Enhancing Trust Modelling for the Internet of Underwater Things

Thesis submitted to the University of Nottingham in partial fulfilment of the degree of Doctor of Philosophy, May 2025

Abeer Almutairi

20405266

Supervised by

Prof Steven Furnell

Dr Xavier Carpent

School of Computer Science

Signature ___________________

Date ______ / ______ / ______

Abstract

The Internet of Underwater Things (IoUT) has gained growing interest from researchers and industry alike, due to its potential for advancing the development of smart cities and underwater intelligent systems. However, the harsh and unpredictable nature of underwater environments, coupled with the inherent limitations of existing technologies, presents significant challenges to establishing a sustainable IoUT. Furthermore, the open nature of such networks renders them highly susceptible to malicious attacks and security threats. Traditional security measures, which are widely implemented in conventional cyber systems, exhibit severe performance constraints in underwater networks, highlighting the urgent need for novel security solutions that meet the unique requirements of underwater networks. Trust modelling has been widely recognised as an effective soft security measure to mitigate the impact of internal attacks. It primarily achieves this by analysing behavioural characteristics between network entities, thereby introducing a layer of defence against malicious activities. In the context of underwater networks, trust establishment between nodes has the potential to significantly enhance overall network security. However, existing Trust Modelling and Management (TMM) often fail to address the complexities of underwater environments, which necessitate new TMM that are lightweight, accurate, and decentralised. In light of these limitations, this thesis investigates and enhances TMM to meet the application requirements of underwater networks while addressing the specific challenges inherent to IoUT. The central research question addressed in this thesis is: To what extent can existing TMM accommodate diverse network topologies within the IoUT and effectively mitigate potential attacks from both the communication and physical domains. In order to answer this question, a comprehensive understanding of the key challenges and potential application requirements for underwater networks is required. To facilitate this investigation, a simulated environment is constructed to analyse the effectiveness of TMM. This study critically evaluates the capabilities of current TMM in detecting malicious activities across various underwater network structures, identifying vulnerabilities, and exposing potential attack vectors. In response to these findings, this thesis proposes a distributed multi-dimensional TMM, referred to as the Mobility-Aware Trust Model (MATMU), designed to enhance the detection of malicious behaviour within the constraints of underwater environments. MATMU expands the metric domain to include mobility-aware metrics, allowing for the assessment of similarities and differences in node movement patterns. Additionally, the model employs a dynamic weighting strategy that integrates metrics from both the communication and physical domains. The performance of MATMU is evaluated through extensive simulations conducted across various underwater scenarios and attack models. The results demonstrate that MATMU effectively mitigates malicious behaviour, exhibiting notable improvements over benchmark models, particularly in terms of faster convergence and enhanced attack detection. These findings underscore the suitability of MATMU for strengthening secure and reliable communication in underwater networks. This thesis also tackles the critical issue of dishonest recommendations within TMM in the IoUT context, which is introduced by malicious entities, aiming to manipulate trust computations by providing false or misleading recommendations, thereby degrading the reliability and stability of the TMM. A novel recommendation evaluation method is introduced, combining filtering and weighting strategies to more effectively detect dishonest recommendations. The proposed model incorporates an outlier detection-based filtering technique and deviation analysis to evaluate recommendations based on both collective outcomes and individual experiences. Furthermore, a belief function is employed to refine recommendations by assigning weights based on criteria such as freshness, similarity, trustworthiness, and trust decay over time. This multi-dimensional approach demonstrates a marked improvement in recommendation evaluation, effectively capturing deceptive behaviours that exploit the complexities of IoUT. The effectiveness of the model is validated through extensive simulations and comparative analyses with existing trust evaluation methods, demonstrating consistently high performance across varying proportions of dishonest recommendations, with the highest accuracy improvement observed when dishonest recommendations constitute up to 45% of the total recommendations. These findings underscore the model's potential to significantly enhance the reliability and security of IoUT networks.

Acknowledgements

First and foremost, I am deeply thankful to God for leading me to pursue my studies at the University of Nottingham. Embarking on this journey far from my family was undoubtedly challenging, but the warm and supportive environment at the university made it much easier to adapt and thrive. It is here, with the patience, encouragement, and support of many individuals, that I had the privilege of undertaking my PhD studies under the expert supervision of Professor Steven Furnell and Dr Xavier Carpent. To them, I extend my deepest gratitude for their invaluable guidance and unwavering support throughout this journey. This experience has not only shaped me into a more confident individual but also a skilled researcher, a contributing author and reviewer, and a part of a vibrant global academic community, for which I am truly thankful. I am also grateful to Dr Ying He for her role in shaping the early direction of this work. I am also immensely grateful to King Abdulaziz University for granting me the opportunity to pursue my studies abroad and for their financial support throughout this academic endeavour.

My heartfelt thanks go to the friends who have made this journey smoother and more enjoyable, their companionship turning challenges into cherished memories. Lastly, and most importantly, I owe a special debt of gratitude to my family, whose support has been beyond measure. My deepest thanks go to my beloved mother, Ayda, whose endless love and prayers have been my guiding light; my wise father, Mohammed, whose dedicated encouragement and guidance have been invaluable; all of my brothers and sisters for their relentless support, with a special thanks to my little sister, Nouf, whose constant encouragement and positivity have uplifted me; and my husband, Abdullah, whose love, patience, and understanding have been the cornerstone of my success. Without their unwavering support, completing this journey would not have been possible.

Contents

List of Tables

List of Figures

List of Abbreviations

Ad Hoc Ad Hoc Network.

Aqua-Sim ng Aqua-Sim Next Generation.

AUV Autonomous Underwater Vehicle.

BM Bad-Mouthing Attack.

BS Ballot-Stuffing Attack.

CATM Controversy-adjudication-based Trust Management.

CFFTM Collaborative Filtering Trust Mechanism.

dB decibels.

DoS Denial of Service Attack.

IoT Internet of Things.

IoUT Internet of Underwater Things.

LOF Local Outlier Factor.

MAD Median Absolute Deviation.

MANET Mobile Ad-hoc Network.

MATMU Mobility-Aware Trust Model.

MTACMM Multi-Domain Trust Assessment in Collaborative Marine MANETs.

PDR Packet Delivery Ratio.

PLR Packet Loss Ratio.

PM Physical Mobility Attack.

QoS Quality of Service.

ROV Remotely Operated Vehicle.

SB Selfish Behaviour Attack.

SF Selective Forwarding Attack.

SSP Sound Speed Profile.

TMM Trust Modelling and Management.

UAC Underwater Acoustic Communication.

UN Underwater Node.

UWSN Underwater Wireless Sensor Network.

WOA World Ocean Atlases.

Relevant Publications

- Almutairi, A., Carpent, X. and Furnell, S., 2024, June. Towards a Mobility-Aware Trust Model for the Internet of Underwater Things. In IFIP International Conference on ICT Systems Security and Privacy Protection (pp. 1-15). Cham: Springer Nature Switzerland.

- Almutairi, A., Carpent, X. and Furnell, S., 2024. Recommendation-Based Trust Evaluation Model for the Internet of Underwater Things. Future Internet, 16(9), p.346.

- Almutairi, A., He, Y. and Furnell, S., 2022, July. A Multi-Level Trust Framework for the Internet of Underwater Things. In 2022 IEEE International Conference on Cyber Security and Resilience (CSR) (pp. 370-375). IEEE.

Chapter 1

Introduction

1.1 Context

The Internet of Underwater Things (IoUT) is an emerging communication system developed to connect various underwater sensing devices that are deployed within underwater environments such as oceans, seas, and rivers. The concept of the IoUT includes many smart underwater objects that collect, distribute, disseminate, and collate data in broad underwater areas by following an ad-hoc fashion and offloading the data to above the water sink nodes, i.e. floating base stations, ships, and such (Khalil et al. 2020). A wide variety of underwater sensors and aquatic vehicles have been developed, which are characterised by their different capabilities and mobilities. The Underwater Node (UN) refers to the utilised entity/sensor/device—fixed and/or mobile— to handle the communication and data collection underwater. Recently, the development of UNs has accelerated, with an overall estimation of growth on the global market of underwater devices by 2026 being forecast to be around $4.3 Billion (Markets 2022). This in turn offers the potential for promising applications that enable extensive control and monitoring of underwater environments, aiding in the prediction of natural phenomena, including catastrophic events that can impact both human and aquatic life. Indeed, the potential applications of IoUT can spread across several domains, including but not limited to underwater oil extraction and monitoring, fish farm production, underwater data centres and storage, surveillance, and environmental monitoring.

Despite the conceptual equivalence with the well-known Internet of Things (IoT), fully distributed underwater wireless communication networks have yet to achieve their full potential (Bello et al. 2022). This is primarily due to the differences in propagation environments, as signal transmission in water presents significantly greater challenges compared to air. According to Zhu et al. (2023), existing underwater wireless communication schemes—including electromagnetic, optical, and acoustic methods—exhibit significant signal propagation impairments that limit their applicability. Among these, Underwater Acoustic Communication (UAC) currently best addresses the IoUT's needs regarding data rate and transmission range. However, the inherent complexity and dynamic nature of the underwater environment, which directly influences the physical communication medium, introduces several challenges that need to be addressed. For instance, UAC is characterised by high propagation delay, as the speed of sound underwater is approximately $1500m/s$, which varies regionally, as well as high attenuation and is easily exposed to noisy environments which, in turn, are highly frequency-dependent. Consequently, UAC is better suited to low-frequency transmissions, which inherently limits the available bandwidth. In addition to these challenges, the continuous movement of water due to currents, combined with the expected mobility of UNs, reduces the stability of connections between nodes. In fact, the dynamic environment and different expectations of mobility of UNs raises the likelihood of miscommunication. Furthermore, UNs are typically battery-powered, and there is currently no efficient method for frequently recharging or replacing these batteries. The nature of underwater network deployments also makes such projects resource-intensive and operationally expensive. Due to these costs, underwater networks tend to be sparsely deployed (Jahanbakht et al. 2021). Given these challenges, and despite the conceptual parallels with terrestrial counterparts, the development of the IoUT must consider significant performance constraints. These constraints highlight the importance of ongoing research efforts, which increasingly focus on improving the efficiency and reliability of underwater networks and their associated domains.

As with other network environments, the use of the IoUT brings associated considerations in terms of security. The security requirements for the IoUT, particularly in sensitive applications such as surveillance, border protection, military operations, and self-defence, are greatly influenced by vulnerabilities inherent in wireless communication. Because of the infrastructure-less and broadcast nature of the IoUT, UNs are less reliable, more prone to failure, and highly susceptible to security attacks, including misleading and selfish attacks, to name a few. Consequently, when an adversarial entity compromises a UN, it can trigger other UNs to misbehave and perform malicious actions. This can lead to erroneous data routing by malicious nodes, ultimately resulting in network failure. When a UN becomes a rogue node, cryptographic approaches and authentication system alone are insufficient to protect other UNs from internal attacks. The influence of the internal attacks is even more alarming due to lack of physical security measures, as UNs are often left unattended in deep oceans. This, combined with the harsh underwater environment, makes it challenging to physically remove compromised nodes. Moreover, most current operations and solutions in the network rely on collaboration between UNs to enhance functionality and prolong the overall network lifetime. This dependency necessitates cooperative and ‘fair’ behaviour among underwater entities, as selfish and unfair behaviour exacerbates the threat landscape. Thus, establishing decentralised monitoring of behaviour among nodes is becoming increasingly important for improving security, trustworthiness, and successful collaboration.

Trust, in the context of enhancing system security, has been widely used in cyberspace. It is mainly driven form social science to represent the concept of two entities (trustor and trustee), where one relies on the action of the other one. Trust Modelling and Management (TMM) refers to the process of analysing the behaviour of the system and the interaction between its entities and measuring confidence towards experienced behaviour, and presenting a general perception about risk or uncertainty associated with the entities to maintain legitimacy (Lenard et al. 2023). This allows for detecting any misbehaviour, making TMM emerge as a complementary component of security mechanisms under the name of soft security measures (Sharma et al. 2020). Trust relationships can be evaluated either through direct evidence derived from personal experiences and interactions or through the intake of a trusted third party serving as a recommender. These two approaches to trust establishment can be employed independently or in combination to facilitate reliable trust formation among entities. However, while TMM has demonstrated significant improvements in security for most existing wired/wireless systems, the distinctive characteristics of the underwater environment hinder the adoption of traditional TMM mechanisms in the IoUT. This thesis focuses on how trust is built and modelled between underwater entities in the context of IoUT.

1.2 Problem Statement

The underwater network, operating under the use of UAC, is characterised as a complex and dynamic system featuring a noisy, high-delay communication channel with sparse and dynamic node deployment. These constraints make the direct utilisation of off-the-shelf security measures highly impractical. Based on the context layout of both the IoUT and the TMM, the central problem addressed in this study is the establishment and management of trust among low-capacity UNs within the context of the IoUT and their communication via UAC in harsh underwater environments. This issue becomes critical due to the unique challenges posed by underwater networks, which impact the accuracy of estimating trust based on unreliable communication. To investigate this problem, the following challenges are considered (noting that further background and supporting details for these issues is presented as part of the discussion in later chapters):

- ***PS.1*: The lack of attack-based underwater datasets and the shortcomings of existing simulation tools in addressing the diverse potential applications of IoUT**

  High costs and logistical challenges associated with real-world testing of security measures—trust models being a key component—necessitate reliance on simulation-based approaches as the most viable practice. While several simulation tools can accurately mimic acoustic communication using real environmental measurements, they fall short in modelling the influence of water currents and other mobility dynamics, which is critical for understanding the behaviour and performance of underwater networks. Due to the lack of proper simulation tools, most current security solutions rely heavily on the assumption of fully stationary UNs or overlook the dynamics produced by water currents and UNs' varying mobility capabilities. This gap undermines the reliability of simulations in replicating real-world conditions, particularly when evaluating security solutions. Additionally, existing literature lacks clarity regarding underlying network configurations in IoUT environments. This ambiguity creates challenges in assessing the applicability of security solutions across a broader view of how underwater networks are constructed. Given the varying capabilities of UNs, proper classification of underwater network topologies must be clearly defined to enable rigorous testing under various conditions, ensuring realistic and thorough evaluations. These constraints highlight the need for enhanced simulation tools and methodologies to bridge the gap between theoretical security measures and practical deployment in IoUT.

- ***PS.2*: Examining the readiness of current TMM with physical misbehaviour**

  While trust models have been extensively studied in various distributed domains to mitigate malicious attacks, their application to underwater networks remains challenging. This difficulty arises from the fact that most well-established TMM rely heavily on stable communication paradigms, allowing the development of high-accuracy TMM based on one or few metrics. However, underwater networks present a far more complex scenario. The well-known theories of trust construction, as well as existing TMM designed for underwater networks, have not been adequately examined beyond the communication-based attacks (i.e. dropping packets, selectively forwarding traffic, etc.). In underwater networks, additional vulnerabilities arise from the lack of proper maintenance, physical security measures, and environmental threats. Malicious nodes can exploit the physical environment, patrolling areas to perform malicious acts. When attacks extend beyond communication-based misbehaviour to exploit the absence of physical countermeasures, to what extent does the current TMM effectively detect and mitigate this misbehaviour? This raises critical questions about how current TMM can adapt to the diverse threat landscape of the IoUT.

- ***PS.3*: The need for lightweight and decentralised multi-metric TMM for IoUT**

  The underwater environment presents a highly complex and dynamic network, making single-metric trust models—such as those based solely on packet loss rate or successful delivery—insufficient for accurately assessing trust. Counting on a single metric can lead to high false positive rates, as environmental factors and channel conditions often obscure the distinction between natural disruptions and malicious activities. This limitation emphasises the necessity for a multi-metric TMM that can more effectively capture the characteristics of IoUT. However, the adoption of multi-metric TMM introduces a fundamental challenge: balancing accuracy with efficiency. A TMM that integrates multiple metrics can improve decision-making, but it also imposes computational and communication overhead, which is particularly problematic in resource-constrained IoUT. This raises the critical question of how to design a lightweight yet robust TMM suited for the IoUT.

- ***PS.4*: The problem of dishonest recommendations**
  TMM is designed to incorporate recommendations into trust evaluations, provided that valid and sufficient recommendations are available to easily detect abnormalities among them. In decentralised systems, where nodes rely on recommendations from others to form trust assessments, malicious entities can inject false or misleading information, undermining the integrity of the entire trust model. Current TMM assumes either that clear and recent evidence is collected directly by the node, which, as demonstrated, is not the case in IoUT due to its unique characteristics, or that a sufficient number of recommendations are received to easily detect misbehaving nodes. However, this is also not the case in sparse underwater networks where the lack of enough recommendations makes it hard to detect such attacks. Addressing this challenge requires mechanisms to ensure the reliability of trust evaluations based on recommendations, particularly in the context of dynamic and resource-constrained underwater networks.

1.3 Aims and Objectives

The purpose of this study is to design, develop and evaluate a novel TMM approach to address the establishment and management of trust between entities under the peculiar characteristics of the underwater environment. Following from this aim, the primary objectives for this project are as follows:

- **Objective 1**: To understand current developments towards IoUT and to identify potential topologies based on application demands of underwater networks. To date, there are no publicly accessible datasets that comprehensively capture attacks on underwater networks. The aim is to construct simulation-based test cases under different topologies driven based on both application demands and current enhancements on IoUT of underwater networks. This objective addresses *PS.1*, where both attacks and network configurations are investigated/tested within an enhanced simulation environment.

- **Objective 2**: To evaluate the applicability of existing trust models to IoUT by examining how direct and indirect aspects of trust influence the suitability of models originally developed for underwater networks or derived from established theories of TMM. The objective includes assessing the ability of these models to mitigate the impact of broad attack domains that include attacks from both communication and physical domains under structural variations of IoUT through extensive simulations. Furthermore, it aims to determine which trust establishment practices can be directly applied, enhanced, or excluded in the context of potential attack scenarios, which addresses the challenges in *PS.2*.

- **Objective 3**: To investigate the impact of utilising metrics from both communication and physical domains in establishing trust among underwater nodes. The aim is to design and develop a trust model that incorporates metrics from both domains and adapts these metrics in real-time to enhance overall trust, taking into account the potential attack domains and TMM requirements. This objective is articulated around *PS.3*, with the main focus on the metrics space to enhance TMM.

- **Objective 4**: To develop a recommendation evaluation process that improves the detection of dishonest recommendations —linked to *PS.4*— in trust establishment within IoUT networks. The focus is to analyse the capabilities of dishonest recommenders to disrupt the performance of the trust model and to design a mechanism for detecting malicious recommendations and penalising dishonest recommenders.

1.4 Thesis Organisation

The main body of the thesis is structured into eight chapters. This section briefly highlights the content of each chapter. Additionally, Figure 1.1 provides a visual representation of the thesis content, illustrating the relationships between the chapters and the associated publications arising from the research (reflecting those in print at the time of submission). The boxes in the diagram represent the domains covered in each chapter, with their grouping indicating the complementary scope of each chapter in relation to the others. The arrows depict the flow between chapters, connecting their content and the corresponding resulting publications. The outlines of each chapter are then organised as follows:

- Chapter 2 provides an overview of IoUT as an emerging area of research and development. It outlines potential applications and architectures as envisioned by current literature, and examines the merits and challenges impacting further advancements in this field. It also explores concepts like underwater localisation to justify the use of physical domain metrics during trust assessment in upcoming chapters. Finally, the discussion highlights security concerns and the heavy burden on existing security measures when

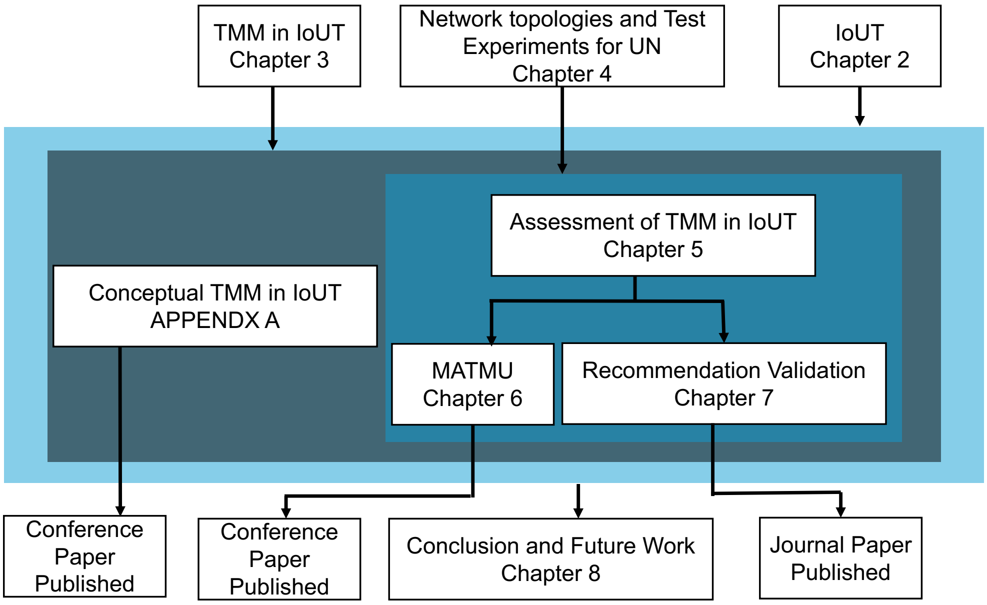


Fig. 1.1. Thesis structure.

applied to underwater networks and IoUT, addressing attacks and extended security demands.

- Chapter 3 provides in-depth background on TMM in the context of Ad Hoc networks, and more broadly on MANET and IoT. It then reviews the relevant state-of-the-art of trust modelling in the domain of underwater networks in general, and examines the few attempts that have been made to address trust in the IoUT, with a suggested taxonomy on the recent developments. The chapter then proceeds to examine and determine the requirements engineering of TMM within the IoUT, considering the network specifications and constraints.

- Chapter 4 presents the foundational concepts and specifications that are then utilised across all test cases in this study. Moving beyond generalised applications of underwater networks, it classifies existing network structures according to their potential applications, enabling a broader range of testing scenarios. This approach allows for a more comprehensive analysis of both network and security-related applications within underwater environments.

This chapter serves as a foundation for the analyses presented in this study by establishing current expectations and advancements.

- Chapter 5 analyses the sustainability of existing trust models in addressing the diverse structures and unique expectations of IoUT. While leveraging the well-established principles of trust from related fields offers potential advantages, directly applying these concepts to IoUT also presents inherent risks. To clarify the boundaries between anticipated outcomes and practical limitations, several models are examined, focusing on both the selected metrics and the underlying deficiencies in trust systems. The findings from this chapter highlight key gaps in both legacy and recent literature regarding trust modelling in IoUT.

- Chapter 6 introduces a new trust model called Mobility-Aware Trust Model (MATMU), focused on metrics for effectively establishing trust between underwater entities. It presents a novel approach that leverages spatio-temporal mobility metrics to evaluate node trustworthiness, with a particular emphasis on mobility trust. This chapter details the processes of trust evaluation and updating used in assessing trust over time in a dynamic manner.

- Chapter 7 introduces a novel mechanism for validating recommendations to improve the detection of dishonest recommendations. Two main methods are introduced for evaluating recommendations: filtering and belief estimation. The chapter details the process involved in evaluating and preferably detecting dishonest recommendations. The findings from the proposed validation mechanism are then evaluated through experimental results with different attack scenarios and against recent benchmarking recommendation models.

- Chapter 8 summarises the key contributions of this thesis, reflecting on its impact in the field of trust for IoUT. It also identifies current limitations and outlines potential ideas for future research.

The thesis also includes a series of appendices that present additional details in support of certain aspects of the main discussion. Each of these is referenced as relevant from the associated chapters.

Chapter 2

The Internet of Underwater Things

2.1 Overview

Recent rapid technological advancements have facilitated exploration of various domains significantly. Nevertheless, despite 70% of Earth's surface being covered by water, the underwater realm remains an incredible and largely unexplored frontier. Underwater wireless sensor networks, marine communication systems, and the recent emergence of the IoUT are some of the new technologies currently revolutionising the development and accessibility of underwater areas. These innovations are expected to pave the way for seamless connectivity and enhanced functionality across underwater networks. However, the underwater environment has unique characteristics that limit the capacity of applications, with security demands being a key area requiring investigation and enhancement.

2.2 Concepts and Properties of the IoUT

The concept of the IoUT has garnered growing interest among both researchers and industries since Domingo (2012) initially mentioned it. As the name implies, this concept is a derivation of the IoT, referring to sensing and intelligent components capable of performing several tasks by combining the Internet and emerging technologies located under oceans, seas, rivers or other major bodies of water. Interest in exploring and monitoring uncharted aquatic environments and oceans is driven by interest from various domains of application, including but not limited to the following (Bello et al. 2022):

- **Military and surveillance domain:** Enhancing marine security operations, including underwater early-warning systems, safeguarding critical infrastructure such as ports and submarines, detecting underwater mines, securing international waters, and improving target detection capabilities.
- **Industrial domain:** Supporting activities such as oil extraction and monitoring, optimising fish farm production, and deploying underwater data centres and storage facilities.
- **Scientific and environmental applications:** Facilitating underwater exploration and improving the detection and prediction of natural disasters, including floods, volcanic eruptions, earthquakes, and tsunamis.

This variety of applications has the capacity to not only expand our understanding of aquatic ecosystems, but also to provide practical solutions to address future challenges and support potential applications.

2.2.1 Architecture and Technologies for the IoUT

Having defined the concept of the IoUT, several studies have explored ideal architectures with which to construct IoUT systems, aiming to maximise usability and potential. Across the literature, however, variations in terms of existing architectures have emerged in relation to deployment methods (static or dynamic), topology models (2D, 3D), and types of technologies employed, such as Autonomous Underwater Vehicle (AUV), Remotely Operated Vehicle (ROV), anchored sensors, buoys, sink nodes, and others (Khalil et al. 2020). Usually, the traditional

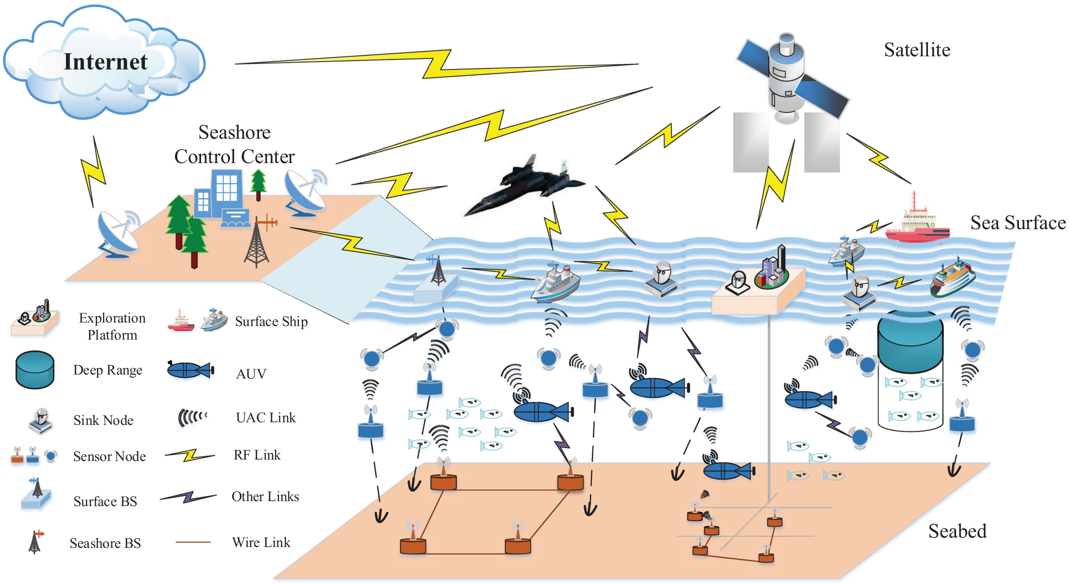


Fig. 2.1. The conceptual model of IoUT.
Sourced from (Qiu et al. 2019).

IoUT infrastructure comprises sensing devices to detect and collect data, and communication components capable of transmitting data both underwater and above water to an onshore station, such as a cloud platform, for further processing. Figure 2.1 illustrates a conceptual IoUT model, which was presented by Qiu et al. (2019). As shown in the figure, the surface communication units (e.g., surface base stations, surface ships, or sink nodes) serve as connection points connecting UNs to above-water systems. These units then offload data to coastal control modules, such as seashore control centres. The underwater portion of the network consists of various sensing and transmission components that can be used to communicate underwater using UAC. This requires underwater devices equipped with acoustic transducers to generate sound waves to transmit data between the nodes. Several underwater entities have been widely developed to facilitate underwater communication, for example, AUV and ROV mobile devices.

Despite advancements to date, clear and validated architectures for the IoUT are still lacking; hence, efforts are ongoing to optimise and enhance IoUT architectures by utilising well-established paradigms in terrestrial networks. For instance, Hou et al. (2021) proposed a futuristic maritime network architecture for mission-critical IoUT systems that integrates with space-air networks. This architecture divides the network into three collaborating domains: the maritime network, the space-air network, and the terrestrial network. The maritime network itself is organised into three layers: the seabed layer, the underwater layer, and the surface layer. The seabed layer comprises IoUT devices deployed at the ocean floor, the underwater layer includes dynamic underwater vehicles, and the surface layer features floating base stations equipped with communication systems such as 3G/4G/5G/WiMAX for transmitting data to the terrestrial network.

Moreover, to address the need for decentralised on-site processing and to overcome the limitations inherent in traditional cloud computing, such as bandwidth constraints, latency, and the inability to handle high data volumes, Bhattacharjya et al. (2021) proposed edge computing. Designed to reduce latency in mission-critical applications, edge computing is employed in an Edge-Drone-based four-layer smart Internet of Underwater Things (EdgeIoUT). This then minimises energy consumption and extends the lifetime of IoUT networks. Additionally, Software-Defined Network (SDN) technologies are emerging as transformative forces in evolving traditional IoUT applications towards software-based, programmable, user-customisable, and service-oriented systems (Luo et al. 2018).

These technologies enable a shift from application-specific, expensive underwater infrastructures (e.g., those designed for environmental monitoring or surveillance) to well-designed networks able to support a wide range of applications. This allows for sensing-as-a-service and infrastructure-as-a-service models, in which underwater networks can share their underlying resources to deliver diverse services to end-users. In similar concepts, Akyildiz et al. (2016) proposed SDN-based architecture for underwater networks called SoftWater, incorporating various underwater communication protocols.

Since different standalone networks often implement distinct protocol stacks, the Transmission Control Protocol/Internet Protocol (TCP/IP) remains the most universally adopted and standardised network architecture. Studies, including that of Jahanbakht et al. (2021), have demonstrated that the IoUT can be effectively modelled using the TCP/IP stack. In this architecture, the physical layer incorporates various hardware technologies, such as acoustic, optical, and radio frequency systems, so as to efficiently handle data transmission and reception in underwater environments. The data link layer is responsible for reliable data transfer with medium access control (MAC) and error control. The network layer is responsible for discovering and maintaining routing for effective data transmission, and the transport layer for flow congestion and reliable data collection. The application layer is mainly concerned with identifying objects, such as the sensor's ID, type, and location, as well as performing specific application-related tasks utilising sensed data and the underlying physical infrastructure.

Although these studies represent a significant theoretical improvement with regard to understanding how to create an efficient and robust IoUT architecture capable of maximising network capacity, critical gaps remain. Specifically, a comprehensive analysis of fundamental requirements, such as security measures between underwater entities, is still lacking, leaving key challenges unresolved.

2.2.2 Communication Channels Underwater

In the IoUT, water serves as the medium for signal transmission, and the unique physical and chemical properties of the aquatic environment introduce significant challenges compared to terrestrial IoT. These challenges stem from variations in water's characteristics, such as salinity, temperature, and pressure, which vary according to geographical regions and environmental conditions. These variations directly affect the behaviour and performance of communication paradigms,

Table 2.1: Characteristics of Transmission Media Underwater.

Criteria	Optical Wave	Radio Wave	Acoustic Wave
Propagation Speed	$\approx 2.25 \times 10^8$ m/s	$\approx 2.26 \times 10^8$ m/s	$\approx 1.5 \times 10^3$ m/s
Bandwidth	$\approx$ 150 MHz	$\approx$ MHz	From 100 to a few kHz
Data Rate	High (Gbps)	Moderate (Mbps)	Low (Kbps)
Reliable Distance	10–150 m	$\leq$ 10 m	$>$ 100 km
Drawbacks	Absorption, scattering, and the required line of sight	Conductivity and permeability issues, high attenuation over short distances	Doppler effect [1]; Multipath fading [2]

[1] Doppler Effect refers to the variation in the frequency of waves whenever the distance between a transmitter and a receiver changes.

[2] Multipath is a phenomenon that occurs when acoustic waves propagate in water and are reflected by sea surface, seabed, or other obstacles (Lurton 2002).

making it more complex to design and implement wireless communication systems for underwater networks. Consequently, the methods and technologies that work efficiently in terrestrial environments often require significant adaptations to function effectively underwater. For instance, radio electromagnetic waves, which typically support long-distance communication (up to hundreds of kilometres) in terrestrial environments, are quickly absorbed and dissipate within a short distance when transmitted in salty water. Similarly, optical waves can be utilised underwater for short-range communications, but require precise alignment between sender and receiver. This limits their practicality for use in long-distance communication between underwater devices.

However, acoustic signals produced by transducers[1] exhibit low absorption in water, making them more effective, and so they are more widely adopted for un-

[1] Acoustic transducers convert electrical energy into sound waves, typically for applications such as sonar and underwater communication, i.e piezoelectric transducers.

derwater communication compared to alternative methods (Pranitha et al. 2020). Consequently, most existing underwater communication solutions prefer to rely on acoustic waves. However, the use of UAC presents its own challenges. One notable limitation is its low propagation speed, with approximately 1.5×10^3 m/s compared to the 3×10^8 m/s propagation speed of radio waves on land. Additionally, UAC suffers from limited bandwidth availability. According to Jiang (2017), the available bandwidth ranges between 2-3 kHz, supporting data rates of about 10 Kbps for long-range transmissions (approximately 20 km). For shorter transmission ranges (less than 1 km), the bandwidth exceeds 20 kHz, supporting higher data transmission rates of up to 100 Kbps. Table 2.1 summarises the distinct characteristics and limitations of existing underwater communication channels (Kao et al. 2017; Khalil et al. 2020).

For the remainder of this study, the UAC will be considered the primary communication medium for the underwater network, given its superior ability to propagate effectively underwater.

2.2.3 Challenges in the Current Development of IoUT

The IoUT is subject to several challenges that need to be considered when proposing its development, as follows:

- **Physical environment's impact on the UAC:** The acoustic channel is affected by variations in water temperature, level of salinity, and other factors, which change in relation to depth. Hence, the propagation speed of sound underwater is a proportional function of these variables (Islam et al. 2022). Consequently, the reliability of the acoustic channel depends on the environmental conditions of the water medium through which the signal is required to travel. This issue is significant, even when both communication entities are stationary, as the dynamic nature of water's momentum causes temporal changes in communication, which differs from the relatively stable radio communications that exist between two stationary nodes in a terrestrial network. It thus follows that the challenge intensifies when the nodes are mobile, making it difficult to estimate the channel's efficiency accurately.

- **Costly infrastructure:** Underwater projects are typically resource-intensive and costly. For example, a mission involving the Bluefin-21—a type of AUV—onboard the Australian Defence Vessel Ocean Shield during a search operation took 370 hours, and was estimated to cost approximately one million dollars (Blinken 2019). Moreover, in light of security incidents such as the attacks on the Nord Stream 1 and 2 pipelines, which have underscored the vulnerability of offshore infrastructure, the Federal Ministry for Digital and Transport has now funded the Argus project, resulting in a total project cost of €3.5 million (north.io 2024). Such high costs are unsurprising given the complexity of underwater acoustic transformers and transceivers, combined with the need to protect UNs hardware from water exposure by using waterproof materials designed to withstand extreme conditions. These factors restrict the widespread deployment of underwater networks, thereby highlighting the importance of sparse network design. This contrasts with the relatively low cost of IoT devices, and the feasibility of deploying dense networks in terrestrial environments. Consequently, additional efforts are necessary to adapt existing communication and management solutions to accommodate the relative costs and correspondence sparse requirements of the IoUT.

- **Limited Device Capability:** Various UNs have been developed with different capabilities, but a key limitation of these devices is their restricted underlying capacity. The majority are constrained by limited storage and processing capabilities, which restricts their functionality. Additionally, the majority of underwater communication devices are battery-powered, and no efficient method currently exists to recharge or replace these batteries frequently. Traditional battery replacement methods are costly, due to the expensive operations involved, further inflating overall expenses. Moreover, the energy consumed for communication purposes is much greater than that needed for the terrestrial sensor network. The reason for this is the high attenuation of the communication channel, the high transmission power required by underwater transceivers, and the high bit error rate. These limitations usually produce a high probability of frequent data packet retransmission, thereby consuming a vast amount of sensor nodes' energy. Approaches such as energy harvesting (e.g., the work proposed in (Alamu et al. 2023)) and duty-cycling (e.g., the work proposed in (Khan et al. 2019)) have been explored to mitigate energy depletion and extend network lifetimes. Nonetheless, this issue remains a significant barrier to the advancement of the IoUT, especially considering the severely energy-constrained nature of underwater environments (Islam et al. 2022).

- **Dynamic Topology:** The nature of aquatic regions is dynamic, resulting in frequent changes in the network topology. Different forces are exerted on any object that drifts underwater, which substantially affects network topology. According to Pompili et al. (2009), UN is exposed to four main forces: the object weight force, the buoyancy force, the fluid resistance force, and the water currents force. The first two forces depend on the volume and density of the UNs and the properties of the water. The resistance force relates to the velocity of the UNs, while the water current force depends on a combination of the diversity of the water current velocity and the UN velocity. The continued movement of the ocean by the water currents force and surface winds creates difficulties that affect the construction and

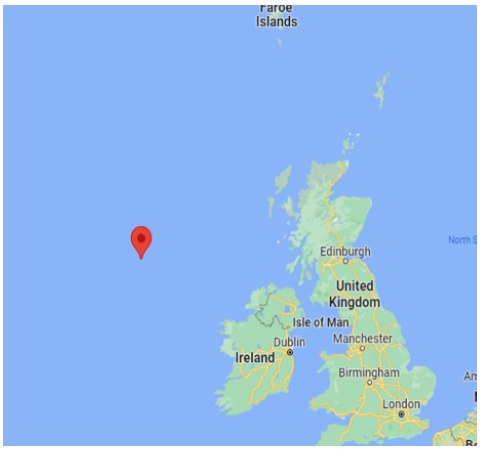


Fig. 2.2. Random location at the North Atlantic Ocean.

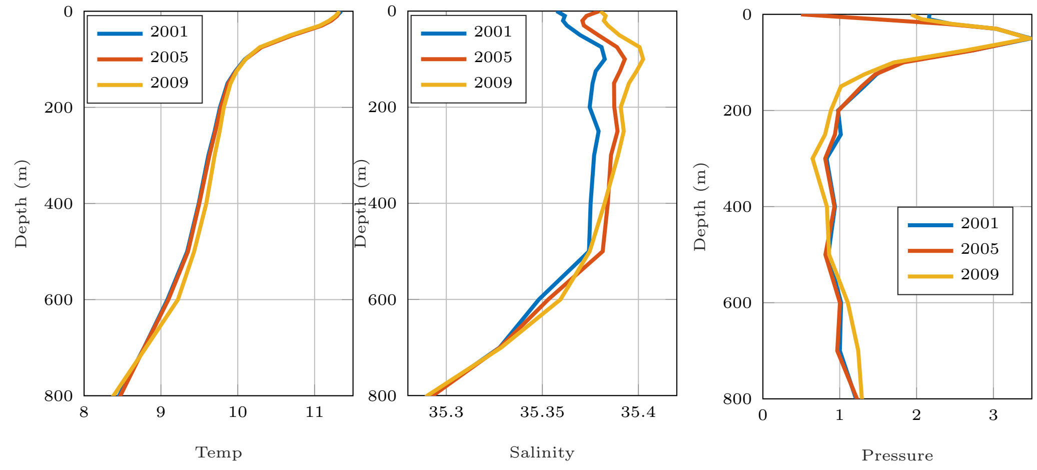
maintenance of the IoUT. Thus, IoUT applications suffer from a complex topology, due to the need to account for frequent changes in topologies.

These extraordinary barriers to designing an efficient IoUT present an opportunity to develop innovative solutions, particularly when reviewing similar, well-studied concepts to maintain functional and secure networks. Nevertheless, significant gaps in possibility, highlighting the need for further advancements and enhancements.

2.2.4 Properties of UAC in the North Atlantic Ocean

This section details the relationship between the speed of sound and the physical characteristics of fluid, based on field measurements in the North Atlantic Ocean. For this demonstration, a random location with the coordinates 56°01'44.4"N 14°23'39.1"W was selected, as shown in Figure 2.2. Data from three different years (2001, 2005, and 2009) were queried from the World Ocean Atlases (WOA) database, hosted by the National Center for Environmental Information (Information 2020). These datasets use actual measurements collected during those years.

While the datasets provide information down to a depth of 5.5 kilometres, the results are depicted for 800 metres of depth. Figure 2.3 illustrates annual vari-



(a) Temperature. (b) Salinity Level. (c) Net pressure.

Fig. 2.3. Temperature, salinity and pressure profiles.

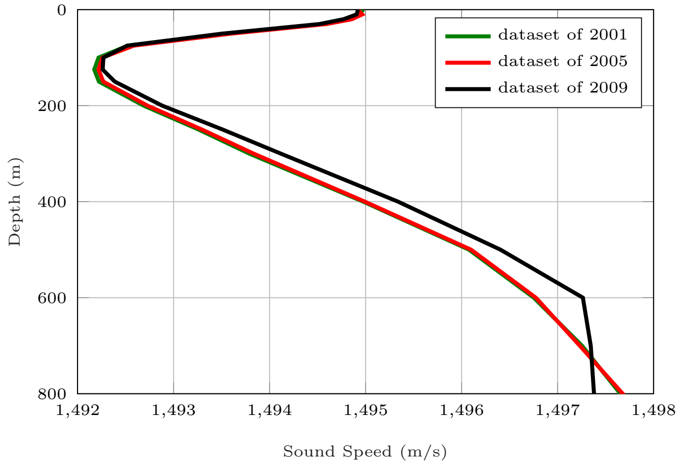


Fig. 2.4. Sound Speed Profile (SSP) on different depths based on measurement obtained from WOA.

ations in temperature, salinity, and net pressure at various depths within the selected area. As depicted in Figure 2.3a, the temperature decreases from approximately 11.5° near the surface, to 8.5° at a depth of 800 meters. The salinity profile (Figure 2.3b) reveals small changes by only a small amount, from 35.3 to 35.4 Practical Salinity Units (PSU) noticeable differences near the surface the datasets, with the general trend showing decreasing salinity at greater depths.

The speed of sound in water is a critical factor for the performance of the UAC, as it serves as the primary constraint on network performance. Speed is not constant, but fluctuates over space and time in response to various environmental factors, with temperature, pressure, and salinity being the most influential. The corresponding Sound Speed Profile (SSP), illustrated in Figure 2.4, highlights variations across each of the different datasets. Near the water's surface, SSP remains relatively stable before it starts to decrease, driven by a rapid drop in temperature while pressure remains almost unchanged. At a certain depth, sound speed reaches its minimum, after which it begins to increase again, as temperature stabilises and pressure continues to rise. The impact of salinity on sound speed varies by geographic location. This behaviour causes sound signals to travel along curved trajectories, rather than straight Euclidean paths, which can result in errors if estimations of propagation speed are improperly calculated or applied.

2.3 Underwater Localisation Mechanisms

Given the nature of the underwater medium, the electromagnetic signals emitted by orbiting satellites utilised in a stable localisation mechanism like Global Positioning System (GPS) experience significant attenuation when passing through water, and are not feasible below a depth of roughly 20 cm underwater (Jouhari et al. 2019). These limitations necessitate the development of alternative localisation techniques to achieve high-accuracy spatial awareness among underwater entities. Various taxonomies have been proposed to classify existing underwater localisation algorithms, such as distributed versus centralised techniques, and stationary versus mobile systems. Despite these classifications, the majority of approaches apply two main phases: location-based information collection and position estimation (Li et al. 2015). The first phase involves observing and collecting information to help determine a node's location, whereas the second phase estimates and optimises the position with high accuracy based on the collected data.

Several techniques in the literature aim to collect location information underwater, relying on different measurements to estimate the distance between reference nodes or GPS Intelligent Buoys (GIBs) and the linked target nodes. These reference nodes establish a coordinate system and emit beacon signals to assist with localising the target nodes (UN to be localised). Methods utilising reference nodes are generally divided into range-based and range-free approaches. Range-based techniques rely on measurements such as Direction of Arrival (DoA), Time of Arrival (ToA), Time Difference of Arrival (TDoA), Received Signal Strength (RSS), Signal Intensity (SI), and Angle of Arrival (AoA) (Nain et al. 2024). The majority of these methods depend heavily on the clock synchronisation of sensor nodes, which can introduce challenges maintaining accuracy. Meanwhile, range-free techniques aim to eliminate dependence on range measurements, which may suffer from high error rates due to node mobility and high propagation delays. For example, some solutions use AUVs with directional beaconing to patrol the region at a constant speed, relying on knowledge of a pre-defined location. Other approaches, such as fingerprinting, require a training phase prior to the localisation process. This method involves a channel signal source capable of transmitting at different frequencies in conjunction with reference nodes (Luo et al. 2021). The collected data can be utilised for localisation in centralised or distributed systems. In centralised systems, information and reference node locations are sent to a central node, which estimates all the node locations, using methods such as triangulation, multilateration, and trilateration. In distributed systems, each node has computational capabilities for self-localisation, and runs algorithms on data collected during the initial phase.

