2. Simulated Attacks: In addition to normal requests, we must also simulate various types of attacks, such as DoS, Normal, and Reconnaissance, as previously discussed. This will provide insights into how the system responds to malicious activities.

5.3.1 Attack Simulation Tools

The tools and applications we used to simulate the attacks are: nmap (a network scanner), masscan (a fast TCP port scanner), nessus (a vulnerability assessment tool), nikto (a web server scanner), uniscan (a web vulnerability scanner), and slowloris (a low-bandwidth denial-of-service attack tool). The tools and the attack types are listed in Table 5.2.

We used some of these tools to simulate some types of attack.

5.3.2 Normal and Attack Data Collection Procedure

In this crucial phase of our research, we focused on the collection of both normal and attack data, which forms the backbone of our analysis and subsequent findings. This data collection was meticulously executed by setting up a controlled environment using a virtual private server (VPS) and implementing various tools to capture and analyze network traffic.

Table 5.2 : Types of attacks and the tools

Attack Type	Tools
Fuzzers	AFL, Peach Fuzzer
Analysis	Nikto, Uniscan
Backdoors	BeEF, Netcatm Shellter, Netcat
DoS	Slowloris, HpING
Exploits	Metasploit Framework
Generic	Nmap, Nessus
Reconnaissance	Nikto, Uniscan
Shellcode	Shellnoob, Shellsploit
Worms	-

Server Setup and Implementation

Our primary server setup involved the deployment of a Gitea website on a Hetzner VPS. Gitea, an open-source forge software package, was chosen for its lightweight framework and ease of cloning for multiple instances. The website hosted on this VPS served as the central point for our attack simulations and data collection. The setup is graphically represented in an accompanying image, which illustrates the structure and workflow of our server configuration.

Expansion to Multiple Servers

To broaden the scope and enrich the dataset, we procured additional servers from Hetzner. Each of these servers was identically configured, with the Gitea web service cloned onto them. This replication was strategic, ensuring consistency in the environment across different servers, thereby enabling a more comprehensive data collection process.

Data Collection and Recording Tools

For the purpose of capturing every incoming and outgoing request to our servers, we utilized "tcpdump,' a robust tool that allowed us to record all network traffic. Tcpdump's ability to capture packet data made it an ideal choice for our needs, ensuring that we had a detailed record of all interactions with the server, both under normal and attack conditions.

Data Analysis

Further analysis of the captured data was conducted using Wireshark, a network protocol analyzer renowned for its detailed data dissection and presentation capabilities. Wireshark enabled us to delve deeper into the nuances of the network traffic, providing us with valuable insights into the patterns and characteristics of both normal and malicious requests. This level of analysis was instrumental in distinguishing between typical network behavior and potential security threats.

Data Collection: Normal Requests

For the collection of normal request data, we employed two primary methods. First, regular SSH (secure shell) connections were established and utilized. SSH, known for its secure data communication capabilities, provided us with a realistic scenario of normal administrative activities that a server typically experiences. These activities included routine server maintenance tasks, file transfers, and system updates, all of which represent legitimate, non-malicious traffic.

Secondly, we simulated normal web browsing behavior. This involved accessing the Gitea web service hosted on our servers using standard web browsers, mimicking the actions of genuine users. This type of traffic is essential to establish a baseline of 'normal' behavior against which potentially malicious activities can be compared.

Data Collection: Attack Simulation Tools

To simulate attack scenarios, we employed a suite of tools, each chosen for their specific capabilities in mimicking different types of cyber attacks. This enabled us to collect data on various attack vectors and also helped in creating a comprehensive dataset that includes a wide range of potential security threats.

Nessus: Used for vulnerability scanning, Nessus helped in identifying weaknesses in our server setup that could be exploited by attackers. This tool is pivotal in simulating scenarios where an attacker first identifies a vulnerability before exploiting it.

Nmap: Renowned for network scanning, Nmap was utilized to simulate reconnaissance activities by attackers. It served to mimic the initial stages of an attack where attackers map out the network for available services and vulnerabilities.

Nikto: This is a web server scanner that was used to simulate attacks that specifically target web servers. Nikto scans for various types of vulnerabilities, including outdated software and potentially dangerous files, thus representing a more targeted approach towards web service exploitation.

Uniscan: This tool was used to perform web-based attacks by scanning for Remote File Include, Local File Include, and Remote Command Execution vulnerabilities. Its inclusion in our simulation process helped in representing attacks that target specific web application vulnerabilities.

Slowloris: Finally, Slowloris was employed to simulate DoS (denial of service) attacks. This tool works by holding as many connections to the target web server open for as long as possible. It represents attack scenarios where the goal is to overwhelm the server resources, rendering it incapable of serving legitimate requests.

5.3.3 Feature Extraction from Captured Network Traffic

The core of our feature extraction process involved two primary tools: Argus and custom Python scripts which are available in our GitHub page:

https://github.com/AbZ1221/ZTA-IDS#real-world-validation

In the following, each code attribute will be covered in detail.

Argus (Audit Record Generation and Utilization System): This tool was instrumental in converting the raw data captured by tcpdump into flow-based network traffic records. Argus is designed to provide a more structured and analyzable view of network transactions. It condenses detailed network traffic data into flow records, summarizing the essential characteristics of each transaction. This summarization was critical in reducing the complexity of our raw data, making it more manageable for feature extraction. However, since not all attributes can be extracted by Argus or any other tools, we need customized Python processing codes to extract the required features which is explained in the following.

Custom Python Processing Scripts: To further refine the data and extract specific features, we developed custom Python scripts. These scripts were tailored to process the flow records generated by Argus, extracting a set of features akin to those used in the UNSW-NB15 dataset. Our Python scripts were designed to be both efficient and versatile, capable of handling large volumes of data while accurately identifying a wide range of features. These features included, but were not limited to, protocol type, packet length, timestamps, source and destination IP addresses, and port numbers.

A significant challenge we encountered in our feature extraction process was the aggregation of data from the diverse tools used – namely, tshark, Argus, and our custom Python scripts. To effectively combine the data generated by these different methods, we implemented a matching and aggregation strategy. This approach involved aligning the datasets based on common identifiers such as source and target IP addresses and ports. We utilized a sequential numeric matching technique to ensure precise alignment. This method allowed us to consolidate the extracted features into a unified dataset, enhancing the comprehensiveness and accuracy of our feature analysis.

5.4 Application of MV

Before passing the collected information to the IDS, we apply the MV method to encode and hide real IP addresses. This step aligns with the principles of Zero Trust, ensuring that every access request is treated as potentially risky until thoroughly verified. The MV method adds an additional layer of privacy and security by masking user or device identities.

5.5 Analysis by the IDS

The encoded information is then passed to the analyser (IDS), where the proposed neural network models evaluate the requests to detect signs of intrusion or malicious activity. The IDS's continuous monitoring and evaluation mechanism plays a vital role in maintaining the integrity and security of the network.

5.6 Evaluation of Accuracy

The final step involves evaluating the accuracy of the proposed IDS using the real-world dumped records. This includes assessing various metrics such as true positive rate, false positive rate, detection time, and overall accuracy.

The evaluation will provide a comprehensive understanding of the system's performance, highlighting its strengths and identifying areas for potential improvement.

The validation process described above ensures that the proposed IDS is not only theoretically sound but also practically applicable in real-world scenarios. By simulating attacks, applying the MV method, and rigorously evaluating the system's accuracy, we demonstrate the viability and effectiveness of this approach in protecting modern network environments against increasingly sophisticated and dynamic threats. The results and performances are discussed in Chapter 6.

CHAPTER 6

Results and Data Analysis

6.1 Introduction

The previous chapter covers the validation of Zero Trust Architecture Intrusion Detection System (ZTA-IDS) Model. This chapter highlights the results of the experiments and the data analysis.

This chapter is organized as follows: Section 6.2 introduces the ZTA-IDS phases, Section 6.3 reports on the IDS training setup; Section 6.4 discusses the data used for the experiments of applying the modified Multi-View (MV) anonymization approach and the proposed IDS; Section 6.5 reports on the Indistinguishability of Modified MV Approach. Section 6.6 discusses the IDS accuracy and compares it with the corresponding work. The aggregation of these two modules can be used as a Zero Trust model, where the network traces are secured and anonymized by the proposed MV method, and IDS can be used by the analyser to identify intrusions more accurately. In Section 6.7the proposed Zero Trust security model is compared with previous works and the needs of Zero Trust important core principles are discussed. Sections 6.8 and 6.9 present the results of the IDS methods for ten and six categories of intrusions correspondingly; In Section 6.10 the effects of Pre-training the CNN using a balanced data sampler is discussed with comparison to the standard pre-training; Section 6.11 reports on the IDS with only two categories of being an intrusion or not. Section 6.13 discusses the run time performance and the ability of online running. Lastly, the conclusion of this chapter is presented in section 6.14.

6.2 Zero Trust Architecture Intrusion Detection System (ZTA-IDS) phases

Our proposed ZTA-IDS solution which comprises three phases, namely Extended MV, IDS and Evaluating Accuracy and Security as shown in Figure 6.1 are described in detail in this chapter. All code implementations related to this thesis are publicly accessible on our dedicated GitHub repository at:

`https://github.com/AbZ1221/ZTA-IDS`.

Detailed documentation on the use of each component, as well as the necessary requirements, are also provided.

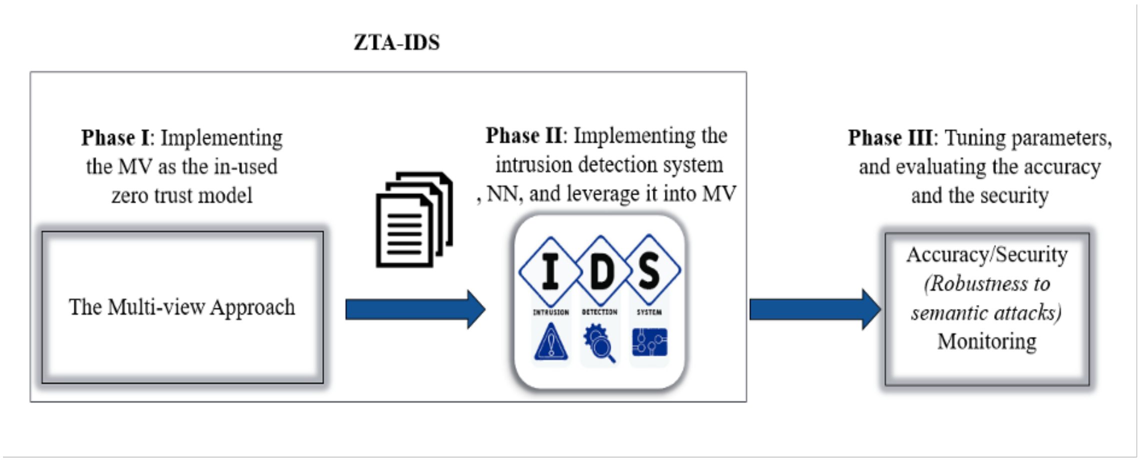

ZTA-IDS
Phase I: Implementing the MV as the in-used zero trust model
The Multi-view Approach
Phase II: Implementing the intrusion detection system , NN, and leverage it into MV
I D S
INTRUSION DETECTION SYSTEM
Phase III: Tuning parameters, and evaluating the accuracy and the security
Accuracy/Security (Robustness to semantic attacks) Monitoring

Figure 6.1 : The phases of ZTA-IDS

6.2.1 Phase I: Implementing the MV as the in-used Zero Trust model

In the first phase, we apply the modified MV model to columns containing data that does not need to be known by the cybersecurity analyst, such as user IP addresses. These columns are anonymized before analysis to ensure privacy and security. The tasks and operations are listed in Table 6.1 briefly. To develop the MV model, the key parameter (K) needs to be determined. Then, we can implement PP(:,K) for IP and non-IP values. We compare the number of instances for different partitioning methods. After this, Random Vector Generation is applied on the output of the previous partitioning methods. In the following, multiple views are generated by applying the pre-fix preserving anonymization on the columns of interest for multiple random times. The used dataset is UNSW-NB15 which is explained in detail in Chapter 4. The complete implementation of the proposed method, encompassing aspects such as data downloading, loading, and pre-processing, in addition to the core methodology, is comprehensively detailed in our GitHub repository at:

`https://github.com/AbZ1221/ZTA-IDS#multi-view-implementation`.

Further analyses pertaining to this implementation are elaborated in Section 6.3 of this chapter.

Table 6.1 : MV tasks and operations

Attribute	**Value**
Environment	Python 3.8
Operations	Encryption & Partitioning
Tasks & Dataset	1. Implementing PP(:,K) for IP and non-IP values 2. Partitioning - compare number of instances for different partitioning methods 3. Random Vector Generation 4. View Generation 5. Preprocessing for IDS 6. Customizing MV to deal with Dataset: UNSW-NB15
Status	Complete

6.2.2 Phase II: Implementing the IDS and leveraging it to MV

In this phase, we leverage the proposed IDS into the implemented MV framework for different combinations of attack types. Even though the UNSW-NB15 dataset includes 10 types of attacks, only the six most common types of attacks are considered in the literature while the others only evaluate in two categorical methodologies, which is the decision of being an attack or non-attack. The types of attacks and the number of occurrences is discussed in Section 5, however, the most common are DoS, Reconnaissance, Normal, Exploits, Fuzzers and Generic. As discussed in Section 5, since the IDS input is the output of MV, an AE feature extractor seems to be beneficial and necessary to be able to add other modules like CNN and LSTM. The proposed IDS neural network comprises four modules, namely autoencoder (AE) feature extractor, convolution blocks, attention mechanism and long short-term memory (LSTM) layers.