A hybrid Bayesian multidimensional scaling-based localisation technique was introduced by Khalil et al. (2021) to use for fully hybrid IoUT networks. This method allows nodes to communicate across different media, such as optical, magnetic induction, or acoustic technologies. The scheme relies on signals of opportunity-based localisation, and its accuracy depends on the precision of range measurements. A centralised footprinting localisation algorithm for the IoUT was proposed by Vegni et al. (2021). This algorithm utilises hybrid wireless networks, including optical wireless signals in the visible range, and IoT devices with radio communication. It follows three stages: database construction, position detection, and position estimation. The system deploys four LED transmitters on the seabed, a central IoT sensor for data collection, processing, and storage, and an on-surface control node (e.g., a ship) for monitoring. Initially, a 3D database is created and periodically updated, representing the underwater space, as a 3D grid with possible locations is stored in the IoT sensor. Localisation involves measuring channel responses between the LEDs and the target and then comparing these to database values. A recursive algorithm then refines the estimated positions to determine the most accurate target location. A simulation of this technique conducted in a $10m^3$ environment showed only a few centimetres of estimation error, primarily caused by water movement.

Localisation mechanisms for oil pipeline monitoring were proposed by Goyal et al. (2022), using range-based measurements, such as AoA, ToA, and RSS. These values were further optimised using the lion optimisation algorithm to enhance area efficiency. Simulation results demonstrated a 28% reduction in localisation error compared to existing approaches, showing lower energy consumption and reduced localisation delays. Motion prediction mechanisms were proposed by Xu et al. (2022) as a way to pre-estimate the speed and location of underwater nodes, enhancing the accuracy of localisation systems. Similarly, Pourkabirian et al. (2023) proposed a hybrid approach, using RSS and AoA for the application of the IoUT. The proposed method accounts for practical challenges and optimises worst-case performance using a minimax approach to handle unknown noise parameters. Additionally, the paper analyses estimation accuracy through mean-square error (MSE), employing semidefinite programming (SDP) to efficiently solve the localisation problem. The findings succeeded in localising 96% of sensor nodes, with less than a 5% positioning error when 25% are reference nodes.

To conclude, it is reasonable to suppose that UNs, with the current technologies and mechanisms available, are capable of estimating the location information with acceptable accuracy, as evidenced by several review papers (Nain et al. 2024; Nanthakumar et al. 2024; Su et al. 2020). This allows most security measures to be established, provided that the UNs are aware of their corresponding locations, enabling measurements from the physical domain to be utilised.

2.4 Existing Simulation Tools

Due to the popularity of underwater networks, the several attempts to develop toolkits and simulation platforms principally focus on simulating and modelling the acoustic channel. Underwater communication simulations are crucial for advancing the design, testing, and optimisation of communication protocols, especially within acoustic networks, given the expensive nature of performing real experiments. Nkenyereye et al. (2024) outlined several specialised tools exist for this purpose. For instance, GODUNOV, a GPU-accelerated simulator, offers high-performance simulations with realistic underwater channel models, while OMNeT++ provides a modular framework, which can be extended for underwater applications. The World Ocean Simulation System (WOSS) integrates established channel simulators, focusing on realistic environmental modelling. MatLab/Simulink and the Acoustic Toolbox (ATB) facilitate underwater acoustic simulations with various toolboxes and customisable modules. DESERT Underwater enhances the NS-MIRACLE simulator for underwater network protocol development, and UANS refers to underwater acoustic network simulations. Within the NS-2/3 family, Aqua-Sim Next Generation (Aqua-Sim ng) is a simulation tool designed specifically for modelling and evaluating UNs (Martin et al. 2017). Built on the NS-3 core simulator, which is excellent for layer-wise communication and protocol design, Aqua-Sim ng features a wide array of specialised components used for simulating the unique challenges of underwater environments, including advanced noise generators, multiple channel models, and trace-driven testing capabilities. These features enable Aqua-Sim ng to accurately simulate the complex acoustic propagation and packet handling processes that characterise underwater networks. Moreover, its modular architecture allows for seamless integration of various protocols and testing scenarios, facilitating the development and evaluation of innovative techniques involved in underwater communications.

2.5 Security in the IoUT

The dynamic nature of the network, together with the resource constraints, in terms of computation capacity and the communication methods of IoUT devices, all impose an additional burden on the existing security solutions that are widely used across similar networks. Given the critical importance of security in modern infrastructures, the security of IoUT needs to be addressed.

The security architecture for the underwater networks is as depicted in Figure 2.5, and was proposed by Yang et al. (2019) to emphasise the security challenges affecting the current advancement towards wireless underwater networks. Within this architecture, the behaviour of malicious nodes can vary, and based on the broader prospects of different attacks, malicious nodes can be classified as performing either passive or active attacks. In a passive attack, the attacker attempts to obtain data by only eavesdropping on signal diversity over the channel, and collecting information that could be fundamental for several other forms of attacks. In contrast, intruders will attempt to perform malicious actions, such as channel jamming during an active attack. Yang et al. (2019) proceeds to explore what each aspect of security demands, whether layer-specific requirements, such

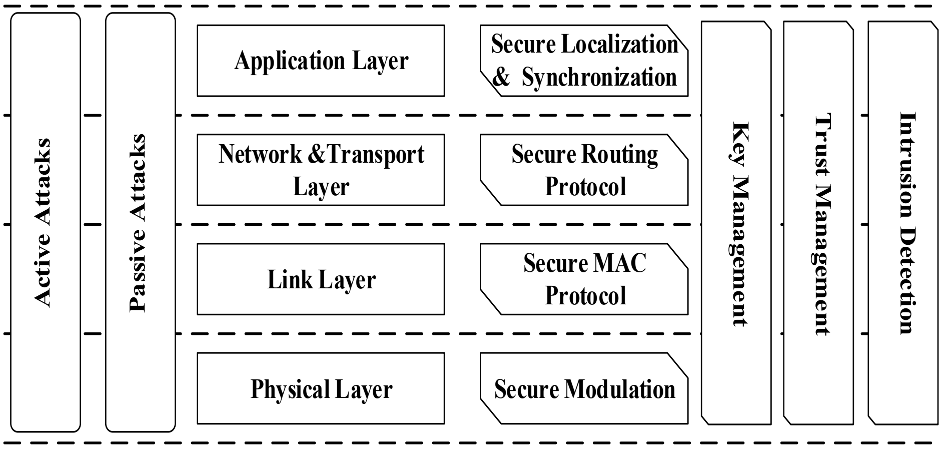


Fig. 2.5. High-Level security architecture for underwater networks.
Sourced from (Yang et al. 2019)

as secure routing methods for the network layer, or overall concepts like intrusion detection and trust management, and how these fit within the constraints imposed by underwater networks.

2.5.1 Security Concerns

The challenges posed by the IoUT extend beyond the reliability and functionality of underwater networks to encompass influential security requirements for several reasons. Firstly, the extreme nature of underwater networks exacerbates existing vulnerabilities. Such networks are often deployed in opaque aquatic environments, in which nodes remain unattended for extended periods. This isolation increases exposure to adversaries and susceptibility to physical attacks. Since applying existing physical security countermeasures is virtually impossible in such environments, underwater networks are especially vulnerable to physical threats. Any capable adversary can potentially capture unattended UNs. Moreover, removing compromised UNs can be prohibitively expensive, especially in hostile or remote environments. Whereas some attacks can be mitigated through remote management, physically accessing compromised nodes is frequently unavoidable. However, this process is both costly and challenging to execute promptly. Secondly, the open nature of underwater communications significantly expands the potential threat landscape. Attackers may eavesdrop on or intercept communications from almost anywhere, making it difficult to isolate the source of threats. Finally, the absence of a standardised model for how UAC behaves across varying environmental conditions complicates the distinction between malicious activities and network abnormalities. This ambiguity burdens the effectiveness of security measures, especially given the constraints associated with the limited data rates in underwater networks (Jiang 2018).

With limited available bandwidth and constrained resources for handling transmissions, any security method that increases message size represents a challenge in underwater networks. This renders most energy-intensive security solutions challenging to implement in such environments. Unfortunately, few real-world experiments have been conducted focusing on security measures in underwater networks. For instance, Dini et al. (2012) conducted a field experiment in Norway to evaluate the performance of a cryptographic suite employed with a FLOOD protocol called SeFLOOD. The cryptographic suite used the AES in CBC-CTS mode for traffic encryption, and SHA-2 as a hash function, to ensure message integrity. Their findings demonstrated that implementing security measures resulted in increased energy consumption and higher error rates. Similarly, Caiti et al. (2013) conducted field experiments using the CipherText Stealing (CTS) technique for encryption, and 4-byte digests derived from truncating the full hash function value for authentication, so as to minimise additional overheads on the channel. The findings indicated an 8% decrease in the delivery ratio when these security measures were incorporated.

Table 2.2: Attacks Classifications and Potential Security Requirements.

TCP/IP Layer	**Attack**	**Attacker Type**	**Desired Security Property**
Physical	Jamming	Active (Internal/External)	Availability
	Signal Eavesdropping	Passive (Internal/External)	Confidentiality
	Tampering	Active (Internal/External)	Integrity
Data Link	Collision	Active (Internal/External)	Availability
	Exhaustion	Active (Internal/External)	Availability
	Unfairness	Active (Internal)	Availability
Network	Selective Forwarding	Active (Internal/External)	Availability, Confidentiality
	Denial of Service	Active (Internal/External)	Availability, Integrity
	Blackhole	Active (Internal/External)	Availability
	Wormhole	Active (Internal/External)	Availability, Confidentiality
	Sinkhole	Active (Internal/External)	Availability, Confidentiality
	Replay	Active (Internal/External)	Authentication, Integrity
	Sybil	Active (Internal)	Availability, Authentication
Transport	SYN Attack	Active (Internal/External)	Authentication
	Flooding	Active (Internal/External)	Availability

2.5.2 Attacks Classification in IoUT

Each TCP/IP layer advances diverse potential threats, with the potential to either destroy UNs or paralyse entire networks, leading to network performance degradation, leakage of secret information, or potential network failure, as depicted in Table 2.2. With the lack of standardised security measures, the majority of the currently proposed countermeasures are geared toward a specific attack or a group of attacks (Yisa et al. 2021). Aside from these attacks, Adam et al. (2024) broadens the scope of potential threats to include device-level security risks, such as physical tampering and unauthorised access, which are relatively straightforward for capable malicious actors.

2.5.3 Categories of Misbehaviour

Within the context of the attack domain, the classification of attacks can be understood as either external or internal attacks. An external attack refers to an attack performed by an external entity with unauthorised access to the network domain. An internal attack implies any attack executed by an internal entity belonging to the same network domain and exploiting network vulnerabilities to maliciously affect the normal operation of the network. In the absence of proper physical security implementations and the challenges of managing a complex network environment, even when security protocols are applied, there remains a risk that nodes may be compromised. A compromised entity can easily be altered and become a source of internal attacks. The intrusion of malicious attacks is expected to affect communication between nodes, and may have a significant impact on network performance. In the context of network misbehaviour, two distinct types have been identified: malicious nodes and selfish nodes. These two terms are defined as follows:

1. **Malicious node:** A malicious node is a compromised node that has been taken over and may disrupt normal network operations. This node can intercept communications and carry out multiple forms of attacks, such as eavesdropping, data tampering, and disseminating false information.

2. **Selfish node:** A selfish node is any node that eludes collaboration with other nodes, with the aim of accomplishing specific tasks. In the context of underwater networks, this can manifest as a node refusing to respond to task-specific requests; i.e. forwarding traffic.

2.5.4 Extended Security Demands for the IoUT

Security solutions for the IoUT should meet several security requirements, each of which acknowledges the limitations of underwater networks. According to Yang et al. (2019), essential security requirements, such as confidentiality, integrity, and availability, are insufficient to maintain an acceptable level of underwater network security. Thus, the anticipated expanded security demands are as follows:

- **Confidentiality:** Protecting the sensitive data produced by UNs and the survivability of different network-related tasks, such as the MAC, routing information, and localisation, among others.
- **Integrity:** Ensuring the received data is not modified, corrupted, or removed in transition by a malicious node or a malfunction upon transmission.
- **Authentication:** The ability of a receiver node to identify the source of the data.
- **Availability:** The robustness of the network to counter different denial of services attacks or UNs failure.
- **Freshness:** Evaluating the age of the data and ensuring the received data is current and fresh.
- **Isolation:** Detecting abnormalities in the network so as to allocate and isolate malicious nodes accurately.
- **Self-stabilisation:** Concerning the ability of UNs to independently recover to their normal state in the presence of attacks, without any outside interventions in real-time. Applying this requirement within the design, each UN can potentially mitigate the influence of malicious nodes, even in the presence of ongoing attacks.

- **Survivability:** As with the self-stabilisation requirement, this mainly focuses on the entire network's ability to function correctly under attacks or failures and restore and maintain essential network services.

The final three requirements; i.e., isolation, self-stabilisation, and survivability, are important in view of the nature of underwater networks, which necessitate nodes be self-dependent and capable of making informed decisions about contiguous abnormalities. The primary challenge, however, lies in balancing the robustness of the implemented security measures with the network's efficiency in terms of energy consumption, packet delivery, and minimising delays. Consequently, there remains a pressing need for lightweight and secure mechanisms for underwater networks (Adam et al. 2024).

2.6 Summary

This chapter established the foundational concepts necessary for understanding the main topics covered in this thesis, integrating the key components relevant to underwater networking and the IoUT. The chapter started by providing background to the IoUT concept relative to the better-known IoT, exploring its fundamental principles and potential applications while emphasising its significance across various human-centred domains.

The chapter examined the communication challenges inherent in the primary physical media used for underwater propagation, identifying key issues faced by underwater networks in general, and the IoUT in particular. As UAC is the primary communication model in this thesis, field data was utilised to analyse sound propagation underwater and determine its dependence on regional and environmental parameters. Localisation mechanisms deemed suitable for underwater environments were also reviewed to support assumptions concerning the appropriate localisation approaches for such networks, as these will be relevant throughout the thesis.

Another critical issue addressed in this chapter was the security of underwater networks. It explored current challenges securing IoUT systems, and analysed the threat landscape, highlighting key attack domains. Finally, the chapter investigated security requirements for underwater networks, as they extend beyond traditional CIA (Confidentiality, Integrity, and Availability) demands. With the proposed security architecture in the literature requiring trust management to be enhanced, along with other security aspects, to meet the security demands of underwater networks, the next chapter will focus on exploring trust management across different domains.

Chapter 3

Trust Modelling and Management

3.1 Overview

Trust has emerged as complementary to security mechanisms to enhance systems' security. Systems that only rely on conventional security solutions, such as cryptography, authentication, and hash functions, face challenges in identifying and mitigating internal attacks launched by a compromised node within the network. Trust modelling significantly improves responses to internal attacks by analysing the behaviour of the system and the interaction between its entities. Thus, demand for it is now notable, especially for decentralised networks. Most recent advanced security solutions suffer low performance when used in IoUT, as discussed in the preceding chapter. Therefore, the concept of trust between UNs in the context of attack detection represents a reasonable improvement for such networks.

3.2 Concepts and Properties of Trust

Security solutions can be classified as either hard or soft security (Rasmusson et al. 1996). In the former, traditional security mechanisms, such as cryptography, authentication, and access control are adopted approaches. In the latter, social control mechanisms enforce normal and secure behaviour over system components, of which trust and reputation systems are examples. Trust, despite its subjective nature, has proven highly valuable in network security due to its applicability to decision-making processes, especially when countering malicious actions. In this section, trust as a security measure will be explored in relevant fields of research.

3.2.1 Trust Definitions

Beyond the dictionary definition of trust as "the belief that someone is good, honest, and will not cause harm, or that something is safe and reliable"—(Cambridge University Press n.d.), each academic discipline offers a unique perspective on the meaning of trust, noting that it is highly dependent on the context in which it is being utilised. In Table 3.1, definitions of trust from selected disciplines were paraphrased to afford a multidisciplinary perspective. While these definitions present various aspects of trust relationships, the majority converge agreeing on key dimensions: behaviour, intention, belief, and disposition. In the social context, trust between individuals can manifest as one person voluntarily relying on another in specific behavioural situations (behaviour), or a willingness to depend on others (intention), or the belief that another person is both willing and able to act in one's best interest (belief), or as the development of expectations about others' trustworthiness over time (disposition) (Carminati et al. 2022).

At its core, trust manifests between two principal stakeholders in a trust relationship: the trustor and the trustee, which can be individuals, agents, or network entities. These roles signify an asymmetry in the relationship, with the trustor being the party that initiates and shapes the trust relationship with the dependent entity, the trustee. A more generalised definition of trust, as provided by (Cho et al. 2015), is as follows:

> "*Trust is the willingness of the trustor (evaluator) to take risks based on a subjective belief that a trustee (evaluatee) will exhibit re-*

Table 3.1: Trust Definition Across Disciplines.

Discipline	**Definition**
Sociology	A mechanism for building cooperation among people to extend human interactions for future collaboration (Luhmann 2018).
Philosophy	A personal and internal experience that supports moral relationships between individuals, where a breach of trust is a distinct violation of ethical conduct, resulting in distrust (Lagerspetz 2013).
Economics	An expectation that applies to situations in which trustors take risky actions under uncertainty or information incompleteness (James Jr 2002).
Psychology	A cognitive learning method developed by an individual through social experiences, shaped by the positive or negative consequences of trusting behaviour. (Rotter 1980).
Computer Science	A belief that an entity can perform reliably, dependably, and securely in a specific situation, leading to varying degrees of trust across different contexts. (Li et al. 2007).

> *liable behaviour to maximise the trustor's interest under uncertainty (e.g., ambiguity due to conflicting evidence and/or ignorance caused by complete lack of evidence) of a given situation based on the cognitive assessment of past experience with the trustee.*"

According to the above definition, trust is established/learned relative to one's own interests and bears a degree of potential risks. This means the trust relationship is more a continuously evolving relationship that draws on experiences (or evidence) to either increase or decrease the level of trust based on the current and previous observations.

Scholars often use trust and trustworthiness interchangeably. However, some researchers, such as Cho et al. (2010), have sought to differentiate between the two. Trust can be understood as a belief probability, with values ranging from 0, indicating full distrust, to 1, indicating full trust. In this context, trustworthiness is a measure of the actual probability that the trustees will act as anticipated.

Table 3.2: Trust Properties.

Property	**Definition**
Subjective	Trust is subject to the evaluation of one node with another. Due to node mobility as well as the possibility of node failure, information is typically incomplete and can change rapidly, causing nodes to experience different trust due to the dynamic change in the topology.
Partial Transitive	If entity i trusts entity j, and j trusts entity k, this does not necessarily mean that entity i completely trusts k.
Asymmetric	Trust between two entities is not necessarily reciprocal. This implies that entity i might trust entity j, but j may not develop a trust relationship with i.
Context-dependent	Trust is dependent on the context of particular situations, e.g., trust in computational power, unselfishness, forwarding, or reporting ability.

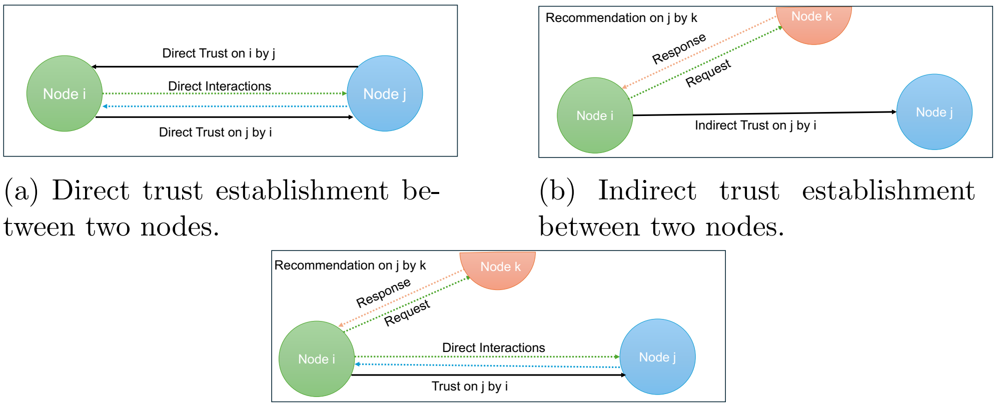
This suggests trust leans more towards being a subjective probability, whereas trustworthiness represents the objective probability of a given trust level.

3.2.2 Trust Properties and Trust Dimensions

Trust relationships are typically defined by four commonly recognised characteristics or attributes: subjective, transitive, asymmetric, and context-dependent. However, notably, not all relationships demonstrate these features, nor do they provide an exhaustive definition of trust. Table 3.2 depicts the interchangeable properties of the TMM system as derived from the social sciences and translated for use in cyber-space (Wei et al. 2022).

In order to build a trust relationship, TMM approaches have been heavily invested in the context of IoT and MANET systems, and generally, the trust relations within these systems have been extracted in the following forms, as depicted in Figure 3.1:

- Direct Trust: This form of trust is based on direct interactions between two entities. The trust relationship between these two entities will result from



(a) Direct trust establishment between two nodes.

(b) Indirect trust establishment between two nodes.

(c) Hybrid trust establishment between two nodes.

Fig. 3.1. Representation of trust establishment schemes.

observations and experience based on past interactions. Mathematically, they can be seen as:

$$T = f(X) \mapsto \{\text{observations, past/current experiences}\}\,.$$

In Figure 3.1a, the trust relationship between node i and node j can only be estimated if and only if there is a direct interaction between them, which then allows the construction and conservation of that trust.

- Indirect Trust: This type of trust relies on others' beliefs about a given entity. A trust relation will be estimated via the full transitivity property of the trust, and relies on a third-party opinion about a particular node. This can be done through a recommendation from either a mutual one-hop node neighbour or a multi-hop node recommendation. Mathematically, this can be represented as:

$$T = f(X) \mapsto \{\text{second-hand neighbour recommendations}\}\,.$$

In Figure 3.1b, the trust relations between node i and node j can be constructed based on k's opinions about j. This mechanism allows for the trust relation to be constructed without a requirement for current evidence.

- Hybrid: This form of trust utilises both direct and indirect trust to determine the evaluation of the overall trust. Mathematically, this can be represented as:

$$T = f(X) \mapsto \left\{ \begin{array}{c} \text{observations, past/current experiences,} \\ \text{second-hand recommendations} \\ \end{array} \right\}.$$

In Figure 3.1c, the trust relations between node i and node j are constructed according to the direct interaction as well as the indirect evidence. This approach is meant to strengthen the validity of trust by combining both methods.

3.2.3 Trust Notations and Representations

To represent the trust relationship between two entities abstractly, n_i and n_j, where n represents a node in the network with identities i and j respectively, Luo et al. (2009) employ the notation of $T_{i,j}$ (or equivalently T_{ij}). This value represents the trust that the trustor n_i has in the trustee n_j. Direct trust can be estimated as T_{ij}^d where the d refers to the direct dimension of trust. Then indirect trust can be represented as T_{ij}^r, which represents the estimated level of trust based on the recommendation obtained via the recommendation request.

In general, trust values are represented numerically or categorically across various models, depending on the application context (Singh et al. 2024). Numerical representations of trust typically utilise either discrete or continuous values. For instance, trust can be quantified using discrete values derived from raw data within predefined ranges. A common example is the range [-1, 1], where -1 signifies distrust, 0 represents neutrality, and 1 indicates complete trust. Some models also adopt threshold-based approaches, such as those commonly found in e-commerce systems. In these systems, user ratings are analysed to evaluate trustworthiness, and an entity is considered trustworthy and reputable if its rating meets or exceeds a predefined threshold. Alternatively, many models adopt probabilistic or similarity-based measures of trust, representing trust as continuous values within the range [0, 1]. In these models, 0 indicates a complete lack of trust, 1 signifies full trust, and intermediate values reflect varying degrees of trust between entities. These trust values, whether discrete or continuous, are critical for enabling systems to make informed decisions. For instance, a high trust score towards another node ($T_{ij} = 1$) implies absolute reliability, allowing unrestricted collaboration, data sharing, or access. On the other hand, a score of $T_{ij} = 0$ signifies complete distrust, triggering actions such as blocking interactions, denying access, or raising security alerts. Intermediate scores, when $0 < T_{i,j} < 1$, represent partial trust and may lead to cautious collaboration, limited permissions, or additional verification steps depending on the applications' policies.

3.3 TMM Process Scheme

Trust encompasses processes intended to extract evidence of trust for a particular node in a network. There are several approaches to computing the trust of entities in the cyber-space domain that differ. This work follows the process of trust as presented in Sharma et al. (2020) and is shown in Figure 3.2, which has been widely adopted in the context of IoT and MANET. This model has four main stages of trust: trust composition, trust computation, trust propagation, and trust update. A detailed explanation of each process is presented in the following sections.

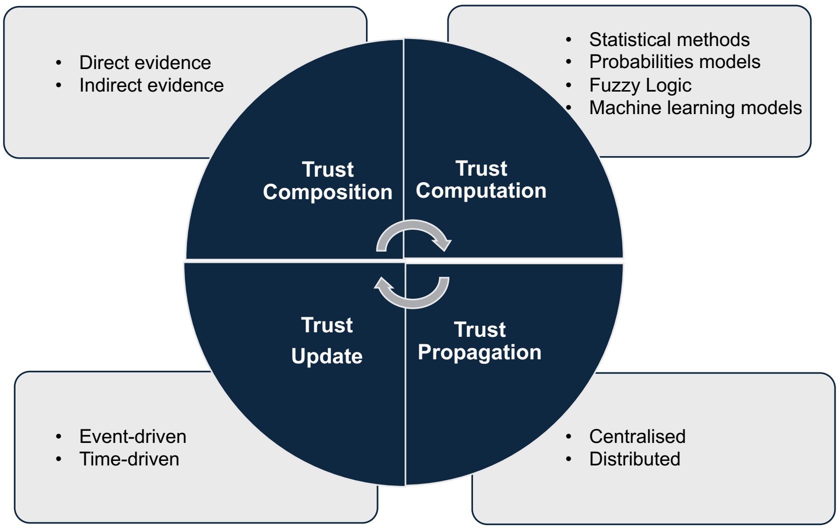


Fig. 3.2. TMM process.

3.3.1 Trust Composition

As stated earlier, for a trust relationship to be established, current/past observations or third-party opinions are required. In the literature, such collections of evidence are referred to as metrics, evidence, or parameters, and the process of creating trust refers to the phase in which trust evidence is collected and utilised to establish trust. During this phase, evidence of trust is gathered through both direct and indirect methods of information gathering. The type of evidence collected plays a fundamental role in establishing the feasibility of TMM, and is therefore crucial. According to Pourghebleh et al. (2019), the collected evidence can be categorised into two types: social trust and QoS trust. In social trust, evidence is derived from social relationships established between two entities. Metrics such as honesty, unselfishness, intimacy, and connectivity are used to measure and evaluate trust based on the nature of these relationships. Conversely, QoS trust measures an entity's ability to provide an acceptable quality of service. Assessment of factors such as energy consumption, competence, reliability, delivery ratio, and packet forwarding comprise part of this evaluation.

3.3.2 Trust Computation

This process handles the appropriate approach to computing trust based on the collected evidence by following approved mathematical models. Many computation models have been investigated for different types of systems, such as weighted sum, inference approaches, regression analysis, etc. (Li et al. 2017). These mathematical models estimate the trustworthiness of an entity, based on one or multiple factors and depending on the trust system.

3.3.3 Trust Propagation

To exchange trustworthiness among nodes, centralised and distributed approaches can be adopted in the TMM system. Applying the centralised method, an entity will be responsible for managing the process of trust for each node. In this method, the centralised entity will be charged for collecting the trust values from nodes, and evaluating, generating, propagating, updating, and revoking trust to all system nodes. The distributed approach assumes each node has its own TMM system, and can perform the entire process of computing, storing, and managing trust locally and autonomously, propagating trust across other nodes. Choosing which methods are more suitable is highly critical for networks with a low communication bandwidth along with the limited computation and storage capabilities of nodes. The trade-off between prolonging network lifetime and adequately establishing and managing trust between nodes is vital when considering different models of trust propagation. A centralised approach shares the trust values between nodes, requiring each node to transfer the collected evidence to a third-party and fully trusted entity (which could be a sink node). A centralised approach is needed to reduce the computation of trust for each node. However, this approach is strongly connected to the robustness of the centralised entity, as well the capacity of the communication to handle the transmission of evidence of trust and the resulting trust value. A fully distributed propagation model can support the scalability of systems, as each node will has its own trust management system. However, the low-resources restrictions of the devices on systems increase the overheads and energy computation, potentially increasing their energy consumption.

3.3.4 Trust Update

This phase focuses on managing changes in trust values over time. The trust update process can apply two primary principles: event-driven and time-sliding window approaches. In the event-driven approach, a set of activities, collectively referred to as an event (e.g., requesting a service or responding to a request), triggers the update process whenever a change occurs affecting that event. In contrast, the time-sliding window approach uses a timer to trigger the trust update after a specified period. The event-driven approach, which updates trust after each occurrence of an event, is then expected to increase the accuracy of trust values. However, it may consume a significant amount of energy and introduce overheads to the network, potentially degrading the network's overall lifetime. On the other hand, the time-sliding window approach, often considered a more suitable timeframe for updating trust, results in far fewer overheads when compared to the event-driven method. However, to trigger the trust update effectively, careful selection of an appropriate interval is necessary.

3.4 State of the Art on Trust within IoUT

Each stage of the trust process requires adaptation to meet the requirements of the IoUT network, which is typically described as highly dynamic, with costly resources and unreliable communication. The underwater environmental impacts

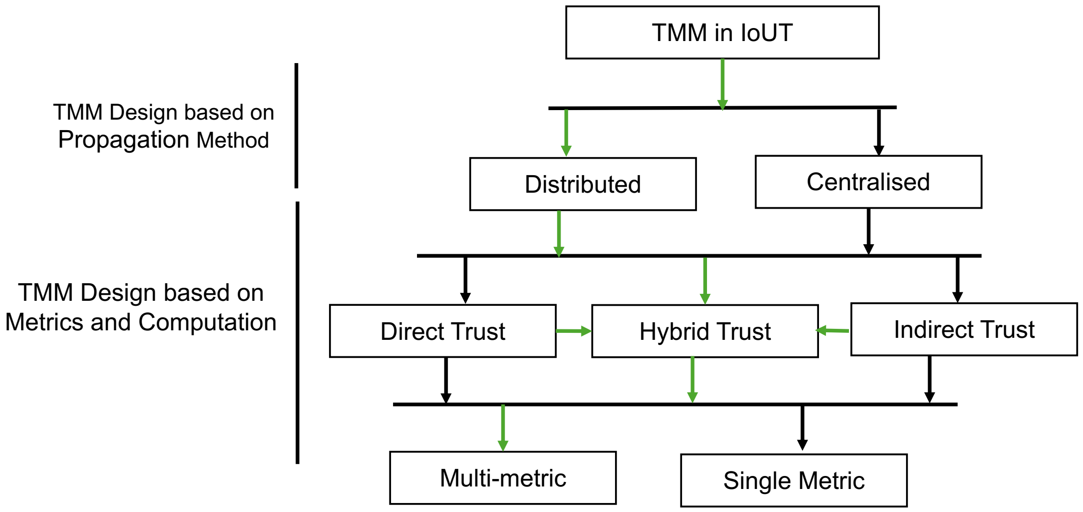


Fig. 3.3. Taxonomy of trust on IoUT.

communication channels, and exhibits different communication behaviours, leading to unexpected packet loss and continuous errors, that require frequent retransmission of signals (Jiang 2017). Figure 3.3, illustrates the trust taxonomy examined in this section to facilitate easy investigation of current state of the art knowledge. The green arrows in the figure represent the main focus of this thesis in relation to the existing literature.

3.4.1 Trust Modelling Design Classification

Following the propagation models of trust, either centralised or distributed, the TMM design varied based on the relevant propagation models. In other words, the trust compositions, computations, and updates will either be handled locally or via a centralised method, which changes the entire design of how the trust is conducted. The following classifications of existing trust models for underwater networks attempt to develop a trust model based on design decisions about how trust is established.

3.4.1.1 Centralised Trust Models

Prior studies have developed methods to address TMM complexity in underwater networks by adopting heavy computation models in a centralised way. For example, Han et al. (2019) propose a machine learning approach for underwater network TMM. The authors first employ K-means clustering to label the data, and then support vector machine (SVM) to compute the trust values for each node via underwater master nodes. These master nodes serve as cluster heads, and are responsible for training the collected data following an unsupervised/supervised model. This model was tested in a network with homogeneous underwater network. Arifeen et al. (2019) proposed an Adaptive Neuro-Fuzzy Inference System (ANFIS) and the Markov decision mechanism to establish trust between homogeneous underwater network. The model utilises a centralised cluster head responsible for computing trust values. Jiang et al. (2020) presents a C4.5 Decision tree mechanism to compute trust values in a centralised entity for UNs. Du et al. (2020) also present a model for computing trust for homogeneous underwater network, employing a centralised approach using an Isolation Forest. Su et al. (2021) presented a model to compute trust based on SVM and Dempster-Shafer theory. Their proposed model works for stationary UNs. Jiang et al. (2024) proposed a mechanism for detecting attacks within the trust model, using the random forest algorithm and an out-of-bag error rate algorithm.

With the centralised approach to TMM, a more sophisticated trust computation approach is proposed, owing to the enhancement of the trust estimation accuracy, although this introduces additional challenges. In particular, the proposed approaches rely on the central entity's availability, as well as constant and stable connectivity between nodes and this entity. This is not optimal given the current state of underwater communications. When nodes experience problems communicating with a central entity, the trust value cannot be computed, and the system becomes exposed to a single point of failure. Furthermore, such a system will not scale well, due to the network congestion caused by a high volume of nodes sending requests to the same entity.

3.4.1.2 Distributed Trust Models

This classification of TMM was developed to allow each UN to adopt a local TMM system to locally collect trust evidence, request recommendations, and estimate the trust of others. Based on this concept of trust, many trust systems have been developed that may serve as distributed trust models for low-capacity UNs. This concept of trust was proposed for, and subsequently explored on, several related systems under the Ad Hoc network. One of the significant contributions of trust modelling and evaluation methods was noted by Sun et al. (2006) as for securing Ad Hoc routing and detecting malicious nodes. The proposed model incorporates the concept of trust as a measure of uncertainty, which can be calculated using entropy. The authors define trust as a continuous variable that does not require transitivity, thereby capturing some of the characteristics of trust in MANET. However, this approach only considers packet dropping as the sole component of direct observations, used to evaluate trust. While this method can be effective in certain scenarios, it may not be adequate for capturing the complex nature of trust in IoUT, where trust can be influenced by a variety of factors that extend beyond packet dropping. Due to its applicability to the scalable nature of IoUT, in the following sections, distributed TMM will be the chief focus of investigation.

3.4.2 Direct Trust Models

This section explores current trust models mainly derived from personal experiences. Although some of these models are considered hybrid, the discussion focuses exclusively on approaches used within the direct estimation of trust. In the context of establishing trust with an emphasis on metrics and trust modelling, advanced trust models for IoUT are examined in two specific contexts: 1) static networks and 2) mobile marine MANETs.

3.4.2.1 Trust for Static Network

Research on Underwater Wireless Sensor Network (UWSN), which is a well-established framework for underwater communication among similar sensors, emphasises the challenges and limitations of existing trust models in underwater environments. In particular, the work of Han et al. (2015) is considered, which, to the best of the author's knowledge, is the first study in the domain of TMM in underwater networks. Han et al. (2015) argues that current TMM metrics are unsuitable for underwater networks and proposes a distributed three-metric, trust-based model to evaluate node trustworthiness using link, data, and node trust. The proposed model employs fuzzy logic and subjective logic methodologies to compute trust, but disregards uncertainty with regard to trust, and does not consider the impact of malicious attacks on trust evidence generation. The authors of this model assume a network of homogeneous static nodes. Different trust-related systems then followed the above approach, and proposed an enhancement to TMM, with the primary focus on addressing the link quality, to differentiate the misbehaviour that appears due to natural environmental limitations from that attributable to malicious attacks. Jiang et al. (2017) propose a cloud-based trust model for UWSNs that addresses trust uncertainty and fuzziness. Their model includes the effect of malicious attacks in trust computation, by analysing attack behaviour in each layer and identifying packet loss, packet error rate, and energy consumption as the primary outcomes of attacks. Similar to the previous study, the authors assume static nodes in their network.

While these models highlight the importance of considering environmental factors and underwater communication limitations, they nonetheless overlook crucial aspects. Specifically, they do not adequately address the diversity of network topologies and node mobility, which are prevalent in underwater networks that tend to be sparse in nature.

3.4.2.2 Trust in Mobile Marine MANET

The dynamic topology of IoUT motivated the adoption of the well-studied concept of MANET, and its security-related solutions including TMM (Govindan et al. 2011). While this appears to be a valid starting point from which to evaluate and explore the limitations of underwater networks relative to MANET, less research has been conducted in this direction.

Lowney et al. (2018) presents a TMM approach for underwater acoustic MANETs, building on the cloud theory model. While the original model was devised for terrestrial MANETs, the authors have adapted it to suit underwater networks. Their proposed model incorporates three significant parameters, namely the expected trust value, entropy to account for uncertainties, and hyper-entropy to capture randomness. However, the study fails to adequately address the influence of the environmental and dynamic characteristics specific to underwater networks. Furthermore, the feasibility of using the proposed model in the realm of the underwater environment has been ignored.

Bolster (2017) attempted to integrate a trust model that was originally developed for MANET called MTFM trust model introduced in (Guo et al. 2011), which was to be adapted to a marine squad of mobile nodes called a Multi-Domain Trust Assessment in Collaborative Marine MANETs (MTACMM). To create a more comprehensive trust framework, the study introduced additional physical metrics, such as Inter-Node Distance Deviation (INDD), Inter-Node Heading Deviation

(INHD), and Node speed (v), along with the existing network-related metrics, such as packet loss rate, signal strength, delay, and throughput. Each node in the marine squad assesses every other node, resulting in $\sim N^2$ assessment vectors at each time. While the adopted trust model has demonstrated robustness and accuracy in MANET, implementing it in marine networks results in certain limitations. One shortcoming has been reliance on Grey Relational Analysis (GRA) to evaluate the trustworthiness of the collected metrics. GRA requires a reference point from which to compare the behaviour of a particular node with its neighbouring nodes. Suppose a node at any instance of time has a shortage of neighbours. In that case, there will not be insufficient reference points to compare the behaviour of a particular node with its neighbouring nodes, rendering the trust framework ineffective. Moreover, changes in the underwater environment have not been considered in the testing and evaluation of the proposed model, which is the primary benchmark used to assess the efficiency of any trust model proposed for the underwater network.

A recent study presented by Jiang et al. (2022) has introduced a TMM framework known as the Controversy-adjudication-based Trust Management (CATM), which was explicitly designed to address the challenges posed by floating nodes in the IoUT. Within this trust model, direct trust is determined by considering network-related metrics and applying a weighted summation. One distinctive aspect of this model is its thorough consideration of the underwater environment's impact on communication as it pertains to establishing trust. However, it is essential to emphasise the efficacy of the CATM model predominantly in environments characterised by high node density. Consequently, it may not be readily applicable to networks with sparse and diverse node distributions, which is a common scenario encountered in IoUT.

3.4.3 Recommendation-based Trust Models

When TMM relies on indirect trust, the received trust about the trustee was subject to rogue and unreliable recommendations. Unreliable recommendations have been categorised into two types: dishonest recommendations and erroneous recommendations (Hua et al. 2021). Dishonest recommendations are a form of subjective unreliability, whereby adversaries intentionally offer misleading advice. Wrong recommendations, on the other hand, represent objective unreliability arising from unavoidable errors in transmission or computation. Wrong recommendations are expected to prevail in unpredictable and complex environments, such as with the application of IoUT. Therefore, existing techniques for filtering dishonest recommendations are expected to account for these challenges.

3.4.3.1 Validation of Recommendation Mechanisms

Addressing the challenge of mitigating the effects of dishonest recommendations in reputation and recommendation-based trust systems remains difficult. Models have been developed specifically to protect the integrity of reputation and recommendation systems from dishonest recommendations. For instance, the Collaborative Reputation Mechanism (CORE) (Michiardi et al. 2002) filters out negative feedback applying the assumption that maliciously provided positive feedback are non-existent, permitting only positive reputation data to spread and thus preventing fake negative opinions. In contrast, the CONFIDANT (Buchegger et al. 2002) model focuses on circulating only negative reviews about other nodes to thwart overselling by malicious entities who provide a fake positive opinion to address bad behaviours. Both approaches, however, struggle with efficiency in detecting unpredictable behaviours in complex networks, as they permanently block either positive or negative feedback. Khedim et al. (2015) outlines two primary methods for evaluating trust in networks: deviation tests and evaluation using trust factors. The deviation test involves comparing individual opinions about an entity against a predefined threshold, to identify and isolate disproportionately negative or positive recommendations. This method is exemplified in the Distributed Reputation-based Beacon Trust System (DRBTS), a distributed security protocol that enables beacon nodes to monitor one another and relay trust information (Srinivasan et al. 2006). The network is modelled as an undirected graph, and deviation tests are used to ensure consistency between first-hand (personal opinion) and second-hand (recommendation) information, thereby preventing the spread of false data. Similarly, the E-Hermes protocol, introduced by Zouridaki et al. (2009), integrates first-hand trust data, enabling nodes to compute independently with second-hand trust data obtained based on other nodes' recommendations. E-Hermes is designed to increase resilience against malicious nodes and recommenders, by implementing a recommender test that validates recommendations against the first-hand trust values computed by the inquiring node. This process ensures that only recommendations closely aligned with independently assessed values are accepted. However, such techniques, especially in dynamic and complex environments like the IoUT, are subject to limitations due to their heavy reliance on personal experiences, which can be problematic given the sparse and mobile nature of these networks. The decay of trust information and the infrequent opportunities for validation in rapidly changing networks make it difficult to maintain accurate trust assessments. Additionally, increased overheads and vulnerability to strategic manipulation by nodes which must adjust their behaviour to evade detection, further challenges the effectiveness of these methods in IoUT environments.

Others approach this issue by introducing so-called trust factors to help judge recommendations. For instance, in Liang et al. (2019), a trust-based recommendation model known as TBRS for vehicular cyber-physical system networks was introduced. This model evaluates recommendations using trust factors, such as contact intimacy, delivery reliability, and position intimacy. These factors are weighted according to Grey relational degrees to enhance judgment accuracy. Additionally, the K-Nearest Neighbour (KNN) algorithm is employed as a filtering mechanism to mitigate the impact of selfish or malicious nodes. However, the effectiveness of the KNN algorithm may be limited by its sensitivity to the choice of k, and by its performance in high-dimensional spaces, which might not adequately reflect the dynamic and complex networks.

Within the same context, two trust evaluation models were presented for vehicular network by Mahmood et al. (2023) and Huang et al. (2017), who introduced several metrics to weight recommendations, defined as familiarity (how well a one-hop neighbour is acquainted with the targeted vehicle), similarity (degree of similar content) and timeliness (the freshness of any reputation segment). Similarly, Shabut et al. (2014) presents a recommendation-based trust model, designed to address the challenges of maintaining reliable packet delivery through multi-hop intermediate nodes in MANETs. The proposed model introduces a defence scheme that employs a clustering technique to dynamically filter out dishonest recommendations from nodes. This approach may attempt to deceive the TMM system through the dissemination of fake recommendations. The core of the model concerns evaluating the honesty of the recommending nodes based on three key factors: confidence based on the interactions, compatibility of information (assessed through deviation tests), and physical proximity between nodes.

Adewuyi et al. (2019) introduce a belief function within the CTRUST model to assess the credibility of recommendations in collaborative IoT applications. This function is not designed to diminish trust in those recommendations, but rather to modulate their influence based on temporal and relational dynamics. The belief function is mathematically derived from several components: the decay of trust over time, existing trust scores between the recommender and the trustor, and the magnitude of change proposed by new recommendations compared to existing trust levels. The operational principle of this belief function is that it assigns minimal weight to recommendations when there are recent and direct observations that may contradict these recommendations, thereby prioritising empirical experience over hearsay. However, a potential issue arises when the calculated belief value becomes negative, due to a large discrepancy between the trust scores provided by the recommender and the existing trust scores of the trustor. This situation is not explicitly addressed in the paper.

3.4.3.2 Underwater Recommendation Models

Several attempts have been proposed to address the issue of dishonest recommendations on the underwater networks. For example, in regard to CATM, as (Jiang et al. 2022) discussed earlier, the authors suggest that trust can be wholly computed based on recommendations, when there is insufficient direct evidence, and it therefore proposed a validation mechanism to counter dishonest recommendations. In this model, they propose a mechanism to evaluate recommendations based on factors such as link stability and node reliability. In Du et al. (2022), a clustered-based trust model is introduced whereby recommendations are computed by the cluster head. Each sensor sends sensory data to the cluster head, and computes trust based on the assumption that data follows a normal distribution. Trust assessments then undergo median filtering to remove outliers and employ collaborative filtering to compute recommendations. This model, while robust, can be susceptible to inaccuracies in node importance assessment and deceptive behaviours among neighbouring nodes, which potentially undermine the reliability of trust recommendations in dynamic underwater environments. A further study by Zhang et al. (2023) introduced a trust model that effectively utilises both collaborative filtering and a variable weight fuzzy algorithm to exclude untrustworthy recommendations and pinpoint dishonest nodes. This model applies various filtering methods informed by deviation tests and the precision of the link quality, employing a preset threshold for each criterion. Based on assessments from recommendations, the link quality filter is vulnerable to manipulation by malicious entities, who might falsify the data to evade detection.

3.5 Trust Requirement for IoUT

Despite several attempts to enhance TMM within underwater constraints, there has been a notable lack of comprehensive requirements engineering for trust models in underwater domains. The mobility among UN, combined with the complexity of the environment, can easily render many well-established TMM approaches obsolete. This highlights a need for more standardised TMM requirements, as TMM in underwater domains is significantly more complex than in other domains. One approach to identifying these requirements is to start by investigating TMM demands in equivalent systems, such as IoT and MANET. However, the unique characteristics of underwater networks must also be validated against such requirements. Consequently, it is essential to extend this chapter to outline the ideal IoUT TMM requirements.

Generally, a TMM should balance security, functionality, and usability, while also considering the resource constraints and dynamic nature of the underwater environment. Based on the trust models explored in Section 3.4 and an analysis of several survey papers on trust models in IoT and MANET domains (Cho et al. 2010; Pourghebleh et al. 2019; Sharma et al. 2020), the requirements of the trust model are classified into platform constraints, utilising parameters, trust persistence, and reliability aspects. The following sections further explore these requirements, to define an ideal TMM to meet the specific demands of IoUT. Each section discusses an issue of relevance and then concludes with a list of resulting requirements.

3.5.1 Platform Constraints

TMM has been a widely studied topic in various domains, including e-commerce and healthcare systems, as well as applied extensively in high-reliability platforms. However, emerging technological paradigms such as IoT and, more recently, IoUT, present unique constraints that challenge the direct use of traditional TMM.

One major challenge associated with IoUT networks, is their reliance on resource-constrained devices, which are limited in power, computational capacity, communication bandwidth, and storage. These limitations create a need for TMM that operates with minimal resource consumption. Lightweight trust, as described by Zhu et al. (2024), involves efficient use of storage, low communication overheads, and simplified computational processes. In IoUT, where energy-intensive communication and scarce power sources heighten constraints, trust models require even greater eminent efficiency to function effectively within the network setting. The large-scale deployment of IoUT networks over sparsely populated areas introduces another challenge: scalability. Network functions, such as routing and localisation, have also been modified to operate in a scalable fashion across a wide area without significant performance degradation. This characteristic highlights the importance of scalable TMM as a way to manage interactions across growing networks. Additionally, the unattended nature of underwater applications poses further constraints. Devices are frequently left to operate independently for extended periods, making physical monitoring or interventions impractical. Under such conditions, the dynamic behaviour of network participants, combined with potential increases in malicious activity, requires mechanisms able to ensure resilience against attacks and with the capacity to adapt to adversarial changes within the network environment. Another important consideration concerns the decentralised architecture of IoUT networks. With no central authority available to manage communication or trust evaluation, nodes must independently monitor and assess the behaviour of one another. This decentralised nature relies on collaborative communication and local decision-making, emphasising the importance of designing trust mechanisms to align with these structural properties. Furthermore, the dynamic nature of IoUT networks, driven by factors such as node mobility, operational constraints (e.g., battery depletion), and environmental influences like water currents, presents additional challenges. Frequent adaptations to network topology require trust models that adapt to nodes joining or leaving the communication range, ensuring trust evaluations remain relevant in fluctuating conditions. Finally, trust evaluation in IoUT networks can be complicated by a lack of sufficient evidence, either due to limited historical interactions or disruptions caused by attacks. In such cases, the absence of adequate information poses a significant challenge to maintaining reliable trust computations. This highlights the need for mechanisms to allow trust evaluations to proceed, even under conditions of limited evidence, or during adversarial disruptions, ensuring the continuity of TMM throughout the network's lifetime. Therefore, existing trust models need to be enhanced to negotiate the limitations of such a platform.

The following are the main requirements that emerge because of the aforementioned specifications and constraints:

- Lightweight: to minimise resource usage, focusing on efficient storage, communication, and computation to suit resource-constrained underwater devices.
- Resilience: to resist malicious attacks and adapt to changes in node behaviour, ensuring accurate trust evaluations without necessitating physical intervention.
- Scalable: to handle large-scale deployments and increasing numbers of nodes efficiently.

- Decentralise: to enable nodes to independently monitor and evaluate trust, without relying on a central authority.
- Adaptability: to adjust to changes in the network topology caused by mobility, constraints, or environmental factors.
- Availability: to function reliably, even with limited evidence or during attacks, ensuring consistent trust evaluations.

3.5.2 Parameter Engineering

The earliest phase of any TMM involves gathering evidence of nodes' past and current behaviour, so as to establish well-informed decisions about the level of trustworthiness of other entities. The significance of any trust model is closely correlated with these parameters. The more reliable and efficient such parameters, the more robust and accurate the whole trust model will be. Moreover, the degree of trust associated with each metric needs to be addressed. For example, a node might have a degree of trust for another node in the network, with regard to accomplishing some tasks but not others. In addition, the mobility of physical nodes and other dynamics (e.g., dynamic environment, objects change their positions) affects associations between IoUT devices and the physical environment, and this may change over time. Based on the metrics used, the core requirements that need to be addressed to maintain an acceptable TMM within the IoUT are as follows:

- Robustness: to incorporate the necessary objective (QoS metrics) or subjective (social metrics) properties in the composition as trust parameters, so that the nodes can make accurate decisions.
- Contextual: to be assigned to a particular context to avoid having absolute trust in an entity.

3.5.3 Trust Persistence

According to Adewuyi (2021), TMM, especially for collaborative IoT, must provide an effective mechanism for persisting trust values, considering the low storage capacity and decentralised architecture of the devices involved. Within the IoUT domain, this characteristic of trust is highly valued for maintaining and storing trust-related data over time. Indicators of efficiency include the complexity of trust computation methods, rate of trust data updates, and time delays when collecting and retrieving trust-related data (Wei et al. 2022). Furthermore, due to the dynamic nature of underwater networks, trust should be updatable, incorporating historical interactions. Therefore, within the context of IoUT, TMM needs to address the following sub-requirements:

- Efficiency: to feature efficient computation methods, a fast rate of trust data updates, and low latency when collecting and retrieving trust-related data.
- Maintain Historical Trust: to store and use historical interactions effectively, while accounting for low storage capacity and decentralised IoUT architectures.
- Updatable: to include a mechanism to age and update trust values, maintaining accuracy over time.

3.5.4 Trust Reliability

In scenarios where trust is key to attack detection, the more accurate and faster the detection, the better this purpose is served. An accurate trust model is paramount to ensure the security of any underlying system. The following are the chief requirements for measuring the accuracy of the trust model.

- Trust Accuracy: to compute trust values that closely reflect a node's actual behaviour, ensuring high accuracy.
- Risk Mitigation: to effectively counter trust-related and malicious attacks, making it difficult for nodes to benefit from malicious actions.
- Convergence: to quickly converge to generate accurate trust values when node behaviour remains consistent, or when a nodes misbehave.

3.6 Summary

This chapter explored trust and the corresponding TMM as a security concept. Given that trust is a well-established concept in many IT systems, with a particular focus on IoT and MANET (due to their similarities with IoUT), definitions and properties of trust were examined. The TMM process for constructing trust relationships was also investigated. With a focus on decentralised TMM, this chapter expanded on prior literature, delving into key concepts related to TMM within the IoUT domain. A review of the main taxonomies of trust was also conducted to facilitate its application within IoUT. Additionally, the current state of TMM in underwater networks was explored to contextualize advancements proposed in this thesis. Finally, the chapter provided a wide range of trust-based requirements that need to be considered alongside the creation of TMM for IoUT. These requirements aim to guide the development of TMM in underwater networks. To explore and construct TMM within the IoUT, an in-depth examination of the network and its specifications is presented in the next chapter. This lays the foundation for the IoUT and facilitates the simulation of various network scenarios, providing a basis for testing and evaluating the TMM.

Chapter 4

Principles of Underwater Acoustic Networks: Theory, Setup, and Network Structure

4.1 Introduction

In the broader context of the IoUT, establishing a functional underwater network is the core focus among many research (Bello et al. 2022). A functional network requires the deployment of interconnected devices in aquatic environments, along with the ability to extract data without compromising the network's operational integrity. While, in theory, the IoUT integrates both above-water and underwater components—and the terrestrial segment is already well-developed—this chapter, along with the rest of the established research, will specifically focus on the underwater network. This focused emphasis is justified, as the underwater network forms the foundational structure necessary for fully harnessing the capabilities of the IoUT.

Due to the potential of acoustic signals to travel effectively underwater, UAC serves as the backbone for establishing underwater networks. However, UAC comes with its own limitations and deficiencies, as discussed in Section 2.2.2. Song et al. (2019) highlight how these limitations on the current development force trade-offs on the corresponding system designers. While there are some standards, such as ISO 13628-6:2006[1], which focus on subsea mechanical operations, the lack of generalisability in existing standards and the proper adaptation of UAC create significant differences and limitations in the application and utilisation of underwater networks.

On the other hand, and beyond real-world experiments, simulation, in general, offers an effective means to study, analyse, and optimise underwater networks. Simulating UAC between UN requires establishing an environment that reflects real-world conditions based on well-established theoretical models and empirical data, as the channel is understood for being noisy and having low available bandwidth and marked propagation delays. Therefore, this chapter proceeds with an investigation into the fundamental theoretical characteristics of UAC and establishes the foundation for the simulation environment that will be utilised for investigations in the subsequent chapters.

Besides the UAC, the possible formation of a network has a non-negligible impact on the performance of the network. For instance, the influence of ocean currents constitutes challenges to the current implementations, with the extent of their impact varying greatly depending on the underlying network's structural topologies. To address the above challenges, a classification of the potential topologies of underwater networks is provided, along with an exploration of possible placement strategies and mobility capabilities to serve as a guideline for future simulations of underwater networks. After establishing both the network and its corresponding communication model, the potential attack model to be utilised in subsequent chapters is introduced and demonstrated.

[1] It provides functional requirements and guidelines for ROV interfaces on subsea production systems for the petroleum and natural gas industries.

4.2 Exploration of the Underwater Environment

The key challenges in establishing efficient underwater networks revolve around the complexities posed by the slow and variable propagation of acoustic waves in water. These challenges manifest in significant propagation delays, frequency-dependent signal attenuation, and the inherently noisy communication channel. In the following sections, the impediments imposed by the underwater environment are examined, specifically focusing on the channel and environmental factors. The key characteristics of these factors are identified, extracted, and systematically modelled to facilitate simulations and testing of a functional underwater communication network. It is noteworthy to mention that, unless stated otherwise, most of the environmental information here is drawn from (Lurton 2002).

4.2.1 Acoustic Propagation Speed Model

The speed of sound in water exhibits substantial variability, unlike the relatively constant speed of sound in air, which is approximately 340 m/s. The velocity of an acoustic wave in water is governed by the properties of the local propagation medium, particularly its density (ρ) and the modulus of elasticity (E), or its relevant inverse quantity in fluid media, the compressibility χ can be represented by the following:

$$c = \sqrt{\frac{E}{\rho}} = \sqrt{\frac{1}{\chi\rho}}. \tag{4.1}$$

In contrast to the density of air, which is around $1.3kg/m^3$, sound propagates more effectively in fluid media. In seawater, the average density is approximately $\rho = 1030kg/m^3$, but this value is influenced by various physical parameters of the medium, such as temperature, pressure, and salinity. In marine sediments, which can be considered a fluid medium as a first approximation, the density ranges from 1200 to $2000kg/m^3$. In water-saturated sediments, where the sound speed is closely related to the speed of sound in the interstitial water, typical values range between 1500 and 2000 m/s.

Del Grosso (1974) provides an empirical estimation of the underwater propagation speed c with respect to varying environmental variables. One form of estimating c based on (t) as the water temperature in Celsius, (s) as water salinity level, and (d) as the water column depth in meters, following Del Grosso equation of computing the propagation speed is summarised by Spiesberger et al. (1991) and can be represented as follows:

$$
\begin{aligned}
c = 1448.96 + 4.591t - 5.304 \times 10^{-2}t^2 + 2.374 \times 10^{-4}t^3 \\
+1.340(s-35) + 1.630 \times 10^{-2}d + 1.675 \times 10^{-7}d^2 \\
-1.025 \times 10^{-2}t(s-35) - 7.139 \times 10^{13}td^3.
\end{aligned}
\tag{4.2}
$$

In order to utilise the Del Grosso formula to estimate the sound speed and the corresponding delay, data derived from WOA database is maintained. This permits the simulation of the propagation speed on different depths d between UN.

4.2.2 Attenuation Model

The reduction in power density of the acoustic signal as the signal travels from a sound source over a larger surface, known as attenuation, is inevitable for underwater communication. It is observed that when a sound signal propagates underwater, the signal energy spreads and is absorbed by mediums. According to Morozs et al. (2020), the acoustic signal path loss can be categorised into several shapes, with the two most influential loss forms as geometric spreading loss and absorption loss.

Geometrical spreading loss refers to the dispersion of acoustic power caused by the expansion of the acoustic wavefront as it propagates outward from the source in all directions. This type of loss intensifies with increasing distance and follows a logarithmic relationship with the ratio of the distance to a reference point.

In the simplest form, in the case of a homogeneous, infinite medium, with a small-domination source radiating in all directions, the energy transmitted is conserved, but is spread over spheres of larger and larger radii, which leads to the acoustic intensity (I) to be decreased with the distance from the source, in inverse proportion to the sphere surface. To measure the decrease in the acoustic intensity between two points (1) and (2) is then inversely proportional to the ratio of the surface of the spheres.

$$\frac{I_2}{I_1} = \left(\frac{4\pi R_1^2}{4\pi R_2^2}\right) = \left(\frac{R_1}{R_2}\right)^2. \tag{4.3}$$

Where R_i denotes the radial distance from the source, the intensity attenuates proportionally to $1/R^2$ and the pressure amplitude diminishes with $1/R$. This characterises the range-dependent amplitude behaviour typical of spherical wave propagation. The spreading transmission loss, considered from the reference unit ($R_{1m} = 1m$) can be expressed in dB as:

$$L_{\mathrm{spr}}(R) = k \times 10 \log_{10}\left(\frac{R}{R_{1m}}\right), \tag{4.4}$$

where k is the spreading coefficient in dB/km (k = 1: Cylindrical, k= 1.5: Practical, k = 2: Spherical), equivalent to the path loss exponent in terrestrial RF propagation models. Spherical spreading loss is commonly expressed as $L_{\mathrm{spr}}(R) = 20 \log_{10} R$.

The absorption loss refers to the conversion of acoustic energy into heat or other chemical reactions due to the seawater being a dissipative propagation medium. The absorption is often the most limiting factor in acoustic propagation. Its amount depends strongly on the propagation medium and the frequency. In seawater, absorption comes from pure water viscosity, relaxation of magnesium sulphate ($MgSO_4$) molecules below 100 kHz, and the relaxation of boric acid($B(OH)_3$) molecules below 1 kHz.

There are different ways to model the absorption of acoustic waves, with the predominant models in the field related to Throp and Francois-Garrison. Francois-Garrison's model is a more complex approach where the absorption decomposed into three terms: contribution of boric acid, magnesium sulphate, and pure water, and can be represented as:

$$L_{\mathrm{abs}}(f) = A_1 P_1 \frac{f_1 f^2}{f_1^2 + f^2} + A_2 P_2 \frac{f_2 f^2}{f_2^2 + f^2} + A_3 P_3 f^2. \tag{4.5}$$

In this formula, the first two terms represent the contributions from the two relaxation processes. Precisely, A_1, P_1, and f_1 shows the contribution of $B(OH)_3$, while A_2, P_2, and f_2 correspond to the contribution from $MgSO_4$. The third term accounts for the viscosity of pure water. More details about the related equations of this model can be found on (Lurton 2002).

Throp empirical model is a more convenient simplified mode for low-frequency sound below 50 kHz. The model is derived from ocean measurement data. The absorption loss coefficient of frequency (f) can be computed according to the Thorp formula (Porter 2011) as:

$$\begin{aligned} L_{\mathrm{abs}}(f) = 0.11 \frac{f^2}{1 + f^2} + \frac{44 f^2}{4100 + f^2} \\ +3 \times 10^{-4} f^2 + 33 \times 10^{-3}. \end{aligned} \tag{4.6}$$

The attenuation then can be formulated as:

$$P(d, f) = L_{\mathrm{spr}}(d) + d_{\mathrm{km}} L_{\mathrm{abs}}(f), \tag{4.7}$$

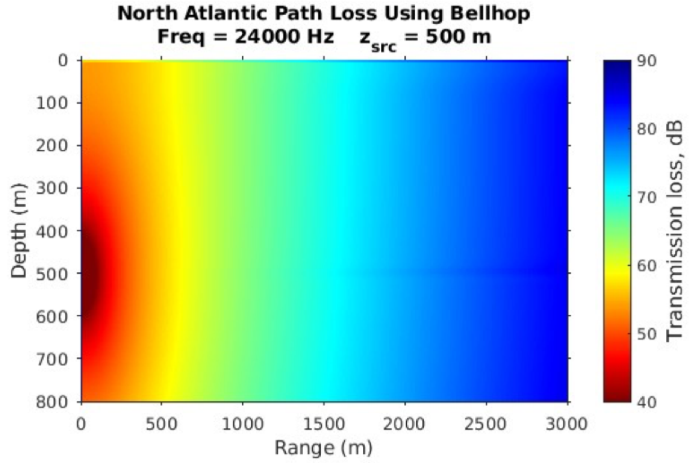


Fig. 4.1. Acoustic signal transmission loss between a sender located in 500m depth and transmitting at 24 kHz to a receiver situated at a depth of 550 (comprising geometric spreading and Thorp absorption).

where $P(d, f)$ is the power loss in dB, and it is a function of distance d and frequency f. $d_{\mathrm{km}} = d^{-3}$ represents the distance in km.

Figure 4.1 represents the path loss between a sender located at 500m depth and transmitting at 24 kHz to a receiver situated at a depth of 550. The path loss here was obtained via Bellhop[2] for illustration.

4.2.3 Noise Model

Underwater noises can be classified into self-generated and ambient noises. Self-generated noises are kinds of noises generated from UNs, such as electromagnetic noises, mechanical noises, and flow-induced vibration noises (Holmes et al. 2010). Ambient noise in underwater communication can originate from two broad categories of outside sources: natural and man-made. Natural sources include noise from wind, waves, and marine life, while man-made sources include noise from shipping, industrial activity, and other human-generated activities. Both types of noise can significantly impact the quality and stability of underwater communication, making it essential to mitigate their effects. The noise can be modelled

[2] Bellhop is a beam tracing model designed to predict acoustic pressure fields in ocean environments as part of the Acoustic Toolbox.

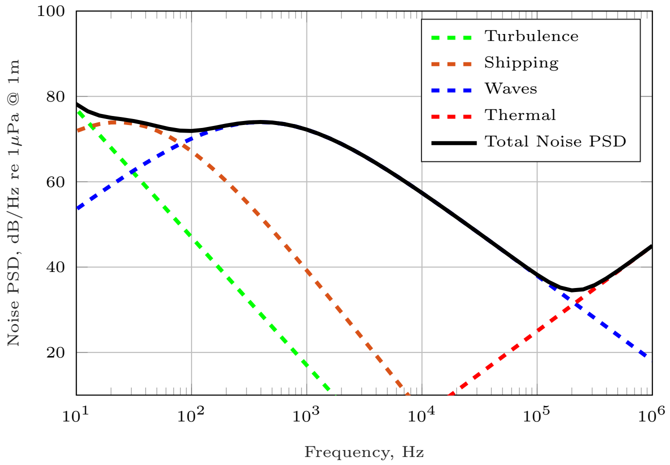


Fig. 4.2. Noise variations with frequencies, wind speed is chosen to be $w = 14$m/s while the shipment noise is considered to be $s_h = 0.5$.

based on four main sources of noise underwater as a function of frequency (f) as follows:

1. Shipping noise: Between 10-100 Hz, shipping is often the main cause of noise ranging between 60-90 dB reμPa/$\sqrt{Hz}$ [3]. The shipping noise can be estimated as follows:

$$N_s(f) = 40 + 20(s_h - 0.5) + 26 \log_{10}(f) - 60 \log_{10}(f + 0.03). \tag{4.8}$$

Here s_h is the shipping factor $\in [0, 1]$ where a value of 0 signifies a low level of shipping activity, while a value of 1 indicates a high level of shipping activity.

2. Wind noise: This source of noise usually affects the surface area over a frequency of a few hertz to a few tens of kHz. It can be estimated as follows:

$$N_w(f) = 50 + 7.5\sqrt{w} + 20 \log_{10}(f) - 40 \log_{10}(f + 0.4), \tag{4.9}$$

where w is the wind speed in m/s.

[3] Unit to express the sound pressure level in decibels (dB), relative to a reference pressure of 1 Pascal (Pa), measured at a distance of 1 meter from the sound source.

3. Thermal noise: Thermal noise is created by molecular agitation and can usually be with high impact beyond 100 kHz. This source of noise can be estimated as:

$$N_{\mathrm{th}}(f) = -15 + 20\log_{10}(f). \tag{4.10}$$

4. Turbulence noise: This noise arises due to the interaction of water with the seafloor or other underwater obstacles, creating vertices and eddies that generate sound waves. The estimation of the turbulence noise level can be computed as follows:

$$N_t(f) = 17 - 30\log_{10}(f). \tag{4.11}$$

The cumulative noise will be estimated in PSD as follows:

$$N(f) = N_s(f) + N_w(f) + N_{th}(f) + N_t(f) \tag{4.12}$$

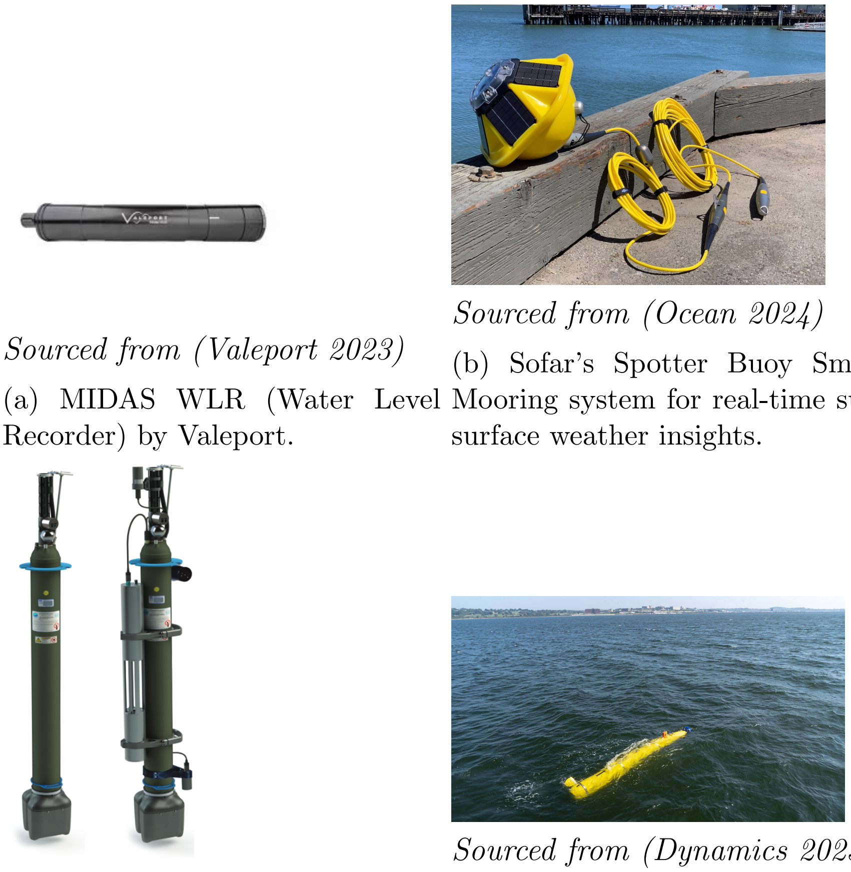
Figure 4.2 illustrates these noises as a function of frequencies where the wind speed is chosen to be 14m/s.

Using a high central frequency allows for the utilisation of a relatively large bandwidth for communication, resulting in a higher Signal-to-Noise Ratio (SNR) due to reduced noise. However, because of the inverse relationship between frequency and acoustic signal absorption, either higher transmission power or a shorter transmission range will be required to compensate for the increased attenuation. In essence, as the central frequency increases, the available bandwidth expands, but at the expense of higher attenuation. The level of attenuation sets the upper limit of the usable channel bandwidth for a given transmission range and power, while noise determines the lower limit, as noise levels rise with decreasing frequency. The communication frequencies for the applications of interest (e.g., industrial and military applications) are typically in the range of 10 Hz to 50 kHz (Cho et al. 2022). To account for this, the Thorp formula for absorption loss is applied in the simulation, along with spherical spreading loss. These choices ensure that the underwater networks are tested in a realistic and suitable environment.

4.3 Exploration of the Network Structure

As previously discussed in Section 2.2.1, the IoUT consists of a variety of UNs with different functionalities and capabilities. Despite the comprehensive theoretical framework of the IoUT architecture, there remains an absence of tangible, on-site applications that fully realise its holistic design. Current real-world applications of underwater networks tend to be more application-specific, with node formations tailored to the needs of particular use cases.

This section seeks to investigate the diverse potential topologies of the IoUT, with a particular focus on the spatial distribution of UNs across various configurations. Existing research has largely overlooked the exploration of different UN formations. Thus, this section serves two key purposes: first, to provide a comprehensive understanding of potential node formation and network topologies and the specific requirements for effectively simulating each configuration, and second, to create standardised testing case studies tailored to the distinct characteristics of each topology. These tests will be fundamental for evaluating various security solutions, ensuring that each is rigorously examined according to the specific requirements and constraints of the network topologies.



Sourced from (Valeport 2023)

(a) MIDAS WLR (Water Level Recorder) by Valeport.

Sourced from (Ocean 2024)

(b) Sofar's Spotter Buoy Smart Mooring system for real-time subsurface weather insights.

Sourced from (Scientific 2024)

(c) Navis BGCi Float.

Sourced from (Dynamics 2023)

(d) General Dynamics Bluefin-21 at ANTX 2018.

Fig. 4.3. Example of UN based on the design and application.

4.3.1 Network Topologies based on Variation of Node Structure

Since the focus is on exploring the underwater network, it is assumed that in each network structure, sink nodes are strategically positioned at sea surface stations endowed with virtually limitless resources. The potential UNs are then categorised based on their respective applications into networks of stationary nodes, anchored nodes, floating nodes, and mobile nodes. These classifications are intended to highlight the differences in each topology structure.

4.3.1.1 Stationary Nodes

These nodes are permanently installed on the seabed, initially configured in a fixed two-dimensional layout designed to remain in place throughout their operational lifespan. This topology is particularly suited for applications in shallow waters, where the limited depth facilitates the transmission of data to the surface-based sink nodes. Such topologies are primarily utilised in underwater monitoring applications, where UNs are deployed to collect geological data, including critical information for disaster prevention, such as earthquake or tsunami warnings. An exemplary device in this category is the MIDAS Water Level Recorder (WLR) by Valeport, as shown on Figure 4.3a, which features a 16MB memory capacity and an energy-efficient design, consuming merely 0.3W of power (Valeport 2023). These formations of UNs are representative of what is referred to as a 2D architecture.

4.3.1.2 Anchored Nodes

These nodes represent a specialised class within seabed monitoring systems, characterised by the deployment of UNs anchored at various depths. These networks typically consist of nodes tethered to the seabed or floating buoys, ensuring that they remain within a designated area despite the presence of water currents (one example presented in Figure 4.3b (Ocean 2024)). This configuration restricts node mobility, thereby maintaining their predefined positions within the network. Unlike the static topology, this approach extends the network's coverage into a three-dimensional space, allowing for enhanced monitoring capabilities. Applications of this topology include scenarios where UNs equipped with depth sensors are embedded in the nodes to extend their applicability across varying depths.

4.3.1.3 Floating Nodes

These nodes are distributed underwater and float, moving in response to water currents, see Figure 4.3c. Drifting buoys, commonly used in oceanographic research, fall under this category. These buoys are extensively employed for large-scale environmental monitoring, tracking ocean currents, and detecting pollution. Equipped with various sensors and communication systems, they gather and transmit data over time. The mobility of these nodes is inherently influenced by the dynamic forces of the surrounding environment, particularly the patterns of water flow.

4.3.1.4 Autonomous Vehicles

This category involves untethered, unmanned, fully mobile underwater vehicles, where scenarios explore the potential of autonomous vehicles patrolling three-dimensional underwater spaces (Jahanbakht et al. 2021). AUVs, in their nature, are equipped with navigation, energy, communication, and built-in sensory systems. Numerous applications leverage the capabilities of these AUVs, which are roughly categorised by Yang et al. (2021) to mainly military and tracking applications. A real mission of cooperative hunting is detailed by Zhao et al. (2022), where a group of AUVs is randomly positioned in formations like line, grid, or V-shape, and their movements towards a designated sink node are observed. Another notable application, as discussed by Zhang et al. (2021), utilises AUVs for monitoring oil leaks. The specifications for these AUVs during this study are modelled based on the Bluefin-21 model developed by General Dynamics, depicted in Figure 4.3d, capable of diving up to 4,500 meters, reaching speeds of approximately 8.334 km/h, and equipped with a total energy capacity of 7 kWh (Dynamics 2023).

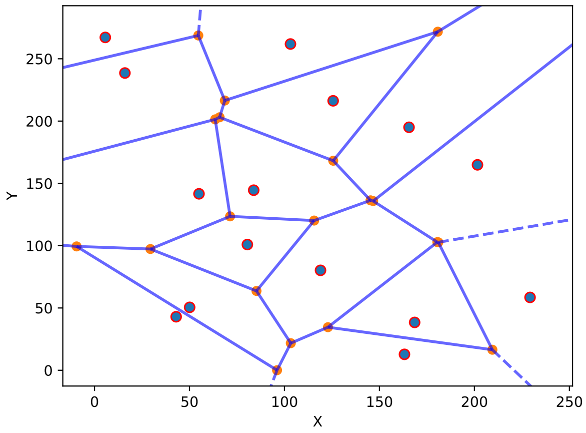


Fig. 4.4. Random placement strategy of 15 UNs with $R_t = 120$ m following the upper and lower bounds in 2D space.

4.3.2 Initial Placement Strategy

The initial placement strategy is paramount for apprehending the behaviour of a connected UNs in a functional underwater network. Deployment techniques, including static and self-deploying approaches, have been extensively studied in UWSN (Tuna et al. 2017). This section focuses on the placement method rather than the deployment technique. The aim is to identify the optimal node placement strategy that enhances network performance in terms of coverage, connectivity, and energy efficiency, regardless of the actual deployment method.

According to Chaudhary et al. (2022), an effective strategy for node placement in a 2D plane is the arrangement of nodes at the vertices of equilateral triangles, forming a triangular tessellation. The distance between adjacent nodes in this configuration, which corresponds to the side length of the triangles, can be calculated as $d = \sqrt{3}R_s$ (Choudhary et al. 2021). Here, R_s denotes the sensing range radius of each UN, which implies the coverage radius of each UN.

In contrast, for a 3D plane, a more sophisticated approach involves positioning the nodes at the centres of dodecahedron (Alam et al. 2008). The distance between nodes in this structure is given by: $d = a\sqrt{\left(\frac{5}{2}\right) + \frac{11}{5}\sqrt{5}}$, where $a = \frac{4R_s}{\sqrt{3}(1+\sqrt{5})}$. Both of these placement strategies are based on the assumption that the relationship between the sensing range R_s and the transmission range R_t satisfies $R_t \geq \sqrt{3}R_s$. This condition ensures that connectivity is maintained as long as the coverage of a given area is sustained. This allows the driving of the upper and lower bounds of the covered area to ensure network connectivity with N denoted as the number of UNs. These bounds are defined as (Boufares et al. 2015):

$$\begin{aligned} v_l &= N \times \frac{4}{3}\pi R_s^3, \\ v_u &= v_l + 2\left(\frac{h \times l}{\pi R_s^2} + \frac{h \times w}{\pi R_s^2} + \frac{l \times w}{\pi R_s^2}\right) + 4\left(\frac{h}{R_s} + \frac{l}{R_s} + \frac{w}{R_s}\right), \end{aligned} \tag{4.13}$$

where v_l and v_u represent the lower and upper bounds of a cubical area, respectively, and h, w, l represent the dimensions of the underwater volume being monitored. For instance, with a transmission range $R_t = 120$ meters, the corresponding sensing range R_s can be estimated as $R_s = \frac{R_t}{\sqrt{3}} \approx 69.28$ meters. The lower and upper bounds of the area that can be covered are then roughly determined based on this sensing range R_s. The lower bound area v_l represents the minimum coverage area under ideal placement conditions and is approximately 20,000,000 m^3 cubic meters. The upper bound area v_u, which accounts for additional factors such as boundary effects, is approximately 25,000,000 m^3 cubic meters. These bounds provide a practical estimate of the area that $N = 15$ nodes can effectively cover, with specific values depending on the actual sensing range R_s and the spatial distribution of the nodes within the network. Restricted by the upper and lower bounds of the area, a random distribution of underwater nodes can be set up under these constraints to maintain the required connectivity. Figure 4.4 illustrates an example in 2D of following the above boundaries to allow for distribution of these nodes in a provided space.

4.3.3 Establishing the Mobility Models

Each scenario in this analysis exhibits variations across multiple dimensions, with the most prominent distinction being the type of node mobility. This section delves into a detailed exploration of the distinct mobility patterns exhibited by anchored, floating, and mobile nodes.

Understanding and modelling the movement of the underwater environment, shaped by the influence of water currents, is essential for replicating realistic conditions in underwater networks. While current network simulation tools, such as those explored in Section 2.4, strive to reflect the characteristics of acoustic channels, they often oversimplify the complex mobility of water currents. This oversight leads to most of the current development relying on the assumption that water currents drift horizontally at a constant speed at each depth. However, in reality, water currents are shaped by numerous geographically specific factors, and the interaction between fluid dynamics and objects is far more intricate than a two-dimensional drifting model suggests. The following subsection explores the established mobility models, addressing both the movement of water currents and the utilised mobility models across all the network formations.

4.3.3.1 The Influence of the Water Current

There are two models with varying levels of realism to explain the drifting effect of the water current on any object: Meandering Current Mobility and the Ekman-based theory. The former explains the effect of meandering sub-surface currents and vortices on objects and suggests that the oceans are a stratified, rotating fluid (Caruso et al. 2008). Hence, vertical movements are almost everywhere, negligible with respect to the horizontal ones. This relaxed assumption, negligible of upwell and downwell influence of the movement, led to the explanation of the two-dimensional drifting with the function ψ, which is a scalar function whose contours represent streamlines of flow. The two components of the divergenceless velocity can be calculated as:

$$u = -\frac{\partial \psi}{\partial y}; \quad v = \frac{\partial \psi}{\partial x}. \tag{4.14}$$

Here u is the zonal eastward component of the velocity field while v is the meridional (northward) one. Following the Hamiltonian ordinary differential equation, the trajectory of the moving object with the current can be estimated as $\dot{x} = -\partial y \psi(x, y, t), \dot{y} = \partial x \psi(x, y, t)$.

On the other hand, the Ekman theory explains the expected drifting as a function of depth, which elucidates the phenomenon where wind induces the movement of a fluid, such as water, at an angle to the wind direction due to the Coriolis force[4], which results from Earth's rotation (E_r) and influences the motion of objects (Price et al. 1987). This leads to Ekman transport, which varies in intensity and direction with depth. The transport is most pronounced at the surface, where wind influence is strongest, and diminishes with increasing depth. Additionally, the angle of transport shifts with depth, deviating from the wind direction at the surface and aligning more closely with it at greater depths (Stull 2015).

The Ekman layer depth is defined as:

$$D = \sqrt{\frac{2K}{f_c}}, \tag{4.15}$$

[4] The Coriolis force is an apparent force caused by the Earth's rotation, which deflects moving objects to the right in the Northern Hemisphere and to the left in the Southern Hemisphere

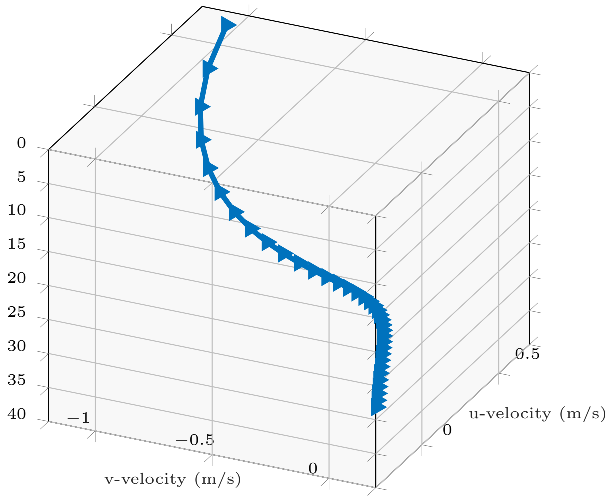


Fig. 4.5. The drift by the Ekman model on an object with depth varies from 0 (on the surface) to 40m below the surface.

where K is the eddy viscosity parameter for measuring the momentum of ocean turbulence, and f_c is the Coriolis force. This will be used in the computation of the equilibrium horizontal ocean-current components $(\overline{u}, \overline{v})$ as:

$$\begin{aligned} \overline{u} &= \left[\frac{w_{fv}}{\sqrt{K f_c}}\right]\left[e^{\frac{z}{D}}\cos(\frac{z}{D} - \frac{\pi}{4})\right] \\ \overline{v} &= \left[\frac{w_{fv}}{\sqrt{K f_c}}\right]\left[e^{\frac{z}{D}}\sin(\frac{z}{D} - \frac{\pi}{4})\right] \end{aligned}. \tag{4.16}$$

Here z is the depth of the water column, while w_{fv} is the friction of velocity of water that can be found from the following:

$$w_{fv} = \frac{\rho_{air}}{\rho_{water}}\alpha_{fv}, \tag{4.17}$$

where α_{fv} is the air friction velocity that can be given by Charnock's relationship that describes the aerodynamic roughness length over a water surface (Garratt 1977). ρ_{air} and ρ_{water} are the density of air and water, respectively.

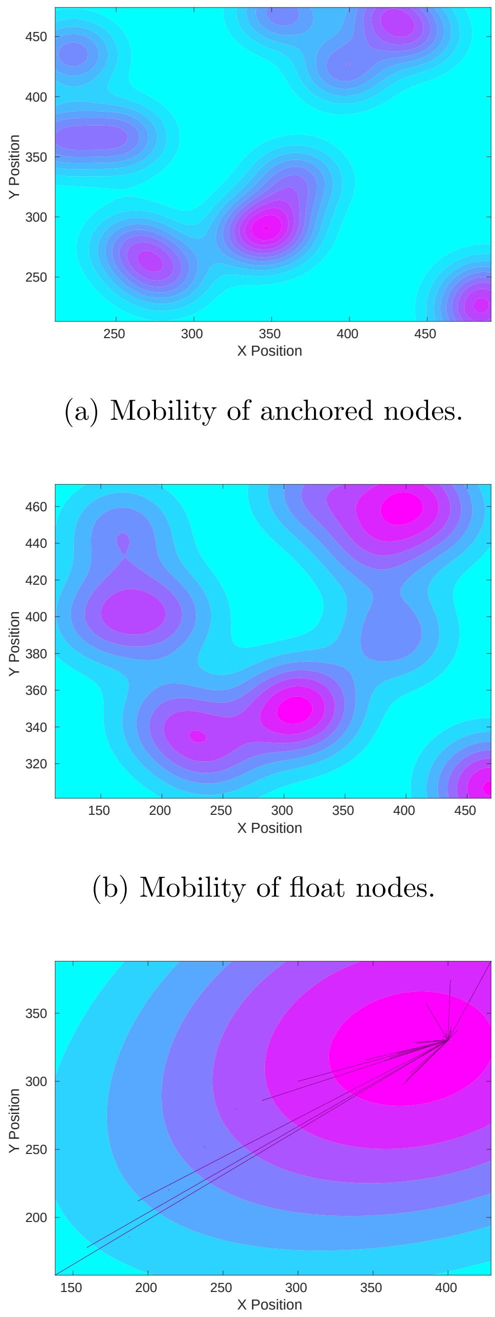
Figure 4.5 illustrates the expected spiral drift by the water current on the speed component based on the variables shown on Table 4.1.

Table 4.1: Parameters of the Ekman Model.

Attribute	**Description and Value**
Eddy Viscosity	$K \approx 0.4\|z\|w_{fv}$ $[\mathrm{m}^2/\mathrm{s}^{-1}0.]$
Latitude (l)	45°
Surface Wind Angle	30°
Surface Wind Speed (w_s)	14 $[\mathrm{m}/\mathrm{s}^{-1}]$
Coriolis Force (f_c)	$f_c = 2E_r \sin(l)$ $[\mathrm{s}^{-1}]$
Air Friction Velocity (α_{fv})	$\alpha_{fv} \approx 0.00044(w_s)^{2.55}$ $[\mathrm{m}^2/\mathrm{s}^{-2}]$

4.3.3.2 Mobility Across Network Structure

Figure 4.6 provides an example of the modelled mobility under the aforementioned conditions. The modelling of water current movement is instrumental in visualising the floating behaviour of nodes, allowing for the estimate and simulate of the drifting effects of the surrounding environment on UNs at varying depths. This modelling approach also facilitates the estimation of mobility patterns for anchored nodes, which, by design, are restricted in movement due to their attachment to either the seabed or the water surface within a specified area. In this study, it is assumed that all UNs are anchored at different depths, with the primary variation between them being the length of the tether connecting the nodes to their anchor points. Scenarios involving anchoring nodes with depths ranging from 50 to 150 meters are explored, enabling a detailed analysis of network coverage and performance under controlled environmental conditions. The differing tether lengths result in variations in mobility, as nodes tethered at greater depths are more susceptible to the influence of water currents compared to those closer to the surface. This allows for the assessment of how mobility, even in a more restricted environment, affects network connectivity and communication efficiency.



(a) Mobility of anchored nodes.

(b) Mobility of float nodes.

(c) Mobility of mobile nodes.

Fig. 4.6. Node trajectories with Kernel Density Estimate (KDE) heatmap showing density areas where nodes frequently occupied positions after 1 hour of the simulation time.

The formation of AUV, as outlined by Yang et al. (2021), typically involves a cohesive group structure that follows a defined trajectory. This formation control is designed to mitigate disruptions caused by unknown ocean topography and marine organisms. To maintain stability within the formation, AUVs continuously share information regarding their speed and relative distances to ensure coordinated movement. This communication enables each AUV in the fleet to maintain the required spacing, based on collision avoidance protocols and pressure considerations. In this study, a team-based mission scenario is examined in which AUVs, whereby AUVs followed a trajectory of visiting places on the area that eventually converged at the sink nodes. The AUVs select a target place and move to the target.

Figure 4.6 represents a 2D visualisation of node movement paths and positional density. The arrows represent the trajectories of nodes in the 2D plane, with each arrow showing the direction and magnitude of movement between consecutive positions. A Kernel Density Estimate (KDE) is overlaid to indicate the density of node positions over time, with darker red areas signifying regions where nodes have spent more time. The KDE was generated using a Gaussian kernel, providing a smoothed estimate of the spatial distribution of the nodes. The effects of water currents in a selected area are simulated, following the specifications outlined in Table 4.1.

4.3.4 Performance Evaluation under Different Topology

This section provides an examination of the performance of the UN topologies under the specified mobility and communication model constraints through simulations conducted using Aqua-Sim ng (Martin et al. 2017). The kinematic and technical specifications of the UNs are modelled based on previously established devices, as detailed in Section 4.3.1. The performance evaluation focuses on providing a general understanding of key evaluation metrics within the established topologies and simulation setup. The following sections detail the simulation setup and the overall performance insights derived from the conducted tests.

Table 4.2: Simulation Parameters to Test the Performance of Network.

Variables	**Value**
Simulation Time	3600 seconds
Number of Nodes	15
Surface Wind	8.5 - 27.2 knots
Communication Model	Acoustic Signal
Propagation Model	Range-based
Initial Energy	Varies (10000-70000) watt
MAC Protocol	Broadcast MAC
Routing Protocol	Vector-based Forwarding Protocol (VBF)
Transmission Range	120 meters
Data Rate	10000 bps
Payload Data	40 bytes
Carrier Frequency	25 kHz
Signal Power	0.2818 Watt
Noise Power	96.7655 $dBre\mu Pa@1m$

4.3.4.1 Simulation Setup

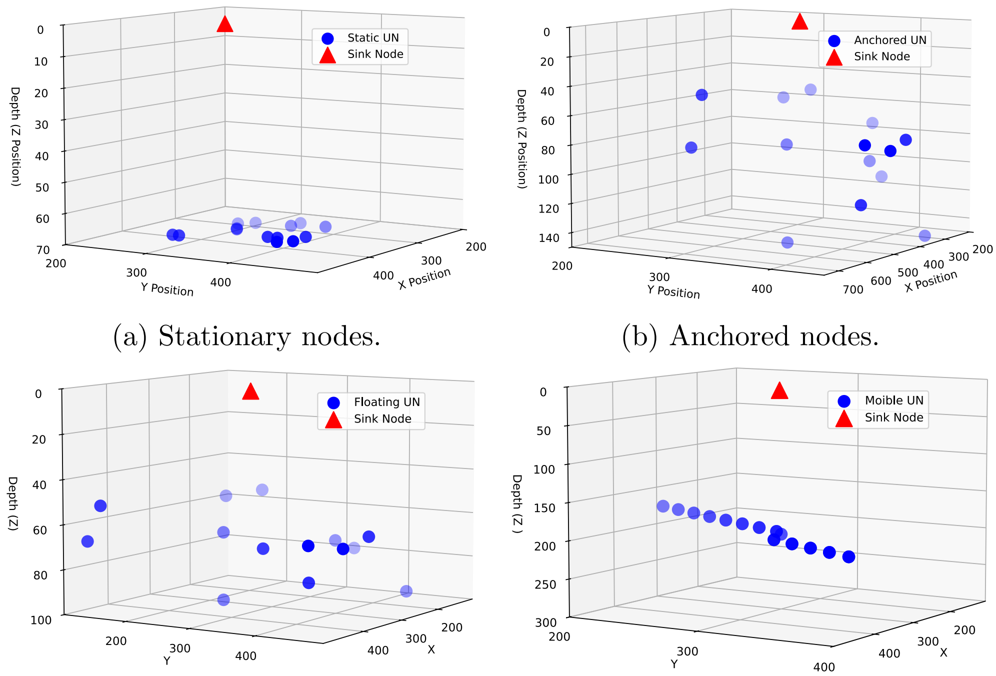
Each UN is presumed to have the essential hardware to facilitate underwater communication, including an acoustic communication medium for signal transmission and reception. Additionally, the UNs are assumed to have onboard energy sources to sustain long-term operations. Furthermore, each UN is equipped with sensors to monitor environmental conditions, such as pressure, temperature, salinity, and motion sensors. The UN are also assumed to possess acceptable processing units, e.g., ARM processors, enabling them to manage communication protocols, sensor data processing, and the execution of control algorithms. Moreover, it is reasonable to assume that each UN is equipped with adequate storage and memory, along with an efficient data management system to handle the data volumes generated during underwater operations. This includes systems for buffering, storing, and transmitting data efficiently. The UN are assumed to be integrated with acoustic localisation systems as explored in section 2.3, allowing them to determine their positions relative to other nodes and the environment.