The proposed IDS was developed using PyTorch and is available in our GitHub repository at:

`https://github.com/AbZ1221/ZTA-IDS#ids-models`. Google Colab was used to train and evaluate the method.

In our approach, MV features are directly fed into the proposed IDS, bypassing conventional feature engineering techniques. Nevertheless, a preprocessing step is crucial to ensure compatibility with the standard input requirements of neural networks. For example, non-numerical attributes are converted into numerical values using a one-hot encoding method. Additionally, it is imperative to normalize all features to avoid potential bias or training divergence in the neural network. Further detailed analyses of these processes are presented in Sections 6.4 to 6.10 of this thesis.

6.2.3 Phase III: Evaluate the security and accuracy of the proposed solution

In this phase, we conduct a comprehensive evaluation of the security and accuracy of our proposed method, designated as ZTA-IDS, specifically targeting semantic attacks. To facilitate this evaluation, it is imperative to simulate and accurately record the attacks, ensuring precise knowledge of each request's label as ground truth. We executed simulations of attacks at various times, meticulously recording both the requests and their corresponding ground truth responses. Although the UNSW-NB15 dataset includes ten types of attacks, we encountered problems in accurately replicating some attack types. The detailed procedure of our methodology is outlined in Table 6.2. This process begins with the establishment of a simple, secure website on a server. Following the simulation of semantic attacks, all requests are dumped to a storage drive. Subsequent pre-processing steps are then applied before utilizing MV and IDS on the requests. The ultimate objective is to enhance the accuracy of the IDS. Detailed evaluation results and their subsequent discussion are comprehensively reported in Chapter 5.

6.3 IDS Training setup

All experiments are deployed in PyTorch and conducted on the Colab platform with a batch size of 32. The AE is trained to minimize the mse criterion as a loss function, which is also known as a reconstruction error. Both encoder and decoder parameters are considered and trained independently. The dimension of the bottleneck features is 64, which is compact enough to compress the input features. We applied the Adam optimizer with a learning rate of 1e-4 and weight decay of 1e-5 to minimize the reconstruction loss.

The model is trained until no further improvement is possible according to the validation results. All data attributes are normalized to numerical values between

Table 6.2 : Zero-Trust security and accuracy

Attribute	**Value**
Environment	Python 3.8
Operations	Evaluate phase 1 and 2 integration using real-world tests
Attack types	Denial of Service (DoS), Reconnaissance, Normal, Exploits, Fuzzers, Generic
Tasks	1. A simple website implementation on a server 2. Semantic attacks simulation 3. Capture any request 4. Apply MV on results 5. Apply IDS on requests and report the results 6. Improve IDS accuracy
Status	Complete

0 and 1. Thus, non-numerical attributes are converted into numerical values using one-hot encoding. The training dataset is set to prevent over-fitting.

The bottleneck features extracted from the trained AE are fed into the CNN for further training and processing while the AE is frozen. As the compact features of the input attributes are available in the bottleneck layer with 64 neurons, the CNN input spatial dimension is 64 and the sequence number equals the size of the mini batch.

The CNN modules are trained with a learning rate of 1e-3, while the Attention and BiLSTM modules are trained with a learning rate of 1e-4 for a maximum 50 epochs. We exploited cross-entropy loss as the classifier loss function and the Adam optimizer.

6.4 Data

We evaluate the proposed model on the UNSW-NB15 dataset [59], which consists of a hybrid of real modern normal activities and synthetic contemporary attack behaviours. This is an upgraded version of the KDD cup dataset which is a more balanced dataset. The UNSW-NB15 dataset contains ten classes of attacks, namely: Normal, Fuzzers, Analysis, Back-doors, DoS, Exploits, Generic, Reconnaissance, Shell Code, and Worms. We use the train and test subsets of the UNSW-NB15 dataset with 175343 and 82337 records respectively.

In the UNSW-NB15 dataset, each record has a 42-dimensional feature, three of which are non-numerical values and require pre-processing to be fed into the neural networks since the input of NN should be a digital matrix. These three features are protocol, service, and state with 133, 13, and 11 symbol attributes, respectively.

One-hot encoding is used to map non-numerical attributes of the data set to numerical feature vectors. In total, the pre-processed input feature size is 196. Then, the features are normalized between 0 and 1 which are used to train the AE unsupervised. The 64-dimensional bottleneck features extracted from the trained AE are applied for the next experiments. As classes of attacks are unbalanced, as presented in Section 6.7,most studies reduce this number by merging some classes together or removing some of them. Binary classification means all 9 classes are merged into 1 class as Intrusion, consequently the classes are Intrusion/Non-Intrusion in this scenario. However, some other works try to merge the classes that are not far from each other. Some other works try to reduce the amount of imbalance by removing the classes with fewer existing items, including Backdoor, Shellcode, Analysis, Worms and sometimes Fuzzers.

In our experiments, we compare the classification results of 10 classes with the corresponding articles. Also, we report the results of removing imbalanced data attributes to ensure a fair comparison with other state-of-the-art methods. On the other hand, to show the advantage of the proposed structure compared to the previous structures, we report the binary classification results using the same data structure, which train and test data are used in reverse and result in fewer training samples.

Moreover, we propose a nearly balanced sampling procedure to enhance the detection of with fewer samples in the CNN module. Because of the sequential nature required to train the LSTM, we cannot use any sampling strategy to train it. The UNSW-NB15 dataset is highly imbalanced, the Normal category has 56000 samples for training while the Worms category only has 130 samples. We reduce the impact of this imbalance using sampling based on a smoothing probability function shown in Equation 6.1.

$$P(\mathrm{cl}_i) = \frac{\#\mathrm{cl}_i - \Big(1 - (\frac{\min \#\mathrm{cl}_i}{\#\mathrm{cl}_i} + \epsilon)\Big)\,\mathrm{median}}{\sum_{j=1}^{10} \#\mathrm{cl}_i - \Big(1 - (\frac{\min \#\mathrm{cl}_i}{\#\mathrm{cl}_i} + \epsilon)\Big)\,\mathrm{median}} \tag{6.1}$$

In this equation, cl_i means ith category (class), so $P(cl_i)$ is the probability of choosing a sample from ith category, calculated using the number of samples in each category ($\#cl_i$) and the median of the number of samples per category. We use a small $\epsilon(0.1)$ to prevent zero addition for the category with a minimum number of samples. According to Section 6.7, the minimum number of samples is 130 associated with the category Worms, and the median is 11378. The proposed sampling strategy keeps the ordering of the number of the categories but make the sampling more balanced by reducing the distance between the number of items per each category. In the following, we report the results of the IDS methods for ten, six, and two categories.

6.5 Modified Multi-View (MV) Approach Results

To validate the modified MV anonymization technique, we applied the whole UNSW-NB15 dataset. Since this work is a balanced version of the *N* Key Vector approach (distinct IP based partitioning method with *N* partitions explained in Section 3.3.2) and mostly affects indistinguishability, we focus on this metric. We believe that applying MV on the other attributes does not have any effect on the results and only gives the ability to send these attributes to the analyser.

Indistinguishability is crucial as it can help prevent intrusions. Whenever an analyst (adversary) receives N different traces with identical attribute values and different attribute values, their purpose can be to detect the real view among all the views. For instance, they may try to observe their injected or fingerprinted flows, or they can launch the semantic attacks on those views, hoping that the real view might respond differently to those attacks. Therefore, the main objective of the MV approach is to satisfy the indistinguishability property, which means the real view must be sufficiently indistinguishable from the fake views under semantic attacks. We use the ε-indisinguishablity property to compare the indistinguishability of different methods [1]. In the following paragraphs, we first review the data properties and configuration, followed by a review of the evaluation conditions and results.

As previously discussed, we used UNSW-NB15. It includes 2540047 records, including 49 attributes. It contains 9 attacks, such as worms, DoS, backdoors, and fuzzers. For simplicity, the source IP attribute is used for evaluation and can be expanded to other attributes. As discussed in [1], "the number of views N is an important parameter that determines both the privacy and computational overhead. The data owner could choose this value based on the level of trust in the analysts and the affordable computational overhead." the number of views should be determined carefully.

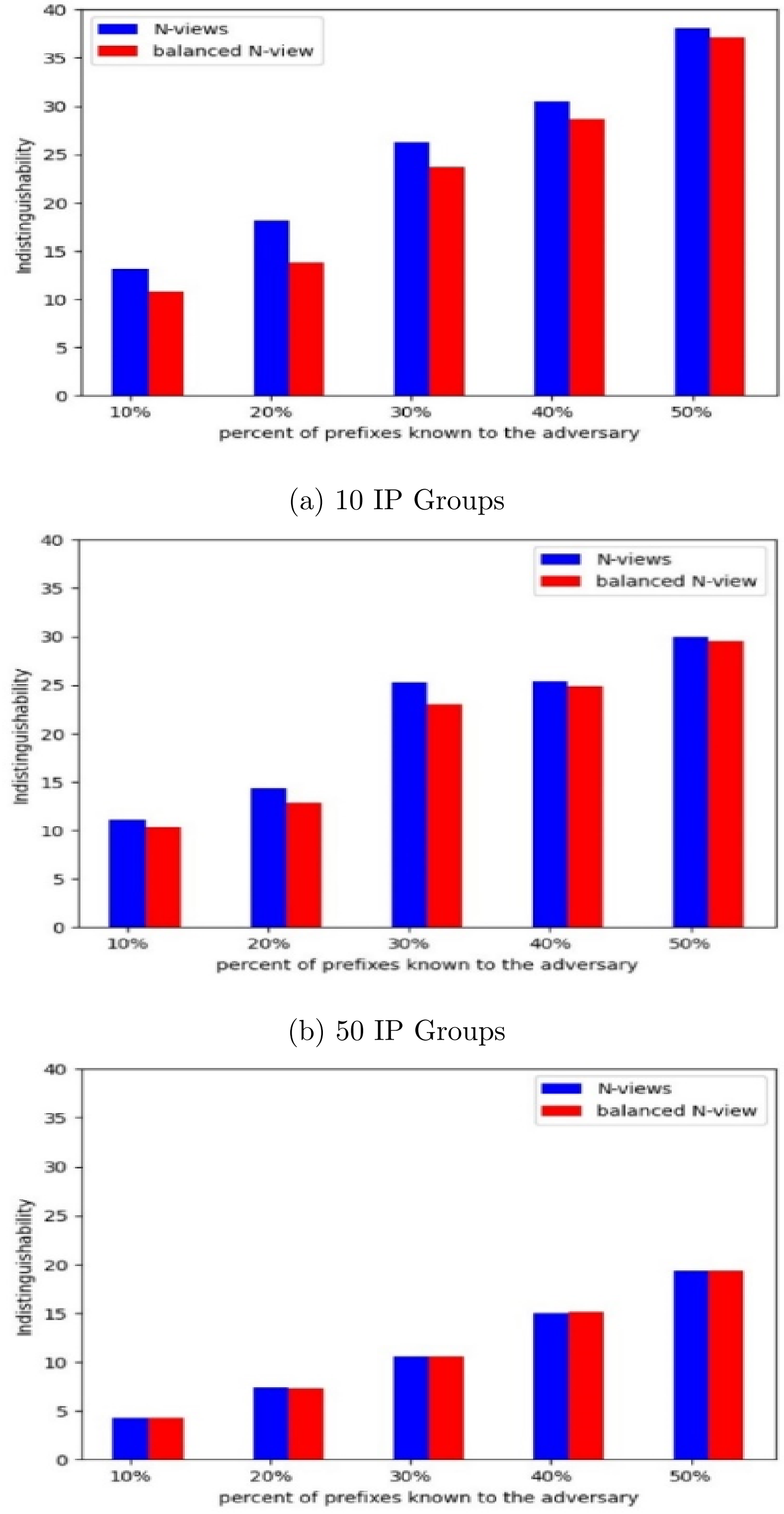
The number of real view candidates is $e^{\epsilon} * N$. The data owner can first estimate the adversary's background knowledge (number of prefixes known to the adversary) and then calculate ϵ[1].

The results of indistinguishability for various numbers of groups are represented as 10, 50 and 136 (the first octet) and different adversary knowledge are shown in Figure 6.2. We cannot achieve the results of [1] since our dataset is different. Obviously, using a balanced partitioning method for a small number of groups outperforms the original method as shown in Figure 6.2, except when the percent of prefixes known to the adversary is high (40% and 50%). Obviously, this improvement reduces as the number of groups increases (Figure 6.2b and Figure 6.2c). The balanced N Key Vector approach does not result in any improvement for 136 IP groups (first octet), although both methods seem to work in the same way.

6.6 IDS Results

To validate the proposed IDS, we use the train and test subsets of the standard UNSW-NB15 dataset [56]. As discussed in Parts II and III-B, we can feed the MV generated data to the model. However, we applied the standard training and testing subsets for IDS validation to compare the results with the related works.

Both the AE and CNN models are trained on a single 3060s GPU and are implemented in PyTorch. The autoencoder is trained with a batch size of 32, an initial learning rate of 1e-4 and mean squared error (MSE) loss, and the Adam optimizer. The models are trained until no more improvement is possible, according to the validation results. All data attributes are normalized to numerical values between 0 and 1. Thus, nonnumerical attributes are converted into numerical values using one-hot encoding. Bottleneck features extracted from the trained AE are fed into the CNN model for classification. The classifier is trained with a batch size of 32, an initial learning rate of 1e-6, and cross-entropy loss for the binary classifier



(a) 10 IP Groups

(b) 50 IP Groups

(c) 136 IP Groups

Figure 6.2 : Indistinguishability comparison of balanced and original N Key Vector approaches for different numbers of groups and adversary knowledge.

and the Adam optimizer.

In terms of IDS performance, since the number of parameters of the proposed neural network is low, it is fast and can be used for low-resource scenarios. The proposed AE model has only 33K parameters and the CNN model has 1.01 million parameters. In total, the IDS model has 1.03 M parameters and it takes 7 milliseconds to infer a sample record on a 1.60 GHz core CPU.

6.7 Zero Trust Model Evaluation

In this subsection, we compare our proposed Zero Trust security model with the four Zero Trust models discussed in Section II, namely Software-Defined Perimeter, Transport Access Control and First Packet Authentication, Fine-grained Big Data Security, and Autonomic Security.

Based on the principles outlined by Modderkolk et al. [57], our proposed approach is able to inspect and log traffic automatically using IDS. It also provides secure access to the resources within the network by anonymizing any fields that do not need to be known by the analyser using the proposed MV approach. Moreover, it can log and inspect all traffic using the combination of both introduced components. As a result, our proposed model ticks all three crucial boxes of important Zero Trust core principles provided in Table 6.3. However, we attempted to create a more efficient authentication system which can be applied on a wider scale using machine learning and artificial intelligence technology.

Since the results of the baseline CNN model using AE features are not available for categorical classification, we deployed it and report the results for comparison.

Undoubtedly, Attention-based CNN-BiLSTM using AE bottleneck features outperformed the other related works using deep learning methods for 10 categories of IDS. The confusion matrix of the proposed method is presented in Figure 6.3.

Table 6.3 : Needs of Zero Trust important core principles

Need	Status
Least privilege and access control	✓
Inspect and log traffic	✓
Ensure secure access	✓

In our proposed model, we tried to make improvements on fine-grained and automatic authorization processes, which can improve efficiency dramatically.

None of the models could block all types of attacks, in spite of their major improvement on traditional cybersecurity philosophies, indicating that new models can offer important improvements. Thus, further experimentation with these models opens the possibility for future enhancements such as latency and the dynamic optimization of other network equipment.

6.8 Ten Categories CNN-BiLSTM Data Classification

The number of items in each category of the train and test sets of the UNSW-NB15 dataset is shown in Table 6.4.

To optimize the hyper parameter settings, we explored the optimal number of layers and neurons for each part of CNN-BiLSTM with the Attention module. The optimal hyper parameters were described previously. To evaluate the effect of each module, the results of BiLSTM, CNN-BiLSTM (with a linear layer in between), and Attention-based CNN-BiLSTM are compared to the related methods in Table 6.5 for ten categories of data classification.

Since the results of the baseline CNN model using AE features are not available for categorical classification, we deployed it and report the results for comparison. Undoubtedly, Attention-based CNN-BiLSTM using AE bottleneck features outper-

Table 6.4 : UNSW-NB15 dataset categories' size in train and test sets [41]

Category	Train	Test
Normal (n)	56000	37005
Backdoor	1746	583
Analysis (a)	2000	677
Fuzzers	18185	6062
Shellcode	1133	378
Reconnaissance	10492	3496
Exploits	33393	11132
DoS (d)	12264	4089
Worms (e)	130	44
Generic (r)	40000	18871
Total (i)	175343	82337

formed the other related works using deep learning methods for 10 categories of IDS. The confusion matrix of the proposed method is presented in Figure 6.3.

As can be seen from Figure 6.3, most errors of Analysis, Backdoor and Exploits attacks are misclassified as DoS. Furthermore, Fuzzers and Exploits are misclassified interchangeably. None of Analysis and Backdoor records are predicted correctly.

Only three records of the Worms class are predicted correctly. Consequently, removing imbalanced data attributes including Backdoor, Analysis, Shellcode and Worms should obviously improve the accuracy. Thus, removing Fuzzers may lead to better accuracy.

Table 6.5 : Accuracy results of ten categories of IDS on test data

Method	**Accuracy**
CNN	78.23%
BiLSTM	77.46%
CNN-BiLSTM	78.76%
Data pre-process with scaling + SVM [58]	75.77%
Decision Tree C5	90.74%
Integrated rule-based [52]	84.83%
CNN-Attention + BiLSTM	87.76%
CNN-BiLSTM [41]	77.16%
Feature Selection + ANN [39]	77.51%
CNN-Attention	88.13%

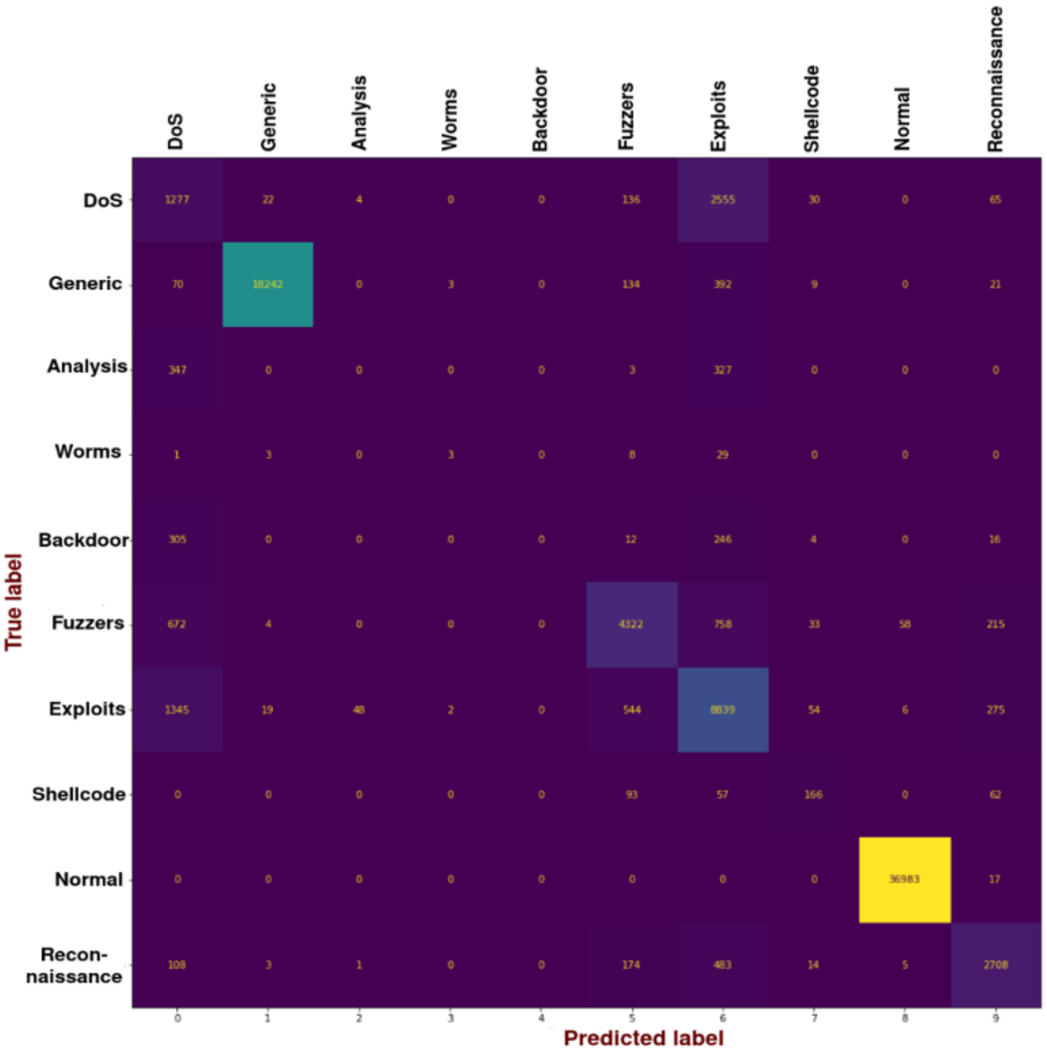
6.9 Six Categories of CNN-BiLSTM Data Classification

Since the recent works removed imbalanced data for their experiment results and most of them reported six categories, we also conduct experiments with the proposed method after removing four imbalanced categories. The results are shown in Table 6.6.

As shown, the introduced method also outperforms the state-of-the-art methods for six categories. The confusion matrix is shown in Figure 6.4.

6.10 10 Pre-train CNN using balanced data sampler

We used a balanced data sampler to pre-train the CNN for later use in CNN-BiLSTM with the Attention module instead of reducing the number of categories. Our aim is to boost the discrimination of the model to learn the data better, even



	DoS	Generic	Analysis	Worms	Backdoor	Fuzzers	Exploits	Shellcode	Normal	Reconnaissance
DoS	1277	22	4	0	0	136	2555	30	0	65
Generic	70	18242	0	3	0	134	392	9	0	21
Analysis	347	0	0	0	0	3	327	0	0	0
Worms	1	3	0	3	0	8	29	0	0	0
Backdoor	305	0	0	0	0	12	246	4	0	16
Fuzzers	672	4	0	0	0	4322	758	33	58	215
Exploits	1345	19	48	2	0	544	8839	54	6	275
Shellcode	0	0	0	0	0	93	57	166	0	62
Normal	0	0	0	0	0	0	0	0	36983	17
Recon- naissance	108	3	1	0	0	174	483	14	5	2708

Figure 6.3 : Confusion matrix of ten categories for CNN-BiLSTM with AE feature extractor

when the number of training samples are imbalanced. Consequently, we applied the trained CNN, which we believe is a better discriminator, in CNN-BiLSTM with the Attention module to enhance the detection of network intrusion. The accuracy results of CNNBiLSTM with the Attention module with and without the balanced sampler are compared in Table 6.7.

As shown, pre-training the CNN using a balanced sampler outperforms standard training and also the other works in terms of accuracy. To illustrate the performance

Table 6.6 : Accuracy results of six categories of IDS on test data

Method	Accuracy
MLP + IGRF-RFE [59]	84.24%
CNN-GRU + RFP [53]	86.25%
Rule-based [52]	84.84%
CNN	82.01%
BiLSTM	83.11%
CNN-BiLSTM	86.28%
CNN-Attention	87.54%
CNN-Attention + BiLSTM	89.79%

Table 6.7 : Accuracy results of ten categories of IDS on test data for the balanced and imbalanced sampler

Method	Accuracy
Proposed method (Standard Sampler)	87.76%
Proposed method (Balanced Sampler)	91.72%

of the proposed method, we also compare the other metrics namely recall, precision and f1-score for the proposed method with the related works in Table 6.8.

The confusion matrix of the proposed method with a balanced sampler for the ten categories is shown in Figure 6.5. Clearly, the number of misclassifications for each category is low when using this method, especially for the Normal category.

6.11 Binary CNN-BiLSTM data classification

The hyper parameters for the binary classification model have been kept the same as the multi-class model.