The transmission parameters, such as signal frequency and power, were adopted from established research, e.g., (Morozs et al. 2020) and (Wang et al. 2023). The OnOff application was configured to generate traffic using the `OnOffHelper` class of Aqua-Sim ng, with a packet socket as the underlying transport. The data rate during the 'On' periods was set to 10 kbps, and the packet size was fixed at 40 bytes (320 bits). Given this configuration, the emission rate during the 'On' period is about 31.25 packets per second. The range-based propagation model utilises Equation 4.2 to compute propagation, which varies with depth. During these tests, the Vector-based Forwarding Protocol (VBF) provided by Mazinani et al. (2018), along with a broadcast MAC approach, is employed to meet both the MAC and network communication demands. More detailed breakdown of the simulation variables can be found in Table 4.2. Unless specified otherwise, all other tests simulated in this thesis adhere to the parameters outlined in this table.

The UNs are initially distributed randomly in each scenario, taking into account the predefined lower and upper bounds, as outlined in Equation 4.13. In other words, the UNs are distributed within a cubic area limit, as shown in Figure 4.7. The exception is the scenario involving mobile nodes, illustrated in Figure 4.7d, where the node formations presented by Yang et al. (2021) are followed, as mentioned earlier. The stationary nodes are set to be distributed in 2D space, as shown in Figure 4.7a, while both anchored and floating nodes are distributed in 3D space, as seen in Figure 4.7b and Figure 4.7c, respectively.

4.3.4.2 Evaluation of Simulated Scenarios

In each scenario, twenty independent simulation runs were conducted, with variations in the initial node deployment to introduce diversity in the network's topol-



(a) Stationary nodes.

(b) Anchored nodes.

(c) Floating nodes.

(d) Mobile nodes.

Fig. 4.7. An example of initial deployment across topologies for 15 nodes.

ogy. The evaluation focused on two key performance metrics: Packet Delivery Ratio (PDR) and end-to-end delay, each defined as follows:

1. PDR is defined as the ratio of the total number of distinct data packets successfully received by the destination node to the total number of data packets transmitted by the source node. In the evaluation of PDR, any duplicated packets are only counted once.

2. The average end-to-end delay represents the time taken for data packets to travel from the source node to the destination node. In cases where multiple duplicates of a packet are received, only the first-arriving packet is considered. This delay can be computed using the following:

$$t_{ij} = \frac{\sum_{i=1}^{n} \left(S_{i(t)} - D_{i(t)}\right)}{N}, \tag{4.18}$$

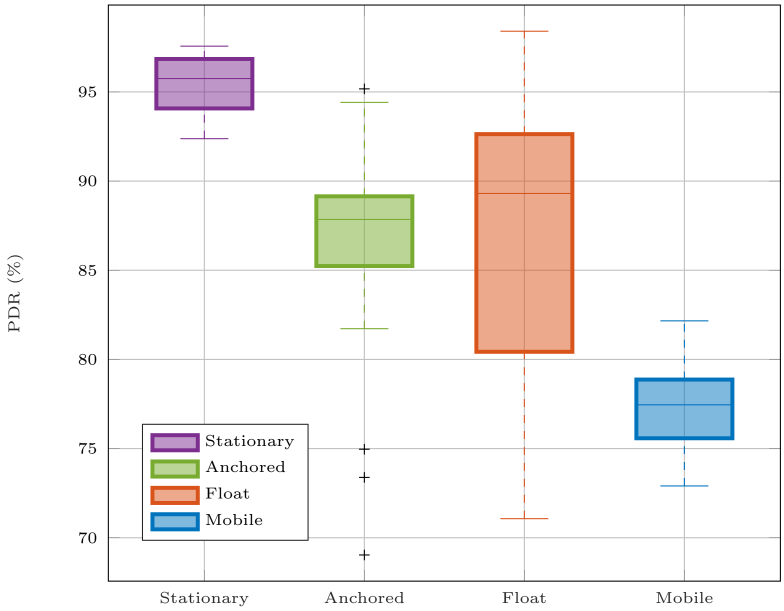


Fig. 4.8. Packet Delivery Ratio (PDR) across trails.

where, N denotes the total number of received data packets. $S_{i(t)}$ is the time at which the source node transmits the ith data packet, while $D_{i(t)}$ is the time at which the destination node receives the data packet.

Figure 4.8 shows the PDR results obtained in each run. In the stationary scenarios, the highest PDR is maintained, with averages ranging between 92% and 98%. However, as mobility is introduced in the other scenarios, PDR tends to decline, with notable drops in PDR observed during multiple runs, reaching as low as 70%. The variation of end-to-end delay with different runs is detected in Figure 4.9. The estimated end-to-end delay exhibits slight variations across different runs and topologies, with the maximum delay reaching approximately 1.6 seconds in the floating network, while the minimum delay across all scenarios is around 0.2 seconds.

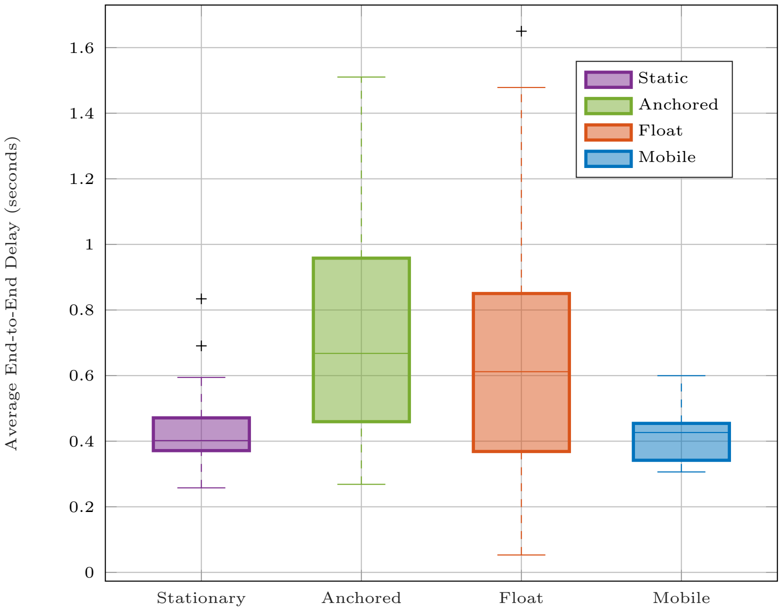


Fig. 4.9. End-to-end delay across trails.

4.4 Establishing the Expected Attack Model

The nature of the IoUT is highly susceptible to various types of malicious attacks. In this study, two primary sources of attacks are examined: physical and communication-related attacks. These two forms of attacks are interlinked and can significantly impact the security of the IoUT systems.

4.4.1 Physical-related Attacks

Zhang et al. (2003) states that, in mobile wireless networks, nodes with inadequate physical protection are susceptible to being captured, compromised, and hijacked. Upon contextualising this threat within the IoUT, the inherent openness of the IoUT networks, combined with the inability to monitor the UNs continuously, introduces a heightened vulnerability to physical misbehaviour compared to similar terrestrial networks. This malicious behaviour is defined as a Physical Mobility Attack (PM). While much of the research has concentrated on communication-related attacks, such as selfish or malicious actions targeting routing and data integrity and availability as explored in Section 2.5.2, limited attention has been given to addressing the potential of intruding nodes to exploit their physical mobility within the network to carry out these attacks. This study aims to address two specific scenarios that illustrate the severity of PM attack:

1. Misbehaving nodes purposefully move to network regions with lower communication demands and overhead in order to conserve energy and evade collaboration. By avoiding collaborative efforts and selecting less congested network segments, these nodes prioritise energy preservation over participating in cooperative tasks or sharing resources with other nodes. This behaviour enables them to optimise their individual resource usage at the cost of network efficiency and collaborative operations.

2. Misbehaving nodes strategically position themselves at the network's core to optimise their ability to capture, monitor, and exert control over the flow of network traffic. By infiltrating the network's central region, these nodes gain privileged access to a significant portion of the network's communication flows. From this advantageous position, they can intercept and analyse network traffic, extract sensitive information, and potentially engage in malicious activities, such as altering data, disrupting communication, or compromising the integrity and confidentiality of the transmitted information.

4.4.2 Communication-related Attacks

In addition to the physical attack, this study explores various other malicious acts, including:

1. Denial of Service Attack (DoS): This attack involves a malicious entity intentionally flooding the network with excessive packets, taking advantage of the IoUT's limited communication window.

2. Selfish Behaviour Attack (SB): This type of attack occurs when an attacker prioritises their own benefit over network collaboration. The attacker may choose to preserve their energy or intentionally disrupt regular network communication, resulting in degraded network performance.

3. Selective Forwarding Attack (SF): This type of attack occurs when a compromised or malicious node selectively chooses which packets to forward and which ones to drop. Through the act of selectively dropping packets, the attacker disrupts the normal flow of information, leading to packet loss and compromised network performance.

4.5 Conclusion

This chapter provided an in-depth understanding of the interaction between environmental factors and network architecture in underwater networks. It served multifaceted purposes, including acoustic theory, simulation methodology, network classifications, and security considerations. To create a realistic simulation environment, background information is provided on UAC with a focus on propagation speed and estimated path loss under various noise sources. These aspects form the backbone of the simulation environment. Due to the limited literature on broader classifications for potential IoUT network formations, this chapter presents a classification based on application requirements for networks comprising stationary, anchored, floating, and mobile components. Based on these classifications, a placement strategy is suggested for both 2D and 3D formations that align with coverage and connectivity needs. The mobility of the water medium, influenced by currents, and the movement of UNs were examined to provide guidelines for potential IoUT applications. Lastly, the attack model within the established network is analysed, exploring both the physical and communication domains. While the attack space is broader than the discussed examples, the severity of these specific attacks is highlighted, showing their potential as primary sources of malicious activities in subsequent chapters. The findings serve as a basis for designing testbeds for operational underwater networks, focusing on the impact of the UAC on overall network performance. The results of this investigation lay the groundwork for future research into more complex underwater networks, where the arrangement of UNs influences network efficiency, scalability, and resilience.

Chapter 5

Evaluation of Trust Mechanisms in the IoUT

5.1 Introduction

This chapter examines the application of TMM within the IoUT, with a particular focus on various network topologies. The analysis is intended to determine the readiness of the current TMM to function effectively under diverse operational conditions, particularly in the detection of malicious behaviour. In order to achieve the aim behind this chapter, TMM in underwater networks is classified into two primary problem domains: the problem of formulating direct trust based on selected metrics and the problem of inspection of recommendation reliability, which in turn supports the establishment of indirect trust. The first problem domain focuses on developing a comprehensive understanding of current approaches for establishing trust between UNs. Subsequently, the performance of existing models is evaluated against pre-established network topologies outlined in Section 4.3.1 and attack scenarios described in Section 4.4. Specifically, the investigation examines whether the TMM can effectively identify malicious behaviour and determines which metrics contribute most significantly to detecting such actions.

Moreover, given the sparse and often unpredictable nature of IoUT networks, establishing trust becomes particularly challenging due to the limited availability of direct evidence regarding the behaviour of neighbouring nodes. This scarcity necessitates the reliance on third-party recommendations, which introduces the issue of dishonest or unreliable recommendations. The second section discusses the impact of this problem and explores how deceptive recommendations can undermine trust formation. Potential strategies to address this challenge are outlined, illustrating improvements designed to strengthen TMM in IoUT.

5.2 Preliminaries of Trust Establishment

To conduct this study, certain assumptions have been established. It is assumed that all nodes must be authorised when allowed into the network and, although they initially cooperate, some nodes may misbehave after joining. In the process of trust establishment, each node evaluates the reliability of its neighbouring nodes based on the information it possesses. The diagram presented in Figure 5.1 illustrates the various entities involved in this trust process. The node responsible for evaluating trust is referred to as the trustor, represented by the green node in the diagram, while the entity being assessed is known as the trustee, depicted as a grey node. Trust evidence denotes the data utilised to estimate the trust of the trustee. The outcome of applying the TMM to the gathered evidence is a trust value, which signifies the level of trust the trustor places in the trustee node. It is anticipated that the trust score, as determined by the model, follows a continuous representation and falls within a scale from 0 to 1. A score of 1 represents complete trust, whereas 0 signifies zero trust in the node. Occasionally, the trustor may face difficulties forming a valid judgment regarding the trustee, often due to insufficient evidence about the trustee's behaviour. In such situations, the trustor seeks input from its neighbouring nodes to obtain their perspectives on the trustee. The response provided by the neighbouring node is referred to as the recommendation about the trustee. In Figure 5.1, the recommender can be

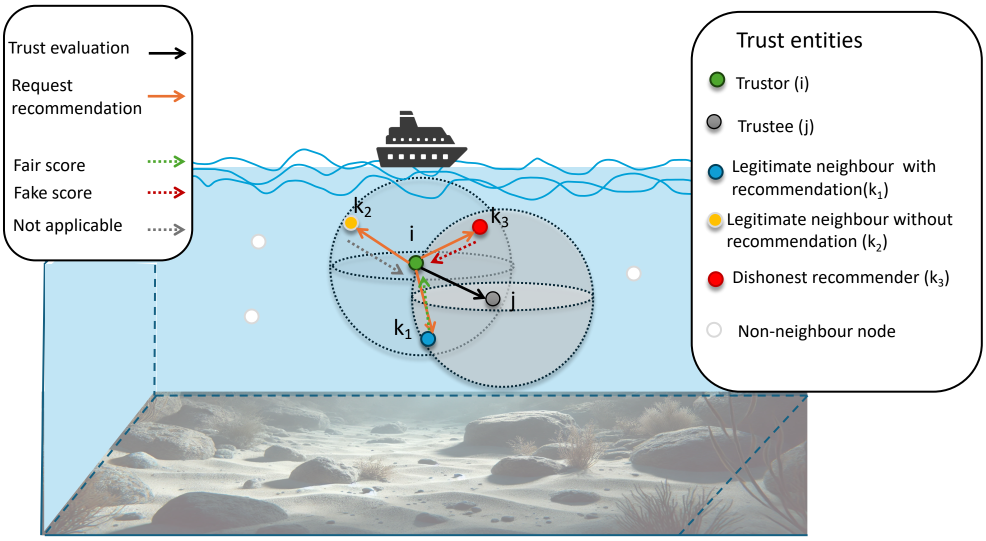


Fig. 5.1. Components of trust evaluation in underwater networks.

fair (depicted as the blue node), a malicious recommender (represented by the red node), or a recommender without any recorded trust information about the trustee (orange node in the diagram).

In the context of this study, the term ‘trust period’ refers to the duration during which the trust score is initiated and updated in accordance with time. It encompasses establishing trust and continuously updating the trust score based on new information and observations over time. The trust period starts when the TMM is initialised and begins evaluating the reliability of nodes. As time passes, the trust score is adjusted or updated to reflect the evolving trustworthiness of the nodes involved.

In this evaluation, the direct and indirect (recommendation-based) aspects of trust are examined individually. To achieve this, when assessing the direct trust metrics, it is assumed that no malicious activity is influencing the indirect trust. In other words, all recommendations received are considered trustworthy. Similarly, when evaluating indirect trust, it is assumed that all entities within the network are capable of accurately estimating and predicting the trustworthiness of others. This assumption allows the analysis to focus specifically on the impact of dishonest recommendations.

5.3 Assessment of Trust Metrics

Trust modelling for the IoUT has evolved beyond single-metric systems, where individual observations, such as packet delivery, were used to estimate the trustworthiness of UNs, as discussed in Section 3.4. Multi-metric TMM has demonstrated increased resilience against various attacks, offering enhanced accuracy in trust assessment. This enhancement introduces several key challenges, primarily in the selection and evaluation of metrics. First, it is necessary to explore how each metric, when analysed individually, contributes to the detection of misbehaviour associated with malicious activity. The effectiveness of a metric depends on how well it aligns with the expected patterns of malicious behaviour, which may vary across different metrics. This analysis provides insights into the role of each metric in identifying attacks. Second, estimating a comprehensive trust score based on a combination of diverse metrics introduces additional complexity. When one metric successfully identifies malicious behaviour, the degree to which this detection influences the overall trust score becomes an essential factor. This raises questions about the relationship between individual metric contributions and their collective impact on trust evaluation.

To investigate these two challenges, existing multi-metric TMM are first categorised into two primary approaches: 1) Composite metric trust assessment, where multiple metrics are used to form a holistic trust score, and 2) Standalone metric trust assessment, where individual metrics are evaluated independently before combining their results. This distinction helps clarify the advantages and limitations of different TMM strategies. In the sections that follow, each category is elaborated upon in detail.

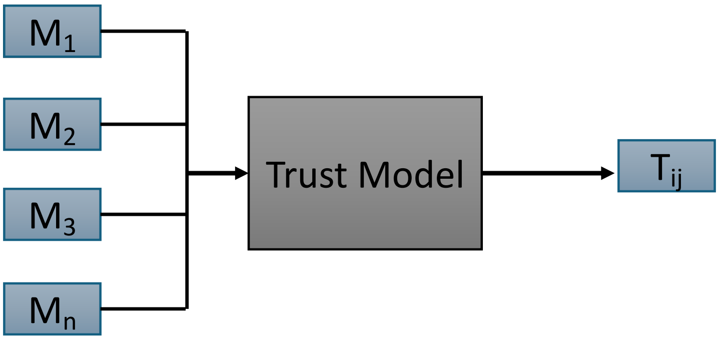


Fig. 5.2. Example of TMM architecture under composite metric trust assessment.

5.3.1 Composite Metric Trust Assessment

In this approach to trust establishment using multiple pieces of evidence, the trust system accepts metrics derived from observations and inputs them into a model that performs analytical processing or applies machine learning to estimate the trust score based on the collected data, as depicted in Figure 5.2. Here, each metric (M) represents corresponding evidence that could be derived from both the communication and physical behaviour of UN.

This architecture of driving the trust has been used on several MANET systems, with MTACMM being a proper example used on marine MANET as shown on Section 3.4.2.2. MTACMM (Bolster 2017) utilises Grey Theory for the trust establishment based on a list of metrics. This approach facilitates cohort-based normalisation, resulting in a Grey Relational Grade (GRG) that reflects the relative trustworthiness of nodes compared to others observed within the same timeframe. MTACMM expand the set of communications metrics like Delay (D), Packet Loss Ratio (PLR), Received Power (P_{RX}), Transmitted Power (P_{TX}), Throughput (S) to include physical metrics like Inter-Node Distance Deviation (INDD), Inter-Node Heading Deviation (INHD), and Node Speed (V). INDD assesses the average spatial separation of a node relative to its neighbours, whereas INHD evaluates the consistency of a node's orientation or heading in relation to others. Node Speed (V) provides information on the velocity of the node. In this section, the underlying computations of the model are examined based on the work of (Bolster 2017). The following equations, therefore, derived from their methodology, are presented here to facilitate an understanding of the model evaluation process with the established context of IoUT.

The Grey Relational Analysis (GRA) acts on the selected metrics to obtain Grey Relational Coefficient (GRC), which in this context can be considered a 'trust vector'.

The GRC is defined as:

$$\begin{aligned} \phi^t_{j,m} &= \frac{\min_j |\alpha^t_{j,m} - \mathrm{g}^t_m| + \rho \max_j |\alpha^t_{j,m} - \mathrm{g}^t_m|}{|\alpha^t_{j,m} - \mathrm{g}^t_m| + \rho \max_j |\alpha^t_{j,m} - \mathrm{g}^t_m|}. \\ \varphi^t_{j,m} &= \frac{\min_j |\alpha^t_{j,m} - \mathrm{b}^t_m| + \rho \max_j |\alpha^t_{j,m} - \mathrm{b}^t_m|}{|\alpha^t_{j,m} - \mathrm{b}^t_m| + \rho \max_j |\alpha^t_{j,m} - \mathrm{b}^t_m|}. \end{aligned} \tag{5.1}$$

Here, $\alpha^t_{j,m}$ represents the observed value of metric x_m for node j at time t, ρ is a distinguishing coefficient typically set to 0.5, and g and b are the 'good' and 'bad' reference sequences respectively, derived from the range of observed metric values, with $g_m = \max_j \alpha^t_{j,m}$ and $b_m = \min_j \alpha^t_{j,m}$.

To improve the detection and classification of misbehaviour, weighting is applied to GRC before generating a scalar trust value. The weighted GRC are calculated as follows (known as GRG):

$$[\phi^t_j, \varphi^t_j] = \left[\sum_{m=0}^{M} h_m \phi^t_{j,m}, \sum_{m=0}^{M} h_m \varphi^t_{j,m} \right], \tag{5.2}$$

where h_m denotes the weighting coefficient for each metric. The final weighted GRC ϕ and φ are then scaled to the range [0,1] using the transformation $y = 1.5x - 0.5$. This scaling process normalises the coefficients, making them compatible for further processing to form a single scalar trust value. The scalar trust value, $T_{ij(t)}$, is computed to minimise uncertainties associated with aligning to the best (g) or worst (b) sequences.

$$T_{ij} = \frac{1}{1 + \left(\frac{\phi_j^t}{\varphi_j^t}\right)^2}. \tag{5.3}$$

5.3.2 Standalone Metric Trust Assessment

In contrast to the previous method, the second approach broadens the concept of trust by calculating individual trust values for each metric, such as trust in packet delivery, trust in energy consumption behaviour and trust in channel delay behaviour, as illustrated in Figure 5.3. In this method, each metric is independently evaluated based on the evidence collected over a defined trust period. This means that each metric is assigned an internal trust score ranging from 0 to 1, representing its specific trust evaluation. The model then uses an aggregation technique to combine these individual trust scores into a single synthesised trust value. This approach enables a more explainable assessment of trust, where distinct aspects of node behaviour, such as energy efficiency or communication performance, are thoroughly evaluated in isolation before contributing to the overall trustworthiness assessment.

While various approaches have been introduced using this method, as discussed in Section 3.4, CATM (Jiang et al. 2022), a recent TMM designed for the IoUT, is utilised. It integrates three main processes: trust calculation, trust recommendation, and trust evaluation, which collectively represent overall trust based

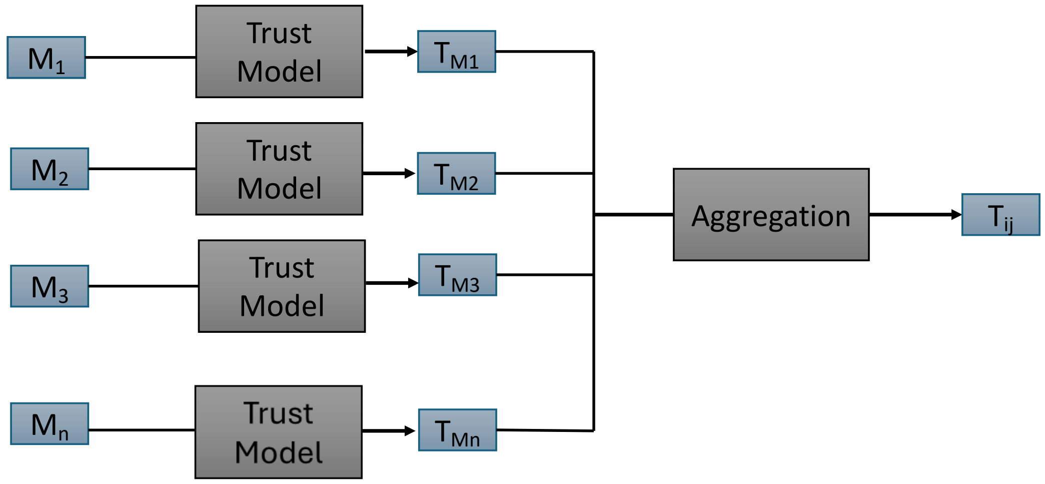


Fig. 5.3. Example of TMM architecture under standalone metric trust assessment.

on either direct trust (T^d) or recommendation trust (T^r) obtained from one-hop neighbours. Direct trust is evaluated using three key metrics: trust based on packet delivery ratio (T_c), trust based on transmission delay (T_d), and trust based on energy consumption (T_e). These dimensions of trust are combined to represent the overall trust that node n_i holds toward node n_j as:

$$T_{ij} = w_1 T_\mathrm{c} + w_2 T_\mathrm{d} + w_3 T_\mathrm{e}, \tag{5.4}$$

where, w_1,w_2, and w_3 represent the weight assign to each trust metric. The details on the model are thoroughly covered on Appendix D.

5.3.3 Simulation and Initial Discussion

Both methods of driving trust are examined using Aqua-Sim ng, under the variations of network topology presented in Section 4.3.1, and the mobility presented in Section 4.3.3. More specifically, a trust layer is developed to operate on top of the established network stack (for more details, see Figure C.1 in Appendix C). The working flow of the established evaluation of each model is as shown in Fig-

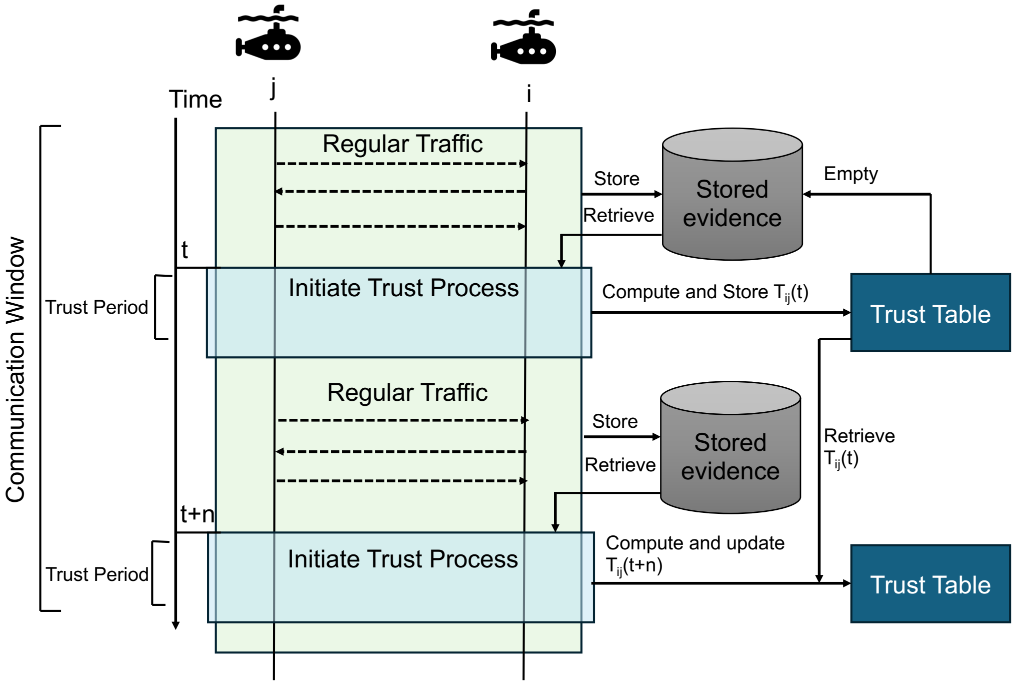


Fig. 5.4. The trust establishment architecture.

ure 5.4. Each UN within the network is activated at time t to initiate the trust evaluation process for its neighbouring nodes. Throughout prior interaction periods, evidence is collected and stored locally. During the trust evaluation phase, this evidence is retrieved to contribute to the trust assessment. The resulting trust score from this process is recorded in a local trust table, which serves as a reference for tracking and updating trust scores in future evaluations. To thoroughly investigate the impact of the variation in this parameter over time, a trust period interval of one minute is established.

The simulated results were obtained by averaging twenty independent random runs within the defined underwater topologies. The choice of 20 runs aligns with established practices in the field, as similar studies typically adopt a comparable number of iterations to ensure consistency and reliability in performance evaluation (Yang et al. 2022).

5.3.3.1 Introducing Attacks to the Simulation

In each test below, a designated trustee node is selected to evaluate the trust scores assigned by each neighbour to the selected node. Across different scenarios, the behaviour of the designated node is altered, and the evaluation by other nodes is observed over time. Each attack scenario presented on Section 4.4 was simulated and planned as follows:

- Denial of Service Attack (DoS): In this case, and at time 1000 seconds, the attacking node starts streaming packets at a high frequency to disturb the network. This represents a typical DoS attack where the goal is to overload the network with excessive data.
- Selfish Behaviour Attack (SB): In this experiment, a scenario was simulated where the attacking node stops cooperating after it has used up 1% of its energy resources. The purpose was to mimic the behaviour of a node that becomes non-cooperative in an effort to conserve its energy.
- Selective Forwarding Attack (SF): This attack was introduced 1000 seconds into the simulation. The attacking node was programmed to selectively allow and block packet forwarding to specific neighbours, effectively simulating the behaviour of a node that disrupts standard communication patterns for malicious purposes. In the subsequent evaluation, the drop frequency is established at a rate of 0.8, meaning that 80% of the incoming packets were dropped probabilistically, simulating a scenario where the attacker selectively forwards only 20% of the traffic to mimic normal behaviour while maximising disruption.
- Physical Mobility Attack (PM): Initiated at the same time as the SF Attack, 1000 seconds into the simulation, this scenario involved an attacking node

changing its physical location within the network following the techniques presented in Section 4.4.1.

5.3.3.2 Anticipating the Estimated Trust Score

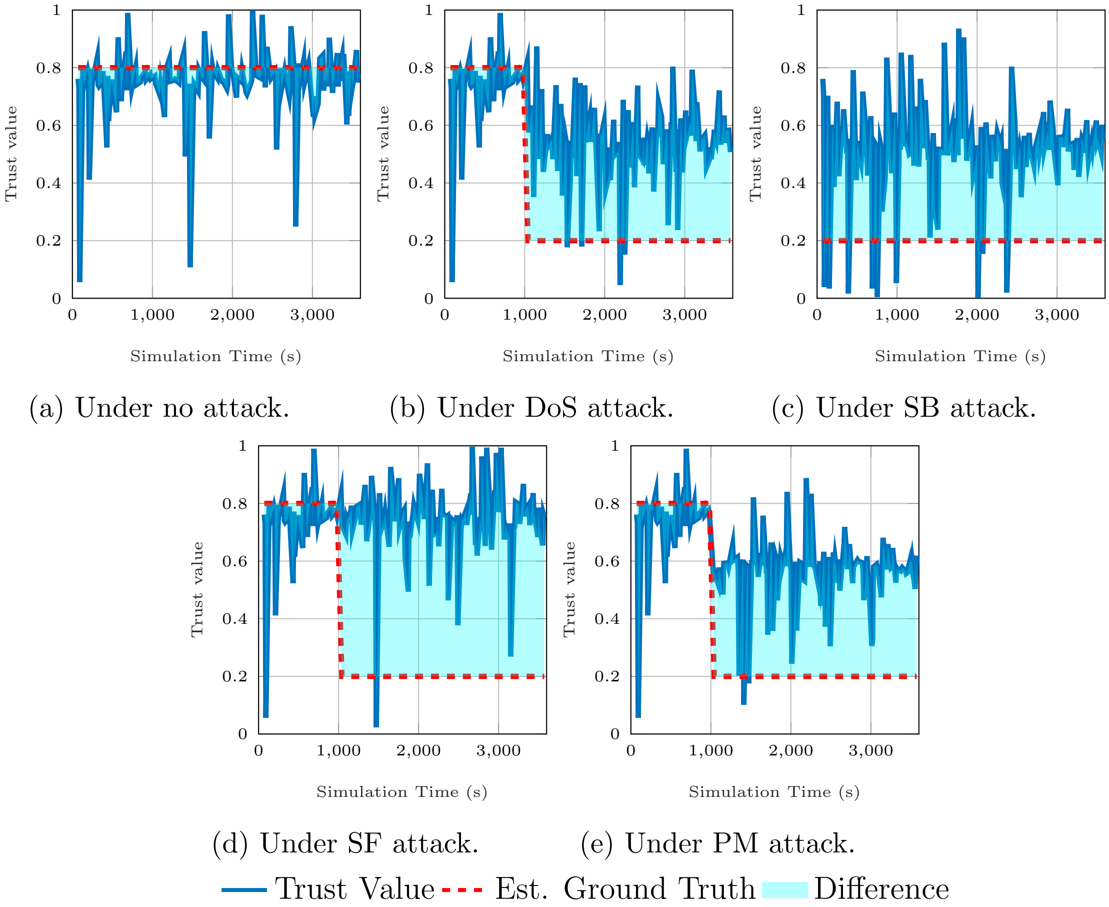
In the following analysis, an estimated ground truth of trust is established to facilitate comparison with the observed outcomes, assigning a trust value of 0.8 for trustworthy nodes and 0.2 for malicious ones. This method accentuates the theoretical basis of estimated trust, serving as boundaries for delineating what constitutes trustworthiness or untrustworthiness within the scope of trust modelling. For the purposes of this analysis of TMM, a more general interpretation of the trust values is assumed, where any value exceeding 0.5 is interpreted as an indication of trust, while values below 0.5 signify distrust. The interval between these thresholds captures the corresponding model's nuanced sensitivity to varying degrees of node behaviour. In practical terms, both a trust score of 0.9 and 0.6 signify some level of trust in an entity, with the former reflecting a higher degree of confidence in the entity's reliability. Similarly, trust values below 0.5 represent varying levels of distrust, demonstrating the model's capacity to quantify different degrees of misbehaviour or malicious intent. Therefore, the selection of 0.8 for trustworthy nodes and 0.2 for malicious nodes is not intended as an absolute or definitive measure of trust. Instead, these values represent the model's anticipated outcomes for trustworthiness and untrustworthiness, serving as reference points for assessing deviations observed during simulation. This allows for visualising the evaluated model's effectiveness in detecting attacks in correspondence to the estimated ground truth.

5.3.3.3 Examining the Performance of MTACMM

With the first approach, the visibility of the cohort-based evaluation is investigated where all the nodes under evaluation (i.e. the process of obtaining trust value), will be examined based on selective metrics. The performance of MTACMM is evaluated under the same topological and configuration sets previously established. A key consideration in this evaluation is the extent to which each metric, in its raw form, contributes to the overall trust assessment.

During the simulation, the model collects evidence based on the selected set of metrics, which is then utilised in the trust evaluation process based on Equation 5.1. Within this method, a decision needs to be made first as to what constitutes a good or bad indication of trust. For example, metrics such as throughput and PLR convey different implications for trust, and their importance must be appropriately defined. Typically, higher throughput is correlated with increased trustworthiness, whereas a lower PLR is considered favourable. As a result, the classification of behaviour as either 'good' (g) or 'bad' (b) is dependent on the performance of each corresponding metric. Specifically, a low PLR signifies good behaviour, while a high PLR suggests negative or untrustworthy behaviour. Similarly, high throughput indicates positive performance, whereas low throughput points to poor behaviour. Therefore, before establishing the computation of trust using GRA, the model needs to manually calibrate each metric to accurately reflect its impact on trust assessment, ensuring that the metrics are properly aligned with the overall trust evaluation process.

Initially, a baseline or 'fair' scenario is established, characterised by fair behaviour with no malicious entities. This scenario is examined across all of the defined topology (Figure 5.5a for stationary scenario, Figure 5.6a, for anchored nodes, Figure 5.7a, for floating nodes, and Figure 5.8a, for mobile nodes). In each of these graphs, the line represents the average trust obtained by all neighbouring UNs



(a) Under no attack.

(b) Under DoS attack.

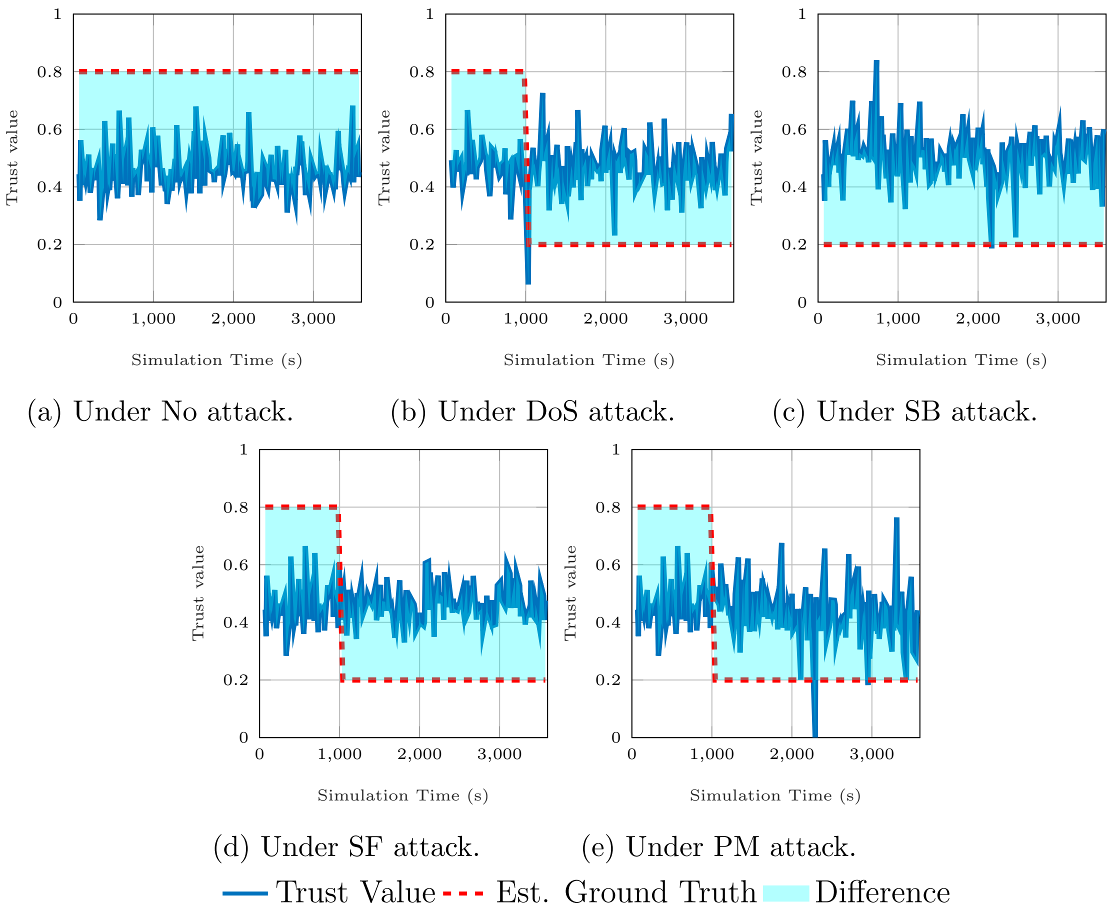
(c) Under SB attack.

(d) Under SF attack.

(e) Under PM attack.

Fig. 5.5. Convergence of the obtained trust to the estimated ground truth using MTACMM in stationary topology.

regarding the designated trustee node across trials. A significant fluctuation in the obtained trust over time is observed. These fluctuations, at some points in time, manifest the trustee node as an untrustworthy entity, even though the nodes being evaluated are intended to be benign. Another key observation for this scenario of no malicious act is the variation in trust across different network topologies. This is evident from the fluctuation of values below the estimated ground truth, where at certain points, the nodes are deemed untrustworthy (with trust values below 0.5). On average, stationary nodes, as illustrated on Figure 5.5a, exhibit higher trustworthiness compared to other topologies, as the obtained trust tends to cluster around the ground truth, albeit with occasional abrupt shifts over time.



(a) Under No attack.

(b) Under DoS attack.

(c) Under SB attack.

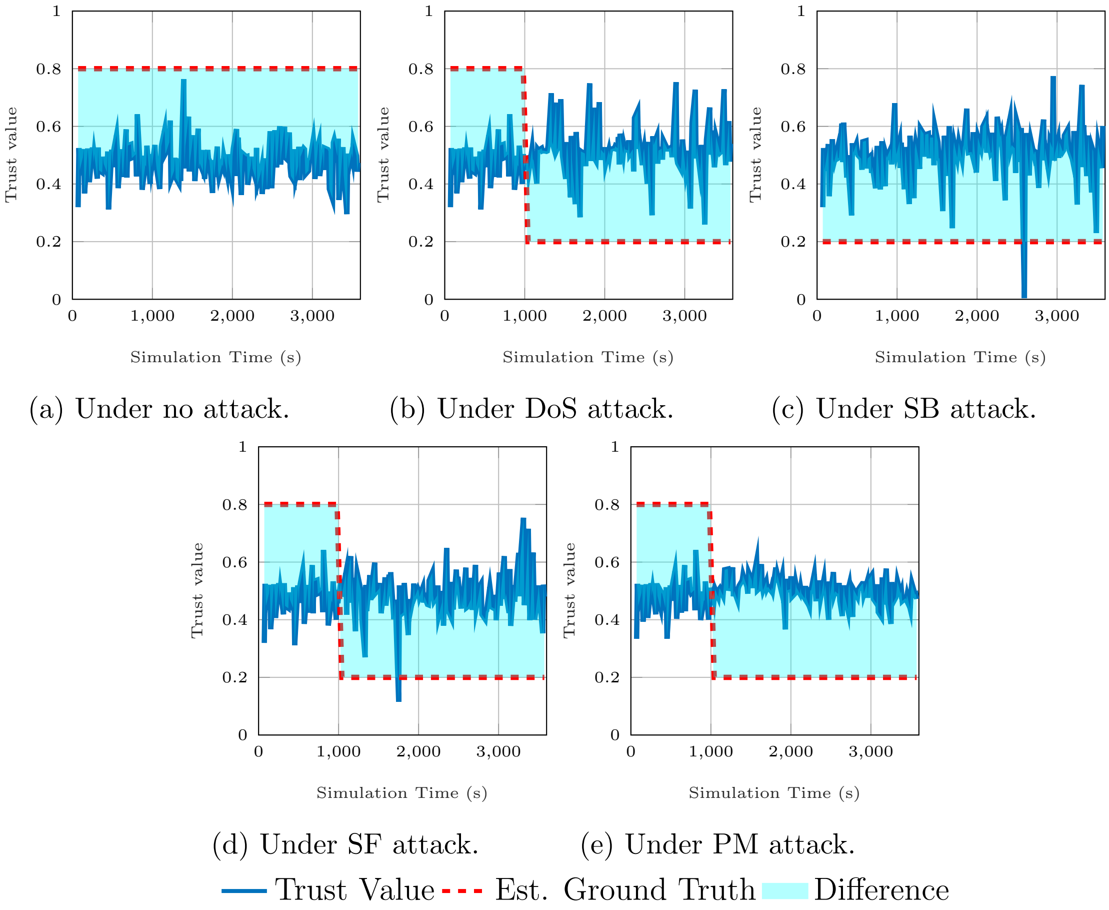
(d) Under SF attack.

(e) Under PM attack.

Fig. 5.6. Convergence of the obtained trust to the estimated ground truth using MTACMM in anchored topology.

The behaviour of the model in the fair scenario indicates that the metrics used to estimate node trust vary inconsistently, even when the nodes exhibit consistent good performance. One possible explanation for these unexpected observations could be the way the model was conducted, where at each point, the metrics used to evaluate the trust will be compared with a reference point in the GRA of the corresponding readings of the metrics, which may not align well with the inherent and continuous behaviour changes in underwater environments. This observation suggests that the model may be more tailored to specific application scenarios, highlighting a potential lack of generalizability across different network topologies.

Nevertheless, despite the inconsistencies in the findings under the fair scenario, the model's performance in the attack scenario is also examined. The introduction



(a) Under no attack.

(b) Under DoS attack.

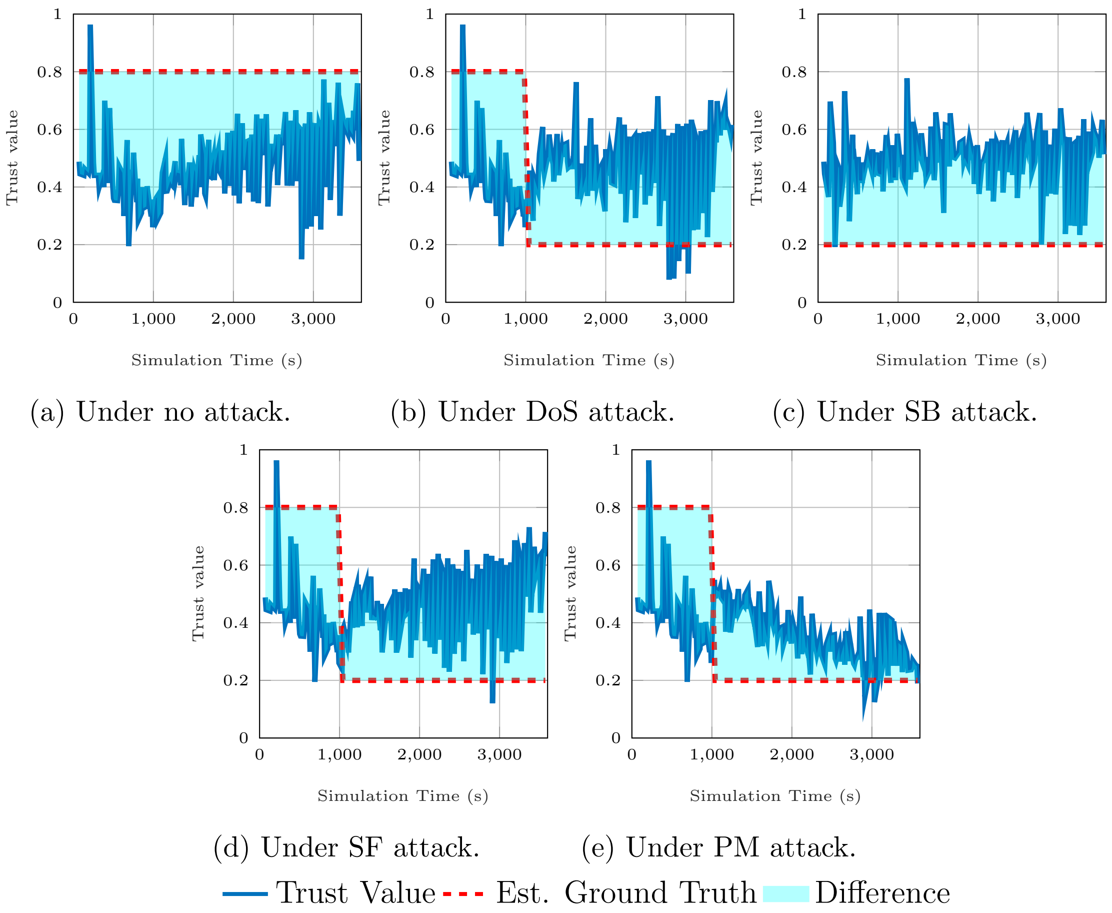
(c) Under SB attack.