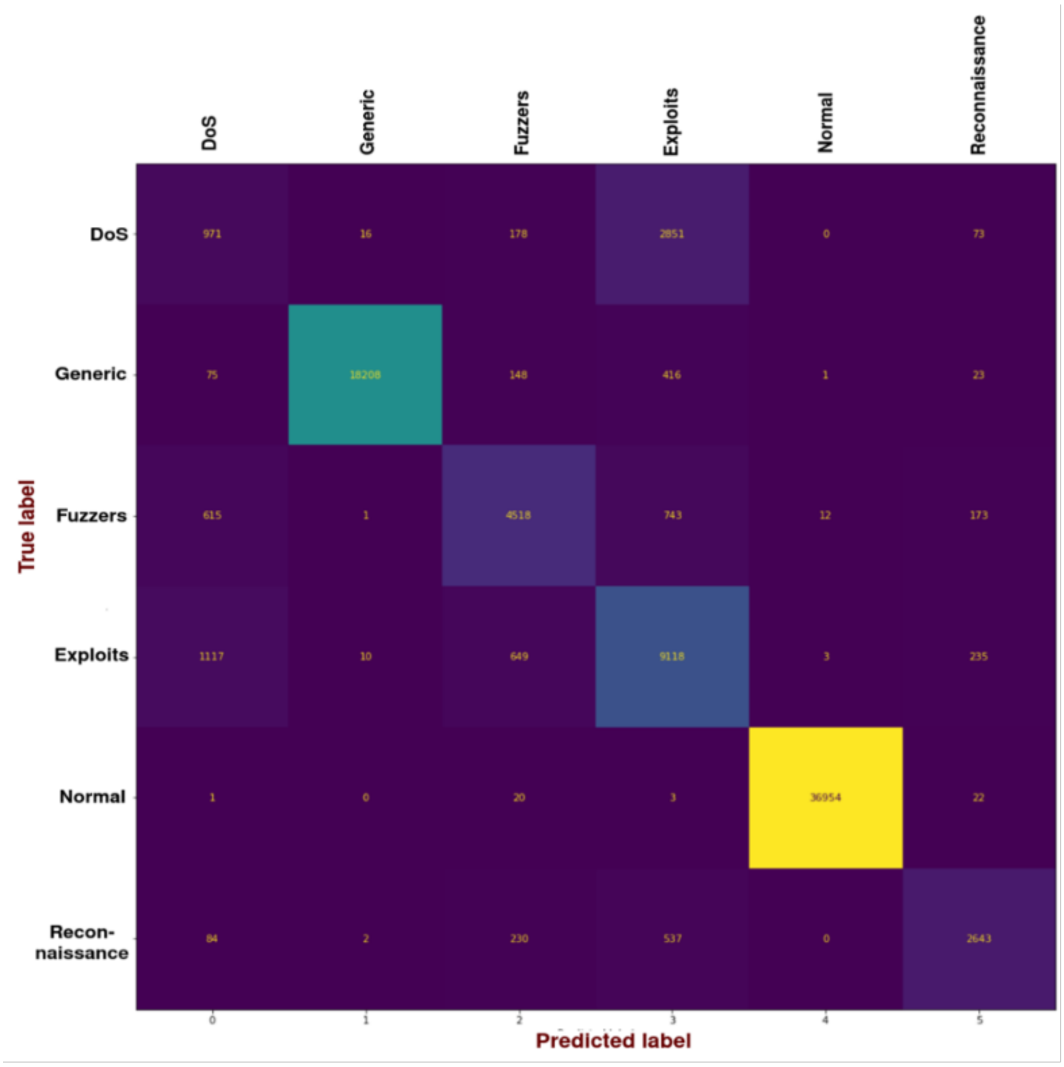


Figure 6.4 : Confusion matrix of six categories for CNN-BiLSTM with the AE feature extractor

To evaluate the effect of each module, the results of BiLSTM, CNN-BiLSTM (with the linear layer in between), and the Attention-based CNN-BiLSTM are compared to the work which is most related to this work, using CNN with the AE bottleneck features for IDS, as shown in Table 6.9. For a fair comparison, our data must be similar. We applied the train and test data interchangeably to ensure a fair comparison with the CNN and AE method.

As can be seen from the three first rows of Table 6.9, using BiLSTM decreases

Table 6.8 : Precision, Recall and F1-Score results of IDS on test data

Method	**Precision**	**Recall**	**F1-Score**
Feature Selection + ANN [39]	79.50%	77.53%	77.28%
Decision Tree C5 [52]	-	75.8%	75.54%
Integrated Rule-based [52]	-	65.21%	68.13%
Proposed method (Balanced)	60.24%	78.5%	62.62%

the accuracy of the model especially in combination with CNN. This can be due to the high dimension of the CNN output, which is fed into the BiLSTM layers. However, using an Attention module on CNN to aggregate the CNN features for feeding into BiLSTM layers outperforms CNN and the BiLSTM models.

Since other binary classification techniques using the original train and test dataset for IDS achieved almost 100% accuracy, further experiments and improvement are not needed.

Table 6.9 : Accuracy results of IDS (binary classification) on test data

Method	**Accuracy**
CNN	92.23%
BiLSTM	90.84%
CNN-BiLSTM	78.93%
CNN-Attention + BiLSTM	93.01%

As shown by the results, CNN and BiLSTM both perform well for IDS using the AE bottleneck features. However, an Attention module is needed to handle the relation between the components of these two structures and combine them.

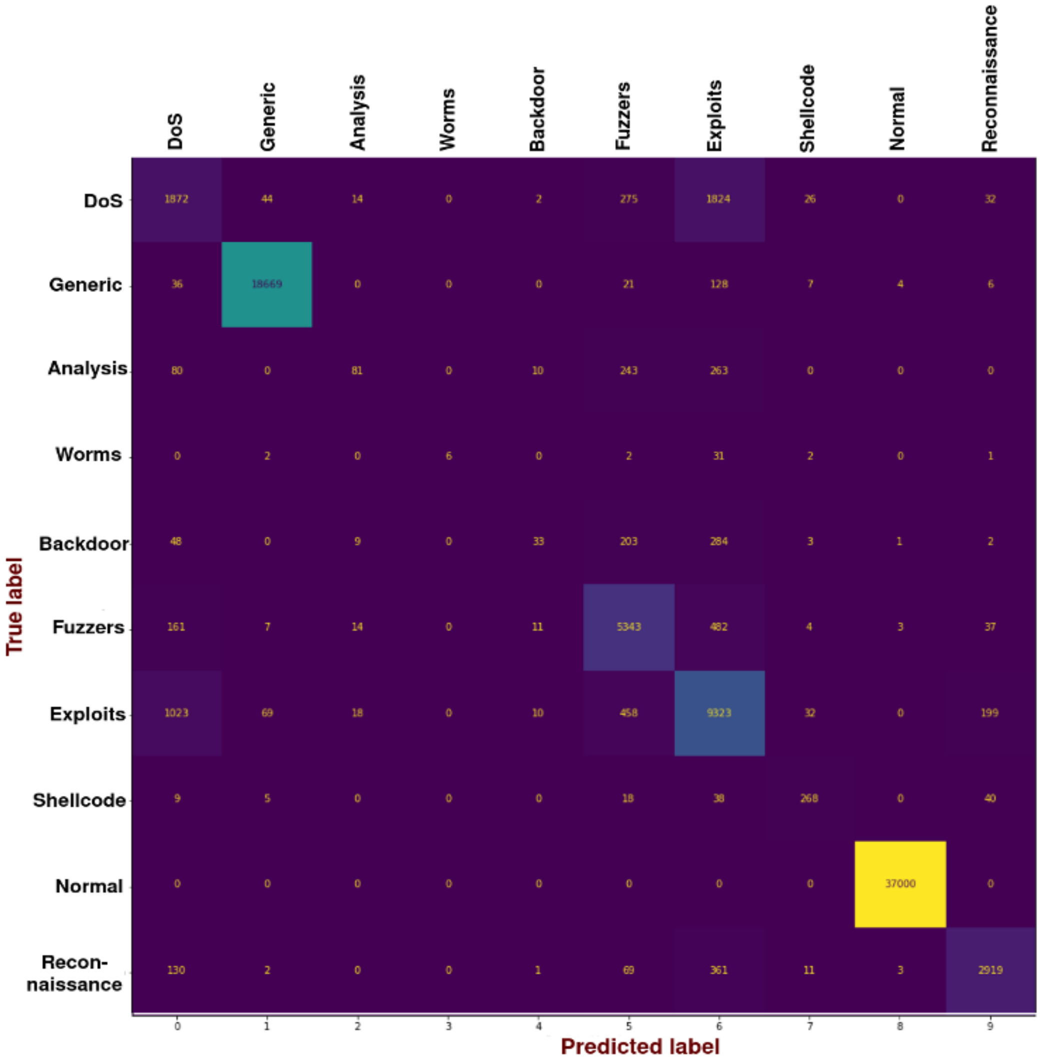


Figure 6.5 : Confusion matrix of pre-trained CNN using balanced sampler + Attention + BiLSTM with AE feature extractor

6.12 Real-World Accuracy Evaluation

Using the simulations and the methodologies described in the previous subsection, we collected 1632 records of all 6 prominent attack/non-attack categories using the attack tools mentioned in Section 7.5.1.2. As previously discussed, all requests dumped and the required features are extracted from the collected requests. Then, MV is applied on the resulting data and fed into IDS for evaluation. The trained IDS models predict the category of the generated data with the accuracy

Table 6.10 : Accuracy Results of the UNSW-NB15 trained models on real-world data

Method	**Accuracy**
CNN-Attention	50%
CNN-Attention + BiLSTM	20%

listed in Table 6.10, which is low and unreliable.

As can be seen from Table 6.10, the model's accuracy degrades on real-world dumped data, especially when utilizing BiLSTM to consider the sequences. The reason for this significant degradation could be due to the way RNN models work. Since, for training the RNN models, we must consider the training data as related sequential records, we are unable to shuffle it during the training or evaluation procedures. This can cause the model to overfit on the training and evaluation data, a phenomenon also referred to as memorizing the data order in machine learning. Consequently, the CNN-Attention model seems to be a better choice for applying the model to real-world data. On the other hand, the much lower accuracy of the CNN-Attention model shows that our trained model does not generalize well. We believe that the data may not be sufficient to train such a deep neural network model. An option would be to adapt the trained model using a small dataset from real-world scenarios, simulated using different tools and servers from the test data.

We used a different server and attack tools from those in the test and collected a total of 500 samples for different categories with corresponding labels. We fine-tuned the trained model for 2 epochs. The accuracy of the adapted model on real-world test data is compared with that of the non-adapted model in Table 6.11. According to the results of Table 6.10 and Table 6.11, the adapted model accuracy improved by up to 80% on real-world data. It can be inferred that one of the drawbacks is the dataset, which is different from a real-world scenario. Hence, it seems that using

Table 6.11 : Accuracy results comparison between adapted and non-adapted model on real-world data

Method	**Accuracy**
CNN-Attention (non-adapted)	50%
CNN-Attention (adapted)	80%

more data which are like a real-world scenario, can improve the results and help the model generalize well.

6.13 ZTA-IDS performance

6.13.1 IDS Performance

Since the number of parameters for the proposed neural network is low, it is fast and can be used for low-resource scenarios. The proposed AE model has only 33K parameters and the CNN model has 1.01 M parameters. In total, the IDS model has 1.03 M parameters, and it takes 7 milliseconds to infer a sample record on a 1.60 GHz core CPU. However, if the number of requests per second overwhelm the system and adversely impact its performance, it can be easily scaled up with adequation more powers to the system.

6.13.2 MV Performance

Since the processes of MV are not complicated and only require a few repetitions for some attributes (K_1, K_2 times of anonymization), it does not consume a lot of CPU resources. As a result, it is faster than real time. Even it seems that dedicated CPUs with low powers are capable of applying this.

6.13.3 Real-Time Performance

In our empirical analysis, we meticulously measured the time efficiency of different components in our model's processing pipeline. The results, obtained using a Google Colab environment with dual processors, revealed the following time metrics per request:

1. MV application: 2.8 milliseconds
2. IDS inference: 1.7 milliseconds
3. Feature extraction, akin to the UNSW-NB15 dataset standards: 0.6 milliseconds

Cumulatively, these components contribute to an overall processing time of approximately 5 milliseconds per request. This translates to a throughput of 200 samples per second, underscoring the model's efficiency in handling requests. Notably, the current setup's performance demonstrates that leveraging more processing cores can significantly enhance the sample processing rate. Furthermore, it implies that incorporating a more powerful CPU could further optimize the inference time, potentially elevating the system's efficiency in real-time data processing scenarios.

6.14 Conclusion

This chapter detailed the evaluation metrics and data and then discussed the experiment results for the modified MV anonymization approach and compared these with previous works. Then, we reported on the IDS accuracy and compared it with the corresponding work. The aggregation of these two modules can be used as a Zero Trust model, where the network traces are secured and anonymized by the proposed MV method, and IDS can be used by the analyser to identify intrusions more accurately. Then, we presented the results of the IDS methods for ten, six, and two categories.

A promising direction for future enhancements of the proposed method in real-world scenarios lies in the adoption of a distillation mechanism. This approach involves transferring the comprehensive knowledge acquired by our current, larger system to a more compact, 'student' model. The essence of this process is to encapsulate the intricate patterns and insights learned by our extensive model into a smaller, more efficient framework.

The primary advantage of this distillation strategy is twofold. Firstly, it allows for the retention of high-level accuracy and learning capabilities, which are characteristics of the larger model, but in a form that is significantly more resource-efficient. Secondly, a smaller model inherently requires less computational power for both training and inference, which can lead to reduced processing times and enhanced applicability in environments with limited computational resources.

By refining our model through this knowledge distillation process, we can potentially achieve a system that maintains high accuracy and efficiency, but with a reduced footprint, making it more suitable for real-time applications and scalable for broader deployment scenarios. The superiority of the proposed method in comparison with homomorphic encryption methods is discussed in 3.2.1

The following chapter covers the importance of Zero Trust in the face of COVID-19.

CHAPTER 7

Zero Trust and Its Importance in the face of COVID-19

7.1 Introduction

The preceding chapter focused on the results and data analysis of Zero Trust Architecture Intrusion Detection System (ZTA-IDS). Expanding on this, in this chapter, we examine the importance of adopting Zero Trust during and after the Covid-19 pandemic.