(d) Under SF attack.

(e) Under PM attack.

Fig. 5.7. Convergence of the obtained trust to the estimated ground truth using MTACMM in floating topology.

of malicious attacks in the stationary nodes scenario, as depicted in Figure 5.5b to Figure 5.5e, reveals a modest shift in the model's behaviour. Specifically, during DoS attacks, the observed fluctuations in trust over time tend to be lower than the anticipated trust, which has the potential to obscure the model's sensitivity to the metrics employed in detecting such attacks. The decline in trust, relative to the fair scenario, demonstrates the model's ability to identify malicious behaviour, with the trust level of the compromised node occasionally falling below 0.5. However, despite this detection, the model sometimes exhibits temporary trust increases for the malicious node, reflecting periodic inconsistencies in its response. This pattern of fluctuation is similarly observed across other attack types. Although the model correctly identifies malicious activity at certain



(a) Under no attack.

(b) Under DoS attack.

(c) Under SB attack.

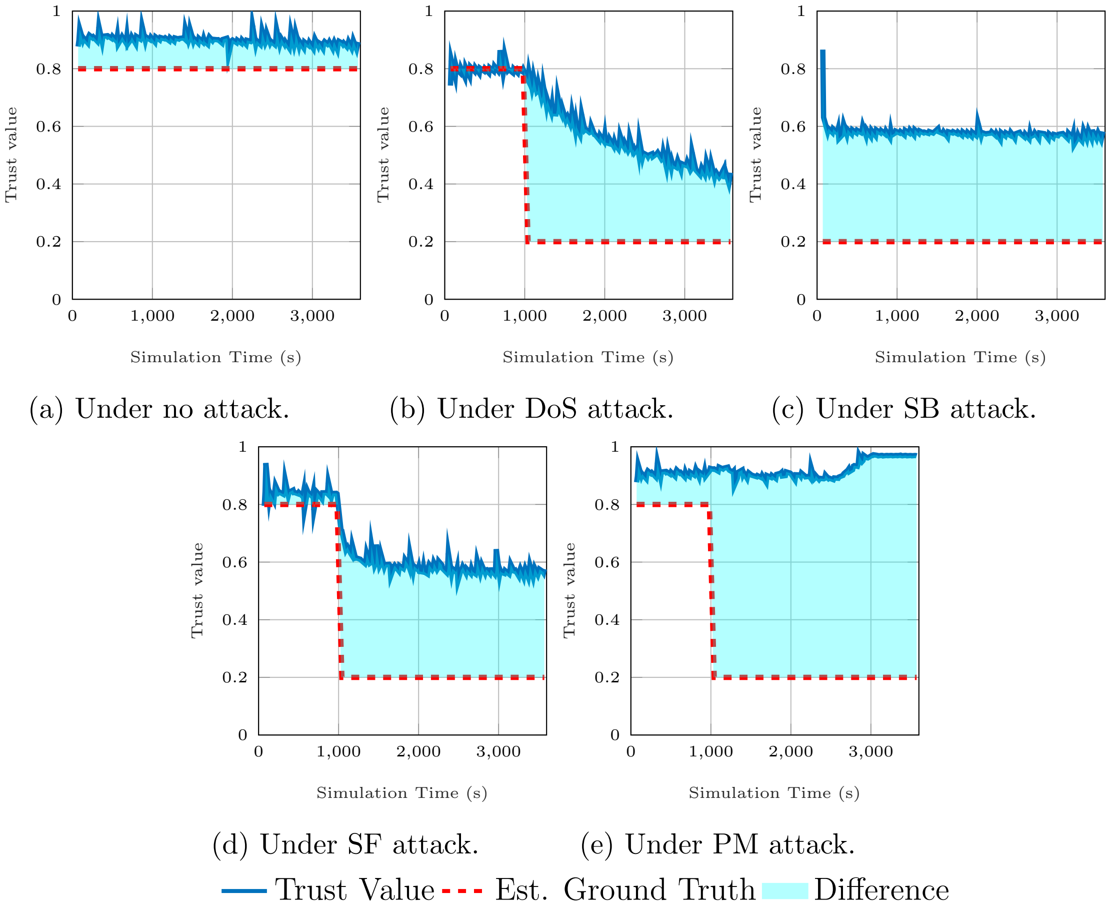
(d) Under SF attack.

(e) Under PM attack.

Fig. 5.8. Convergence of the obtained trust to the estimated ground truth using MTACMM in mobile topology.

points, it tends to revert to higher trust levels, even while the attack persists, with some attacks evading consistent detection. Figure 5.5e demonstrates quite timely detection of a PM attack, where a drop in trust was observed at the time of the attack (t = 1000s). However, the consistency of the detection changes over time. Nevertheless, the metrics used from the physical domain must account for this cause of detection.

The other network topology scenarios demonstrated in Figure 5.6, Figure 5.7, and Figure 5.8, demonstrate a similar pattern of fluctuations in observed trust over time suggest where the model's stability is frequently disrupted, likely due to variations in the value of collected evidence.



(a) Under no attack.

(b) Under DoS attack.

(c) Under SB attack.

(d) Under SF attack.

(e) Under PM attack.

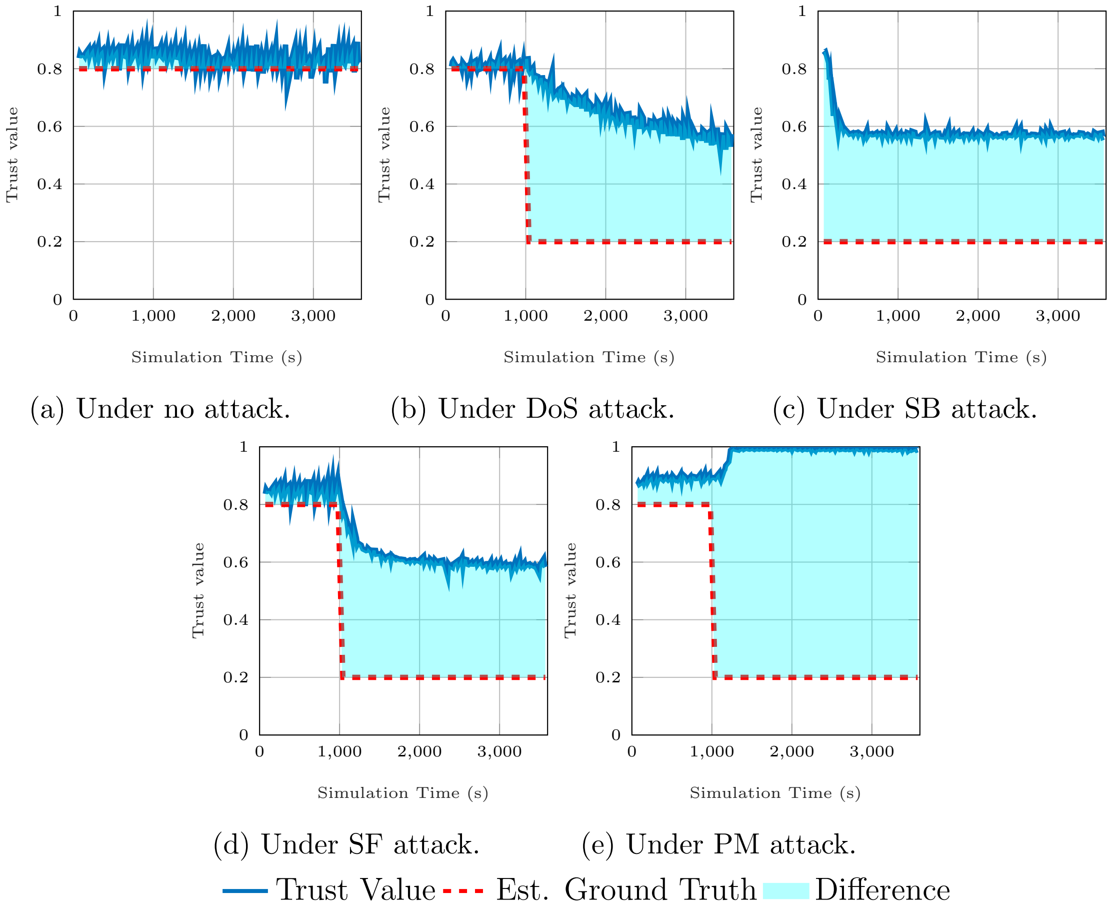
Fig. 5.9. Convergence of the obtained trust to the estimated ground truth using CATM in stationary topology.

5.3.3.4 Examining the Performance of CATM

Following the second approach, the process begins by examining CATM, a more recent and advanced TMM for the IoUT, to be utilised in the comparison. The results for examining CATM over time within each topology and attack are illustrated in Figure 5.9, Figure 5.10, Figure 5.11, Figure 5.12, representing the stationary, anchored, floating, and mobile topologies, respectively. In the fair scenario, the model's performance aligns closely with the expected outcomes. Compared to the previous approach, this trust evaluation model demonstrates reduced fluctuations over time. The introduction of consistent trust computation for each metric leads to greater convergence and reduced instability over time, ensuring a more reliable and stable trust assessment. However, upon introducing malicious attacks, a significant divergence is observed between the model's obtained results and the anticipated trust values for the attacks like DoS, SB, SF and certainly with PM. Although the model, by design, does not account for malicious actions influenced by the environment—unlike the previous model—it is nonetheless valuable to examine the model's behaviour under a PM attack. Specifically, a gradual and noticeable decline in trust towards malicious nodes is observed across all topologies under DoS, beginning at 1000 seconds. However, a persistent discrepancy between the expected trust values and the observed trust levels remains, albeit diminishing over time. During SB, the initial trust towards the malicious entity is high; however, upon the initiation of the attack—triggered by energy depletion—a decline in trust is observed. A similar trend is evident in the case of SF, where trust in the malicious node declines abruptly at the onset of the attack. However, in all these scenarios, the influence of the trust metrics on the overall trust assessment remains questionable, as evidenced by the discrepancies between the obtained overall trust and the anticipated values.

Interestingly, during PM, trust scores towards malicious nodes, at certain time points, exhibit an increasing trend. This phenomenon is primarily attributed to the nature of the PM attack, wherein the malicious node deceptively conceals its harmful intention and strategically positions itself in a manner that, while potentially degrading overall network performance, simultaneously fosters increased trust within its immediate 1-hop neighbourhood. This technique also exploits the fact that the CATM model relies on either direct or indirect trust in cases where insufficient evidence is available to compute direct trust. As a result, during certain phases of this attack, nodes propagate inflated trust scores about the malicious node, further complicating trust evaluation.

To elucidate the underlying factors contributing to the variations in detection across different attacks, the corresponding trust metrics are investigated under



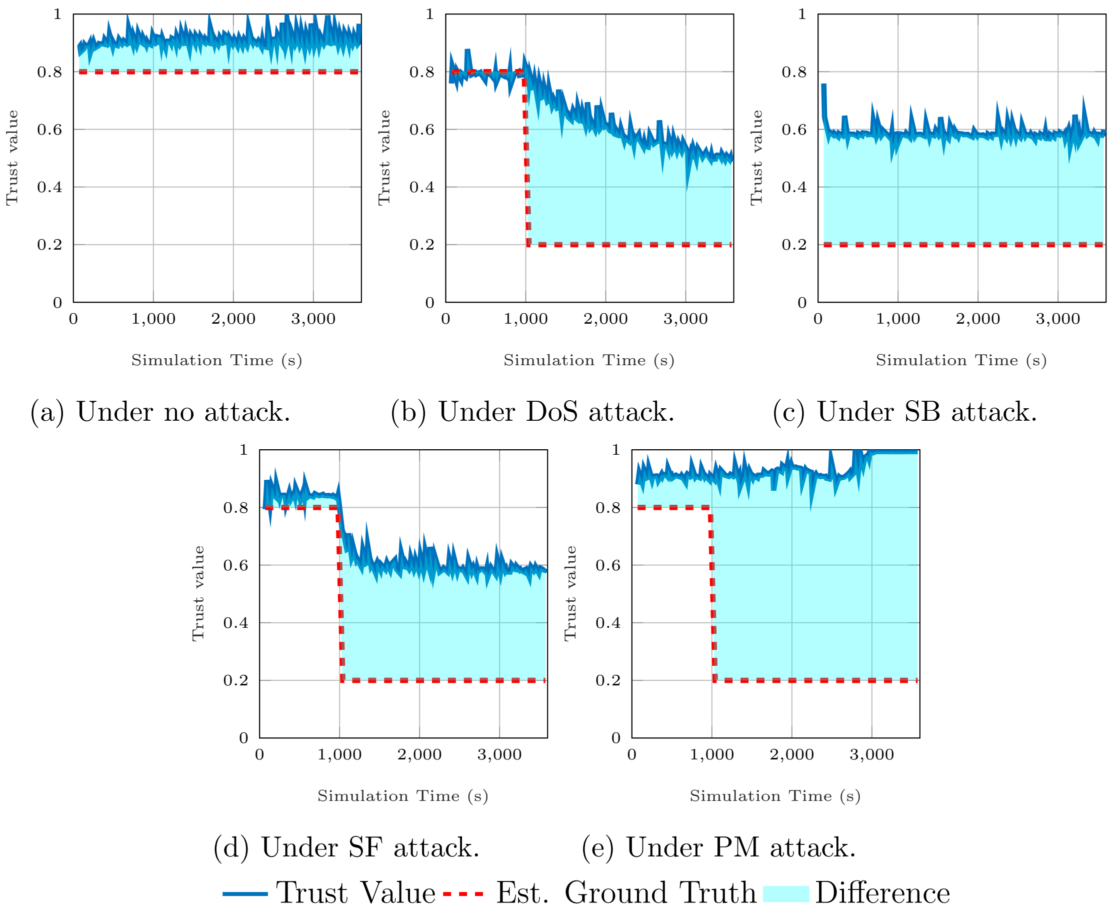
(a) Under no attack. (b) Under DoS attack. (c) Under SB attack.

(d) Under SF attack. (e) Under PM attack.

Fig. 5.10. Convergence of the obtained trust to the estimated ground truth using CATM in anchored topology.

each of the above scenarios. Figure 5.13 illustrates the behaviour of each proposed metric across different network topologies. The primary goal of these tests is twofold: (1) to determine whether these metrics can effectively identify malicious nodes by reducing the trust score of such nodes and (2) to assess the impact of each trust metric on the overall trust score.

The results presented reflect the mean values obtained from twenty trials for each individual metric, with the overall average of total trust T_{ij}, as utilised across all metrics, also indicated by Equation 5.4. The metrics used in this evaluation are derived from the corresponding trust equations: T_c, as shown in Equation D.2, represents the trust based on packet delivery ratio; T_d, defined in Equation D.4,



(a) Under no attack.

(b) Under DoS attack.

(c) Under SB attack.

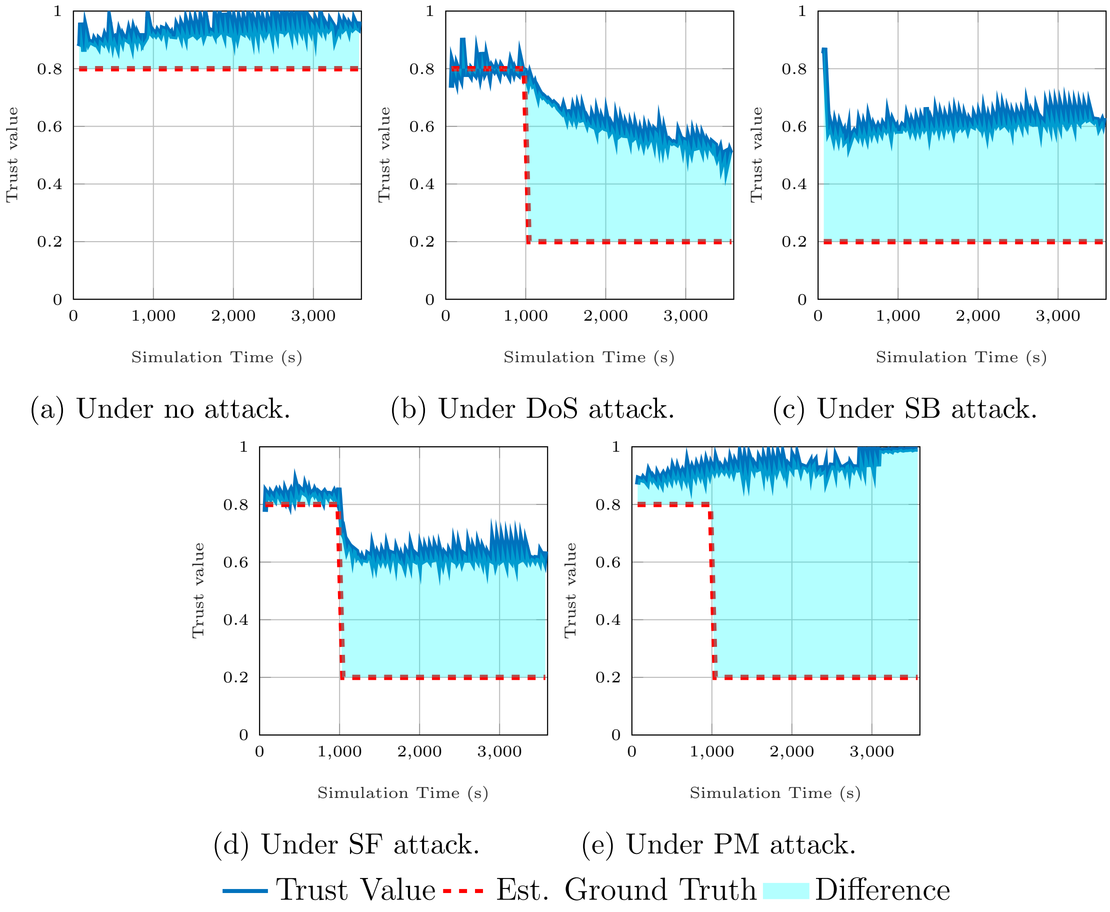
(d) Under SF attack.

(e) Under PM attack.

Fig. 5.11. Convergence of the obtained trust to the estimated ground truth using CATM in floating topology.

evaluates trust based on transmission delay; and T_e, as depicted in Equation D.5, reflects trust based on energy consumption;

In the fair scenario, as depicted in Figure 5.13a, slightly lower values are consistently recorded by both T_c and T_d, which is expected due to the inherent features of underwater networks where packet loss and channel congestion are prevalent. More specifically, in the beta-based TMM employed in Equation D.2 to define T_c, the ratio of successful communication to total communication stresses the impact of the underwater environment's noise and the challenge of maintaining a stable channel. Similarly, T_d accounts for the propagation speed of the channel, reflecting these conditions as well. The energy trust measured by T_e is the highest, indicating a fair and aligned energy consumption as modelled by Equation D.5.



(a) Under no attack.

(b) Under DoS attack.

(c) Under SB attack.

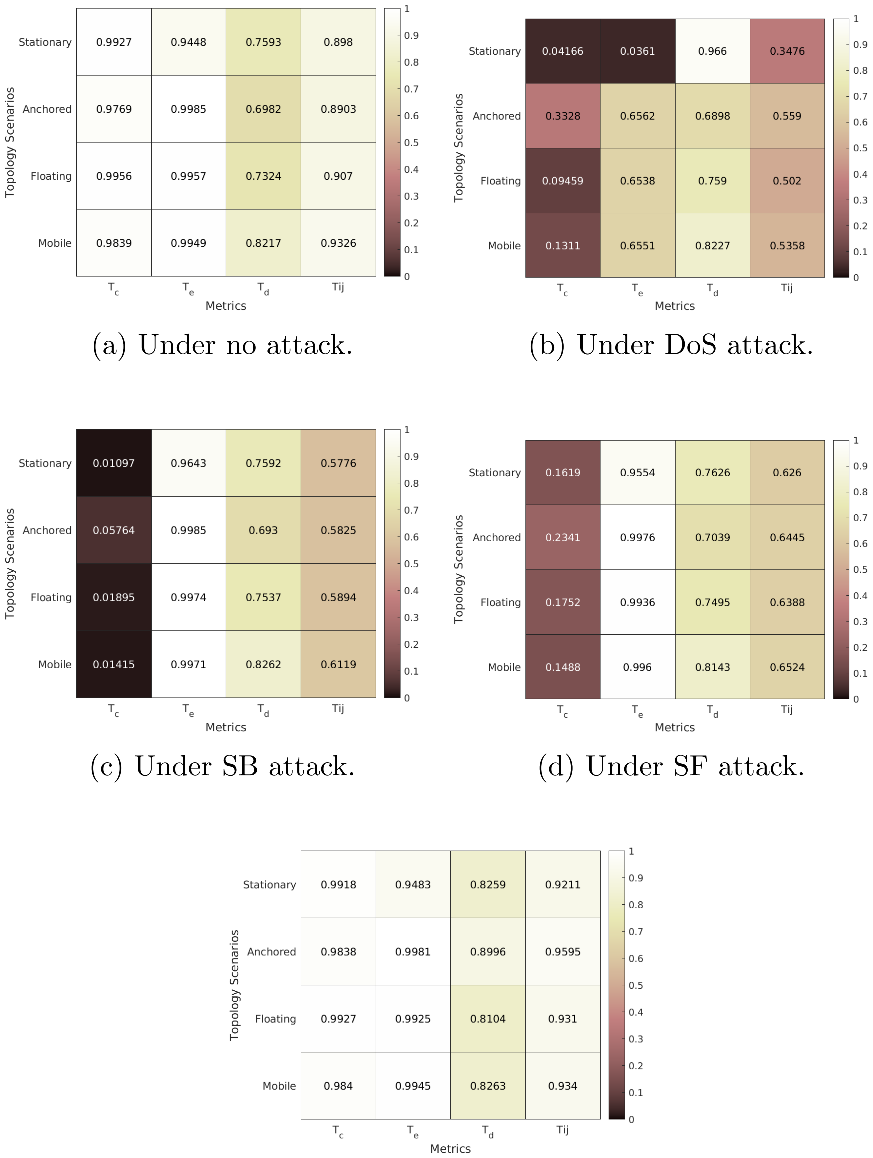
(d) Under SF attack.

(e) Under PM attack.

Fig. 5.12. Convergence of the obtained trust to the estimated ground truth using CATM in mobile topology.

In other words, the ratio of energy consumed to the initial energy is consistent. Another observation related to the network topology shows slight variations in the evaluated trust across different topologies, highlighting the necessity of examining each topology to understand how the corresponding TMM behaves.

Following the observation of the fair scenario, each scenario was evaluated against the previously established attacks. The results for the DoS, SB, SF, and PM attacks are depicted in Figure 5.13b, Figure 5.13c, Figure 5.13d, and Figure 5.13e, respectively. In this analysis, the fluctuations of each trust metric in response to these attacks were examined. The DoS attack, observed across all topologies, significantly impacts most trust metrics, with both T_c and T_e displaying high sensitivity to the attack, as evidenced by a decrease in the associated trust levels.



(a) Under no attack.

(b) Under DoS attack.

(c) Under SB attack.

(d) Under SF attack.

(e) Under PM attack.

Fig. 5.13. Metrics assessment using CATM.

Conversely, the variable T_d exhibits lower sensitivity to this attack due to the manner in which its corresponding value is derived, which is contingent upon the delay attributes of the received packets. Nevertheless, the overall trust values across the topologies, except for the stationary topology, were found to be above average (with 0.5 assumed as the baseline trust threshold), indicating distrust only in the stationary topology toward the nodes under observation.

In the case of the SB attack, the primary variable affected is T_c, which exhibits a significant decline due to the failure of packet delivery, a characteristic of the SB attack. It is noteworthy that T_e did not effectively address this attack, as the energy consumption by the node was lower than expected. The equation presented in Equation D.5 appears to be more sensitive to higher energy consumption, as seen in the DoS attack, rather than the lower consumption anticipated in a SB attack. Consequently, the overall trust during this attack remains above the average, resulting in the SB nodes being trusted across all trials conducted for this scenario. Similar trends were observed in the SF attacks, where the malicious nodes selectively forwarded packets to some nodes but not all. Discrepancies in evaluating T_c were noted across scenarios; however, the overall trust obtained was higher than in the SB attack. Finally, the evaluation was extended to consider the PM attack introduced in the attack model. The potential impact of such an attack on the trust metrics was investigated. The results were strikingly similar to those of the fair scenario, where no attack was present, indicating that the TMM failed to detect the PM attack.

5.3.4 Summary

The above sections focused on evaluating the performance of two methodologies of trust evaluation under various topological configurations and attack models, with the objective of determining their suitability for dynamic IoUT networks. The two models were tested and analysed through a series of predefined case studies. Although MTACMM demonstrated sensitivity to PM attack, specifically under stationary topology, its significant fluctuation in trust values revealed a topology-dependent behaviour, which makes it potentially unsuitable for handling the diverse range of IoUT topologies. This was illustrated by testing the model across different topologies and attack scenarios, where the performance of metrics varied significantly based on the specific topology/attack combination.

In contrast, CATM exhibited more stable measurements over time, with greater consistency across different scenarios. However, its sensitivity to attacks, particularly the PM attack, was often misrepresented, indicating that while it is less topology-specific, it may not effectively capture certain attack impacts in a dynamic IoUT environment. Moreover, within this approach, the influence of each trust dimension on the overall trust can obscure malicious activities, making them less detectable. Specifically, when malicious behaviour is not evident within individual dimensions, the positive representation of those dimensions can overshadow and mitigate the effects of the negative ones. Employing dynamic and adaptive methods to represent the influence of each dimension on the overall trust could greatly enhance the TMM ability to accurately reflect trust and improve its resilience to such malicious acts.

5.4 Assessment of Indirect Trust Methods

In this section, the indirect aspect of trust will be explored further. The validity of recommendations poses significant challenges, especially in the complex nature of underwater networks where communication can be intermittent and unreliable. The open nature of the IoUT, which lacks a comprehensive security infrastructure, increases the likelihood of attacks. One issue is that the network's sparsity renders most advanced recommendation systems ineffective in this context. This becomes particularly problematic when there is no timely, relevant evidence, and due to the decaying nature of the trust, there will be high differences between the stored trust score and the recommendations. Additionally, there is a significant risk of mistakenly identifying honest nodes as dishonest due to link quality issues, leading to nodes being misjudged within the network and assigned incorrect trust levels due to communication misrepresentations.

5.4.1 Recommendation-related Attacks

In the context of trust-based recommendation systems, certain malicious activities are aimed at manipulating trust metrics. The following list details the most prevalent types of attack:

- Bad-Mouthing Attack (BM): A malicious node undermines the reputation of another node by providing false negative feedback. This form of attack is particularly challenging to detect when the attacker has a history of providing accurate and unbiased recommendations.
- Ballot-Stuffing Attack (BS): In this attack, a malicious node artificially boosts the trust level of another node by giving fake positive feedback, setting the stage for more complex and collaborative attacks.
- Selfish Recommender Behaviour: This involves a node refusing to participate in the network's trust establishment by not responding to recommendation requests, thereby withholding necessary collaboration with peers.

For the remainder of this discussion, the term *dishonest recommendation* refers specifically to recommendations made by malicious entities within the network to facilitate either BM or BS attacks. The focus on these two types of attacks is deliberate, as they have a significant and direct impact on trust metrics, as well as a pronounced ability to manipulate the TMM, leading it to make erroneous decisions. Further details regarding the behaviour of these attacks are provided in the following section.

5.4.2 Challenges of Dishonest Recommenders

Dishonesty in recommendation is not always characterised by overtly negative behaviour or a generally low level of trustworthiness. Instead, a recommender,

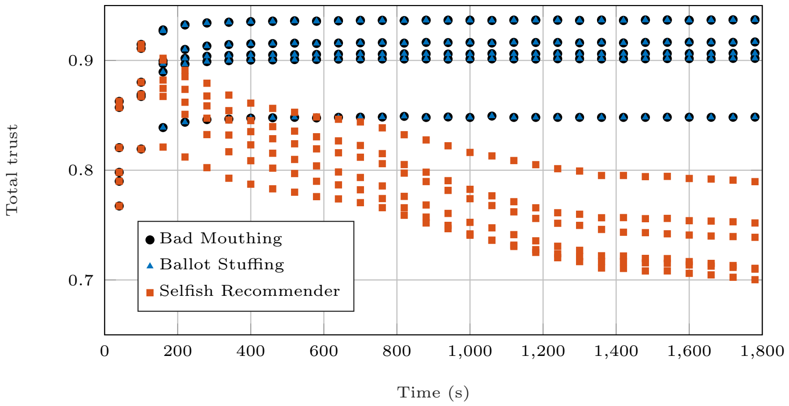


Fig. 5.14. Trust scores for dishonest recommenders over time obtained by CATM.

or node in this context, may exhibit seemingly normal behaviour in most interactions but engage in deceptive practices by deliberately providing false recommendations about others. This subtlety makes detecting dishonesty challenging because the overall behaviour of the node might not immediately arouse suspicion. For instance, the graph in Figure 5.14 illustrates the average trust score towards a dishonest recommender who continues to perform attacks as described in Section 5.4.1, following CATM. In cases of trust manipulation, such as BM and BS, the trust towards the recommender might naively overlook the misbehaviour, resulting in a high trust score over time for the dishonest recommender. On the other hand, a selfish recommender can be detected if the frequency of its recommendation requests and responses significantly deviates from typical communication patterns. If the trust evaluation of the recommender takes into account the frequency of communication, unusual behaviour can be identified. Specifically, a selfish recommender who often refuses to provide recommendations will exhibit a lower frequency of outgoing recommendations compared to the norm. This deviation can be quantified and monitored over time, leading to a decrease in the trust score of the selfish recommender as their non-cooperative behaviour becomes evident.

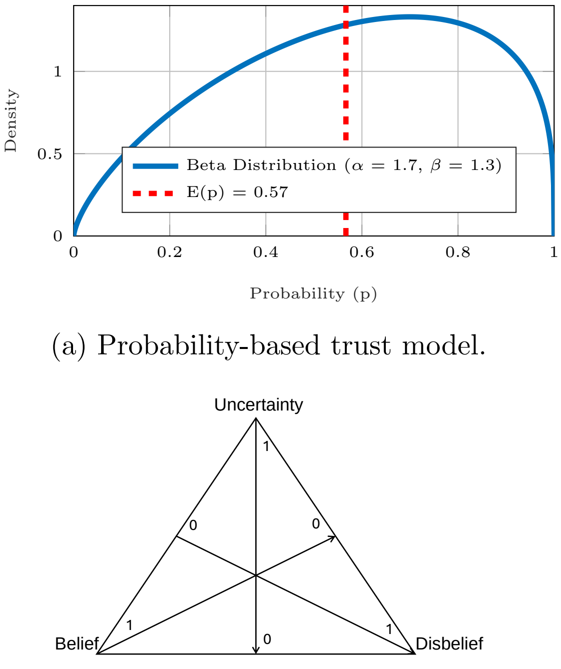
Existing methods of detecting dishonest recommendations are classified according to their aims to (1) Node-centric dishonesty detection and (2) Value-centric dishonesty detection. The former approach validates the received recommendations based on the characteristics of the recommender, while the latter focuses on assessing the values of the recommendations themselves. The following subsection provides a more detailed explanation of each method when applied to underwater networks.

5.4.3 Examining Node-Centric Dishonesty Detection

This method of validating recommendations is based on the social concept that individuals are less reluctant to accept recommendations from known and familiar sources. Nodes are more inclined to accept and act upon recommendations from a node in which they have confidence. In other words, this method leverages the level of confidence in the trust values assigned to different nodes based on the communication period with time. High confidence in a node typically suggests reliability, assuming that the source node consistently provides successful interactions. To explore the meaning behind this measurement in evaluating recommendations, analysis begins by examining how confidence in well-established TMM is quantified and, therefore, utilised to accept the recommendations.

5.4.3.1 Methods of Quantifying Confidence

Trust can be interpreted through various advanced theoretical frameworks. Among these, two widely recognised models have gained significant traction in the literature: the beta-based reputation model (Josang et al. 2002), derived from Bayesian inference, and a model grounded in logic theory (Jøsang 2001). These foundational paradigms offer powerful tools for assessing trust and reputation, enabling the quantification of trust while seamlessly accounting for uncertainty. Figure 5.15 illustrate the representation of each model measurement to construct the trust.



(a) Probability-based trust model.

(b) Belief-based trust model.

Fig. 5.15. Fundamental trust models with confidence measures.

Beta-based Model employs the beta probability distribution to assess the trustworthiness of entities within a system by analysing their observed actions, such as successful or failed packet deliveries. In this model, trust is quantified as a probability value ranging between 0 and 1, representing the likelihood of favourable outcomes based on historical interactions. The beta distribution is defined by two parameters, α and β, which correspond to the counts of positive and negative outcomes, respectively. Specifically, $\alpha = s + 1$ and $\beta = f + 1$, where s represents the number of successful communication attempts and f represents the number of failed attempts, provided that $s, f > 0$. The resulting trust value reflects the ratio of positive interactions relative to the total interactions. Figure 5.15a illustrates the beta distribution for various probabilities p, given fixed values for good (successful) and bad (failed) experience factors.

Drawing on the beta-based trust model, numerous studies across various domains have explored methods of validating recommenders using confidence measure. (Li et al. 2009; Shabut et al. 2014, 2018; Zouridaki et al. 2005). Within a beta distribution framework, confidence is deduced from the variance in historical interactions. For instance, if node i engages in both positive and negative interactions with node k at time t, the confidence is calculated by utilising the deviation from the mean, emphasising the stability and predictability of their interactions over time, such as:

$$c_{ik} = 1 - \sqrt{\frac{12\alpha_{ik}\beta_{ik}}{(\alpha_{ik} + \beta_{ik})^2(\alpha_{ik} + \beta_{ik} + 1)}}, \tag{5.5}$$

where α_{ik} represents the aggregated positive observations when a node forwards packets, and β_{ik} represents the aggregated negative observations when a node drops packets. One notable work, presented by Shabut et al. (2014), focuses on evaluating the honesty of recommending nodes based on three key factors: confidence derived from interactions, compatibility of information (assessed through deviation test), and closeness between nodes. In their model,the confidence value, denoted as V_{ik}^{conf} is derived as: $V_{ik}^{conf} = 1 - \sqrt{12\sigma_{ik}}$, where σ_{ik} is the beta distribution variance between i and k. This formula ensures that the confidence value lies within the interval [0, 1], where 0 indicates no previous interactions (hence no confidence), and 1 signifies complete confidence based on substantial positive and negative interactions.

Subjective Logic Model functions based on subjective perceptions of the world and employs opinions to signify these perceptions. The fundamental idea is to represent trust through belief, disbelief, and uncertainty (Gong et al. 2020). The trust value T is given by: $T = (b, d, u)$, in which b stands for belief, d is disbelief, and u accounts for the uncertainty related to the trustworthiness of a node, and $b + d + u = 1$. In the case of absence of belief or disbelief, a base rate (r) is defined as the prior probability. The relationship between these variables is shown in Figure 5.15b, where each vertex of the triangle represents a pure state of either complete belief, complete disbelief, or total uncertainty, with intermediate points indicating varying degrees of each.

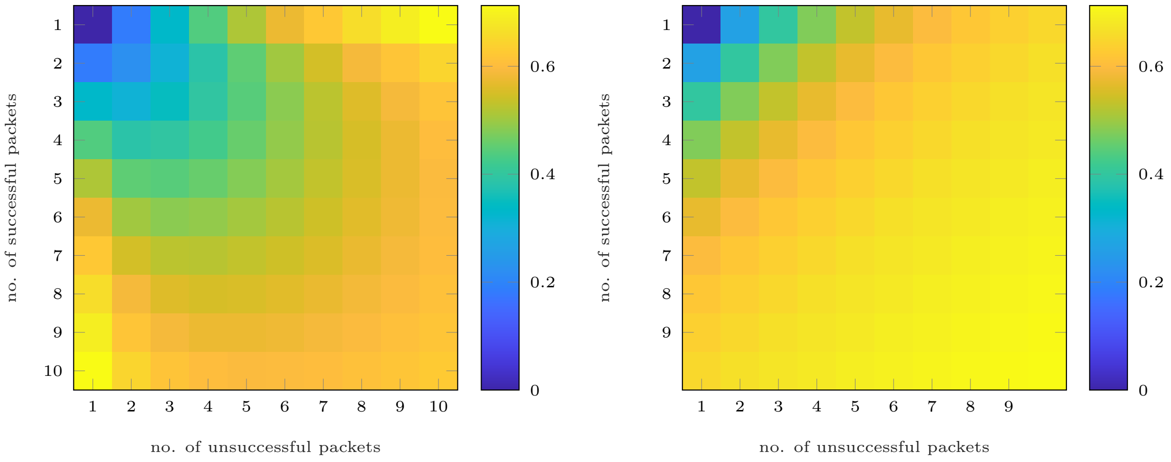
Similar to the beta model, several studies aimed to evaluate recommendations through the measurement of confidence. Within the subjective logic models, confidence can be derived from the uncertainty of the subjective logic. In both models proposed by Huang et al. (2017) and Mahmood et al. (2023), the weight for recommender $\alpha(v_i, v_j, t)$ would be considering the uncertainty of the subjective logic, while the term 'familiarity' is used in this work, its representation is equivalent to the confidence level using subjective logic. Therefore, 'familiarity' will be replaced with 'confidence' in their work to avoid confusion. In both works, they introduced the confidence in subjective logic to measure how much vehicle i (rater) is familiar with vehicle k (rate) as:

$$c_{ik} = 1 - u_{ik}, \tag{5.6}$$

where u_{ik} represent the uncertainty of the subjective logic. The higher the confidence value, the more prior knowledge the rater has about the rate. In other words, a recommendation received from a recommender to whom it has a high c_ik is deemed acceptable and considered valid.

5.4.3.2 Assessing Confidence Among Different Interactions

Theoretically, IoUT are expected to experience higher loss rates compared to other types of networks. To effectively utilise the confidence measure previously established within well-known trust models, it is essential to evaluate the behaviour of these models—specifically, Equation 5.5 in beta-based trust and Equation 5.6 subjective logic—by testing various combinations of positive and negative interactions before applying them to underwater networks. In particular, this section investigates how confidence is constructed when these models are used to evaluate recommenders, where recommendations from nodes with high confidence are typically accepted.



(a) Confidence values under Beta Distribution.

(b) Confidence value under Subjective Logic.

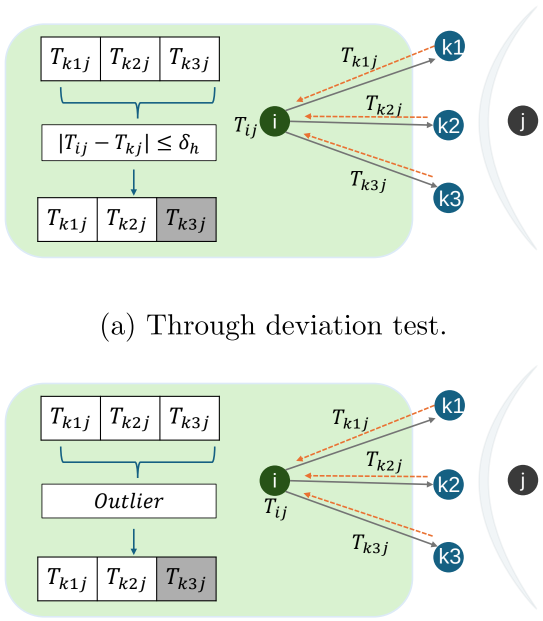
Fig. 5.16. Confidence values for varying interactions.

Both models were tested using a range of simulated interactions, varying from 1 to 10 positive and negative interactions. The results, illustrated in Figure 5.16, reveal how confidence evolves in each model and highlight key differences in their behaviour. Initially, when only one interaction occurs, the confidence level is zero, as entities are merely initiating communication and have not yet established trust. As the number of interactions increases—both successful and unsuccessful—confidence levels in both models also increase. Figure 5.16a demonstrates the confidence distribution in the beta-based trust model. The observed changes in colour from blue to yellow are not exclusively dependent on successful interactions but are also influenced by the accumulation of unsuccessful ones. This non-linear and counterintuitive behaviour suggests that confidence can increase even in the presence of negative interactions, where the peak confidence is observed when either positive or negative interactions increase in isolation. Such behaviour may pose significant challenges, particularly for IoUT, as it risks inaccurately growing confidence in nodes that are unreliable or even malicious. Figure 5.16b illustrates the confidence derived from subjective logic. This model exhibits a more uniform and predictable confidence distribution, with a consistent colour gradient from blue to yellow. Confidence increases directly with the number of interactions, demonstrating a steady and linear rise. Unlike the beta-based model, subjective logic peak confidence occurs when both positive and negative interactions increase simultaneously.

However, in both models, the accumulation of interactions—irrespective of whether they are positive or negative—results in an increase in confidence. This phenomenon raises concerns regarding the suitability of the proposed method of validating recommenders through confidence measures in IoUT, which are characterised by high packet error rates. In such environments, confidence may be artificially elevated for nodes involved in frequent negative interactions, which could be attributed to either unreliable nodes or malicious activity. As a result, applying these models to evaluate recommenders in underwater domains could lead to the acceptance of recommendations from entities that are not truly trustworthy. This analysis highlights a fundamental limitation in the application of the existing methods proposed in Section 5.4.3.1 within underwater networks. It underscores the need for alternative approaches or modifications to ensure that the recommenders are accurately validated in environments prone to high communication error rates.

5.4.4 Examining Value-Centric Dishonesty Detection

In the field of recommendation systems, one way to mitigate the impact of dishonest recommendations is to lean towards personal opinions and utilise the deviation from a personal experience-based approach (Khedim et al. 2015). This method compares recommendations against individual user experiences to identify and disregard those that deviate significantly. For instance, by applying a defined threshold, if the absolute differences between the recommendation and one's own experiences fall outside this threshold, it can be ignored. This approach can effectively isolate both unfairly negative and positive recommendations, offering a


T_{k1j} T_{k2j} T_{k3j}
$|T_{ij} - T_{kj}| \leq \delta_h$
T_{k1j} T_{k2j} T_{k3j}
T_{ij}
i
k1
k2
k3
j
T_{k1j}
T_{k2j}
T_{k3j}
Outlier

(a) Through deviation test.

(b) Through outlier test.

Fig. 5.17. Elimination of dishonest recommendation.

more personalised assessment. However, it faces challenges in situations with limited personal experience and is prone to the decay of trust over time, especially when there is a lack of direct interactions with certain nodes. Another method involves majority-rule and outlier detection to eliminate opinions considered outliers among others, with one example shown by Iltaf et al. (2013). While simple and effective in certain contexts, the majority rule-based method falls short in scenarios with collaborative attacks, as a dishonest majority can ultimately manipulate the TMM to carry out dishonest recommendation attacks without being detected.

Figure 5.17 shows the process of eliminating dishonest recommendations using the above-described methods. When node n_i assesses the trustworthiness of node n_j and requests recommendations from its neighbours, it begins by filtering out dishonest recommendations. According to Figure 5.17a, node i relies on its own evaluation of node n_j, using a validation threshold to identify and remove dishonest feedback. On the other hand, Figure 5.17b applies outlier detection techniques to identify and eliminate recommendations that significantly deviate from the majority, targeting dishonest recommenders considered outliers. The outlier test employs one of the existing outlier detection techniques to identify recommendations that are significantly different from the majority (Wang et al. 2019). A recommendation is flagged as an outlier based on a statistical measure that quantifies how much an observed trust score deviates from a central tendency (like the mean or median) of the dataset.

While employing each technique individually presents limitations, their combined efficacy is explored through a structured set of test cases. This section elaborates on the potential outcomes of utilising each method based on logical test cases designed to present their corresponding evaluations.

5.4.4.1 Experimental Setup

Simulation data is collected to evaluate various methods of detecting dishonest recommendations. An environment is set up where, at a given interval of time, a node n_i evaluates the trust score towards a trustee n_j and then generates a score T^d_{ij}, representing the direct trust node n_i has of node n_j. The indirect trust process is initiated by requesting recommendations from 1-hop neighbours about n_j. To focus on the validation of recommendations under dishonest recommendations, it is assumed that all nodes are cooperative in responding to the recommendation requests. Additionally, the number of nodes responding to the recommendation request is adjusted across three different sets. The configurations for these recommendation sets are established as follows: Set 1 (resp. 2, 3) corresponds to receiving 5 (resp. 10, 15) recommendations about n_j. These sets represent small, medium, and large neighbourhood sets, respectively. In each set, a proportion of malicious recommendations is included that engage in recommendation attacks presented in Section 5.4.1. The proportion of dishonest nodes is set at one-third of the total population. This proportion is chosen to represent a high percentage of malicious nodes, which allows the testing of the different methods in a challenging scenario.

To differentiate the methods used for detecting dishonest recommenders, each method is initially tested separately (in **Case 1** and **Case 2**). Subsequently, the performance of combinations of these methods is evaluated (in **Case 3** to **Case 6**). The testing scenarios are defined below:

1. **Case 1**: Applies the deviation test, D, to flag recommendations deviating significantly from direct trust as dishonest.
2. **Case 2**: Applies the outlier test, O, to detect recommendations that are statistical outliers.
3. **Case 3**: First applies the outlier test O, then uses the deviation test D on filtered recommendations to check for dishonesty, represented as $O \rightarrow D$.
4. **Case 4**: Applies the deviation test D, then analyses flagged recommendations with the outlier test O to confirm genuine outliers represented as $D \rightarrow O$.
5. **Case 5**: Combines outlier and deviation tests by intersection, $O \cap D$, considering only recommendations flagged by both methods as dishonest.
6. **Case 6**: Combines both tests by union, $O \cup D$, considering recommendations flagged by either methods as potentially dishonest.

Let $R = \{T^r_{k_1j}, T^r_{k_2j}, ..., T^r_{k_nj}\}$ be the set of recommendations received at a given time interval. Define the set of honest recommendations as $H \subseteq R$ and dishonest recommendations as $\bar{H} \subseteq R$ where $\bar{H} = R \setminus H$. The deviation test identifies recommendations based on the trust score T_{ij} computed by node n_i towards trustee

Table 5.1: Evaluation Metrics of Eliminating Dishonest Recommendations.

Category	O	**Set 1**				**Set 2**				**Set 3**			
		a	p	r	s	a	p	r	s	a	p	r	s
D	N/A	0.76	0.68	1.00	0.52	0.76	0.67	1.00	0.51	0.76	0.67	1.00	0.52
O	MAD	0.85	1.00	0.53	1.00	0.82	0.99	0.47	0.99	0.80	1.00	0.41	1.00
	LOF	0.92	0.80	0.98	0.88	0.84	0.99	0.50	0.99	0.81	1.00	0.43	1.00
$O \rightarrow D$	MAD	0.96	1.00	0.88	0.93	0.97	1.00	0.90	1.00	0.79	1.00	0.36	1.00
	LOF	0.96	1.00	0.89	1.00	0.96	1.00	0.89	1.00	0.85	1.00	0.54	1.00
$D \rightarrow O$	MAD	0.61	0.27	0.15	0.86	0.61	0.23	0.07	0.87	0.63	0.20	0.03	0.93
	LOF	0.66	0.50	0.04	0.97	0.61	0.23	0.07	0.88	0.63	0.16	0.02	0.92
$O \cap D$	MAD	0.83	1.00	0.50	1.00	1.00	1.00	0.90	1.00	0.80	1.00	0.40	1.00
	LOF	1.00	1.00	1.00	1.00	0.78	1.00	0.83	1.00	0.87	1.00	0.60	1.00
$O \cup D$	MAD	0.66	0.49	1.00	0.49	0.67	0.50	1.00	0.51	0.67	0.50	1.00	0.50
	LOF	0.68	0.51	1.00	0.52	0.69	0.51	1.00	0.52	0.68	0.51	1.00	0.52

n_j. A recommendation is flagged if the deviation of $|T_{ij} - T^{r}_{k_n j}|$, exceeds a defined threshold δ_h. In this analysis, $\delta_h = 0.05$ is set to align with the tested work for the ultimate threshold by Chen et al. (2012). Two methods of outlier detection are examined in this analysis: a simple method based on Median Absolute Deviation (MAD) and a more sophisticated yet computationally intensive method such as Local Outlier Factor (LOF). MAD identifies outliers by comparing each data point to the median, using the median's resistance to extreme values. A point is classified as an outlier if its deviation from the median, scaled by MAD, exceeds three times the median absolute deviation. LOF detects outliers by assessing a point's local density relative to its neighbours. Points with a significantly lower density than their surroundings receive a higher LOF value, indicating they are outliers.

The results of these test cases are evaluated over 100 trials using accuracy (a), precision (p), recall (r), and specificity (s) metrics. Each trial examines all cases across the three sets established earlier, varying the number of neighbours. This analysis seeks to offer a detailed understanding of the strengths and weaknesses of combining these tests in different sequences and structures.

5.4.4.2 Results and Discussion

In the evaluation of all the cases, see Table 5.1, each shows distinct performance characteristics in identifying dishonest recommendations within a network. In the analysis of different cases, Case 1, D, although distinguished by its high recall, which effectively identifies most dishonest recommendations, shows varying specificity and precision, resulting in an overall accuracy of 76%. Case 2, O, demonstrates balanced performance in smaller node sets, but its recall decreases as the node set size increases. Notably, within this case, the LOF method outperforms MAD, achieving an accuracy of 85.43% compared to 82.44%. Case 3, $O \rightarrow D$, exhibits improvement in both accuracy and precision as the number of neighbours increases, indicating that filtering followed by testing recommendations enhances overall reliability, with the LOF method reaching an accuracy of 92.5%. Conversely, Case 4, $D \rightarrow O$, underperforms relative to Case 3, suggesting that the sequence in which tests are applied plays a crucial role, with outlier filtering before direct dishonesty testing being more effective. Case 5, O $\cap D$, is notable for its high precision and specificity. Lastly, Case 6, $O \cup D$, is characterised by high recall, capturing a broader range of dishonest cases, but this comes at the expense of precision, leading to a higher likelihood of false positives, thus emphasising the trade-offs between maximising dishonesty detection and maintaining accuracy. Overall, it is evident that the combination of the two methods leads to improved performance, particularly in Case 3 and Case 5, with Case 3 demonstrating superior results compared to Case 5.

5.5 Conclusion

This chapter evaluated the performance of existing TMM for potential application in the IoUT, considering the unique characteristics and challenges posed by underwater network topologies. Specifically, it addressed the establishment of both direct and indirect trust under the inherent constraints of underwater environments. In the realm of direct trust establishment, existing mechanisms from the literature were classified into composite and standalone categories. Within each category, recent TMM developed for underwater networks were selected and their performance assessed under pre-defined network configurations. The conducted assessment, based on selected attack scenarios, revealed critical limitations in the ability of these models to effectively detect malicious behaviour by lowering the trust scores of compromised nodes. The findings emphasised the necessity of resilience to malicious attacks and consistent evaluation metrics to fulfil the requirements of a decentralised underwater TMM. This underscores the importance of aligning trust metrics with the specific nature of misbehaviours to ensure accurate detection. Furthermore, adaptive and resilient trust construction mechanisms are essential, particularly given the challenges in distinguishing between malicious actions and performance degradation caused by the temporal constraints of underwater networks.

The issue of dishonest recommendations, a significant challenge due to the sparse topology of underwater networks, was also explored. Solutions from similar terrestrial systems were examined and categorised into approaches focused on recommender nodes and those based on trust values. For the former, existing solutions were explored, and gaps in their applicability to underwater scenarios were identified. For the latter, which relies on trust values, the effectiveness of outlier detection methods and approaches favouring personal experience in identifying malicious recommendations was evaluated. The analysis demonstrated that a hybrid approach combining these methods showed improved accuracy in the test cases conducted.

Chapter 6

Mobility-aware Trust Model

6.1 Introduction

The previous chapter demonstrated through experimental evaluation that existing TMM, whether designed specifically for the underwater domain or based on fundamental theories, have yet to achieve the necessary clarity and accuracy under varying network topologies and underwater constraints. In particular, in Section 5.3, trust metrics and mechanisms proposed in recent TMM for underwater networks are evaluated against established attack domains in both physical and communication contexts. The assessment revealed their shortcomings in addressing threats that arise in both the physical and communication domains. This highlights the need for enhanced mechanisms to establish trust across various types of underwater networks.

Within the context of potential UNs misbehaviour, in Section 2.5.3, misbehaviour is categorised to be driven by selfish or malicious entities. Selfish node behaviour can manifest as a refusal to cooperate. This may include actions such as dropping received traffic without forwarding it or, in a more deceptive way, relocating to the network's periphery to minimise participation. Conversely, malicious nodes may strategically reposition themselves within the network for more intrusive objectives, such as intercepting, monitoring, and manipulating traffic at critical network points. In underwater networks, a highly capable adversary could exploit these dynamics by withdrawing from the network to offload data to rogue external sinks before reintegrating. Both forms of misbehaviour represent a notable departure from the expected or common behaviour of nodes within the network. Since trust is primarily established based on specific metrics, it becomes essential to broaden the scope of these metrics to effectively address such behaviours and account for anomalous activities. This raises the question of how each metric influences the overall trust evaluation process, with the objective of refining the TMM by prioritising and increasing the weight of metrics that are particularly effective in detecting and signalling potential misbehaviour.

This chapter explores the role of physical domain metrics in identifying such anomalies. The chapter introduces a model called Mobility-Aware Trust Model (MATMU), a distributed multi-metric TMM designed for IoUT. The model incorporates spatio-temporal mobility metrics to evaluate the trustworthiness of underwater nodes, thereby presenting a novel approach to trust assessment in underwater networks. From the physical domain, mobility-based trust is introduced, along with other essential trust metrics based on QoS, to evaluate UNs from several dimensions. After defining the metrics to capture both physical and communication-related threats, constructing trust based on multiple metrics becomes challenging, as not all metrics have the same impact on specific types of attacks. To overcome this issue, MATMU introduces a dynamic, real-time weighting mechanism to adjust the influence of each metric on the overall trust evaluation. The proposed MATMU is designed to address the key requirements previously identified in Section 3.5 outlines the requirements for an ideal TMM for IoUT. Specifically, the proposed TMM is designed to align with these predefined requirements, with a particular focus on being lightweight, decentralised, and achieving fast convergence. While not all requirements were explicitly tested, they are inherently addressed through the model's specifications and design considerations discussed throughout the chapter.

6.2 Proposed MATMU

Figure 6.1 illustrates the methodology employed within the proposed TMM, structured around two core components: direct and indirect trust processes. The direct trust process utilises direct interactions as primary evidence to assess trustworthiness, whereas the indirect trust process incorporates feedback from neighbouring nodes to evaluate a given node's reliability. The model is built on data received and stored during the operation of UNs. Data collection begins during the communication process, where information is gathered by analysing received packets and extracting evidence from sensory inputs, as well as from exchanged and forwarded control or data messages, and the relevant information obtained during communication and localisation phases. These pieces of evidence are then integrated into the model through three main categories: mobility-aware trust, which reflects evidence from the physical domain; communication-aware trust, which reflects evidence from communication activities and the underlying channel; and node-aware trust, which represents the capacity and capabilities of the nodes themselves. These categories of evidence are subsequently utilised to form a trust metric that is employed in the evaluation of direct trust.

During the trust evaluation process, nodes request evidence from their neighbours in the form of recommendation requests. This process can be carried out in different ways. For instance, a dedicated recommendation packet can be used for each particular node. However, to minimise network overhead, this work employs a single aggregated request packet containing information about all neighbouring nodes. Upon receiving recommendations, the model initiates a validation process to assess the integrity and reliability of the received trust information. Although dishonest recommendations pose a potential risk, at this stage, it is assumed that all recommendations are derived from trustworthy sources. While this assumption may lack realism—an issue further discussed in Chapter 7—it is

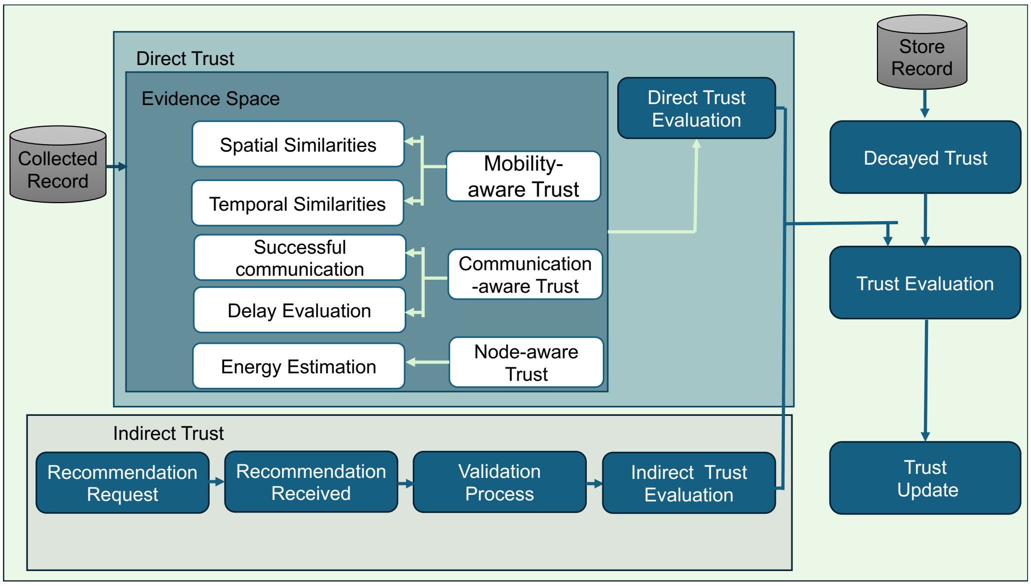


Fig. 6.1. Proposed MATMU model.

introduced to simplify the analysis by ensuring that all recommendations used to construct indirect trust originate from honest recommenders. This approach enables a more focused examination of the attacks detailed in Section 4.4, specifically those affecting the network's operational performance, without the added complexity of dishonest recommendations. Accordingly, the current approach to indirect trust is formulated based on the CATM method for recommendation validation. For more details about the recommendation validation process provided by CATM, see Appendix D and Equation D.11. Following the validation of recommendations, the indirect trust is computed based on the verified evidence.

The overall trust score is then formulated as a function of direct trust, indirect trust, and historical trust, ensuring that past interactions also contribute to the trust assessment. The model is designed to continuously update the estimated trust over time, enabling dynamic adaptation to changes in the network's behaviour and conditions.

6.3 Direct Trust Metrics

Direct trust is formulated mainly based on direct interactions and experiences, which are evaluated using specific metrics to determine the extent to which the behaviour of UNs is deemed trustworthy or untrustworthy. Fundamentally, trustworthy nodes are characterised by their ability to maintain 'consistent', 'reliable', and 'non-anomalous' behaviour over time (Aaqib et al. 2023). Concentrating on direct interactions, it is logical to rely on metrics within the communication domain to determine patterns and assess the quality of prior interactions. While theoretically correct, this approach is complicated by the dynamic and continuously moving nature of underwater environments, compounded by potential threats such as those discussed in Section 4.4.1. Thus, to account for these complexities, both communication and physical domains are examined in establishing direct trust among UNs. The subsequent sections provide a comprehensive overview of the utilised metrics.

6.3.1 Mobility Similarities and Differences Metric

There are various indicators that can be extracted from the physical domain and utilised in trust evaluation, such as local sensory data measurements, the number of relocations, signal strength variations, and node proximity changes. Among these, this work prioritises mobility as a primary indicator, as it closely aligns with the misbehaviour patterns examined in PM attacks. It is reasonable to infer that nodes demonstrating similar mobility patterns, influenced by regional water currents, may be considered consistent in their movement behaviour. Any abrupt deviation from these expected patterns could signal potential misbehaviour. Likewise, when a node within a collaborative network shifts its mobility path unexpectedly, it could signal potential misbehaviour. These behavioural anomalies can be identified through an analysis of similarities and deviations in the mobility patterns among nodes. The concept of detecting similarities in patterns has been applied in different contexts. For instance, McKenzie et al. (2021) utilises this concept in urban data science to identify similar regions based on the movement of people, while Duong et al. (2012) proposed its application in clustered MANET environments to determine similarities using mobility metrics.

Two aspects of similarity are introduced to account for both spatial and temporal changes over time. Spatial similarity is defined to measure the extent to which nodes within a defined area exhibit comparable movement patterns. Environmental factors, such as ocean currents, often lead to nodes moving in coordinated ways. Additionally, nodes with predetermined mission paths will display aligned movements. Detecting spatial similarities facilitates the identification of nodes that share behavioural patterns, allowing trustworthiness assessments based on peer behaviour. Temporal mobility is also defined to account for temporal change, indicating behavioural consistency across sequential time periods. This study defines the temporal metric to measure shifts in a node's movement over time. The detailed definitions of these metrics are outlined below.

6.3.1.1 Spatial Similarity

At time t, the spatial similarity between two nodes n_i and n_j can be evaluated based on their velocities. Let $\vec{v_i}(t) = (v_{ix}(t), v_{iy}(t), v_{iz}(t))$ and $\vec{v_j}(t) = (v_{jx}(t), v_{jy}(t), v_{jz}(t))$ represent the velocities of n_i and n_j, respectively. The euclidean distance between n_i located at (x_i, y_i, z_i) and n_j at (x_j, y_j, z_j) is defined as $d_{ij(t)} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$. If $d_{ij(t)}$ between n_i and n_j is less than R_t, where R_t denotes the transmission range, the spatial similarity can be computed as follows:

$$M_{\text{ss}} = \frac{1}{2}\left(\frac{\vec{v_i}(t)\cdot\vec{v_j}(t)}{||\vec{v_i}(t)||||\vec{v_j}(t)||}\frac{\min(||\vec{v_i}(t)||,||\vec{v_j}(t)||)}{\max(||\vec{v_i}(t)||,||\vec{v_j}(t)||)}+1\right). \tag{6.1}$$

Here the symbol $||.||$ represents the norm of the vector as $\sqrt{(v_x(t)^2+v_y(t)^2+v_z(t)^2)}$. The spatial similarity between two nodes at time t is calculated here through a combination of cosine similarity, which assesses the directional alignment between two velocity vectors, and magnitude ratio similarity, which addresses variations in speed. The cosine similarity value ranges from [-1, 1], where a score of 1 indicates perfectly aligned movement, 0 signifies perpendicular movement, and -1 represents opposite directions. To account for differences in speed, the magnitude ratio similarity is incorporated, ensuring that significant speed disparities reduce the overall similarity. Because trust is represented in a continuous scale from 0 to 1, the above equation is transformed and normalised to allow the results similarity score within the range [0, 1], where a score of 1 indicates identical motion in both direction and magnitude, while 0 denotes either completely opposing movement or extreme differences in speed.

6.3.1.2 Temporal Mobility Change

At times t and t', node n_j with velocities equivalent to $\vec{v_j}(t)$ and $\vec{v_j}(t')$ respectively, can address the temporal change in velocity of n_j over two consecutive periods of time as follows:

$$M_{\text{ts}} = \frac{1}{2}\left(\frac{\vec{v_j}(t)\cdot\vec{v_j}(t')}{||\vec{v_j}(t)||||\vec{v_j}(t')||}\frac{\min(||\vec{v_j}(t)||,||\vec{v_j}(t')||)}{\max(||\vec{v_j}(t)||,||\vec{v_j}(t')||)}+1\right). \tag{6.2}$$

The above equation follows the same principle of measuring similarity between two nodes in the previous definition of spatial similarity, but instead examines the consistency of a node's behaviour across different time intervals.

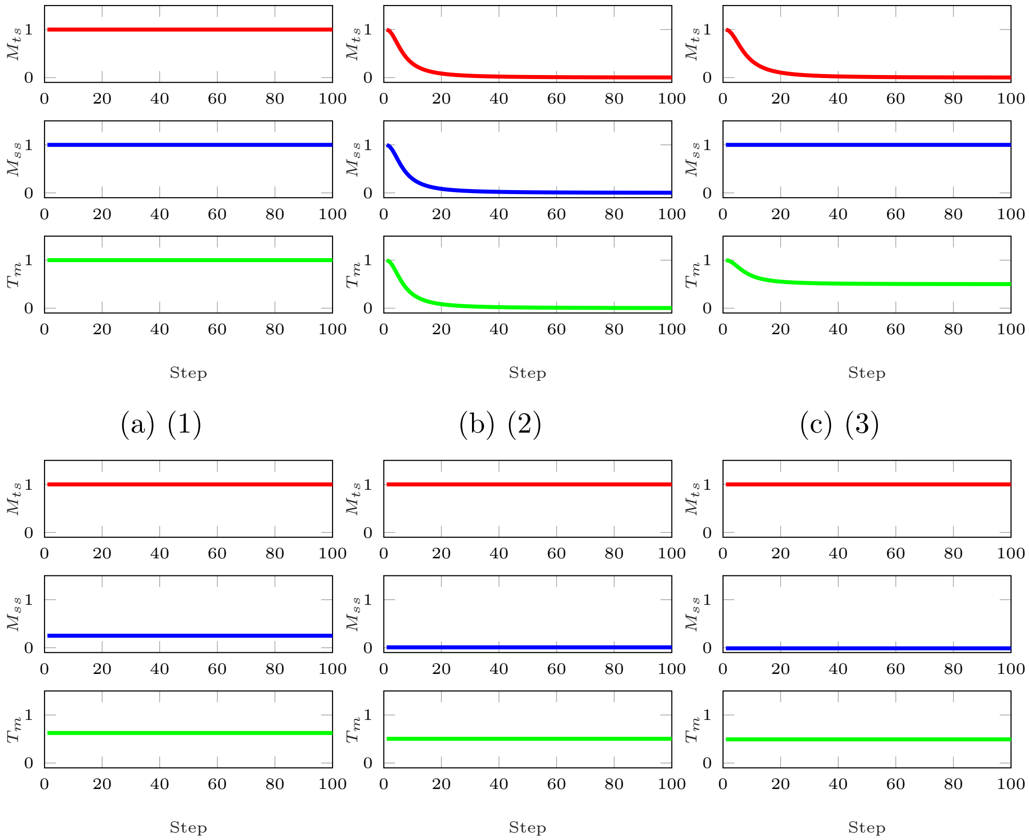
6.3.1.3 Trust based on Mobility

The trust formulation based on mobility will then be the weighted sum of both spatial and temporal similarity metrics, determining these variables' importance. Thus, the trust based on node mobility can be approximated as follows:

$$T_{\mathrm{m}} = wM_{ss} + (1-w)M_{ts}, \tag{6.3}$$

where w in the range between 0 to 1 to scale the importance of each factor.

In order to validate the obtained mobility trust, M_{ss}, M_{ts}, and the resulting T_m between two nodes n_i with velocity vector $\vec{v_i}(t)$ and n_j with velocity vector $\vec{v_j}(t)$ were examined over a series of movement steps. The initial velocities of the objects were set to be the same, where both $\vec{v_i}(t)$ and $\vec{v_j}(t)$ have identical velocities. Changes in velocity were introduced through controlled modifications, such as acceleration increments, directional shifts, or speed variations, to simulate different motion scenarios and evaluate their impact on spatial and temporal similarities as well as the mobility metric. The scenarios presented in Figure 6.2 were analysed, and M_{ss}, M_{ts}, and the resulting T_m were demonstrated for each case. The resulting values were consistently maintained within the range of 0 to 1, aligning with the perception of trust on a scale from 0 to 1. Notably, identical similarities ($T_m \simeq 1$) were only achieved when both nodes exhibited consistent drift over time, as shown in Figure 6.2a. The behaviour when n_j accelerated, as shown in Figure 6.2b, was examined, revealing changes in similarities and a subsequent reduction in the obtained T_m. Figure 6.2c illustrates a scenario where node n_j oscillated its orientation while maintaining speed. This change



(a) (1)

(b) (2)

(c) (3)

(d) (4)

(e) (5)

(f) (6)

Fig. 6.2. Examination of mobility similarities on scenarios as (1) both nodes maintain constant velocities. (2) the observed node accelerates over time. (3) the observed node changes direction over time. (4) both nodes maintain similar directions but with different speeds. (5) both nodes maintain similar speeds but in opposite directions. (6) one node remains stationary while the other is moving.

was mainly captured by M_{ts}. Additionally, the scenario in which both n_i and n_j exhibited changes in velocity was examined, as depicted in Figure 6.2d and Figure 6.2e, representing cases where both nodes moved in the same direction but at different speeds and at the same speed but in opposite directions, respectively. Both cases demonstrated a decline in maintained T_m. Lastly, the scenario where one node remained stationary while the other moved, as seen in Figure 6.2f, was assessed. A notable observation is that behavioural changes cause T_m to adjust accordingly. The extent to which T_m is influenced by both M_{ss} and M_{ts} depends on the relative impact of each similarity vector, which can be set according to specific application requirements by adjusting w. During this evaluation, equal weight is used where $w = 0.5$.

6.3.2 Communication-related Metric

In this approach, the percentage of successful communication and the transmission delay are considered the primary indicators of communication trust as follows.

6.3.2.1 Trust based on Successful Communication

A successful attack scenario within a communication network can lead to packet loss, making the ratio of successful to unsuccessful transmissions a commonly used metric for indicating misbehaviour. Focusing on a single aspect of observation, i.e., successful versus unsuccessful communication, allows the adoption of methodologies from other trust evaluation systems that rely on a single metric. In this work, the well-established subjective logic theory, previously explained in detail in Section 5.4.3, was employed. Specifically, trust based on communication, denoted as T_c, represents the expected behaviour of node n_i towards node n_j. This expectation is evaluated using the parameters of belief (b) and uncertainty (u), expressed as:

$$T_c = b + ur. \tag{6.4}$$

Here, r represents a pre-defined constant that signifies the willingness of n_i to believe n_j. In essence, knowing the successful communication between n_i and n_j in the interval between t_a and t_b as s_{ij} and the unsuccessful ones as f_{ij}, during the evaluation period, the components of the subjective logic notions can be estimated as:

$$\begin{aligned} b_{ij} &= \frac{\sum_{t_a}^{t_b} s_{ij}}{\sum_{t_a}^{t_b} s_{ij} + \sum_{t_a}^{t_b} f_{ij} + 2}, \\ u_{ij} &= \frac{2}{\sum_{t_a}^{t_b} s_{ij} + \sum_{t_a}^{t_b} f_{ij} + 2}, \end{aligned} \tag{6.5}$$

where b_{ij} represents the belief, u_{ij} is the corresponding uncertainty of the obtained b_{ij}.

6.3.2.2 Trust based on Transmission Delay

Despite successful communication, the inherent characteristics of underwater environments make it challenging to distinguish between malicious and non-malicious behaviour effectively. This is primarily due to channel-related issues that can often result in communication drops caused by natural factors rather than malicious intent. To address these challenges, the proposed TMM incorporates channel influences by using delay as an additional metric. Transmission delay is an effective measure for evaluating the trustworthiness of a connection, as shorter delays indicate a more dependable and reliable connection. However, in the context of underwater communication, propagation delays can be both significant and variable. This variability necessitates the consideration of environmental impacts when integrating the delay metric into the TMM. Consequently, the proposed TMM accounts for environmental factors, including depth, temperature, and salinity, which influence the delay metric. The trust based on delay can be calculated following CATM as:

$$T_{\mathrm{td}} = \left(\frac{1}{p} \sum_{i=1}^{p} \frac{d_{ij}/c}{\mathrm{Ed}_{ij}} \right). \tag{6.6}$$

The d_{ij} is the corresponding distance between the sender node n_i, and the receiver node n_j, Ed_{ij} is the expected delay in time to deliver packets (p) successfully. The underwater propagation speed c is estimated following Equation 4.2.

6.3.3 Energy Consumption Metric

Given that UNs start with a fixed energy level and consume energy during operation, their trust level, based on energy, follows an inverse relationship. If a UN consumes more energy, its trust score decreases. This relationship is closely tied to the reliability of the UN; a node with less energy is less reliable for consistent functionality (Kesari Mary et al. 2022). Malicious nodes are expected to consume either significantly higher or lower energy than surrounding nodes, making energy depletion a useful metric for detecting malicious behaviour. In the proposed TMM, energy consumption is included as an indicator for estimating node trust. Let RE_i be the reported remaining energy of node n_i. For each interval, the remaining energy values are reported as RE_i^0, RE_i^1, RE_i^2, ..., RE_i^m for intervals $0, 1, 2, ..., m$. Let EC_i be the energy consumption of n_i estimated based on the reported remaining energy. The energy consumption can be defined as:

$$\begin{aligned} EC_i^0 &= 0, \text{ for } m = 0. \\ EC_i^1 &= RE_i^0 - RE_i^1, \text{ for } m = 1. \\ EC_i^m &= \vartheta(RE_i^{m-1} - RE_i^m), \text{ for } m > 1. \end{aligned} \tag{6.7}$$

Here, ϑ represents a factor that adjusts the difference between consecutive reported remaining energy values RE_i^{m-1} and RE_i^m to reflect the energy consumption dynamics more accurately.

The trust based on energy consumption can then be examined as follows:

$$T_e = 1 - \frac{EC_i^m}{E_i}, \tag{6.8}$$

where E_i is the initial residual energy of the node, and $E_i > 0$.

6.4 Direct Trust Evaluation

The trustor node calculates multiple trust dimensions—namely T_m, T_c, T_e, and T_{td}—based on evidence collected prior to the trust evaluation period. Each trust dimension reflects the trustee's behaviour during the communication period, producing a value within the range [0,1]. These values indicate the degree of trust in each specific dimension, with values closer to one signifying high trust and values closer to zero indicating low or no trust. The following subsections detail the approach employed to derive a single trust value from the four established trust dimensions.

6.4.1 Direct Trust Aggregation

Direct trust evaluation represents an aggregation of each individual dimension to yield a singular, comprehensive trust score. Although various methods for aggregation exist, as discussed in Section 3.3, this study employs a weighted sum approach due to its simplicity and straightforward nature as a fusion function. The direct trust, denoted as $T^d_{(ij)}$, between trustor node n_i and trustee node n_j, is computed as follows:

$$T^d_{ij}(t) = w_m T_m + w_d T_{\text{td}} + w_c T_c + w_e Te, \tag{6.9}$$

where w_m, w_d, w_c, and w_e represent the weight associate with T_m, T_{td}, T_c, and T_e, respectively. The resulting value represents the direct trust score estimated by the trustor on the trustee, a value closer to 1 indicates entirely trustworthy behaviour, while 0 signifies a lack of trust.

6.4.2 Dynamic Weighting Strategy

In Equation 6.9, the weights w_m, w_d, w_c, and w_e are assigned to each trust dimension, reflecting their relative importance in the overall direct trust calculation. In essence, each weight highlights the estimated influence of the corresponding metric on the evaluated trust score. Assigning weight or, alternatively, selecting the optimal weights for each metric is challenging. Manually assigning weights or distributing them equally across metrics may lead to inefficiencies, as this approach does not consider the varying significance of each metric.

Given the linear nature of the model, each weighted value contributes cumulatively to the overall trust, which can increase the trust score if key metrics show positive values. However, this approach can be problematic, especially when facing sophisticated attacks, as outlined in Section 4.4.1. Certain types of attacks can mask malicious behaviours, making it harder to detect deviations in trustworthiness. This issue underscores the need to increase the weight assigned to specific metrics based on contextual threats and the behaviour of nodes. In response, MATMU introduces a dynamic weighting strategy for each trust dimension T_i, where i represents dimensions T_m, T_c, T_e, and T_{td}. Upon computing each T_i, the corresponding weight w_i is adjusted to reflect the priority of that specific trust dimension in real time. Initially, all trust variables are assigned equal importance. However, if any of the computed trust metrics fall below predetermined thresholds—set according to application requirements—the weighting strategy dynamically adjusts, emphasising the importance of those specific metrics that indicate potential vulnerability or risk. This adaptive approach ensures that the TMM is responsive and more sensitive to changes in critical metrics. In this work, the softmax function has been integrated as a weighting strategy following:

$$w_i = \frac{e^{(\theta_i - T_i)/\nu}}{\sum_{m=1}^{M} e^{(\theta_i - T_i)/\nu}}, \tag{6.10}$$

where θ_i represents the threshold value of each corresponding trust dimension. During the rest of this research, each threshold is set to be 0.5, although the actual setting for each threshold needs to be aligned with both the application and the underlying environment to reflect the importance of each metric. T_i refers to the trust value associated with the θ_i category. The ν is a scaling factor that determines the sensitivity of the weights to the difference between the threshold and trust dimension. The above weighting method satisfies the requirement for w_i values to represent probabilities by scaling each input value relative to the sum of the exponentials of all inputs, ensuring that the total sum of the w_i equals one, while also constraining each w_i to lie within the range of 0 to 1.

6.5 Trust Update

Trust in nature is a continuous value where past behaviour influences current evaluations. This characteristic of trust is incorporated into the model to capture the impact of historical trust scores on the current evaluation over time. Suppose trust was computed at an earlier time, t_i, and new evidence emerges or a trust

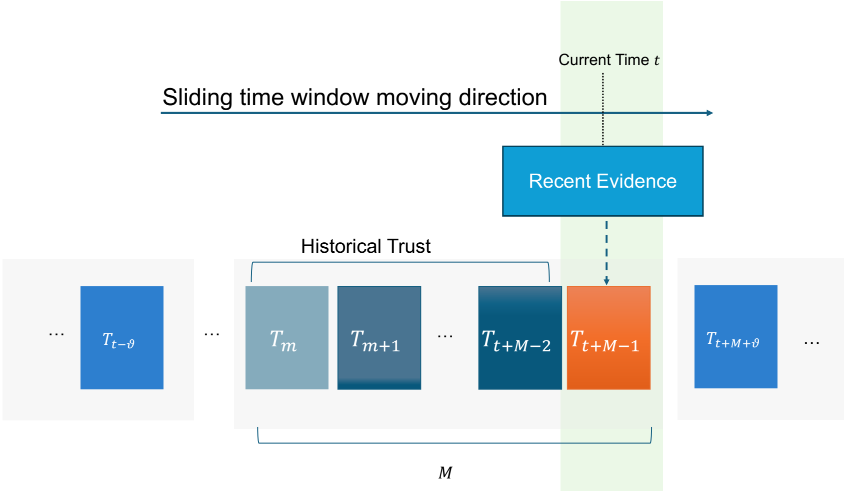


Fig. 6.3. The time sliding window with M slots.

evaluation is triggered at the current time, t_i; in this case, the new computed trust must also account for the last recorded trust values. This information constitutes 'historical trust', which represents the accumulated trust computed in previous periods.

To enable continuous updating of trust scores while incorporating historical information, a sliding time window approach is employed, as illustrated on Figure 6.3. This technique allows for integrating past trust values in the computation of the most recent trust score. The trust score, denoted as $T_{ij}^{d}(t)$, represents the direct trust of node n_j as evaluated by its neighbour n_i at the current time t. The sliding time window operates by utilising a window of size M with m slots, represented as $m = 1, 2, \ldots, M$, progressing sequentially with time. Each slot contains a trust value obtained during that interval of time. As time advances, the previous slots within the window shift and become older historical records. The decayed weight for each trust score is computed as:

$$\Lambda_{slot}^{t} = \frac{e^{t/M}}{\sum_{m=1}^{M} e^{m/M}}. \tag{6.11}$$

Here, t represents the current time and m iterates over the slots within the time window. The numerator $e^{t/M}$ calculates the exponential decay factor for the current slot, resulting in a higher weight for more recent slots, while the denominator $\sum_{m=1}^{M} e^{m/M}$ normalises the weights, ensuring that their sum totals to 1.

In cases where no interactions occur, this mechanism enables the natural decay of the trust value. At each trust evaluation period, the current trust score reflects both recent evidence (or lack thereof) and the cumulative influence of past interactions. This approach is particularly relevant when a node disconnects for an extended period and later re-establishes a connection. In such situations, it is prudent to assign a lower level of trust to this node upon its return, given the uncertainty introduced by its absence. The degree of reduced trust is directly tied to the duration of inactivity: the longer the period without interaction, the lower the trust score will become. This progressive decay ensures that trust scores dynamically adjust based on both the recency of interactions and historical trust patterns.

6.6 Trust Process

This section elaborates on the process underlying the TMM proposed in this chapter. The proposed TMM initiates by collecting evidence, which involves allocating storage to retain information about each connection upon receiving a packet from a node. This evidence is subsequently used to assess the UN's trustworthiness and to make informed decisions regarding potential collaboration. While the focus of this study is not on the decision to collaborate per se, the objective is to evaluate trust by assigning a trust score to each trustee node based on its behaviour across the proposed trust dimensions and to examine the model's ability to detect malicious nodes by lowering their corresponding trust scores. This evaluation involves both direct and indirect trust.

Algorithm 1 MATMU Process

```
Input: List of neighbours ($N$), Trust time period ($t_p$), Metric trust threshold($\theta_i$), Final trust weight ($a \in [0, 1]$)
Output: trust scores for each $n_j \in N$ and trustor $n_i$ denoted as ($T_{ij(t)}$)
Initialise evidence collection variables
while True do
    for each Packet received by Trustor do
        Collect evidence for each element on $N$
    if $t = t_p$ then
        for each $n_j \in N$ do
            if Obtained sufficient evidence then
                Compute $T_m$, $T_{\text{td}}$, $T_c$, $T_e$
                for each Computed Trust Metric $T_i \in \{T_m, T_c, T_e, T_{\text{td}}\}$ do
                    Initiate $w_i = \{w_m, w_c, w_e, w_{\text{td}}\}$
                    if $T_i \leq \theta_i$ then
                        Update the importance of $T_i$ (Equation 6.10)
                        Set $w_i$ for others metrics on $T_i$
                Evaluate Direct Trust Score $T_{ij}^d$ based on $T_i$ and $w_i$ ( Equation 6.9)
                Request recommendation
                Evaluate Indirect Trust Score $T_{ij}^r$ ( Equation D.11)
                Compute $T_{ij} = aT_{ij}^d + (1 - a)T_{ij}^r$
            else
                Decay the trust (Equation 6.11)
            Update trust table
    Update $t_p$
```

Let N be the set of neighbouring nodes of a trustor such as $N = \{n_1, n_2, \ldots, n_N\}$. During each trust evaluation period t_p, the trustor node begins by iterating through each neighbouring node recorded in the collected evidence. As outlined in Section 5.2, it is assumed that, initially, all nodes are trustworthy; therefore, all neighbouring nodes are included in the trust table. This trust table lists all neighbouring nodes along with their corresponding trust scores. The trust table is regularly reviewed and updated based on both current evidence and the decaying influence of historical trust, as explained in Section 6.5.

The model incorporates both direct and indirect trust. At each t_p, the trustor examines each trust dimension and calculates the direct trust according to Equation 6.9. The model evaluates neighbouring nodes based on direct trust and also requests recommendations from surrounding nodes, employing the recommenda-

Table 6.1: Simulation Parameters for Direct Trust Evaluation.

Parameter	**Value**
Simulation Time	3600 seconds
Number of Nodes	20, 50, 100
Attack Rate	1% and then 10%
Threshold (θ_i)	0.5
Initial Weight (w_i)	0.25
Trust Period (t_p)	60 - 600 seconds

tion system previously proposed by CATM. The final trust score of a trustee is a function of both direct and indirect trust, as well as any previously recorded trust. In case there are no updated interactions recorded between the last trust establishment period and the current t_p— for instance, if a node moves away or encounters issues preventing proper interactions— the trust process utilises the last established trust $T_{ij(\Delta t)}$, but with a time-decayed effect, as defined in Equation 6.11. This means that without current evidence the trust score decreases with time. Algorithm 1 outlines the process used to establish trust between nodes.

6.7 Results and Discussion

The TMM was implemented in the Aqua-Sim ng simulator, following the process previously illustrated in Figure 5.4. In each topology established in Section 4.3.1, the specifications outlined in Section 5.3.3 were initially followed to provide an identical layout for testing and validation. The results, averaged over twenty trials conducted in each topology, are presented throughout this section. Each trial involved random nodes within the network topology initiating data packet generation, following an intermittent on/off pattern with a traffic generation set to 0.25 per second. In each scenario, attacks, as described in Section 5.3.3.1, were simulated to evaluate the trust model's effectiveness in detecting misbehaving nodes. Additional details on the simulation parameters are provided in Table 6.1

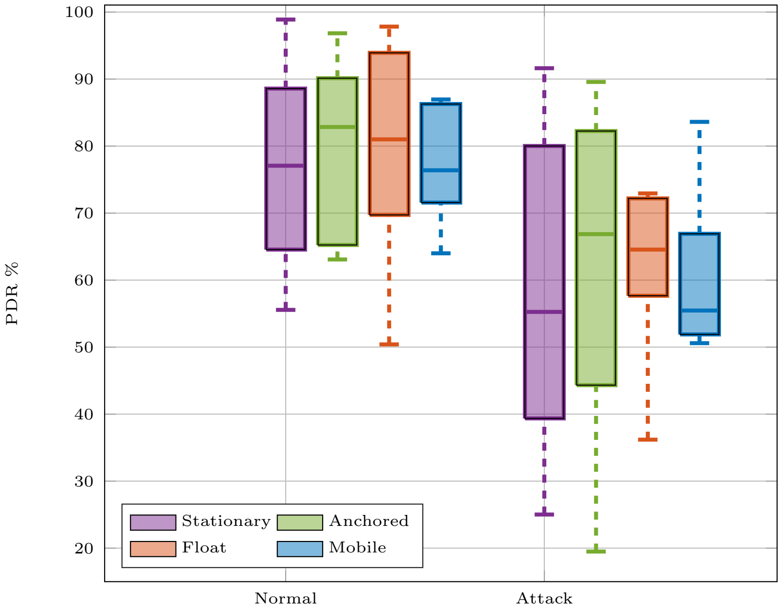
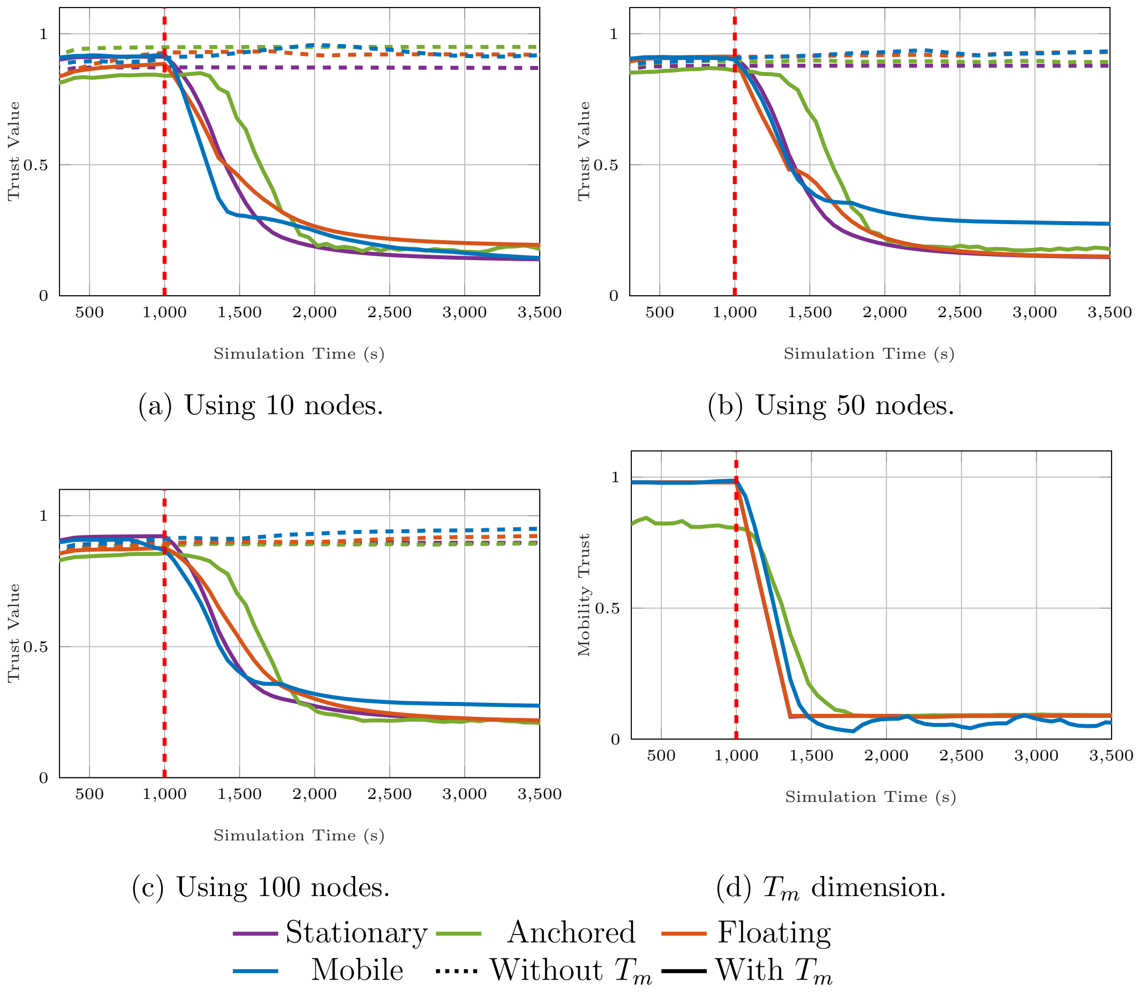


Fig. 6.4. Evaluation of PDR under PM attack.

6.7.1 Evaluation of PDR under PM Attack

To demonstrate the impact of the PM attack discussed in Section 4.4.1 on the network's performance, a series of simulated experiments were initially conducted for each case scenario presented in Section 4.3.1. In these scenarios, nodes collaborated fairly to forward data to the above-water sink node, providing a benchmark for the regular operation of such a network, considering environmental and link quality limitations. Subsequently, selected malicious nodes initiated a PM attack, while continuing to function normally alongside legitimate peers.

The aim of these experiments was to assess the overall network performance by analysing the PDR. In this context, the PDR refers to the ratio of packets successfully received at the above-water sink node. In Figure 6.4, overall network performance is compared in terms of successful packet delivery to the sink node in two scenarios: the fair scenario with no malicious nodes and the attack scenario with designated nodes performing the PM attack. In the fair scenario, where no malicious nodes interfere with packet delivery, the PDR for all node types is relatively high. As expected, a notable drop in the PDR was observed in the



(a) Using 10 nodes.