The rest of the chapter is organized as follows: sections 7.2 and 7.3 review some of the related works and investigate the impact of COVID-19 on existing IT infrastructure, respectively. In section 7.4, the impact of COVID-19 on existing IT infrastructure is detailed. Section 7.5 explains how a ZT network functions. Sections 7.6 and 7.7 detail the objectives and the results, respectively. Section 7.8 concludes this chapter.

The contents of this chapter have been published in the International Conference on Advanced Information Networking and Applications.

The content of this chapter is available at the following link: `https://link-springer-com.ezproxy.lib.uts.edu.au/chapter/10.1007/978-3-031-28694-0_17`

7.2 Zero Trust and the Covid-19 pandemic

The COVID-19 pandemic spread across the globe in an unprecedented manner, resulting in a work-from-home situation for which organizations were unprepared. Businesses had to deal with portable device shortages and insufficient bandwidth, with virtual private networks (VPNs) buckling under the strain of increased use and IT departments pushing the limits to maintain efficiency to enable employee access to corporate resources/applications so they could work at full capacity. The pandemic caused a profound re-modification and reorganization of human-to-human interaction [60]. Before the pandemic, interest in Zero Trust was being driven by a need to modernize how the information security stack works. The traditional perimeter-centric security model is not compatible with the way businesses are working today. The pandemic forced organizations to consider ZT because so many employees had commenced remote work that the organizations' networks were no longer trustworthy. An increased attack surface is a concern and increased mobility was already occurring but was accelerated in the pandemic. ZT offers encouraging solutions, but requires reasonable re-architecture, re-modification, and re-investment. This means the work-from-home scenario is a continuous arrangement for staff whose job profiles enable them to fulfil their professional obligations without commuting to and from their job site. These factors have seen the risk of cybercrime and cyberattacks increase. With older technologies like VPNs making headlines related to breaches in security, a new approach to empowering distributed teams while ensuring optimal data security emerged, namely ZTA. To investigate this issue, a survey was conducted to obtain insights which could lead to the development of an IDS using ZTA. More than 40 of the survey respondents were chief information security officers from various IT companies, banks and government in India and Saudi. The survey's objectives were to answer the following questions:

1) Why did businesses experience similar pain points to enable secure remote work?

2) How could a ZTA assist the business continuity in a pandemic outbreak?

3) Why are organizations committed to adopting a ZT security architecture?

4) Which are the key initiatives to enable ZT security adoption in an organization?

7.3 Background

As reported in [11], between 2014 and 2016, over 200 executive board members of 80 companies were interviewed and asked to answer the question "How do we secure increase dynamic architecture in an environment without a perimeter?". The answers revealed that the policy of bring your own device (BYOD) was valuable but posed onerous risks. Setting up a centralized and scalable mobile device management system using access controls (LDAP/AD) was reported to be the most important challenge. This suggests a more-risk based approach to cybersecurity is needed in today's dynamic technological environment.

According to [3],ZT treats all network traffic as untrusted, continuously confirming users and endpoints using secure cloud data. The benefit of ZT is that it is a highly flexible infrastructure that can be integrated with the cloud to enhance organizational security. To ensure network safety amid new cybersecurity threats, cybersecurity professionals should embrace additional philosophies alongside a ZT mindset.

According to [4], one of the largest debates of our generation is addressing the issue of who we can trust with our data? It has never been more important to develop security models that keep users safe. The author reported that approximately 60% to 80% of network misuse comes from within the network. ZT therefore, offers a solution to both issues, with its ability to increase the micro-segmentation of a network offering more visibility of overall traffic through the inspection of users and devices which connect the network. The work in [61], proved that working remotely, especially for employees with minimal cybersecurity resources, increased the risk for personal and organizational data to be compromised. In [62], the authors elucidated the challenge pointing out the use of AI and poorly secured technologies deployed in response to COVID-19 challenges increased the risk of cybercrimes due to the high volume of data being generated and shared. The work in [61] proved that working remotely, especially for employees with minimal cybersecurity resources, increased the risk for personal and organizational data to be compromised. In [62], the authors elucidated the challenge pointing out the use of AI and poorly secured technologies deployed in response to COVID-19 challenges increased the risk of cybercrimes due to the high volume of data being generated and shared.

In [40], a two-stage ensemble classifier including hierarchical rotation forest and bagging classifiers, along with a hybrid evolutionary algorithm for feature selection was proposed for network intrusion detection (NID). In [63], a four-way ensemble classifier comprising a support vector machine, linear regression, naïve Bayes, and decision tree was proposed which utilized a combination of feature selection methods.

Since ML methods require feature extraction and parameter tuning [10], DL methods have become a trend in solving AI problems such as image, language, and speech processing and NID [64][50]. A two-stage deep neural network (NN) was proposed for NID which included a deep sparse autoencoder (AE) as the feature extractor and a shallow NN classifier [41].

In [65], a sparse auto-encoder was proposed for feature extraction, however, support vector regression was used as the classifier instead of a shallow NN. The AE bottleneck features were shown to be effective in enhancing NIDS, giving the ability to feed any type of attributes to the NID model. Plus, bottleneck features were shown to be robust against noise.

In [66], a recurrent neural network (RNN) was proposed for NID to consider the changes of the input in real-time applications. Also, deeper RNN models were used for NID which outperformed previous works. Since long short-term memory (LSTM) cells hold long-term dependencies and prevent the vanishing gradient problem, the work in [67], extended RNN models to LSTM and bi-directional LSTM (BiLSTM) for NID.

In [68], a convolutional neural network (CNN) classifier using a two-stage feature extraction including principal component analysis (PCA) and a feature engineering method to select the most relevant features have been proposed for NID. In [82], the CNN models were used in combination with other classifier methods including RNN, LSTM, and gated recurrent unit (GRU), which proved the power of CNN.

In [59], IGRF-RFE was introduced as a hybrid feature selection method for multi-class network anomalies using a multi-layer perceptron network. It was a feature reduction model based on both the filter feature selection and the wrapper feature selection methods. The filter feature selection method was the combination of information gain (IG) and random forest importance, to reduce the feature subset search space. Recursive feature elimination was a wrapper feature selection method to clear redundant features recursively on the reduced feature subsets.

In [53], a NID model that fused a CNN and a gated recurrent unit was proposed to tackle the low accuracy problems of existing ID models for the multiple classification of intrusions and the low accuracy of class imbalance data detection. They applied a hybrid sampling technique combining adaptive synthetic sampling and repeated edited nearest neighbours for sample processing to solve the positive and negative sample imbalance issue in the dataset. Feature selection was carried out by combining RF and Pearson correlation methods to address the feature redundancy problem.

In [52], an IDS was proposed to detect five categories in a network: Probe, Exploit, DOS, Generic and Normal. This system was based on a misuse-based model, which acted as a firewall with some extra information added to it. Moreover, unlike most related works, they considered UNSW-NB15 as the offline dataset to design their own integrated classification-based model for detecting malicious activities in the network.

7.4 Impact of COVID-19 on Existing IT Infrastructure

In [61], the authors emphasized that cybercrime is among the greatest threats for most organizations. The problem's magnitude was further elucidated by the financial burden, which they report was $3 trillion in 2015 and was projected to be over $6 trillion every year by 2021. The damage caused by cybercrime is profound, including data destruction, reputation attacks, hindering company progress, the loss of intellectual property, embezzlement, increasing mitigation costs, and the cost of damage control in cases when such attacks occur. Therefore, a safe cyber security for organizations is necessary.

The average ransom payment demanded by cybercriminals carrying out ransomware attacks went up by 33% in the first quarter of the 2021 to $111,605 compared to the previous quarter. Phishing attempts have also increased, with Google's Threat Analysis Group noting 18 million COVID-19-related phishing and malware Gmail messages each day in April 2021.

Most companies with remote teams must address new cybersecurity concerns and points of vulnerability. They are working diligently to fortify their network perimeter, implementing the latest in hardened routers, next generation firewalls, and IDSs. However, a lot of IT departments weaken their own security by launching a website, home banking systems, ERP/ERM systems that provide access to other networks and computers behind firewalls which enables attackers to penetrate the system. The major reason for such diluted security is the lack of acceptance and the trial of different methods other than the traditional security and legacy systems.

Nowadays, as some users work from home and the network structures of all the organizations are changing, the traditional way of working is moving towards the cloud and there is rise in the use of Software as a Service (SaaS). Many organizations are also embracing flexible working conditions, with staff connecting from multiple devices in various locations, leading to a declining traditional network cycle/perimeter which is causing a decline in security. Additionally, hackers are attempting to find unaddressed vulnerabilities in newly deployed remote work infrastructure. Hence, businesses have felt a compelling need for an advanced and dependable security solution.

Organizations are being driven by the stress the pandemic was putting on their infrastructure, particularly on VPNs. Before the pandemic, VPNs were good enough to satisfy most companies' work-at-home demands, which were occasional. The legacy system model works on the principle of trust, where it considers the elements inside a particular network as harmless. Today, employees often work from home, which poses a huge threat to the security of network. With employees now working remotely, they are effectively creating a hacker's playground, with new, vulnerable endpoints and access points being exposed. However, all is not lost since the ever-advancing technology space has realized the importance of ZTA, which enhances cybersecurity and safety.

7.5 How a Zero Trust Network Functions

With legacy security models such as VPNs failing in relation to security, ZTA is gaining popularity for its optimal data security and empowering redistributed teams. The idea behind ZT networks is not new. ZTA is based on the basic philosophy that no technology user should be trusted [69]. It is self-explanatory, which means trust no network. ZTA is a security tenet that asks organizations not to automatically allow access from inside/outside into the network structure, but instead to verify each request trying to connect.

ZTA offers cybersecurity paradigms that are more focused on users, assets, and resources as opposed to traditional paradigms that were more static and network-oriented. By evoking the ZT principle in restructuring workstation cyber-system security, ZTA blocks the major loopholes exploited by hackers and malicious intruders. In the work in [70], the authors noted the tactic to move the firewall from outside to inside the system architecture as ZTA makes it harder to target the system internally, which is the most-used approach in cybercrimes. The checking and recording of traffic within a network enable effective system monitoring, further securing the system. They cite four methods that help achieve a feasible and effective ZT strategy for an effective scalable security infrastructure:

1) Identity authentication which validates credibility to be allowed in a network.

2) Access control which regulates the layering degree an authenticated user can access.

3) Continuous encryption-based diagnosis to offer monitoring and feedback services that help trace a threat with ease to an origin point and potential damage.

4) Mitigation, which is critical in reducing the occurrence of damage through threat identification and prevention strategies.

The goal of ZTA is to prevent unauthorized access to data along with creating enforcement mechanism for control. To achieve this vision, several technical elements are necessary, and it is important to note a single commercial tool or technology will not be able to deliver all capabilities. According to the National Institute of Standards and Technology (NIST), the logical elements of ZT include the policy engine, policy administrator, and policy enforcement point. Several data sources are necessary to provide input to these policy-based mechanisms which will feed the trust algorithm that ultimately determines whether to grant/deny access to information resources based on the level of evaluated trust of the endpoint/user combination. ZT models, according to NIST, assume an attacker is present on the network and an enterprise-owned network infrastructure is no different to any non-enterprise-owned network. NIST categorizes the types of input as: access request; user identification, attributes, and privileges; asset database and observable status; resource access requirements; and threat intelligence (Figure 7.1). The authors in [71], noted that the scalable approach is able to thwart threats to cybersecurity using a multistage strategy, hence it is more reliable than a traditional network architecture.

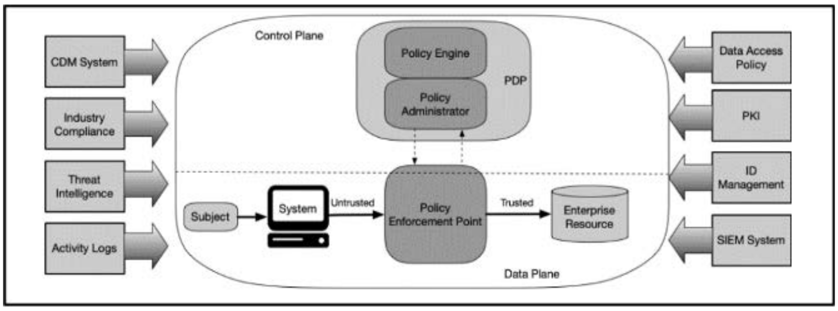


Figure 7.1 : NIST, 800-207, ZTA 2nd Draft

Innate trust is removed from the network under ZTA principles which means someone does not necessarily have access to everything on the network, even though they are connected to that network. Inherent trust is removed and declined, so devices and users are denied access until they are verified on the basis of pre-defined parameters. The pre-requisite for gaining access is authorization and crossing the security set level. This works best to avoid breachers who witness an attacker and move laterally into the network in cases where everything is trusted in the network. Treating the network as hostile has many advantages. By leveraging micro-segmentation and granular perimeter enforcement based on end-user characteristics like location, role, and permissions, a ZTA only gives people access to the specific resources they need. Therefore, the strategy secures layers with fine-grained segmentation, stringent system access controls, strict data retrieval management, and a sophisticated data protection strategy [70]. All users access a system through devices which must be authenticated while the information degree accessed is tightly controlled and restricted to a need-to-know basis [71]. The encryption ZTA strategy protects the information from internal and external intrusion and maintains a continuous monitoring and adjustment process that maintain proprietary interfaces in check. Thus, this study evaluates the effectiveness of adopting ZTA in the COVID-19 era to reduce the risk of cyber threats and incidence.