(b) Using 50 nodes.

(c) Using 100 nodes.

(d) T_m dimension.

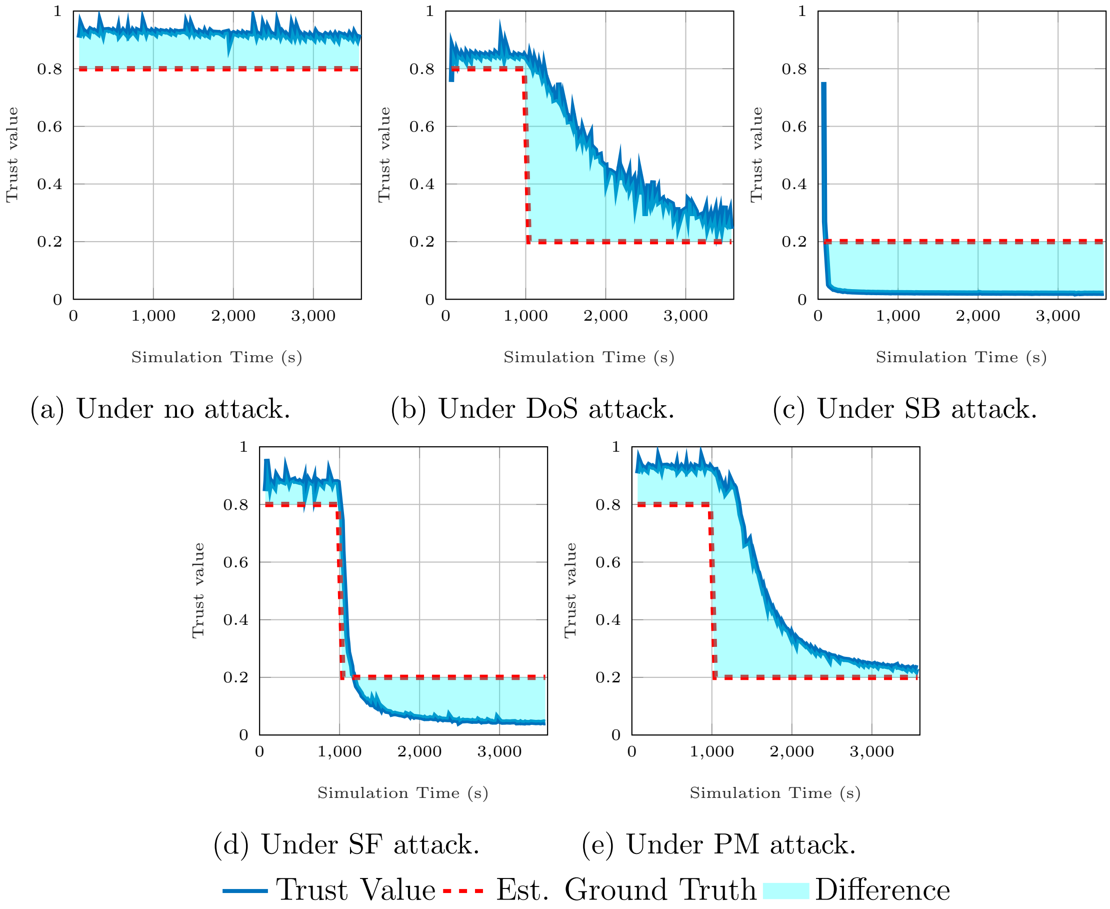
Fig. 6.5. Evaluation of the influence of mobility metric on the overall trust during PM attack.

attack scenario compared to the fair scenario during tests with the PM attack. This confirms earlier descriptions of how exploiting the lack of control over nodes within the network enables malicious nodes to reallocate resources and cause disconnections, thereby avoiding collaboration. This, in turn, degrades network performance and prevents nodes from effectively transmitting data to the entity above water. Subsequent evaluation tests were conducted to assess the ability of MATMU to effectively detect and account for the decline in network performance by appropriately reducing the trust score of the attacking nodes.

6.7.2 Feasibility of Mobility Metric

To evaluate the effectiveness of the suggested mobility metric in detecting PM attacks, the established topologies are examined using 10, 50, and 100 nodes scattered randomly in an underwater area. The PM attack was set to commence 1000 seconds after the initial deployment. This facilitates a thorough inspection of the TMM both before and after the attack. Similar to the assessment made on Section 5.3.3, the trust period is configured to occur every 1 minute.

Figure 6.5a, Figure 6.5b, and Figure 6.5c represent the trust score of the attacking nodes as examined by other nodes in the network with the mobility metric (T_m) and without the mobility metric. Among the number of trials of assessing PM attack detection, the model exhibited robust performance across various node types on detecting the PM through lowering the corresponding trust of malicious nodes upon including the T_m. Statistically, the model achieved high mean accuracy levels— 0.9375 for stationary nodes, 0.8758 for anchored, 0.9150 for floating, and 0.9690 for mobile— coupled with perfect precision in each case. Recall rates were also noteworthy, ranging from 0.8267 in anchored nodes to 0.9581 in mobile nodes, demonstrating the model's effective detection capabilities across different operational environments. The computed T_m obtained using Equation 6.3 is also examined for the attacking nodes across various case scenarios. In Figure 6.5d illustrates a visualisation of the average T_m values evaluated by all neighbouring nodes towards the attacking nodes. Initially, the mobility metric exhibited a high degree of consistency with their peers and produced acceptable values above the average of 0.5, hovering around 1, indicating high spatial and temporal mobility similarities. Nevertheless, the T_m metric experienced a substantial decline shortly following the establishment of the attack by the malicious nodes. This sharp decline in the mobility trust value signifies that the proposed metric for assessing mobility similarities and differences is adept at recognising and capturing abrupt



(a) Under no attack.

(b) Under DoS attack.

(c) Under SB attack.

(d) Under SF attack.

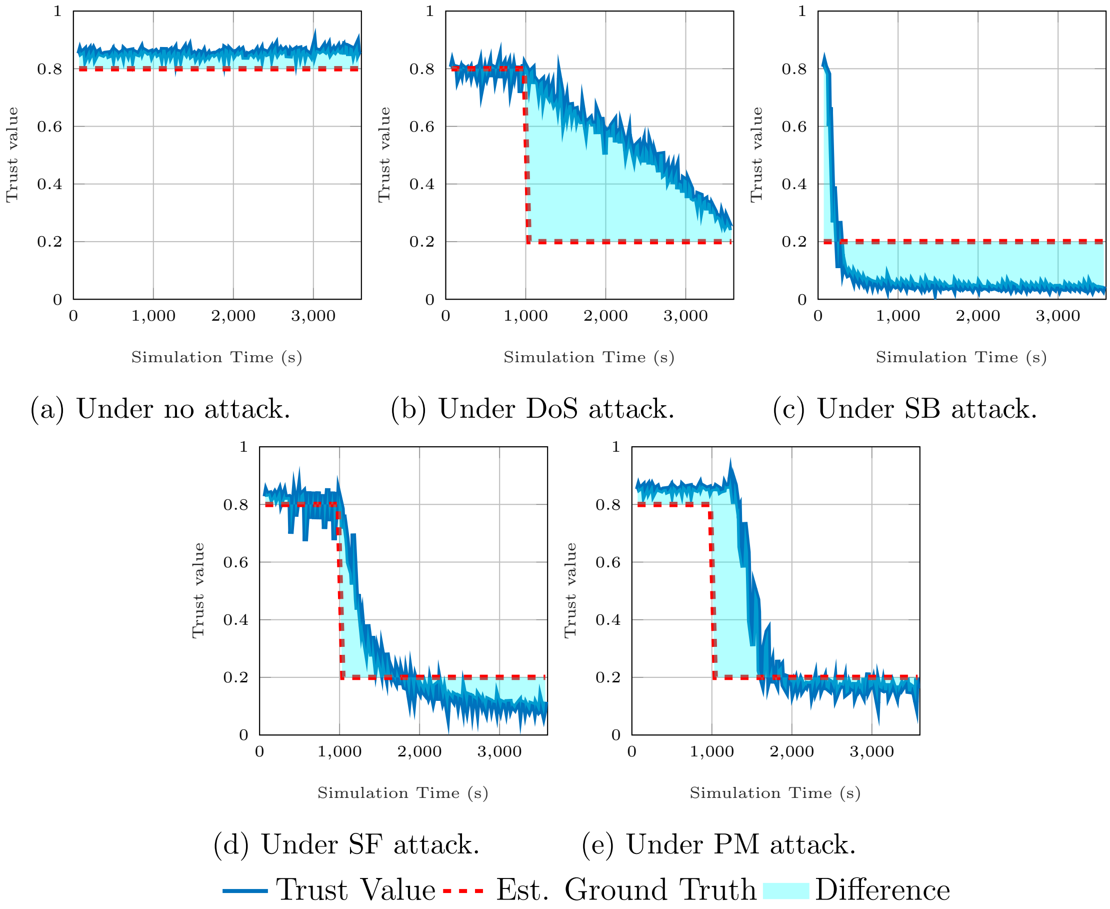
(e) Under PM attack.

Fig. 6.6. Convergence of the obtained trust to the estimated ground truth using MATMU in stationary topology.

changes in the mobility patterns of malicious nodes. When a malicious node exhibits abnormal mobility behaviour, such as sudden and drastic alterations in its movement patterns, the proposed metric can promptly identify these deviations and reflect them by reducing the trust in mobility.

6.7.3 MATMU Convergence over Time

Following the estimated ground truth established earlier in Section 5.3.3.2, the same set of attacks is initialised as in Section 5.3.3.1. The primary goal of this evaluation is to assess the trust model's ability to minimise the convergence area between the expected trust values and the maintained trust values in a timely



(a) Under no attack.

(b) Under DoS attack.

(c) Under SB attack.

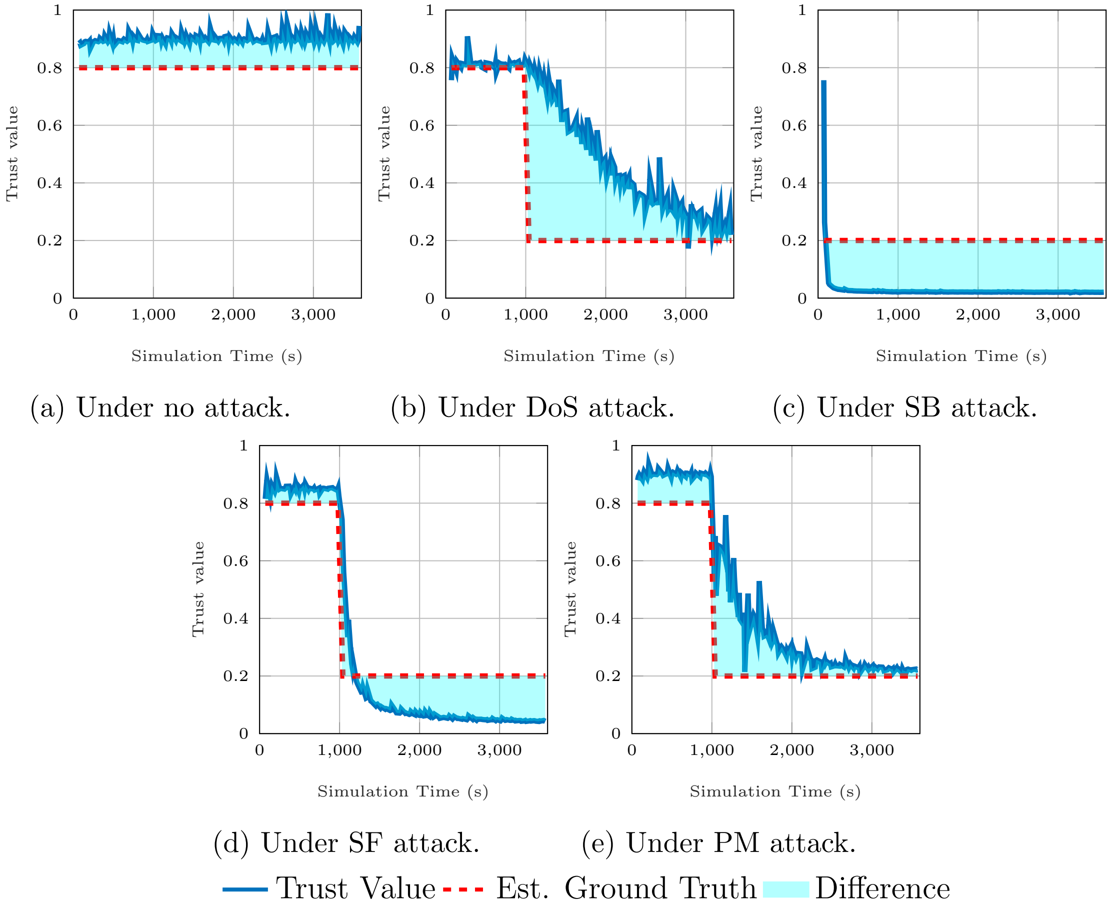
(d) Under SF attack.

(e) Under PM attack.

Fig. 6.7. Convergence of the obtained trust to the estimated ground truth using MATMU in anchored topology.

manner. The results obtained under MATMU for stationary, anchored, floating, and mobile scenarios are presented in Figure 6.6, Figure 6.7, Figure 6.8, and Figure 6.9, respectively.

In a fair scenario, free from malicious attacks, the proposed MATMU consistently maintains a high trust score that closely aligns with the expected values across all scenarios presented in Figure 6.6a, Figure 6.7a, Figure 6.8a, and Figure 6.9a. This indicates that the proposed trust model effectively estimates high trust toward non-malicious nodes within the network, as reflected by the trust scores obtained. Furthermore, since these values represent the average trust across all neighbouring nodes, the model demonstrates consistent and nearly convergent trust scores. In attack scenarios, the red dotted line represents the estimated



(a) Under no attack.

(b) Under DoS attack.

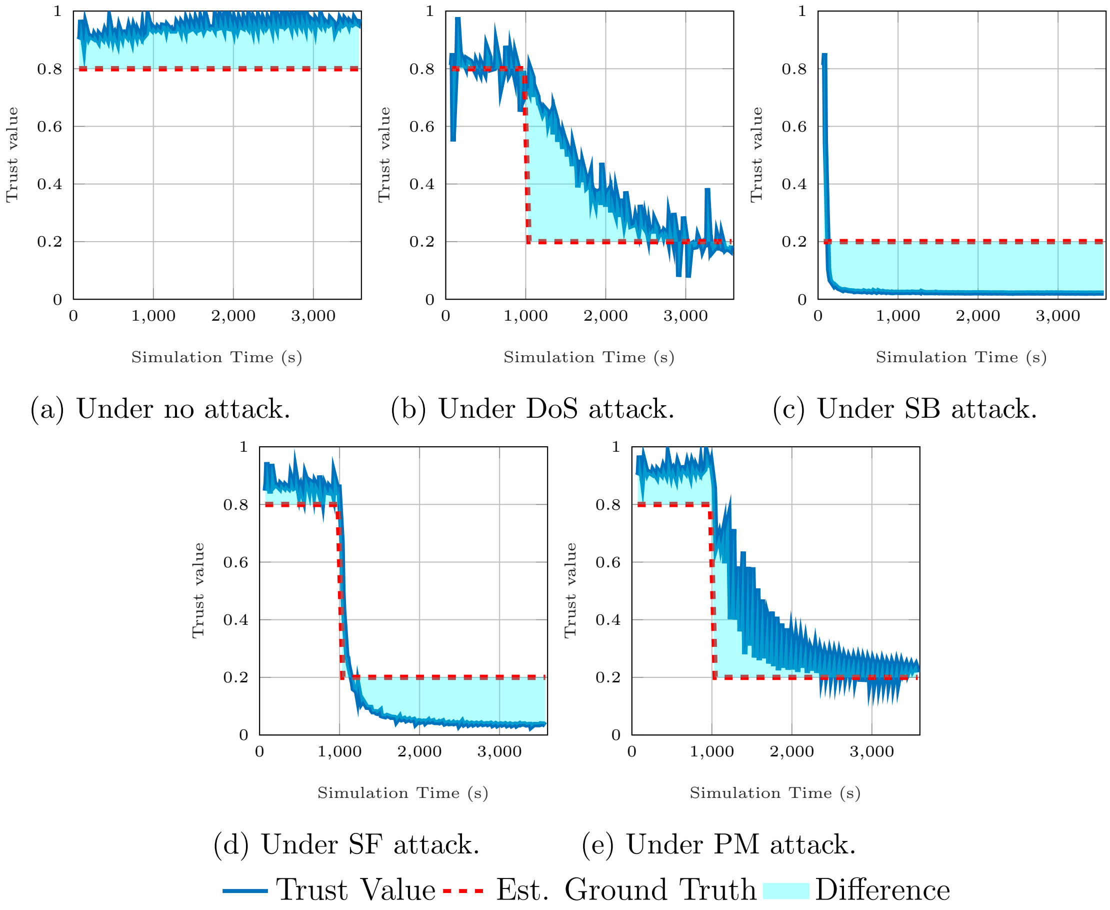
(c) Under SB attack.

(d) Under SF attack.

(e) Under PM attack.

Fig. 6.8. Convergence of the obtained trust to the estimated ground truth using MATMU in floating topology.

trust score directed toward malicious nodes. For communication-based attacks, such as DoS, SB, and SF attacks, only minimal divergence is observed between the estimated ground truth and the trust score produced by the model. This result can be attributed to the dynamic weighting mechanisms integrated within the model, which adjust the importance of each trust metric in real-time based on available evidence. Such an approach enhances the model's capability to detect attacks in a timely manner. Importantly, upon including PM attack to the network, the TMM is capable of detecting the attack in all cases.



(a) Under no attack.

(b) Under DoS attack.

(c) Under SB attack.

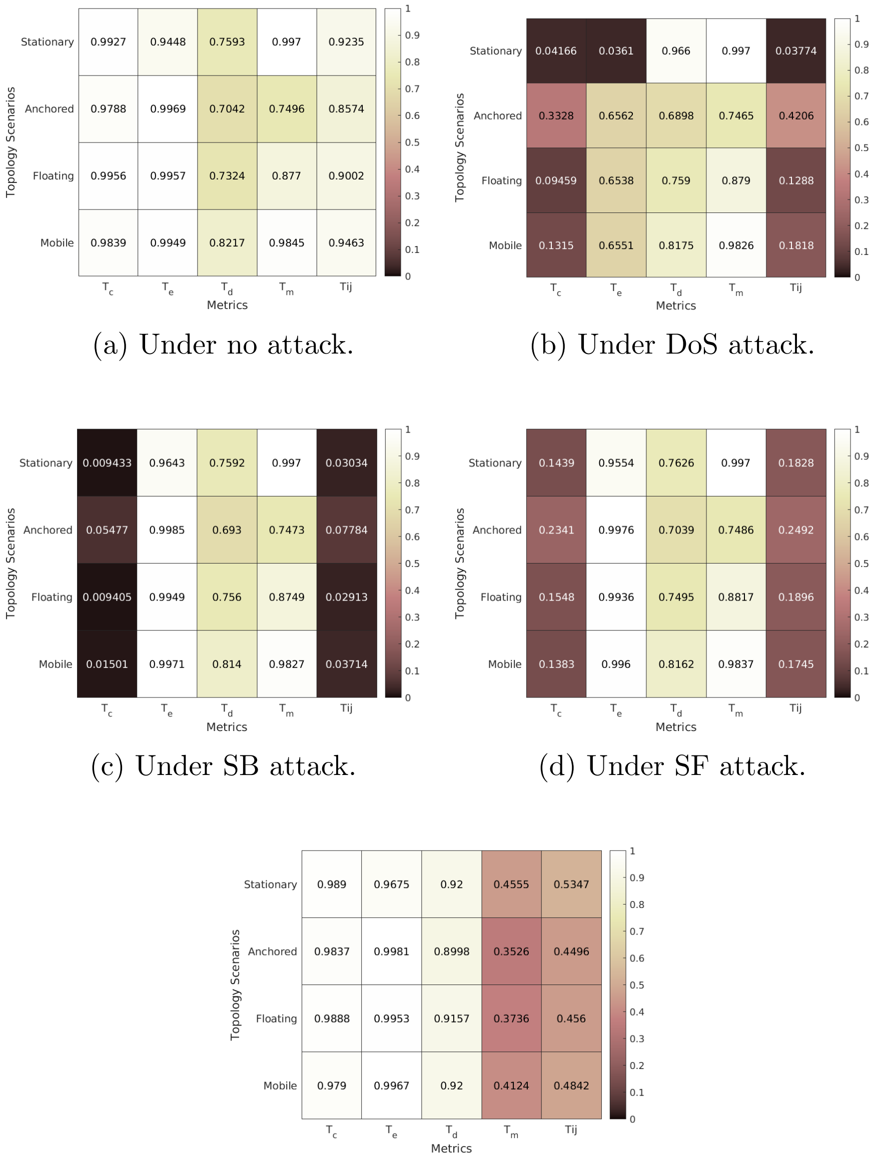
(d) Under SF attack.

(e) Under PM attack.

Fig. 6.9. Convergence of the obtained trust to the estimated ground truth using MATMU in mobile topology.

6.7.4 The Influence of Dynamic Weighting Strategy on the Overall Trust

To revisit the primary evaluation objectives established in the previous chapter, particularly regarding Section 5.3.3, the focus is on (1) determining whether these metrics can effectively identify malicious nodes by reducing their trust scores and (2) assessing the impact of each trust metric on the overall trust score. The previous experiments demonstrated that MATMU successfully achieved the first objective by identifying misbehaving nodes through the reduction of their associated trust scores. The next step is to conduct a more detailed analysis of the second objective: analysing how each individual metric contributes to the overall



(a) Under no attack.

(b) Under DoS attack.

(c) Under SB attack.

(d) Under SF attack.

(e) Under PM attack.

Fig. 6.10. Metrics assessment using MATMU.

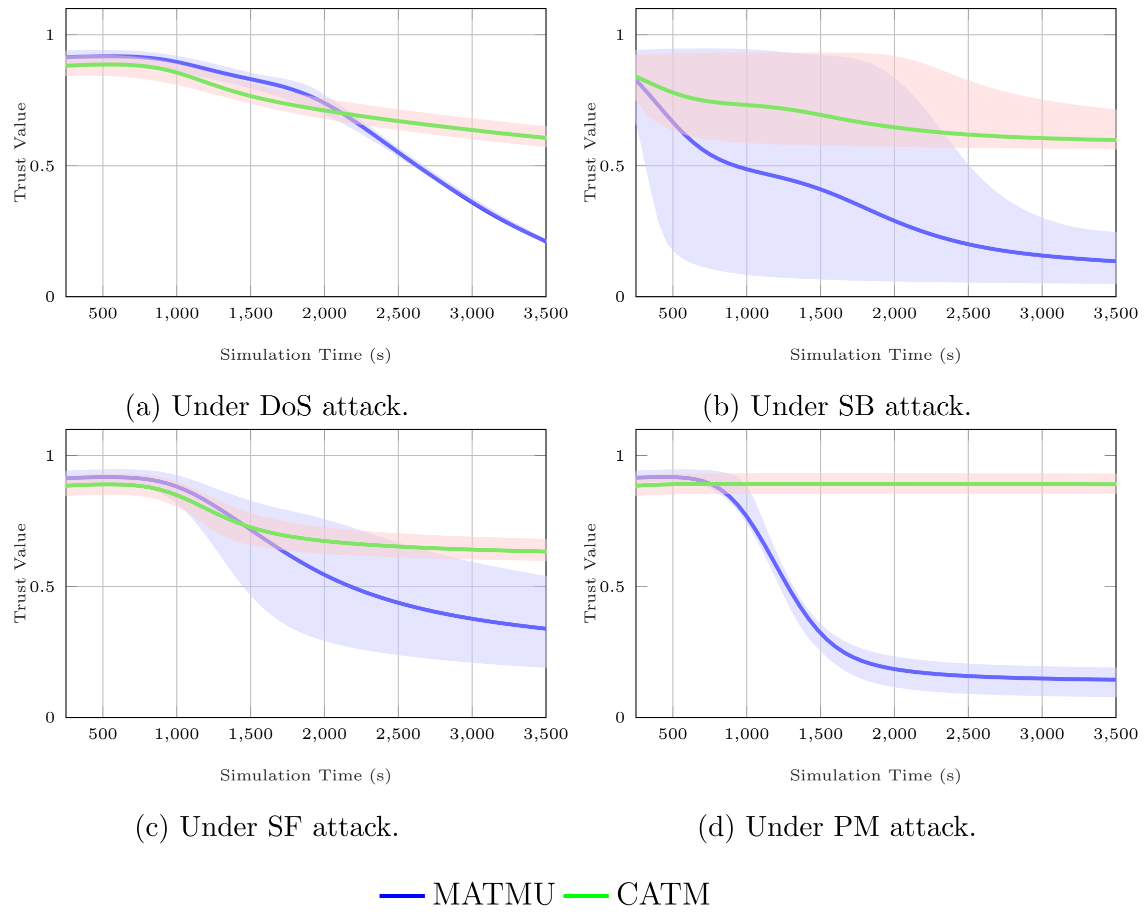
trust. Specifically, this analysis aims to demonstrate the role of the real-time dynamic weighting methods employed in this work in enhancing overall trust.

Figure 6.10 presents the average accumulated results for each metric across the various network topologies and attack scenarios. Specifically, the analysis focuses on how each trust dimension utilised in the proposed model influences the corresponding trust scores, providing a deeper understanding of their internal impact. In the fair scenario, each trust dimension contributes equally to the overall trust score, as demonstrated in Figure 6.10a. However, the subsequent figures, Figure 6.10b, Figure 6.10c, Figure 6.10d, and Figure 6.10e reveal how the metrics perform under different attack scenarios. Unlike the earlier analysis presented in Figure 5.13, the current model demonstrates its ability to integrate lower trust metrics into the overall trust score through the dynamic weighting strategy. This adaptability is particularly evident in the PM attack scenario, where the metric T_m stands out as the only dimension showing sensitivity.

All of the experiments above were conducted under the same attack conditions, similar to those tested in the previous chapter, which also highlighted the challenges of being captured using existing TMM. This issue is clearly addressed by the proposed model. However, the severity of the attacks significantly influences both the detection time and the corresponding reduction in the trust score of the compromised nodes. In other words, the simulated nature of the attacks can be adjusted to be less detectable, and examining the models under these conditions seems reasonable. To investigate this further, different malicious test cases will be conducted to examine the effect of convergence time in the following sections.

6.7.5 Evaluating the Resilience of MATMU versus CATM Against Low-Profile Attack Strategies

The previous tests were conducted using the attack scenarios detailed in Section 5.3.3.1 to provide a comparative evaluation of the proposed model against previously established models outlined in Section 5.3. This section analyses the performance of the model under less detectable attack behaviours. The term ‘low-profile’ attacks in this context refers to a strategy where the severity of an attack is intentionally diminished in order to develop a new sets for testing network vulnerabilities. Consequently, each attack simulated is adjusted to align with these criteria. Specifically, the parameters of the SB attack were modified



(a) Under DoS attack.

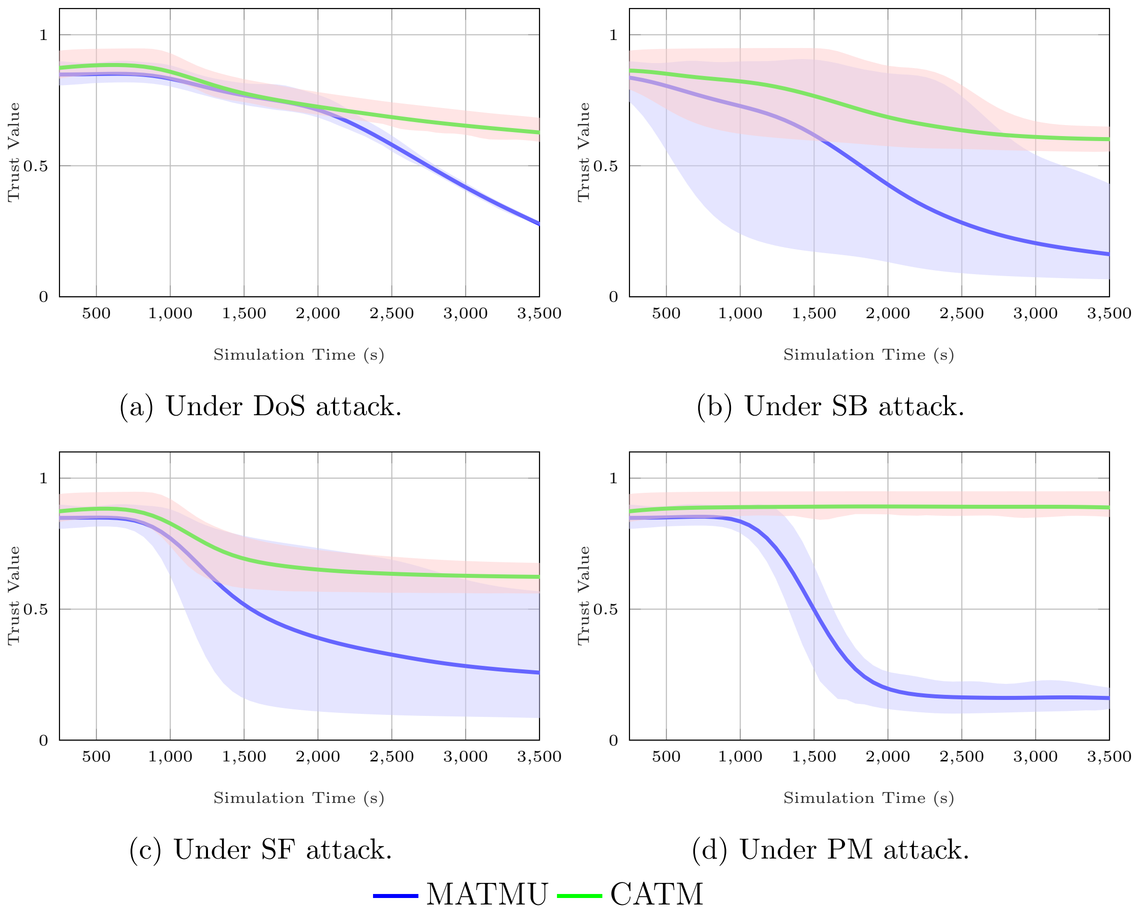
(b) Under SB attack.

(c) Under SF attack.

(d) Under PM attack.

Fig. 6.11. Trust evaluation for stationary nodes under different attacks.

to initiate after 10% energy consumption, rather than the previous threshold of 1%. This change establishes varied starting points for the attack, based on the percentage of energy used. For the SF attack, the intensity was reduced from 0.8 to 0.5, and a dynamic mechanism for switching between blocked nodes was introduced. Likewise, within the DoS attack, the packet streaming frequency decreased, rendering it less detectable. The duration of the attacks—SF, PM, and DoS—remained consistent at 1000 seconds throughout the simulations to maintain uniform conditions across all scenarios. For each test, the trust scores assigned to malicious nodes were monitored and analysed at the conclusion of each trust evaluation period. Figure 6.11, , Figure 6.12, Figure 6.13, and Figure 6.14 offer a comprehensive visualisation of trust evaluations under a spectrum of adversarial conditions, as discerned across multiple trials. Within each graph,



(a) Under DoS attack.

(b) Under SB attack.

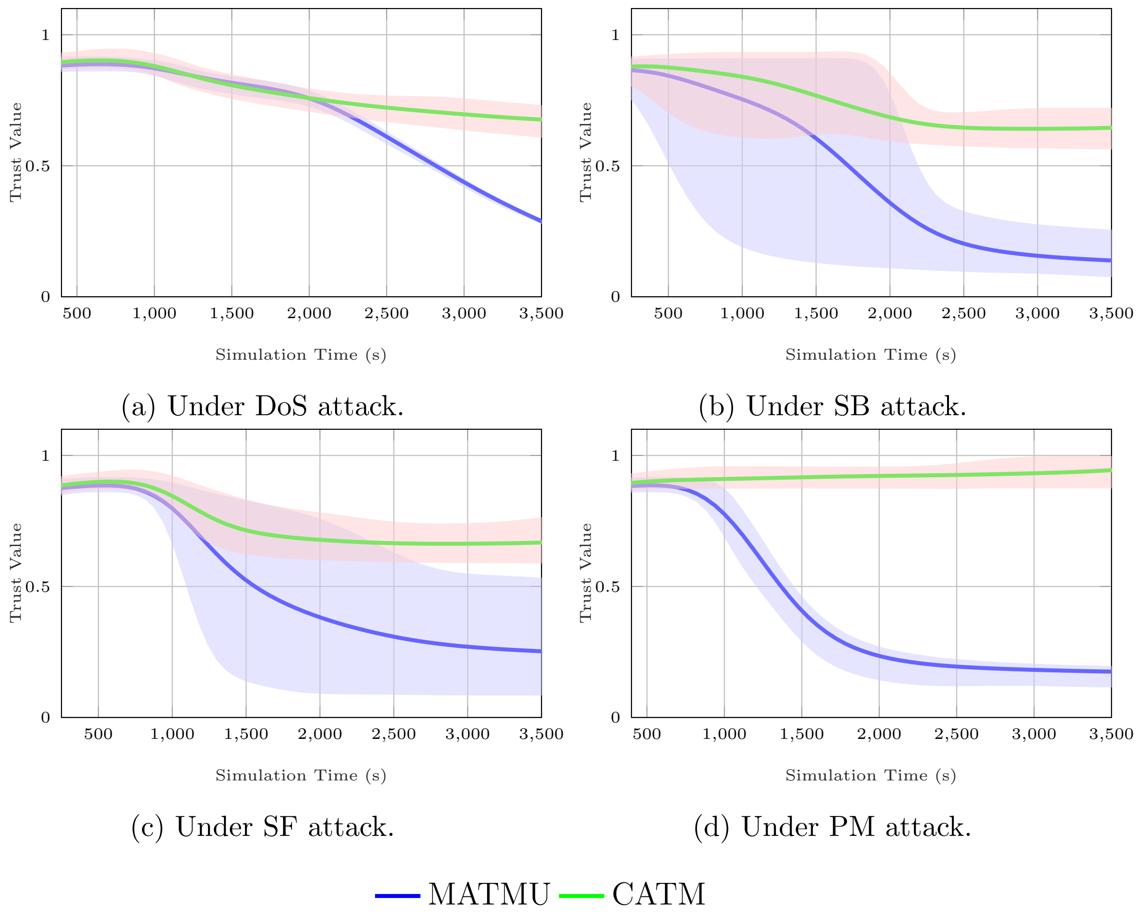
(c) Under SF attack.

(d) Under PM attack.

Fig. 6.12. Trust evaluation for anchored nodes under different attacks.

the line plot aggregates the average trust scores from 20 trials. The blue shaded area in these graphs encapsulates the spreading of the results from experiments where MATMU was employed. Conversely, the red-shaded area represents the outcomes achieved using CATM, to provide a comparative analysis of its efficacy in identical conditions.

Figure 6.11 offers an in-depth analysis of the trust scores obtained in scenarios involving stationary nodes. Figure 6.11a showcases a steady decrease in trust scores in response to a DoS attack. The blue shaded area indicates an adaptive, albeit variable, model response as the attack unfolds compared to CATM. During the SB Attack case in Figure 6.11b, MATMU successfully detected nodes that failed to cooperate, as evidenced by the trust scores. The variation in the



(a) Under DoS attack.

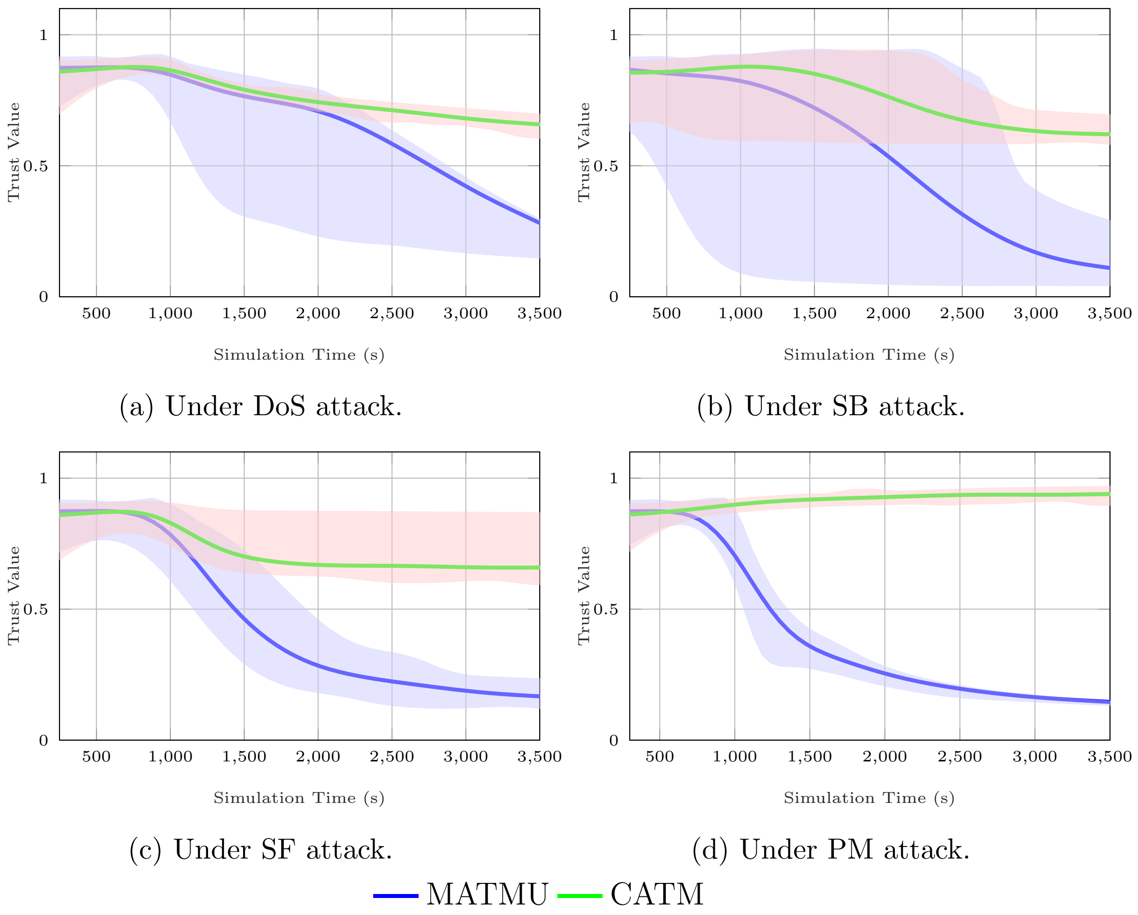
(b) Under SB attack.

(c) Under SF attack.

(d) Under PM attack.

Fig. 6.13. Trust evaluation for floating nodes under different attacks.

onset time of the attack was a direct result of the specific conditions set for the attack. Particularly, a node positioned at a higher energy consumption spot in the network depleted its energy faster than others, triggering the SB behaviour earlier. In both early and later onset cases, MATMU was adept at identifying the attacking nodes and correspondingly lowering their trust scores. A key factor enhancing MATMU's ability to detect the attack more rapidly compared to CATM was the utilisation of a dynamic weighting strategy, where MATMU swiftly adjusts its trust evaluations in response to changing behaviours, thereby speeding up the detection of SB attacks. This is also noticeable in the case of SF attacks in Figure 6.11c that delineates a more tempered descent in trust under a SF attack, a subtlety that is potentially captured with greater sensitivity by MATMU where MATMU shows superiority over CATM. In the case of a PM attack, as



(a) Under DoS attack.

(b) Under SB attack.

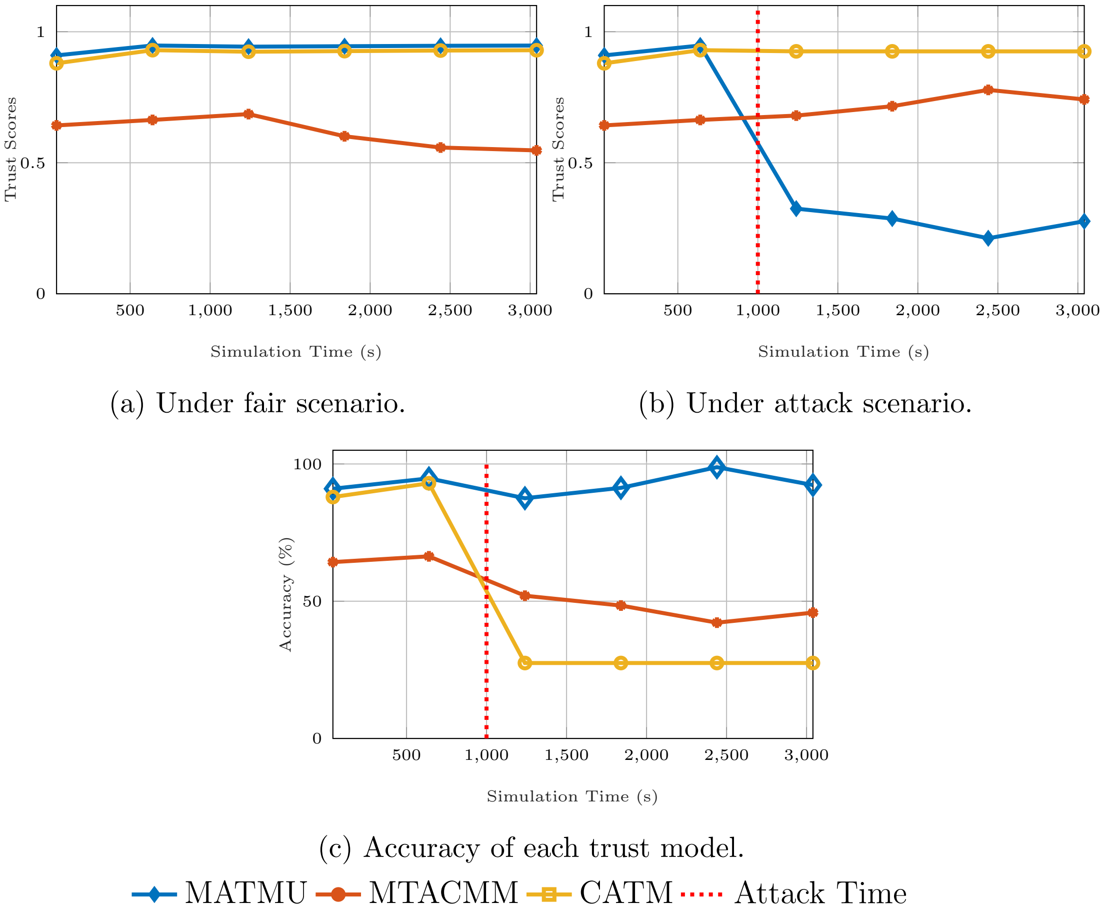
(c) Under SF attack.

(d) Under PM attack.

Fig. 6.14. Trust evaluation for mobile nodes under different attacks.

shown in Figure 6.11d, the attack was recognised almost immediately after incorporating the mobility metric. The sharp decline of the attacking node's trust score signifies the significant impact of physical mobility on network trust. This indicates that solely considering factors like successful communication, delay, and energy consumption is insufficient in addressing this particular type of attack.

Figure 6.12, Figure 6.13, and Figure 6.14 illustrate the obtained trust scores in the case of anchored nodes, floating nodes, and mobile nodes, respectively. When communication-related attacks were tested, the trust scores in MATMU dropped considerably compared with CATM. This demonstrates that MATMU is sensitive to such attacks and adjusts the trust scores accordingly. In contrast, CATM, lacking the proposed mobility similarity metric, proved insufficient in identifying



(a) Under fair scenario.

(b) Under attack scenario.

(c) Accuracy of each trust model.

Fig. 6.15. Comparison with benchmarked TMM.

PM attacks in all the listed scenarios and was ineffective in detecting such attacks. Throughout the simulation period, the nodes were unable to detect these attacks, and high trust scores were maintained for the attacking nodes. This is primarily because, in these types of attacks, malicious nodes maintain their normal behaviour towards other nodes while maliciously altering their formation to paralyse the network performance. This means MATMU, incorporating the mobility metric, effectively addresses this issue by promptly detecting such attacks and accurately adjusting the trust scores of the attacking nodes.

6.7.6 Comparison under Physical Attack

MATMU was compared with both CATM and MTACMM, as illustrated in Figure 6.15. To provide a more equivalent comparison, the simulation parameters outlined in (Bolster 2017) were followed. More specifically, this comparison test was conducted using only floating nodes, and the trust estimation period was set to 600 seconds to initiate the trust evaluation every 10 minutes. Two distinct tests were conducted to evaluate the effectiveness of each model in detecting PM attacks. The initial experiment assessed trust levels in a normal situation without any attacks in the network, while the second experiment introduced attacks and observed the trust scores toward the malicious nodes. As depicted in Figure 6.15a, all three models maintained trust scores above 0.5 in the absence of attacks. Notably, MATMU and CATM consistently maintained high trust scores, hovering around one, while MTACMM maintained a nominal trust score specific to its trust mechanism. Figure 6.15b shows the resulting trust scores after the attack occurred. Clearly, MATMU responded quickly to the attack, causing a sharp drop in trust scores for the attacking nodes. In contrast, both CATM and MTACMM exhibited no sign of an attack. The second experiment aimed to assess the accuracy of the models in identifying and detecting attacks. The accuracy of each trust model was quantitatively evaluated using the Root Mean Square Error (RMSE) measure, which quantifies the discrepancies between expected and actual trust scores, particularly in the context of a PM attack. The simulation results, as presented in Figure 6.15c, offer insights into the performance of the models when faced with PM attacks. Notably, MATMU consistently outperformed both CATM and MTACMM in effectively detecting malicious behaviour. This superior performance can be attributed to the incorporation of mobility metrics, which enhance the model's ability to detect and respond to attacks more accurately and promptly.