7.6 Survey Objectives-Adoption of ZTA

We conducted a survey across multiple mediums (email, SMS, and web surveys), incorporating qualitative and quantitative data to offer a qualifying and justifiable argument about the role of ZTA in reducing cyber threats during the coronavirus pandemics.

7.6.1 Sampling and Sample Size

The survey involved 14 companies sampled through the purposeful sampling method and then grouped into two categories. The first category of seven companies, those using VPNs, is the primary architecture in cybersecurity.

7.6.2 Data Collection and Analysis

The survey's data collection process entails asking selected individuals from the identified companies to answer relevant questions aimed to determine the degree of cybersecurity and safety offered by ZT networks compared to VPN strategies. The data collection was undertaken by sending the survey tool to the information technologists or persons responsible for maintaining cybersecurity in these organizations. Once they completed the questionnaire, they returned them via email for data extraction and analysis. The qualitative data were analyzed thematically, and the quantitative data were analyzed using SPSS version 22 to yield descriptive and appropriate inferential statistics.

7.7 Survey Results

When staff are working from home, it is more difficult and complex to safeguard a company's network perimeter than when staff are logging in from a single location. Figs. 2 and 3 show the results of the first and second questions of the survey. In Figure 7.2, at-risk devices mean unknown, unsanctioned or non-compliant endpoints, and the top challenge of companies in terms of securing access to applications and resources is the complex manual process. So, the ability to react quickly is an important feature for practical usage. In Figure 7.3, application-specific access is based on a user's identity, device posture, and group membership, and most in most companies it was the chosen method.

Network security is contingent on a strong perimeter security model which has a strong outer defence. To ensure the safety of the network, the perimeter needs to be impenetrable using methods such as VPN, network segmentation and firewalls. However, security is not guaranteed. Figure 7.4 illustrates the results of the third

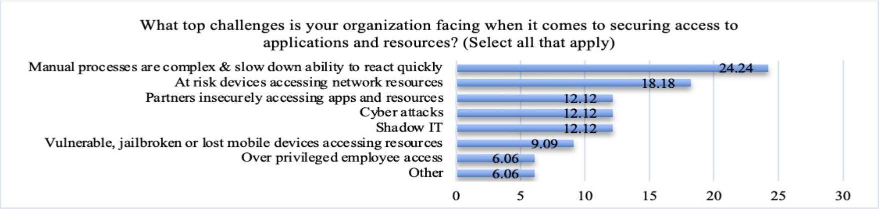


Figure 7.2 : Results for the first question in the survey

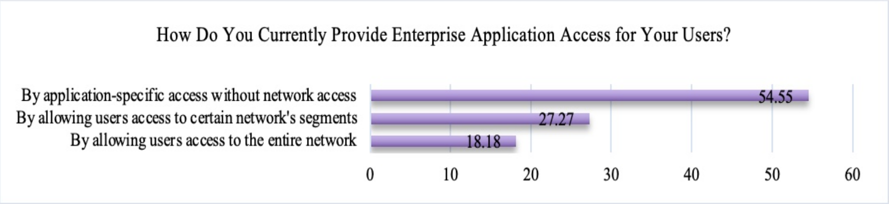


Figure 7.3 : Results for the second question in the survey

question.

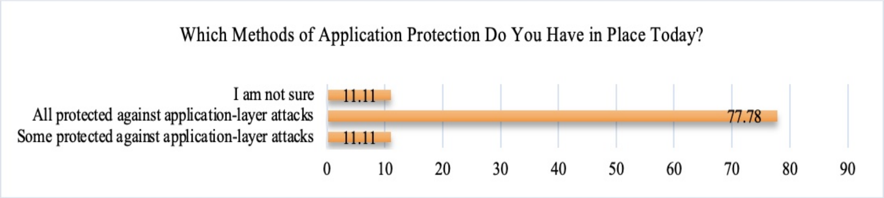


Figure 7.4 : Results for the third question in the survey

As shown in Figure 7.4, no one chose the “None of our corporate applications are protected against application-layer attacks” option, and the majority indicated that their corporate applications are protected against application-layer attacks with a Web Application Firewall (WAF). Thus, WAF is the most-used method for protection against application-layer attacks. Figure 7.5 shows the results for the fourth survey question In Figure 7.5, the results for the fourth survey question are presented, where ’entity verification’ encompasses users, devices, infrastructure, and similar components.

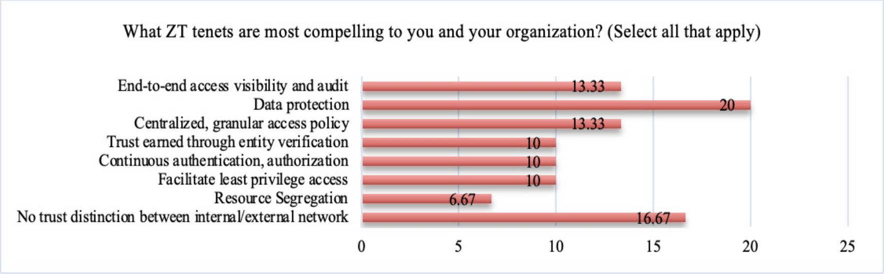


Figure 7.5 : Results for the fourth question in the survey

As ZTA eliminates trust from the network, it amplifies the trust in the users, devices, and services which is possible via undeterred authorization, authentication, and encryption. Its efficiency arises from its principle of authenticating each user connecting to the server regardless of where the access request is generated from. For effective use, the authentication and authorization levels and access policies should be well-defined, taking into consideration all circumstances. The trust degree depends on the data value magnitude and impact of the performed action. Implementing ZTA on traditional systems is difficult and it needs to be installed in phases with iterations. Once the new approach foundation is laid over the legacy system, it is easier to further build on. Establishing a strong identity for users and devices or deploying modern authentication across the organization can be time-consuming.

The survey results show that 66% of the respondents were neutral about adopting ZTA while 33% were in the satisfied to extremely satisfied spectrum. Around 60% of the respondents plan to implement ZTA capabilities on-premises and SaaS. Figure 7.6 shows the results of the fifth question.

Many organizations have implemented VPNs, however these organizations may experience data breaches in the absence of regular patching, updates, and the implementation of MFA for remote access accounts. To address these challenges, IT teams attempt to make access secure. Around 50% of companies provide applica-

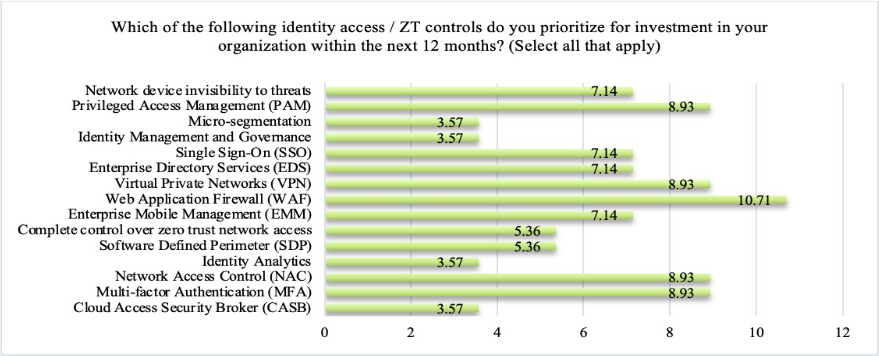


Figure 7.6 : Results for the fifth question in the survey

tion access by allowing users to access only a certain network segment, however, 38% of organizations allow application-specific access, without network access post-authorization scrutiny. Figure 7.7 shows the results of the sixth question. While

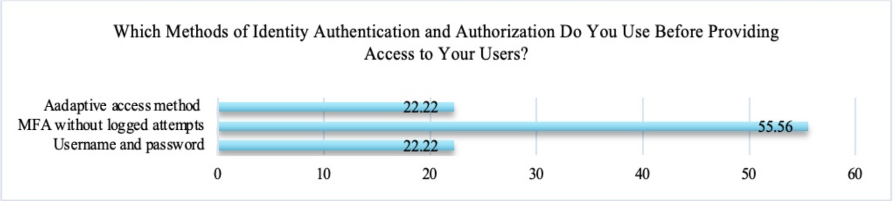


Figure 7.7 : Result for the sixth question in the survey

transitioning to a new architecture, it is not suitable to decommission traditional security controls before ZT controls have been implemented and tested. Due to the nature of ZTA, it may leave systems exposed to considerable risk if they are not properly configured. Thus, it is vital not to dismantle VPN until ZTA performs satisfactorily. VPN can manage the potential threats if needed. Systems may be hosted using a traditional architecture or they might not support the features of ZTA. So, some environments may need to manage both a traditional perimeter and a ZTA. This could involve using a split VPN tunnel to the legacy application or an authentication proxy. Figures 7.8, 7.9, 7.10 and 7.11 present the results for 7th-10th questions.

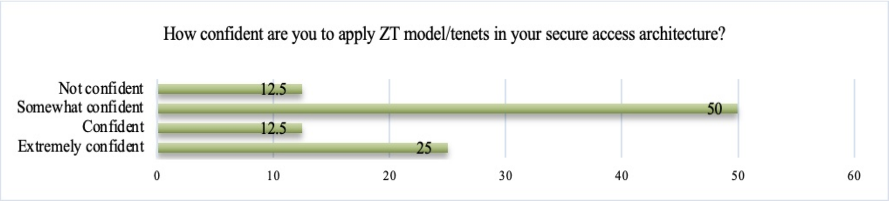


Figure 7.8 : Results for the seventh question in the survey

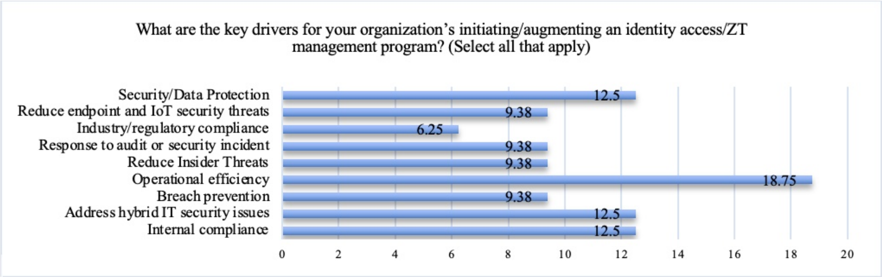


Figure 7.9 : Results for the eighth question in the survey

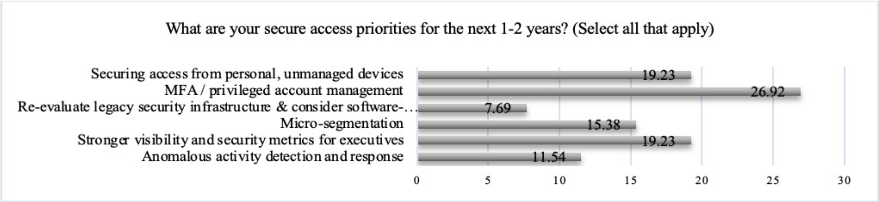


Figure 7.10 : Results for the ninth question in the survey

Figure 7.8 demonstrates that over 85% of the surveyed participants express confidence in applying the ZT model to some extent. This highlights the limitations of traditional approaches and the potential benefits of using this model. As shown in Figure 7.9, regulatory compliance could be HIPAA, GDRP, PCI DSS, etc. Furthermore, it shows that the most important management key driver for companies is operational efficiency. As shown in Figure 7.10, the least appropriate secure access

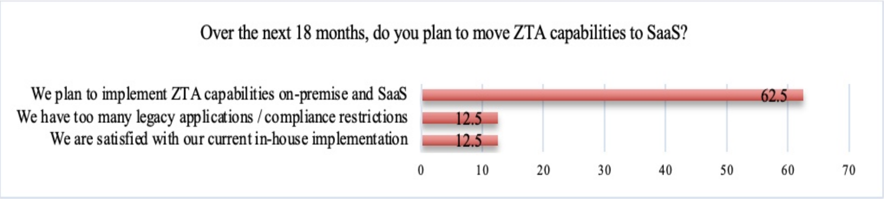


Figure 7.11 : Results for the tenth question in the survey

priority for companies is re-evaluating legacy security infrastructure and considering software-defined access. Furthermore, no one choose “Significant – we plan to solely use SaaS-based ZT access capabilities” option.

7.8 Conclusion

There is always a potential risk when accessing a corporate network or data, even when stringent security protections are in place. COVID-19 taught businesses to be resilient and to prepare for uncertainty. Hence reliability on ZTA in terms of cybersecurity will enhance future alertness and adaptability. To summarise, the disadvantages of perimeter-based security models are as follows:

1) Perimeter security largely ignores the insider threat.

2) The impenetrable fortress model fails in practice.

3) Network segmentation is time-consuming and difficult.

4) Defining a network perimeter is difficult in a remote-work, BYOD multi-cloud world.

5) VPNs are often misused and exacerbate the further issues.

ZT attempts to mitigate these shortcomings by implementing the following principles:

1) Trust flows from identity, device-state, and context, not network location.

2) internal and external networks should be treated as untrusted.

3) Act like somebody already breached because they probably are.

4) Each device, user, and application must be authenticated, authorized, and encrypted.

5) Access policies should be dynamic and built from multiple sources.

The literature is replete with evidence supporting the superiority of ZTA over traditional VPNs in providing the maximum level of cybersecurity and safety. However, ZTAs come with challenges that introduce complexities when implementing and utilizing them for securing personal and business data against intrusions. In an era where the reliance on cyberspace has surged, driven in part by complications arising from the pandemic, it is prudent to examine the feasibility of ZTA in safeguarding cyber interactions and transactions from malicious attacks. Ultimately, implementing ZTA is the best choice for businesses that want to give remote access to users while maintaining security. In the following, we discuss possible directions for future work in detail and conclude this thesis.

CHAPTER 8

Conclusion and Future Works

8.1 Introduction

This chapter concludes this thesis by summarizing the research findings and providing recommendations for future work in the realm of Zero Trust Architecture-Intrusion Detection System (ZTA-IDS). The journey undertaken in this thesis has been one of exploring the intersection of Zero Trust models with network security, specifically in the context of IDS. Beginning with an examination of the limitations of conventional trust-based security systems, this research has illuminated the pressing need for innovative approaches in combating rising cyber threats. As we reflect on the insights garnered and the challenges encountered, this chapter aims to synthesize these experiences, drawing a comprehensive conclusion on the research conducted and outlining a path forward for future explorations.

In Chapter 1, we embarked on this exploration by establishing the pressing need for a paradigm shift in network security approaches, highlighting the limitations of traditional trust-based systems in the face of escalating cyber threats. The subsequent chapters meticulously dissected the Zero Trust framework, examining its components, mechanisms, and implementation challenges. This journey involved a deep dive into the technical intricacies, presenting a novel ZTA-IDS model, crafted to address the specific challenges of intrusion detection in an increasingly complex cybersecurity landscape.

The innovative ZTA-IDS model, as elucidated in the thesis, marks a significant advancement in intrusion detection methodologies, balancing the often-conflicting demands of security and accuracy. Through rigorous experimentation and analysis, the model demonstrated its robustness against various types of semantic attacks, while maintaining high accuracy levels. These findings not only reinforce the viability of the Zero Trust approach but also underscore the potential for its broader application in safeguarding network infrastructures.

This chapter aims to provide a cohesive overview of the research outcomes, emphasizing the contributions of this work to the field of network security. Looking ahead, the chapter will delve into potential future works, exploring avenues for the further refinement and application of the Zero Trust model. It seeks to chart a roadmap for future researchers and practitioners, envisioning advancements that can fortify the resilience of network systems against evolving cyber threats. The journey of enhancing network security is ongoing, and this thesis aims to contribute a significant milestone to that journey.

This chapter is organized as follows: Section 8.2 presents an overview of the existing literature, Section 8.3 presents the technical aspects of Zero Trust, Multi-View (MV), and the IDS model, Section 8.4 details the MV methodology and the IDS model, Section 8.5 provides the model validation, Section 8.6 details the experiment procedure, Section 8.7 discusses Zero Trust and its importance, Section 8.8 provides a brief conclusion.

The next section summarizes the contributions of this thesis to the existing literature.

8.2 Overview of Existing Literature

This section revisits the literature on MV technology for network trace anonymization and the Zero Trust model. It highlights the initial study of the MV method and acknowledges its drawbacks. This overview sets the stage for understanding the need for improved models in network trace anonymization and the importance of Zero Trust in a robust network security framework. The discussion aligns with the thesis's initial chapters, providing a backdrop against which the subsequent innovations were developed.

8.3 Technical Aspects of Zero Trust, Multi-View (MV), and the Intrusion Detection System (IDS) Model

In Chapter 4, we extensively explored the technical aspects of Zero Trust models, with a special focus on IDS and the integration of MV approaches in network monitoring services. This exploration was crucial in understanding how Zero Trust models can be effectively applied in modern network security environments, particularly in the context of the increased reliance on network security post the Covid-19 pandemic.

The chapter began by highlighting the importance of adopting Zero Trust models in response to the evolving security landscape. We then provided a detailed explanation of how IDS can be integrated within these models to enhance network security. This was particularly significant in outsourcing scenarios, where maintaining the integrity and accuracy of IDS is essential. We also delved into the MV approach to defense, which involves analyzing network traffic from multiple perspectives. This approach is key to improving the detection capabilities of IDS, making them more resilient to sophisticated cyber threats.

Furthermore, the chapter discussed the technical specifications and effectiveness of proposed network IDS solutions, emphasizing their accuracy and reliability in third-party settings. The integration of Zero Trust principles into these systems was explored to illustrate how they can bolster security measures in outsourced network environments. Lastly, we detailed the adversary model to demonstrate the types of threats these systems are designed to counteract, underscoring the necessity of robust and adaptable IDS in contemporary network security.

Overall, this chapter provided a comprehensive overview of how Zero Trust models, coupled with advanced IDS and MV strategies, form a formidable defense against the increasingly complex landscape of network security threats. This set the stage for the next chapter, where we detail the research methodology employed in our study, building on the technical foundation laid out in this chapter.

8.4 MV Methodology and the IDS Model

In Chapter 5, we highlighted the research methodology employed in our research. The methodology has been designed to address the research gaps identified in Chapter 2. At the beginning, we first selected our research methodology. Then we explained the proposed schema, which is the ZTA-IDS schema, followed by the phases of ZTA-IDS. We also defined the criteria on security and accuracy. In addition, we introduced our proposed IDS and described each module in it. Furthermore, we explained the implementation of ZTA-IDS in detail. Finally, the challenges and solutions are listed and explained. Overall in this chapter, we covered the methodology and in the following chapter, we highlighted the experimental performance and data analyses.

8.5 Model Validation

This section is dedicated to the validation of the proposed models in a real-world context. To assess the adaptability and robustness of our models, we conducted an additional experiment using a different server and attack tools to those used in the initial test phase. A total of 500 samples were collected across various categories, each with corresponding labels, to create a test set that closely mirrors real-world scenarios.

The trained model underwent a fine-tuning process for 2 epochs, and its performance on this real-world test data was meticulously analyzed. The comparison between the adapted model and the non-adapted model is presented in Table 7.4, showcasing the changes in accuracy. Notably, the adapted model exhibited a significant improvement in accuracy on real-world data, with an increase from 50% to 80%. This marked enhancement underscores the model's capability to adapt and maintain high performance in diverse and realistic environments.

However, the results also highlight a critical limitation: the discrepancy between the dataset used for initial training and real-world scenarios. This disparity suggests that incorporating data more representative of actual network environments could further improve the model's effectiveness and generalization capabilities. The findings from this model validation process offer valuable insights into the potential improvements and adaptations needed for the model to perform optimally in real-world settings. It emphasizes the importance of diverse and realistic data in training robust and effective IDS.

8.6 Experiment Procedure

This section delves into the details of the experimental procedures employed in the thesis, with a specific focus on the application and analysis of the UNSW-NB15 dataset. It outlines the experimental setup, describing the process and criteria used to assess the performance of the proposed intrusion detection models. Crucially, this section highlights the impressive results obtained: our proposed approach, particularly the Attention CNN with Bi-directional Long Short Term Memory (CNNBiL-

STM) model, demonstrated superior performance in comparison to state-of-the-art approaches. The model achieved classification accuracies of 89.79% and 91.72% for 6 and 10 categories, respectively, on the test set of the UNSW-NB15 dataset. These results are pivotal in demonstrating the efficacy and robustness of the proposed model, marking a significant advancement in the field of network intrusion detection.

8.7 Discussion on Zero Trust and Its Importance

Here, we re-emphasize the philosophy and necessity of the Zero Trust model in modern network security, as introduced in the early chapters of the thesis. This section synthesizes the discussions on the foundational principles of Zero Trust, illustrating its critical role in ensuring a secure and resilient network environment, especially in the context of intrusion detection.

8.8 Conclusion

As we reach the conclusion of this thesis, it is essential to reflect on the journey that began with the introduction of the philosophy and need for a Zero Trust model in Chapter 1. Our research embarked on exploring advanced methodologies in network trace anonymization, addressing the drawbacks of existing models. We successfully developed and implemented a deep neural network model, integrating an autoencoder and CNN for feature extraction and classification, which has shown significant progress in network intrusion detection.

Through meticulous experimentation, particularly using the UNSW-NB15 dataset, our findings have been nothing short of promising. The proposed methods for network trace anonymization have demonstrated superior performance over the original N Key Vector approach, especially for a smaller number of groups. Moreover, our innovative deep neural network models, including the Attention CNN with Bi-directional Long Short Term Memory (CNNBiLSTM), achieved remarkable classification accuracies of 89.79% and 91.72% for 6 and 10 categories, respectively. These results not only outperformed previous works but also validated the efficacy of our approach in real-world scenarios.

However, the validation process revealed some limitations, particularly the discrepancy between our dataset and real-world scenarios. The adapted model's performance on real-world test data, improving accuracy from 50% to 80%, underscored the potential for further enhancements. This has affirmed that our initial objectives, as planned in Chapter 1, have been successfully met through our experiments and results.

As we advance towards the culmination of this thesis, the focus shifts to future work in enhancing and extending the capabilities of our current models and systems. The field of network security, especially in IDS and IPS, demands continuous evolution to counter sophisticated cyber threats. Our roadmap for future development includes:

- **Comprehensive Attack Simulation:** Implementing a variety of attack methods across multiple servers and clients to generate more authentic training data.
- **Model Fine-Tuning with Collected Data:** Utilizing the data from these simulations to fine-tune our IDS model, ensuring it remains adaptive to evolving attack patterns.
- **Integration with Firewall or IPS Systems:** Combining our ZTA-IDS with existing firewall or IPS systems to create a comprehensive application for real-time network security management.
- **System Validation through Web Services:** Validating the integrated sys-

tem within a web service environment to assess its practical efficacy in a live setting.

In conclusion, this thesis represents a significant stride forward in the quest for a more secure and resilient network environment. By integrating advanced techniques and continuously refining our approach based on real-world data, we pave the way for future advancements in cybersecurity.

8.9 Future Work

In this chapter, we rigorously validated our ZTA-IDS solution, highlighting the advancements in network security within the Zero Trust framework. Despite the progress made in this area, challenges remain, notably in privacy preservation, real-world application, and the risk of overfitting and poor generalization in machine learning models. Our method innovatively combines MV anonymization with advanced neural network models, addressing these limitations to offer a robust solution for modern networks.

The cross-domain adaptation remains a significant hurdle in machine learning, as models often underperform when applied to real-world data, a phenomenon known as domain shift. Our work addresses this by bridging the gap between theoretical models and practical application, aiming to enhance the adaptability and effectiveness of machine learning across various domains.

Bibliography

[1] M. Mohammady, L. Wang, Y. Hong, H. Louafi, M. Pourzandi, and M. Debbabi, "Preserving both privacy and utility in network trace anonymization," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, 2018, pp. 459–474.

[2] J. Kindervag *et al.*, "Build security into your network's dna: The zero trust network architecture," *Forrester Research Inc*, vol. 27, 2010.

[3] D. Puthal, S. P. Mohanty, P. Nanda, and U. Choppali, "Building security perimeters to protect network systems against cyber threats [future directions nnaderi]," *IEEE Consumer Electronics Magazine*, vol. 6, no. 4, pp. 24–27, 2017.

[4] P. Assunção, "A zero trust approach to network security nnaderi," in *Proceedings of the Digital Privacy and Security Conference*, vol. 2019. Porto Protugal, 2019.

[5] E. U. A. for Cybersecurity, *ENISA threat landscape report 2018 – 15 top cyber-threats and trends.* European Network and Information Security Agency, 2019.

[6] P. Kampanakis, “Security automation and threat information-sharing options,” *IEEE Security & Privacy*, vol. 12, no. 5, pp. 42–51, 2014.

[7] W. Newhouse, S. Keith, B. Scribner, and G. Witte, “National initiative for cybersecurity education (nice) cybersecurity workforce framework,” *NIST special publication*, vol. 800, no. 2017, p. 181, 2017.

[8] A.-I. Z. T. P. Team, “Zero trust cybersecurity current trends,” 2019. [Online]. Available: https://www.actiac.org/documents/zero-trust-cybersecurity-current-trends

[9] M. C. Paulk, B. Curtis, M. B. Chrissis, and C. V. Weber, “Capability maturity model, version 1.1,” *IEEE software*, vol. 10, no. 4, pp. 18–27, 1993.

[10] N. Tohidi and C. Dadkhah, “A short review of abstract meaning representation applications nnaderi,” *Modeling and Simulation in Electrical and Electronics Engineering*, vol. 2, no. 3, pp. 1–9, 2022.

[11] C. Griffy-Brown, D. Lazarikos, and M. Chun, “How do you secure an environment without a perimeter? using emerging technology processes to support information security efforts in an agile data center.” *Journal of Applied Business & Economics*, vol. 18, no. 1, 2016.

[12] J. Srinivas, A. K. Das, and N. Kumar, “Government regulations in cyber security: Framework, standards and recommendations,” *Future generation computer systems*, vol. 92, pp. 178–188, 2019.

[13] O. Adeyinka, “Internet attack methods and internet security technology,” in *2008 Second Asia International Conference on Modelling & Simulation (AMS)*. IEEE, 2008, pp. 77–82.

[14] B. Daya, “Network security: History, importance, and future,” *University of Florida Department of Electrical and Computer Engineering*, vol. 4, 2013.

[15] C. DeCusatis, P. Liengtiraphan, A. Sager, and M. Pinelli, "Implementing zero trust cloud networks with transport access control and first packet authentication," in *2016 IEEE International Conference on Smart Cloud (SmartCloud)*. IEEE, 2016, pp. 5–10.

[16] D. Eidle, S. Y. Ni, C. DeCusatis, and A. Sager, "Autonomic security for zero trust networks," in *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*. IEEE, 2017, pp. 288–293.

[17] 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.

[18] R. Heartfield and G. Loukas, "A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks," *ACM Computing Surveys (CSUR)*, vol. 48, no. 3, pp. 1–39, 2015.

[19] P. Bienias, A. Warzyński, and G. Kołaczek, "Application and preliminary evaluation of anontool applied in the anomaly detection module," in *2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*. IEEE, 2020, pp. 119–123.

[20] Y. Li, A. Slagell, K. Luo, and W. Yurcik, "Canine: A combined conversion and anonymization tool for processing netflows for security," in *International conference on telecommunication systems modeling and analysis*, vol. 21. Citeseer, 2005.

[21] A. J. Slagell, K. Lakkaraju, and K. Luo, "Flaim: A multi-level anonymization framework for computer and network logs." in *LISA*, vol. 6, 2006, pp. 3–8.

[22] S. Keele *et al.*, "Guidelines for performing systematic literature reviews in software engineering," *ver. 2.3 EBSE Technical report, Software Engineering Group, Keele University and Department of Computer Science, University of Durham*, p. 65, 2007.

[23] L. Coppolino, S. D'Antonio, V. Formicola, G. Mazzeo, and L. Romano, "Vise: Combining intel sgx and homomorphic encryption for cloud industrial control systems," *IEEE Transactions on Computers*, vol. 70, no. 5, pp. 711–724, 2020.

[24] X. Jin, H. Zhang, X. Li, H. Yu, B. Liu, S. Xie, A. K. Singh, and Y. Li, "Confused-modulo-projection-based somewhat homomorphic encryption—cryptosystem, library, and applications on secure smart cities," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6324–6336, 2020.

[25] H. Pang and B. Wang, "Privacy-preserving association rule mining using homomorphic encryption in a multikey environment," *IEEE Systems Journal*, vol. 15, no. 2, pp. 3131–3141, 2020.

[26] J. H. Cheon and J. Kim, "A hybrid scheme of public-key encryption and somewhat homomorphic encryption," *IEEE transactions on information forensics and security*, vol. 10, no. 5, pp. 1052–1063, 2015.

[27] M. Kim, H. T. Lee, S. Ling, B. H. M. Tan, and H. Wang, "Private compound wildcard queries using fully homomorphic encryption," *IEEE Transactions on Dependable and Secure Computing*, vol. 16, no. 5, pp. 743–756, 2017.

[28] G. Qiu, X. Gui, and Y. Zhao, "Privacy-preserving linear regression on distributed data by homomorphic encryption and data masking," *IEEE Access*, vol. 8, pp. 107 601–107 613, 2020.

[29] M. Yuan, D. Wang, F. Zhang, S. Wang, S. Ji, and Y. Ren, "An examination

of multi-key fully homomorphic encryption and its applications," *Mathematics*, vol. 10, no. 24, p. 4678, 2022.

[30] G. Alagic, Y. Dulek, C. Schaffner, and F. Speelman, "Quantum fully homomorphic encryption with verification," in *Advances in Cryptology–ASIACRYPT 2017: 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part I 23*. Springer, 2017, pp. 438–467.

[31] J. Dyer, M. Dyer, and J. Xu, "Practical homomorphic encryption over the integers for secure computation in the cloud," *International Journal of Information Security*, vol. 18, pp. 549–579, 2019.

[32] T. K. Saha, M. Rathee, and T. Koshiba, "Efficient private database queries using ring-lwe somewhat homomorphic encryption," *Journal of Information Security and Applications*, vol. 49, p. 102406, 2019.

[33] J. Xu, L. Wei, Y. Zhang, A. Wang, F. Zhou, and C.-z. Gao, "Dynamic fully homomorphic encryption-based merkle tree for lightweight streaming authenticated data structures," *Journal of Network and Computer Applications*, vol. 107, pp. 113–124, 2018.

[34] D. Catalano and D. Fiore, "Boosting linearly-homomorphic encryption to evaluate degree-2 functions on encrypted data," *Cryptology ePrint Archive*, 2014.

[35] R. Bellafqira, G. Coatrieux, D. Bouslimi, G. Quellec, and M. Cozic, "Proxy re-encryption based on homomorphic encryption," in *Proceedings of the 33rd Annual Computer Security Applications Conference*, 2017, pp. 154–161.

[36] A. Ahmed and H. S. Ogalo, "From hrm to e-hrm: Contemporary

developments from scholarly work," *Annals of Contemporary Developments in Management & HR (ACDMHR)', Print ISSN*, pp. 2632–7686, 2019.

[37] Ç. K. Koç, F. Özdemir, and Z. Ö. Özger, *Partially Homomorphic Encryption.* Springer, 2021.

[38] M. A. Will and R. K. Ko, "A guide to homomorphic encryption," 2015.

[39] S. M. Kasongo and Y. Sun, "Performance analysis of intrusion detection systems using a feature selection method on the unsw-nb15 dataset," *Journal of Big Data*, vol. 7, pp. 1–20, 2020.

[40] B. A. Tama, M. Comuzzi, and K.-H. Rhee, "Tse-ids: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system," *IEEE access*, vol. 7, pp. 94 497–94 507, 2019.

[41] K. Jiang, W. Wang, A. Wang, and H. Wu, "Network intrusion detection combined hybrid sampling with deep hierarchical network nnaderi," *IEEE access*, vol. 8, pp. 32 464–32 476, 2020.

[42] J. C. Mogul and M. Arlitt, "Sc2d: an alternative to trace anonymization," in *Proceedings of the 2006 SIGCOMM workshop on Mining network data*, 2006, pp. 323–328.

[43] P. Mittal, V. Paxson, R. Sommer, and M. Winterrowd, "Securing mediated trace access using black-box permutation analysis." in *HotNets*, 2009.

[44] F. McSherry and R. Mahajan, "Differentially-private network trace analysis," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 4, pp. 123–134, 2010.

[45] S. R. Pokhrel and J. Choi, "Federated learning with blockchain for

autonomous vehicles: Analysis and design challenges," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4734–4746, 2020.

[46] J. Daemen and V. Rijmen, "Aes proposal: Rijndael," 1999, aES Algorithm Submission. Available online: http://csrc.nist.gov/archive/aes/rijndael/Rijndael-ammended.pdf.

[47] T. Mayberry, B. Erik-Oliver, and C. Agnes Hui, "Efficient private file retrieval by combining oram and pir," *In NDSS*, 2014.

[48] W. Stallings, *Cryptography and Network Security: Principles and Practice*, 7th ed. Pearson, 2017, this book provides comprehensive coverage on network security, including discussions on cryptographic techniques and privacy-preserving methods across network protocols and applications.

[49] J. Xu, J. Fan, M. H. Ammar, and S. B. Moon, "Prefix-preserving ip address anonymization: Measurement-based security evaluation and a new cryptography-based scheme," in *Proceedings of the 10th IEEE International Conference on Network Protocols.* IEEE, 2002, pp. 280–289, this paper introduces the CryptoPAn scheme for IP address anonymization, detailing its prefix-preserving properties and implications for network data analysis.

[50] M. Abolghasemi, C. Dadkhah, and N. Tohidi, "Hts-dl: hybrid text summarization system using deep learning," in *2022 27th International Computer Conference, Computer Society of Iran (CSICC).* IEEE, 2022, pp. 1–5.

[51] M. Ghezelji, N. Tohidi, C. Dadkhah, and A. Gelbukh, "Personality-based matrix factorization for personalization in recommender systems," *International Journal of Information and Communication Technology Research*, vol. 14, pp. 48–65, 2022.

[52] V. Kumar, D. Sinha, A. K. Das, S. C. Pandey, and R. T. Goswami, "An integrated rule based intrusion detection system: analysis on unsw-nb15 data set and the real time online dataset," *Cluster Computing*, vol. 23, pp. 1397–1418, 2020.

[53] B. Cao, C. Li, Y. Song, Y. Qin, and C. Chen, "Network intrusion detection model based on cnn and gru," *Applied Sciences*, vol. 12, no. 9, p. 4184, 2022.

[54] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of management information systems*, vol. 24, no. 3, pp. 45–77, 2007.

[55] J. Kim, J. Kim, H. L. T. Thu, and H. Kim, "Long short term memory recurrent neural network classifier for intrusion detection," in *2016 international conference on platform technology and service (PlatCon).* IEEE, 2016, pp. 1–5.

[56] N. Moustafa and J. Slay, "Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set)," in *2015 military communications and information systems conference (MilCIS).* IEEE, 2015, pp. 1–6.

[57] M. Modderkolk *et al.*, "Zero trust maturity matters: Modeling cyber security focus areas and maturity levels in the zero trust principle," Master's thesis, 2018.

[58] D. Jing and H.-B. Chen, "Svm based network intrusion detection for the unsw-nb15 dataset," in *2019 IEEE 13th international conference on ASIC (ASICON).* IEEE, 2019, pp. 1–4.

[59] Y. Yin, J. Jang-Jaccard, W. Xu, A. Singh, J. Zhu, F. Sabrina, and J. Kwak, "Igrf-rfe: a hybrid feature selection method for mlp-based network intrusion

detection on unsw-nb15 dataset," *Journal of Big Data*, vol. 10, no. 1, pp. 1–26, 2023.

[60] H. S. Lallie, L. A. Shepherd, J. R. Nurse, A. Erola, G. Epiphaniou, C. Maple, and X. Bellekens, "Cyber security in the age of covid-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic," *Computers & security*, vol. 105, p. 102248, 2021.

[61] T. Ahmad, "Corona virus (covid-19) pandemic and work from home: Challenges of cybercrimes and cybersecurity," *Available at SSRN 3568830*, 2020.

[62] A. Hajj and M. Rony, "Cyber security in the age of covid-19: An analysis of cyber-crime and attacks," *International Journal for Research in Applied Science & Engineering Technology*, vol. 8, no. 8, pp. 1476–1480, 2020.

[63] S. Krishnaveni, S. Sivamohan, S. Sridhar, and S. Prabhakaran, "Network intrusion detection based on ensemble classification and feature selection method for cloud computing," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 11, p. e6838, 2022.

[64] N. Tohidi and R. B. Rustamov, "A review of the machine learning in gis for megacities application," *Geographic Information Systems in Geospatial Intelligence*, pp. 29–53, 2020.

[65] D. Preethi and N. Khare, "Sparse auto encoder driven support vector regression based deep learning model for predicting network intrusions," *Peer-to-Peer Networking and Applications*, vol. 14, pp. 2419–2429, 2021.

[66] M. Almiani, A. AbuGhazleh, A. Al-Rahayfeh, S. Atiewi, and A. Razaque, "Deep recurrent neural network for iot intrusion detection system," *Simulation Modelling Practice and Theory*, vol. 101, p. 102031, 2020.

[67] I. Lee, D. Kim, and S. Lee, "3-d human behavior understanding using generalized ts-lstm networks," *IEEE Transactions on Multimedia*, vol. 23, pp. 415–428, 2020.

[68] I. Al-Turaiki and N. Altwaijry, "A convolutional neural network for improved anomaly-based network intrusion detection," *Big Data*, vol. 9, no. 3, pp. 233–252, 2021.

[69] V. Stafford, "Zero trust architecture," *NIST special publication*, vol. 800, p. 207, 2020.

[70] X. Yan and H. Wang, "Survey on zero-trust network security," in *Artificial Intelligence and Security: 6th International Conference, ICAIS 2020, Hohhot, China, July 17–20, 2020, Proceedings, Part I 6.* Springer, 2020, pp. 50–60.

[71] K. Uttecht, "Zero trust (zt) concepts for federal government architectures," *Department of Homeland Security (DHS) Science and Technology Directorate (S&T), Lexington, Massachusetts*, 2020.