6.8 Conclusion

This chapter has introduced a novel distributed mobility-aware TMM for the IoUT, referred to as MATMU. The proposed TMM specifically focuses on constructing trust based on multiple metrics. In addition to communication-based metrics such as communication trust, delay trust, and energy trust, the model incorporates trust metrics derived from the physical domain. Specifically, it integrates mobility metrics to assess node behaviour by analysing similarities and differences in mobility patterns. Both spatial and temporal mobility changes are factored into the estimation of suspicious behaviour, particularly in alignment with the nature of PM attacks. This expands the metric space to include mobility trust as a key element. Within the selected set of metrics, the influence of a dynamic weighting strategy was demonstrated, enhancing overall trust accuracy. This adaptive strategy enables the use of a straightforward fusion method, such as a weighted sum, to avoid computational complexity while ensuring the reliability of trust establishment.

Under the previously established attack models, the proposed TMM exhibited high accuracy in detecting malicious entities by reducing the corresponding trust values associated with such entities. These results were validated across various network formations, demonstrating the model's stability in stationary, anchored, floating, and fully mobile topologies. Furthermore, the proposed model was evaluated against benchmarked related works, showcasing superior accuracy in detecting malicious activities. However, these evaluations were conducted under the assumption that no malicious entities were actively attempting to disrupt the TMM through dishonest recommendations. Consequently, the results presented in this chapter assume that all recommendations received are fair and truthful and, therefore, the related estimation of T_{ij}^{r} neglects the influence of malicious recommender. Addressing the issue of dishonest recommendations—where entities may provide fake or biased feedback—remains a significant challenge. Such behaviour could lead to the misrepresentation of trust and undermine the system's reliability. This limitation constitutes the next focus of this thesis.

Chapter 7

Recommendation-based Trust Evaluation Model

7.1 Introduction

In the preceding chapter, the focus was on establishing trust based on direct evidence, which allows trust to be built on tangible and observable interactions. However, the sparse nature of underwater networks often makes direct evidence insufficient. In such situations, it becomes necessary to rely on feedback from other nodes in the network. While this reliance is useful for making decisions about the trustworthiness of an entity, it also introduces vulnerabilities. Malicious actors can manipulate this feedback to misrepresent trustworthiness, as discussed in Section 5.4.1. Addressing the challenge of effectively evaluating and verifying received recommendations is, therefore, a key concern in this chapter.

Several mechanisms for detecting dishonest recommendations have been discussed in Section 5.4, where these methods are classified based on the received trust values themselves. The decision-making process may either be strictly based on the trust values received at a specific time or rely on the trustworthiness of the recommender. As stated earlier, the sparsity of the network restricts the ability to validate and evaluate received recommendations. With limited evidence available at a given time, it becomes difficult to avoid relying on recommendations, and even more challenging to assess their validity due to insufficient prior knowledge for decision-making. For instance, in the case of the depart and return scenario, when a UN temporarily leaves the network, its trust value decays over time to comply with the requirements of TMM. The longer the absence, the lower the trust value it retains. However, when requesting recommendations from others, ensuring the validity of the received recommendation about that node becomes challenging.

This chapter explores a recommendation evaluation process that aims to detect dishonest recommendations in the context of trust establishment among IoUT nodes. A new evaluation mechanism is proposed for validating received recommendations within the TMM. It includes a detection mechanism that evaluates recommendation trust values, focusing on improving the identification of inconsistencies in recommendation data over time. A belief function is also introduced to refine the weighting of recommendations, taking into account factors like freshness, similarity, trustworthiness, and the natural decay of trust over time. The effectiveness of this approach is assessed through a detailed evaluation process, which includes conceptual analysis and simulations to compare its performance with existing methods.

7.2 Proposed Recommendation Evaluation Model

TMM is constructed using direct and indirect trust, where the former is formulated based on direct observation, while the latter relies on the opinions of others. In this chapter, a new evaluation mechanism for received recommendations is proposed to enhance the indirect trust evaluation as illustrated in Figure 7.1. The process begins with a filtering mechanism that combines two distinct phases. The first phase, outlier detection, identifies recommendations that significantly deviate from the majority opinion. The second phase, deviation exemption, examines the differences between the recommendations and direct trust values, filtering

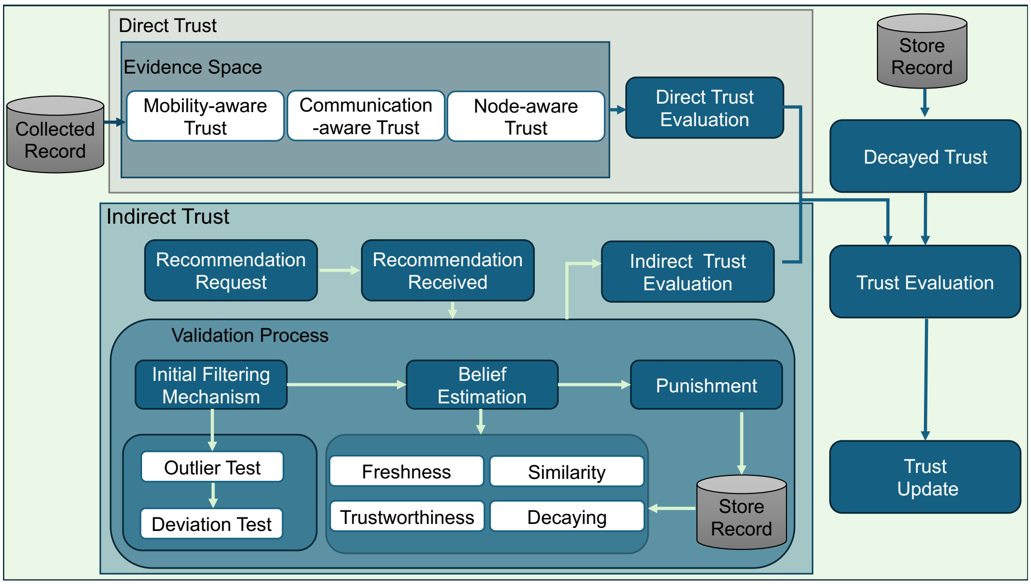


Fig. 7.1. Proposed recommendation evaluation model.

out those that show substantial divergence. This dual-phase approach aimed to enhance the accuracy and reliability of the recommendation evaluation process, as demonstrated in the preliminary examination detailed in Section 5.4.4 where the combination of mechanisms for filtering out recommendation values showed effective performance. Building on the work proposed by Adewuyi et al. (2019), the issue is further addressed by introducing a belief method aimed at weighting recommendations. The proposed belief method employs several key metrics: freshness, trustworthiness, similarity, and trust decay in constructing a belief toward each recommender in relation to other recommenders. Each of these factors contributes uniquely to the evaluation process.

The recommendation evaluation process is then applied as part of the validation of received recommendations to construct indirect trust. In other words, each recommendation is assessed to determine its contribution to the trust assigned to the trustee (T^r_{ij}). The resulting value is aggregated with the direct trust (T^d_{ij}) and stored in the trust record for further updates following: $T_{ij} = aT^d_{ij} + (1-a)T^r_{ij}$.

7.2.1 Indirect Trust Evaluation

Based on the proposed TMM earlier, MATMU, it is anticipated to compute and periodically update trust values among nodes. Assuming n_i wants to compute the trust score toward node n_j. The trustor n_i computes the direct trust of the trustee n_j following the previously developed method on Section 6.2. The workflow of the proposed approach is as follows:

Step 1: n_i sends a recommendation request to its 1-hop neighbours regarding node n_j. This is done by transmitting a request recommendation packet. n_i then sets a timer to receive recommendation responses about n_j. Nodes that do not cooperate are considered selfish nodes.

Step 2: Upon receiving recommendation responses from all neighbours, n_i begins the evaluation process, as outlined in Figure 7.1, to assign a corresponding weight to all the received recommendations. This step involves analysing the received forwarding response messages to gauge the recommendations.

Step 3: n_i applies the necessary punishments to nodes that misbehave and stores this information for the next trust evaluation process.

Step 4: n_i uses the assigned weights to evaluate the recommendations. Hence, the indirect trust will be estimated as follows:

$$T^{r}_{ij}(t) = \sum_{k=1}^{K} B_{ij\leftarrow k} T_{kj}, j \neq k, \tag{7.1}$$

where, $T^{r}_{ij}(t)$ represent the indirect trust computed by n_i towards n_j, based on recommenders n_k, with the number of recommendations received being $K \geq 2$. $B_{ij\leftarrow k}$ here is the final weight mechanism based on the belief n_i produced to each recommendation received by n_k.

7.2.2 Establishing Belief Assumptions

The following premises are the driving force to model the belief function ($B_{ij\leftarrow k}$).

1. Let the change between the current recommendation on n_j provided by recommender n_k be known as $T^r_{kj(t)}$ and let the previous observed trust computed by n_i towards n_j be known as $T_{ij(t-\epsilon)}$, where ϵ represent the proportion of time elapsed between constructing direct trust and receiving recommendations. The smaller the absolute proportion of change $|T^r_{kj(t)} - T_{ij(t-\epsilon)}|$, the easier it is for n_i to accept the recommendation. Thus, $B_{ij\leftarrow k} \propto \Big(|T^r_{kj(t)} - T_{ij(t-\epsilon)}|\Big)$

2. The longer the time t that has passed since the last session of interaction between n_i and n_j, the more open n_i will be in accepting n_k's recommendation. This is because of the trust **decay**; the longer the time goes by, the smaller the proportion of historical trust based on the last observation left. Thus, $B_{ij\leftarrow k} \propto (1 - d_{T(ij)})$.

3. The greater the **trustworthiness** of n_k, represented by the value of T_{ik}, the more likely one is to trust the recommendation received from n_k. While this assumption alone is not necessarily correct, as explained in Figure 5.14, more sophisticated attackers might exploit this by intermittently or even consistently behaving correctly to mask their malicious actions. Nevertheless, a lower trust rating of recommenders can still indicate a potentially bad recommendation. Thus, $B_{ij\leftarrow k} \propto T_{ik}$.

4. The opinion computed by n_k is more valuable if it is recent, as newer opinions are generally more relevant than older ones. The **freshness** Γ_{kj} is defined as the duration between the present time and the moment when $T_{kj(t)}$ was formed. Thus, $B_{ij\leftarrow k} \propto \Gamma_{kj}$.

5. In dynamic environments, it is often observed that nodes increasingly favour others exhibiting similar behaviours over time (Cho et al. 2010). Trust, drawn from social constructs, is interpreted through the lens of interaction consistency between two entities (Marche et al. 2020). A relationship's strength is presumed to be stronger with more consistent interactions over time. Furthermore, the relative duration or depth of a relationship compared to other peers indicates the level of similarities between entities. This is particularly evident in collaborative networks like IoUT, where all nodes are initially assumed to collaborate equally to achieve the needed coverage and maintain connectivity (Mohsan et al. 2023).

   Let $s_{ik(t)}$ represent the degree of **similarity** between n_i and n_k, the more similarity, the more likely for n_i to believe the opinion provided by n_k, and therefore, $B_{ij \leftarrow k} \propto s_{ik(t)}$.

7.3 Recommendation Evaluation Process

Let $R_j(t) = \{T_{k_1j}, T_{k_2j}, \ldots, T_{k_Kj}\}$ be the set of recommendations received upon a request from node i to nodes: k_1.. k_K about j, where K represents the number of 1-hop neighbours willing to respond to the recommendation request at time t. The following subsections highlight the main process involved in the proposed recommendation evaluation model.

7.3.1 Initial Filtering Process

In this work, a two-step filtering process is introduced that integrates the majority rules approach with the deviation technique to ensure thorough scrutiny of recommenders before subjecting them to a penalty process. The first subset of $R_j(t)$ is defined as $S_{1j} \subset R_j$, representing a list of recommendations detected as outliers. A second subset, derived from S_{1j}, denoted as $S_{2j} \subset S_{1j}$, contains all recommendations that deviate from the computed trust by i. Based on the analysis demonstrated in Section 5.4.4, particularly in examining the resulting accuracy of both majority rules and deviations from the personal experience process, it is concluded that utilising both methods can effectively enhance the accuracy of eliminating dishonest recommendations. This chapter focuses on $O \rightarrow D$ relations due to their high performance across different neighbour sets compared to other methods of examination. Upon isolating S_{1j}, which includes potential outliers, each element within this list is compared against personal experience using the deviation test to formulate a potentially smaller list as S_{2j}. A notable issue arises when $T_{ij(t)}$ becomes outdated due to an extended period since the last valid interaction between n_i and n_j. Assuming the time passes since the last timely observation is ϵ and the stored trust is $T_{ij(t-\epsilon)}$, this outdated value fails to provide a reliable baseline for testing deviation. To address this issue, a deviation test that incorporates the temporal decay of trust is proposed, defined as follows:

$$|T_{ij(t-\epsilon)} - T^{r}_{kj(t)}|(1 - d_{T(ij)}) \leq \delta_h, \tag{7.2}$$

where $T_{ij(t-\epsilon)}$ denotes the trust score stored by n_i regarding node n_j at a previous time instance $(t - \epsilon)$, δ_h is the predefined trust deviation threshold, and T^{r}_{kj} denotes the trust recommendation received from node n_k. The variable $d_{T(ij)}$ is the decay factor, with more recent T_{ij} values resulting in a lower $d_{T(ij)}$ compared to older values.

Both outlier detection methods exhibit satisfactory accuracy, with the MAD method achieving approximately 90.59% accuracy and the LOF method achieving 92.5% accuracy. The selection of the appropriate outlier detection method is highly contingent upon the computational complexity requirements of the specific application in which it is deployed. The complexity of each method is estimated by first applying the outlier detection mechanism followed by the deviation test (denoted as $O \to D$) as follows. Let n denote the size of the recommendation set $|R|$, and let m denote the size of the filtered recommendation set $|R'|$, where $m \leq n$. The computational complexity of the MAD method is estimated to be $O(n \log n)$, whereas the complexity of the LOF method is approximately $O(n^2 \log n)$ (Domingues et al. 2018). In the proposed filtering approach, the overall complexity is estimated to be $O(m) + O(n \log n)$ for MAD, or $O(m) + O(n^2 \log n)$ for LOF. For the remainder of the chapter, the initial filtering is employed using the MAD due to its simplicity (referred to as: `isOutlier`).

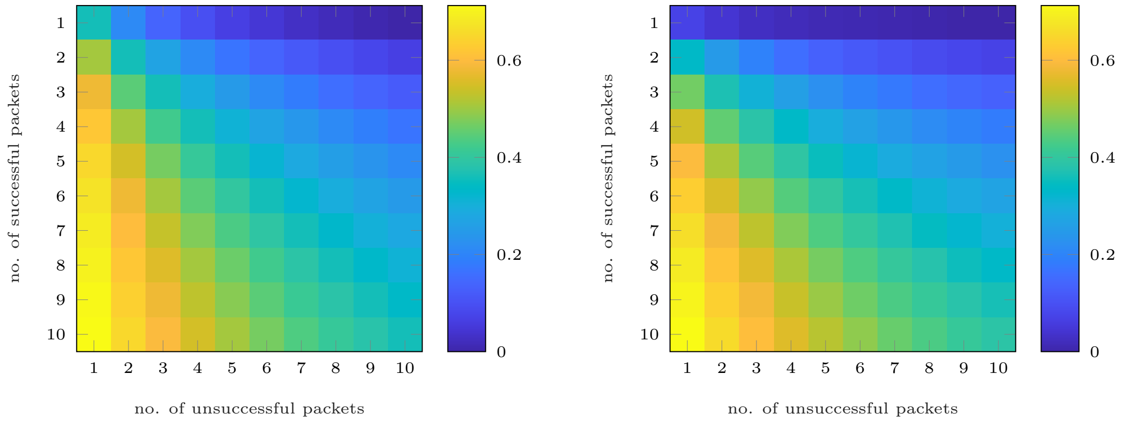
Following this, a monitoring system is initiated where each node's frequency of appearing as in S_{2j} is tracked. A counter is set up for each node to record the number of times it appears on the outlier list. If a node's appearance as an S_{2j} exceeds a predefined threshold (v_h), indicative of consistently poor or suspicious recommending behaviour, a penalty is then applied to its trust score, $T_{ki}(t)$ as outlined in Section 7.3.4.

7.3.2 Definition of Similarity

This section explores the established definition of similarity and provides a discussion on how it is derived in existing TMM.

7.3.2.1 Confidence Level as a Measure of Similarity

In Section 5.4.3, the challenges associated with using confidence as a validation metric for recommenders, as discussed in the literature, were examined in the context of IoUT. Specifically, it was demonstrated that the confidence value derived from both the beta-based model and subjective logic theory tends to increase even when unsuccessful communications rise. To effectively validate a



(a) Expectation of beta distribution.

(b) Expectation of subjective logic.

Fig. 7.2. Confidence values for varying interactions as a function of expectations.

recommender, confidence should be linked exclusively to positive behaviour, i.e., successful interactions.

From the perspective of similarity between two UNs, confidence can be defined in terms of expected behaviour over time. A node that consistently engages in successful interactions is more likely to maintain similar behaviour in the future. Considering the two confidence measurement approaches previously discussed in the literature, where confidence is quantified based on communication outcomes, these measures can instead be adapted to incorporate behavioural expectations. Based on this refined definition of confidence, it can be evaluated using the beta-based trust as follows:

$$c_{ik} = \frac{\alpha_{ik}}{(\alpha_{ik} + \beta_{ik})}. \tag{7.3}$$

Here, the expectation of node k's behaviour is represented by the mean of the beta-based model. Similarly, confidence can also be derived from subjective logic using both the belief (b) and the uncertainty (u) as:

$$c_{ik} = b_{ik} + u_{ik} r. \tag{7.4}$$

In this equation, confidence in a node is determined by the belief formed through successful interactions, combined with an uncertainty measure.

The results of both models, utilising Equation 7.3 and Equation 7.4, are illustrated in Figure 7.2. In Figure 7.2a, unlike previous observations in Figure 5.16a, confidence increases with more positive interactions and decreases with more negative interactions, aligning well with the established definition. A similar trend is observed in Figure 7.2b, where the colour gradient illustrates a more linear decline in confidence as unsuccessful packets increase. Additionally, in this model, confidence exhibits a gradient increase with the accumulation of successful interactions, with low confidence assigned for initial interaction. In this study, confidence between nodes is defined based on the consistency of successful interactions, following the approach presented in Equation (7.4).

7.3.2.2 Familiarity as a Measure of Similarity

While the defined confidence in the previous section measures the degree of consistency by taking into account the uncertainty within a continuous interaction between two entities, it does not capture the depth of interactions compared to other neighbours. In a collaborative network, it is meaningful to expand the concept of similarities to include the behaviour of neighbours as a form of familiarity. Therefore, the notion of similarity is further expanded to define the concept of familiarity as a measure of closeness between nodes, determined by the duration of adjacency between the target node and the recommended node. This concept emphasises that nodes should give more weight to suggestions from neighbours with whom they have a long-term relationship, as opposed to those with a short-term association. This approach to defining relational familiarity highlights the importance of the frequency and duration of interactions in evaluating the closeness between different nodes, as noted in (Zhang et al. 2023):

$$f_{ik} = \frac{S_i^k}{S_i^K} \tau^{\frac{1}{S_i^k}}, \tag{7.5}$$

where S_i^k is the number of successful communications between i and k, and S_i^K is the number of successful communications i has with all K neighbours. $\tau \in [0, 1]$ is the adjustment factor of the number of communications.

7.3.2.3 Similarity Computation

In this model, the concept of similarity is articulated based on two aspects: confidence based on subjective logic as well as familiarity among neighbours. Both variables allow for a balanced consideration of both abstract trust (based on subjective logic) and empirical evidence (based on direct interactions), and they are integrated using a weighted sum as follows:

$$s_{ik(t)} = wf_{ik} + (1 - w)c_{ik}, \tag{7.6}$$

where $s_{ik}(t)$ represents the similarity score between i and k at time t, f_{ik} denotes the familiarity score between the nodes, c_{ik} is the confidence score derived from subjective logic, and $w \in [0, 1]$ is the weight assigned to the familiarity and confidence scores.

7.3.3 Belief Process

Based on the assumptions highlighted in Section 7.2.2, the belief function is introduced, as outlined in Algorithm 2, to evaluate the credibility of recommendations by integrating several key factors: freshness, similarity, trustworthiness, and decay. Each factor contributes to the overall belief score, which is used to assess the belief in recommendations.

The freshness factor, Γ_{kj}, emphasises the importance of the recency of the recommendation. More recent recommendations are given a higher weight because they reflect the most up-to-date interactions and behaviours. In a dynamic environment where entities continuously evolve, older recommendations may no longer accurately represent the current trustworthiness or behaviour of a trustee. For instance, if a trustee's behaviour has significantly changed over time—whether improving or deteriorating—prioritising fresh recommendations prevents outdated assessments from skewing the trust evaluation.

Mathematically, the interval between conducting the opinion and sending the recommendations adheres to a power-law distribution as stated by Mahmood et al. (2023), and then it can be expressed as:

$$\Gamma_{kj} = \eta_{sc}(t - t_{T_{kj}})^{-\varepsilon}, \tag{7.7}$$

where η_{sc} is a predetermined scaling constant, and ε is the exponent of the power law. This formula is used to underscore the significance of the recency and freshness of the trust evaluation made by k.

Another factor influencing the weight of a recommendation is the decay time of the trustor's opinion on the trustee. The decay factor, $d_{T(ij)}$, accounts for the gradual reduction in the trust score T_{ij} stored by node i over time. This process reflects how the persistence of past trust assessments influences the acceptance of new recommendations. The decay factor helps determine the extent to which a trustor incorporates new opinions into its trust evaluation. If the trustor has maintained trust in a trustee for an extended period, it may be less inclined to adjust its trust level based on new recommendations. In contrast, if the previously established trust has diminished over time, the trustor is more likely to consider and integrate recent recommendations into its assessment. The decay is computed using the technique provided by MATMU.

The overall belief score, $B_{ij \leftarrow k}$, is computed by integrating the aforementioned factors as:

```
Algorithm 2 Belief Process
Input: List of Received Recommendations $R_j(t) = \{T_{k_1 j}, T_{k_2 j}, \ldots, T_{k_n j}\}$
Output: $B_{ij \leftarrow k}$ of each Recommender $k$
for Each $T_{kj}$ in $R_j(t)$ do
    Computes $s_{ik(t)}$
    Compute $\Gamma_{kj}$
    Compute $d_{T(ij)}$
    Find $T_{ik}$
    Compute $B_{ij \leftarrow k}$ (Equation (7.8))
```

$$B_{ij \leftarrow k} = \frac{s_{ik(t)}\Gamma_{kj}(1 - d_{T(ij)})T_{ki(t)}}{\sum_{k'}^{K}\left(s_{ik'(t)}\Gamma_{k'j}(1 - d_{T(ij)})T_{k'i(t)}\right)}, k \neq k'. \tag{7.8}$$

Here $s_{ik(t)}$ is the similarity factor, Γ_{kj} is the freshness factor, $d_{T(ij)}$ is the decay factor for the trust score that the trustor i has towards the trustee j, and $T_{ki(t)}$ is the trustworthiness of the recommender k at time t.

7.3.4 Penalty Process

The penalty process aims to adjust the trust value of a node based on its history of violations. The trust value for node k obtained by i at time t, denoted as $T_{ik}(t)$, is subject to a penalty factor that diminishes the trust based on the number of violations recorded for the node. This adjustment is represented by the following:

$$T_{ki} = T_{ki}(t)\left(1 - \frac{1}{\exp^{\lambda}}\right). \tag{7.9}$$

In this equation, λ represents the number of violations where the node has been alarmed during the recommendation evaluation process. The penalty factor, $\left(1 - \frac{1}{\exp^{\lambda}}\right)$, decreases the trust value exponentially as the number of violations increases. This model ensures that each additional violation results in a progressively smaller decrement in the trust value, reflecting an exponential penalty mechanism, as shown in Figure 7.3.

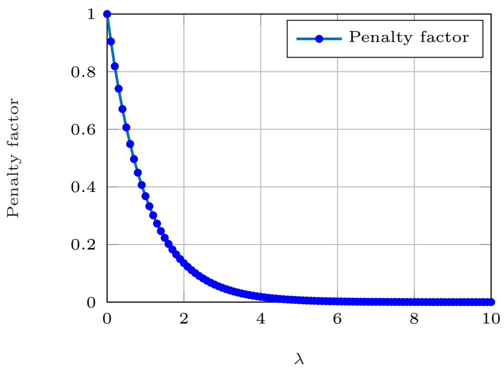


Fig. 7.3. Penalty factor of trust as a function of λ.

Violations are determined based on two criteria: the count of appearances in the initial filter and the belief score of the nodes. Penalties are applied if a node is detected in the filtering process and has a belief score lower than the average belief score of all recommenders. This method ensures that nodes frequently flagged by the filter and with relatively low belief scores are appropriately penalised.

7.3.5 Recommendation Evaluation Workflow

Algorithm 3 outlines the process of the recommendation evaluation. Upon receiving the list of recommendations R_j, which contains all recommendations received from K nodes regarding j, node i initiates the process of evaluating the received recommendations.

The initial filtering mechanism begins by applying the outlier detection function, `isOutlier`, to identify potential outliers, which are then stored in S_{1j}. Only those identified as outliers are further examined against the previously stored trust score that node i has for node j, denoted as T_{ij}, while considering the duration for which this trust has been maintained using $d_{T(ij)}$. At this stage, a counter $c_k(t)$ is maintained to record the number of times node k has been detected in both lists as a result of the initial filtering phase. The belief function

Algorithm 3 Recommendation Evaluation Process

```
**Input:** List of Received Recommendations $R_j = \{T_{k_1j}, T_{k_2j}, \ldots, T_{k_Kj}\}$, Trust deviation threshold ($\delta_h$), Violation threshold ($v_h$), Counter ($c_k$)
**Output:** $B_{ij\leftarrow k}$, $T_{ki}$, $c_k$
for Each $T_{kj(t)}$ in $R_j(t)$ do
    if $T_{kj(t)}$ isOutlier then
        $S_{1j} = S_{1j} \cup \{k_j\}$
for Each $T_{kj}$ in $S_{1j}$ do
    if $|T_{ij(t)} - T_{kj(t)}|(1 - d_{T(ij)}) > \delta_h$ then
        $S_{2j} = S_{2j} \cup \{k_j\}$
        $c_k = c_k + 1$
Evaluates the Belief Function(Algorithm 2)
if $c_k \geq v_h$ and $B_{ij\leftarrow k} < \frac{1}{|K'|}\sum_{k'\in K} B_{ij\leftarrow k'}$ then
    Penalty on $T_{ki}$, based on ( Equation 7.9)
Evaluate Indirect Trust Score $T^r_{ij}$ ( Equation 7.1)
```

is then applied to all nodes in the initial set R_j, following the procedure outlined in Algorithm 2.

Next, a penalty assessment is conducted based on the condition that a node has been repeatedly flagged during the initial filtering phase beyond a specified violation threshold and, at the same time, has a belief score lower than the average belief score computed across all recommendations. Nodes meeting these conditions are subjected to a penalty, leading to a reduction in their trust score.

7.4 Conceptual Analysis

In this section, the performance of the recommendation evaluation model proposed in this chapter is evaluated using CFFTM presented by Zhang et al. (2023) as a benchmark. For clarity, the notation has been standardised, with trustor (i), trustee (j), and recommender (k) following the conventions outlined in Section 3.2.3.

Property 1: Rejecting Recommendations from Malicious Recommenders and Accepting Recommendations from Honest Recommenders

The capability of node i to reject recommendations from malicious nodes and accept those from honest nodes is examined.

Claim 1: The proposed recommendation model can reject recommendations from malicious nodes.

Rationale: CFFTM employs several conditions to process collected recommendations. However, it is demonstrated that node i might inadvertently accept recommendations from malicious node k. In the CFFTM model, each recommendation is represented as T_{kj}, accompanied by two additional values, $Link_{k \to j}$ and Com_{jk}. Here, Com_{jk} denotes the communication status between the recommender node and the trustee node, while $Link_{k \to j}$ signifies the link status when the trust score is obtained. These values are gathered by the recommender and included in the packet sent to the trustor, who will use them to assess the recommendation.

The CFFTM model begins by filtering unreliable recommendations from the original trust list, such that for each recommendation T_{kj}, unreliable values are identified as those where $|T_{kj} - T_{ij}| > \delta_h$. Subsequently, the model assesses the link quality using the $Link_{k \to j}$ data provided by the recommender about its interaction with trustee j. If the link state $Link_{k \to j}$ is below a certain threshold, the recommendation is deemed erroneous due to a link error. Otherwise, it is considered dishonest. A new list of recommendations is then generated, excluding both erroneous and dishonest recommendations.

CFFTM is effective in rejecting dishonest recommendations as long as T_{ij} remains current and up to date. However, since the values of δ_h and the link state threshold are not specified within CFFTM, relying solely on the deviation test could still lead to the acceptance of dishonest recommendations. This limitation might restrict the model's applicability to specific situations, depending on the density and environmental conditions of the underwater network. One example would be when a significant amount of time has passed since the direct interaction between i and j. Due to the decay property of trust, the stored trust will also decrease. Malicious nodes, aware of this, can engage in BM on j, distributing lower scores, which are then accepted by CFFTM. Since T_{ij} degrades over time due to the sliding time window with an exponential decay mechanism, the model may end up filtering all honest recommendations as dishonest if there is no interaction in the current period. For example, at a previous time t, node i estimates the trust score towards node j to be 0.9. Without interactions for subsequent periods, the trust score decays with a factor $\alpha = 0.8$ over a window size of 5. The trust score decays as follows: 0.9, 0.72, 0.576, 0.4608, 0.3686. A dishonest recommender k propagates $T_{kj} = 0.2$ (BM j where the actual T_{kj} should be > 0.5). In this scenario, i will accept k's recommendation given that $|T_{kj} - T_{ij}| \leq \delta_h$, leading the model to accept the malicious recommendation.

To address this issue, the proposed recommendation evaluation model in this chapter employs multiple metrics to collectively determine the weight of each recommendation rather than rejecting them outright. Dishonest recommendations are identified by obtaining lower weights, rendering them ineffective. This is achieved by utilising the proposed initial filtering method and the belief function to enhance the weighting mechanisms affecting the evaluation of trust. Initially, the prevailing opinion is prioritised to filter out potential malicious outliers. However, due to the complexity and unpredictability of underwater communication, where an honest recommender might be treated as an outlier, these outliers are further scrutinised against individual opinions obtained by node i. While this approach alone may still allow some advanced collaborative attacks to succeed, the belief method acts as an additional layer of filtering, serving as a secondary protection against dishonest recommendations that bypass the proposed initial filtering method. This method adjusts the weight of recommendations by considering several factors that collectively work to reduce the influence of untrustworthy sources, outdated stored opinions, and nodes with little similarity to the trustor over time. The similarity factor ensures that recommendations from nodes with historically similar behaviour to the trustor are given more weight, thereby prioritising reliable sources. If a malicious node has a history of interactions that diverge significantly from the trustor's interactions, the similarity factor between them is low, and recommendations from that node will carry less weight. The freshness factor prioritises recent interactions, diminishing the influence of outdated recommendations, which is critical in dynamic environments where trust levels fluctuate over time. If a malicious node provided a recommendation in the past but has not interacted recently, the freshness factor will lower the weight of that recommendation. Finally, the inherent trustworthiness of the recommender node at the current time, influenced by historical behaviour, ensures that consistently honest nodes are trusted more, while those with a history of dishonesty are penalised. If a node has repeatedly provided false recommendations, its current trustworthiness will be low, making its recommendations less influential.

Claim 2: The proposed recommendation model can accept recommendations from honest nodes.

Rationale: Both CFFTM and the proposed model succeed in accepting recommendations from honest recommenders if the nodes have a valid and current personal opinion on j. However, if T_{ij} is outdated and new information becomes available, CFFTM might misclassify honest recommendations as dishonest. Suppose $T_{ij} = 0.4$ and all neighbours send recommendations for j in the range of 0.8 to 0.9. Since the deviation test is the primary filter and if $\delta_h \leq 0.3$, then all honest nodes will be deemed dishonest, and their recommendations will be rejected.

With the proposed filtering method, outliers among recommendations are first identified, followed by the application of the proposed deviation test. If honest recommendations constitute the majority (most recommendations fall between 0.8 and 0.9), they are accepted according to the majority rule. If dishonest recommendations dominate, hence, T_{kj} is flagged as an outlier, then the subsequent deviation test ensures fair filtering, favouring personal experience while considering the decay in Equation 7.2. Moreover, with the belief function, the weighting is further scaled based on the behaviour of recommenders, as described in the previous claim.

Property 2: Punishment for Malicious Recommendations and Avoiding Punishment for Accurate Recommendations

The capability of node i to correctly punish k for engaging in dishonest recommendation acts is examined.

Claim 3: The proposed recommendation model correctly punishes dishonest recommenders and avoids unnecessary punishment for accurate recommendations.

Rationale: In the CFFTM model, the $Link_{k \to j}$ is used to filter recommendations, identifying either dishonest recommenders or those struggling with communication. The model then identifies dishonest recommendations based on a threshold, such that those with good link quality yet providing recommendations that deviate from personal trust will be treated as malicious and punished accordingly. This approach can lead to inaccurate penalties since the recommenders provide the $Link_{k \to j}$ and may forge these values to avoid punishment. It is argued that this decision should rely exclusively on the information obtained by i about k. Therefore, the proposed model employs a threshold of a number of violations based on the initial filter as well as evidence of lower belief compared to others to penalise a node. This punishment is reflected in the trust score towards k,

Table 7.1: Simulation Parameters for Recommendation Evaluation.

Attribute	**Value**
Simulation Time	1800 seconds
Number of Nodes	20
Attack Rate %	10%-50%
Scaling constant (η_{sc})	40
Power law's exponent (ε)	1.05 (Huang et al. 2017)
Violation Threshold (v_h)	2
Recommendation request	each 60 seconds

addressing the issue illustrated in Figure 5.14, where a dishonest recommender could continue being perceived as trustworthy despite compromising the accuracy of the TMM through the propagation of dishonest recommendations. Revisiting the previous examples of discrepancies in the recommendations, it is observed that an honest recommender might be flagged as dishonest despite having an accurate and good $Link_{k \to j}$. This misclassification would result in an honest node being unfairly punished as a malicious recommender.

7.5 Experiments and Performance Evaluation

In this section, the methodology and experimental setup employed to rigorously evaluate the performance of the proposed recommendation evaluation process within an underwater network environment are outlined. Within this trust-based model, the MATMU model is employed to compute the direct trust of each node. The proposed recommendation evaluation techniques were applied to assess indirect trust based on 1-hop recommendations. In all the experiments, the threshold for violation was set at 2, as it was found to be reasonable and improved the model's accuracy. Details of the simulation parameters and specific configurations are presented in Table 7.1.

Each trial examined the effectiveness of the proposed recommendation evaluation model, varying the percentage of malicious recommendations from 0% to 50%, where 0% indicated no attacks. The trust establishment process is analysed under both BM and BS recommendation attacks, where a proportion of neighbouring nodes propagate dishonest recommendations. Specifically, when BS attacks are examined, malicious nodes are also introduced into the network and exhibit one of the attack behaviours previously explored in Section 4.4. The dishonest recommenders involved in BS attacks issue misleading recommendations to boost the trustworthiness of these malicious nodes by increasing their trust scores. On the other hand, when BM attacks are examined, all nodes in the network behave fairly, except for the dishonest nodes intending to perform BM attacks. These nodes engage in bad-mouthing honest nodes to undermine their trustworthiness. In each test, the initial trust evaluation process begins at 40 seconds into the simulation and is updated periodically every 60 seconds.

The evaluation methodology is designed to assess the effectiveness, attack resistance, and comparative performance of the proposed approach. Initially, the emphasis was on evaluating the feasibility of the recommendation evaluation process in detecting misbehaving nodes, which was essential for establishing the baseline performance of the model. Subsequently, the attack resistance of the network was examined by simulating scenarios both with and without the proposed recommendation filtering method. This analysis was reasonable for demonstrating robustness in mitigating malicious behaviour. Finally, a comparative analysis against existing methods was conducted to evaluate the relative performance of the proposed approach. This comparison was justified by the need to benchmark the proposed method against current models, thereby providing a clear understanding of its advantages and limitations.

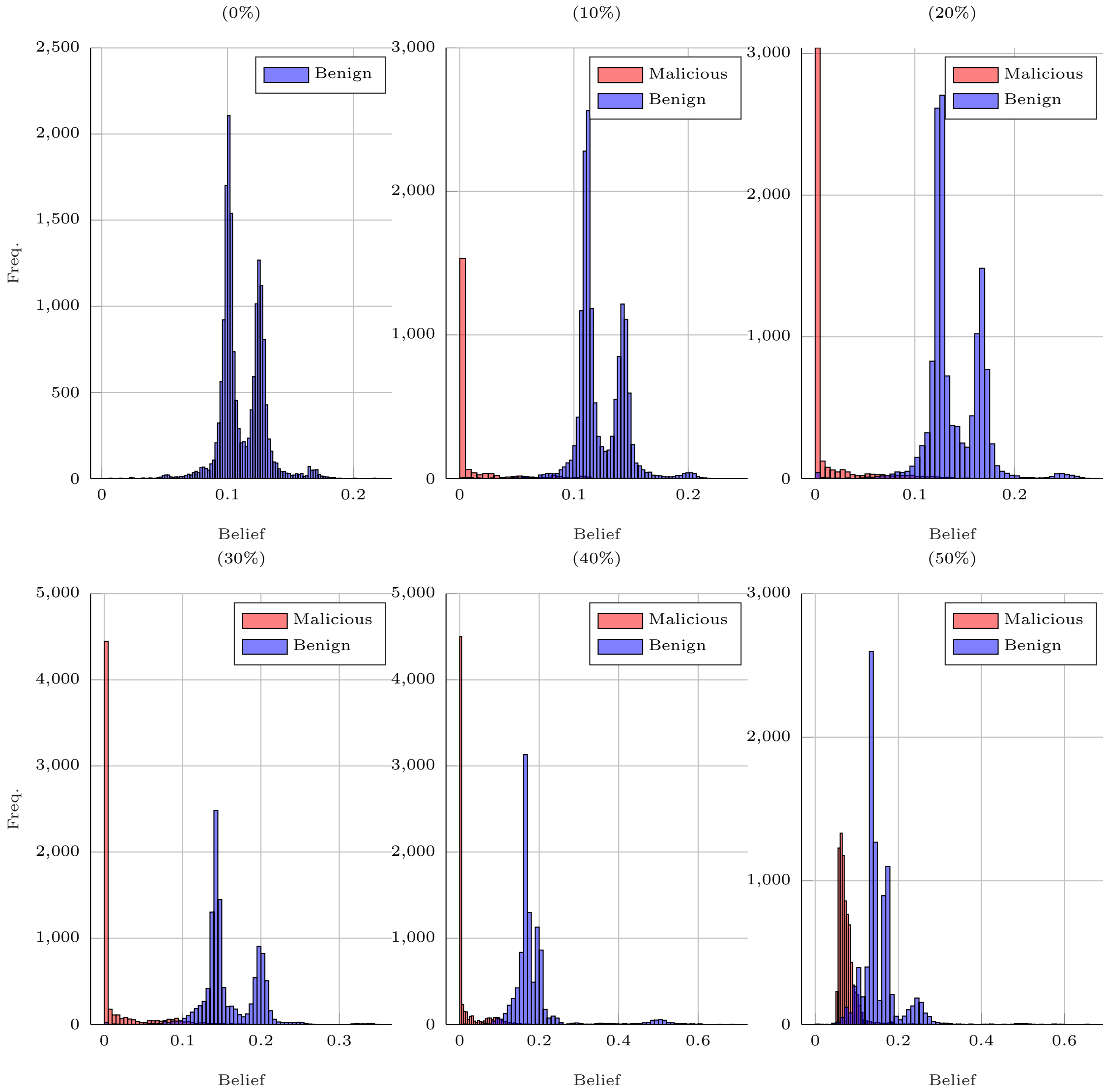


Fig. 7.4. Evaluation of belief ($B_{ij\leftarrow k}$) over varied proportion of BS attack.

7.5.1 Effectiveness Evaluation of the Belief Method

The impact of different attack percentages on the belief score ($B_{ij\leftarrow k}$) obtained in the proposed model among benign and malicious nodes is examined. The belief scores of the nodes were measured, and the results are presented in Figure 7.4 and Figure 7.5, which represent the frequency distribution of obtained belief scores for both benign and malicious nodes across all experiments under BS and BM, respectively. Figure 7.4 shows how frequently each belief value appears among the nodes, effectively illustrating the impact of varying attack intensities on the belief assignments to benign and malicious nodes. At 0% attack, only benign nodes are present. The histogram for benign nodes (shown in blue) is relatively narrow, indicating a tight distribution around the mean belief. The belief score obtained here reflects the weight assigned by the model to each recommender among the received recommendations, with the variation reflecting the number of recommendations received by nodes and the associated weight assigned to them. This suggests that overall, benign nodes have consistent belief scores under normal conditions, with slight variations depending on the number of recommendations and the factors considered in the computation of the belief model, such as similarity, trustworthiness, and freshness.

As the attack percentage increases from 10% to 50%, the graph displays both benign (blue) and malicious (red) nodes. There is a noticeable shift in the belief assigned to malicious nodes compared to benign ones, with malicious nodes typically receiving lower belief values. This indicates that the proposed model is capturing their altered and potentially disruptive behaviour. The distinction between the blue and red distributions diminishes as the attack percentage rises, highlighting the increasing influence of malicious actions. In the worst-case scenario, with 50% of nodes being malicious, the proposed model reduces the belief scores for the malicious nodes but still faces difficulty in fully separating the distributions compared to less intensive attacks. This challenge arises because when half of the network consists of malicious nodes, their collective influence becomes significant enough to obscure the clear distinction between benign and malicious behaviour.

Similarly, Figure 7.5 illustrates the belief scores acquired during BM attacks. The proposed model effectively recognises malicious nodes by lowering their belief scores. However, its performance declines in scenarios where half of the recommendations (50%) are malicious, though it still exhibits slight resistance favouring benign nodes. This behaviour is expected, as malicious recommendations attempt

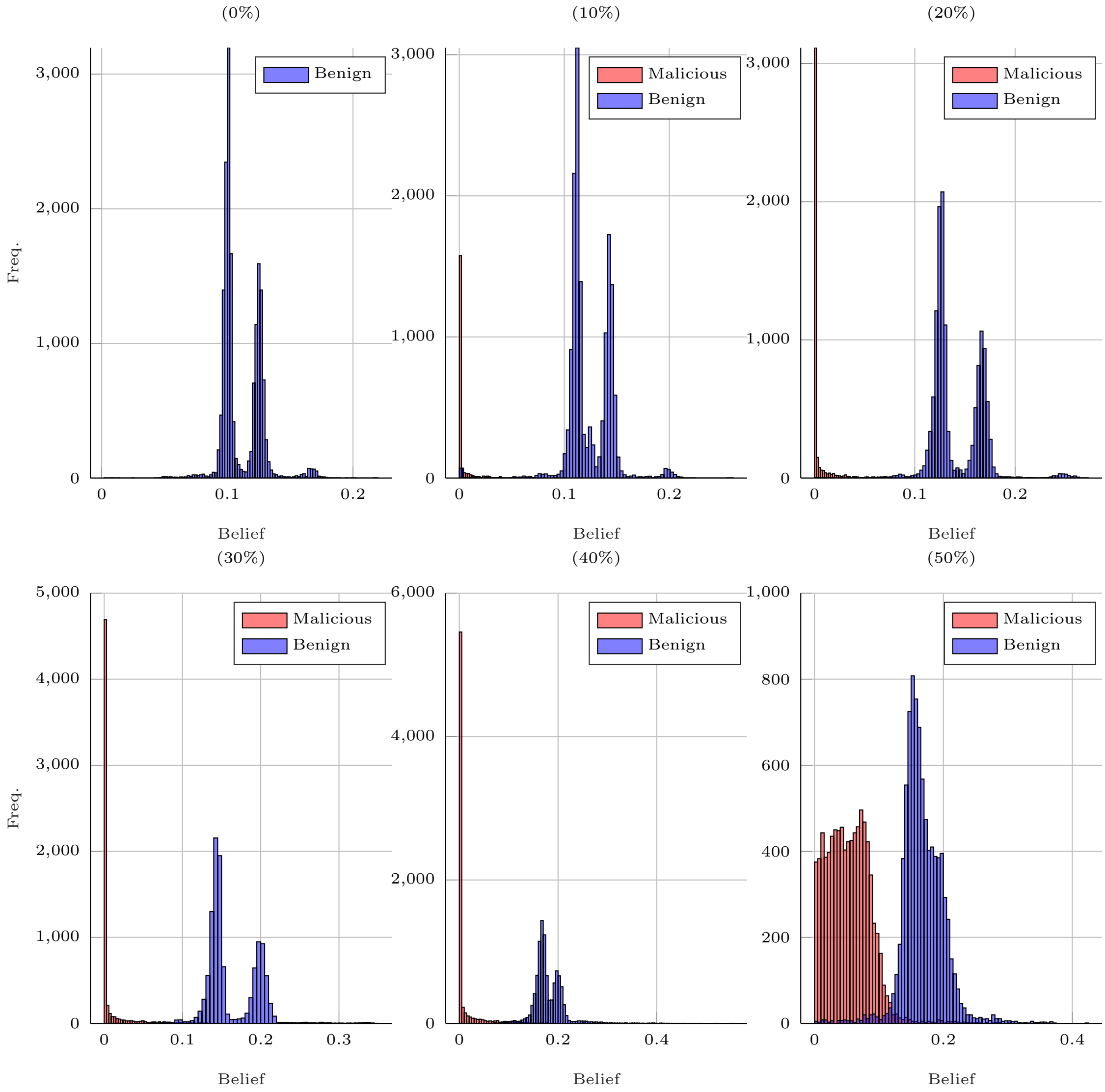
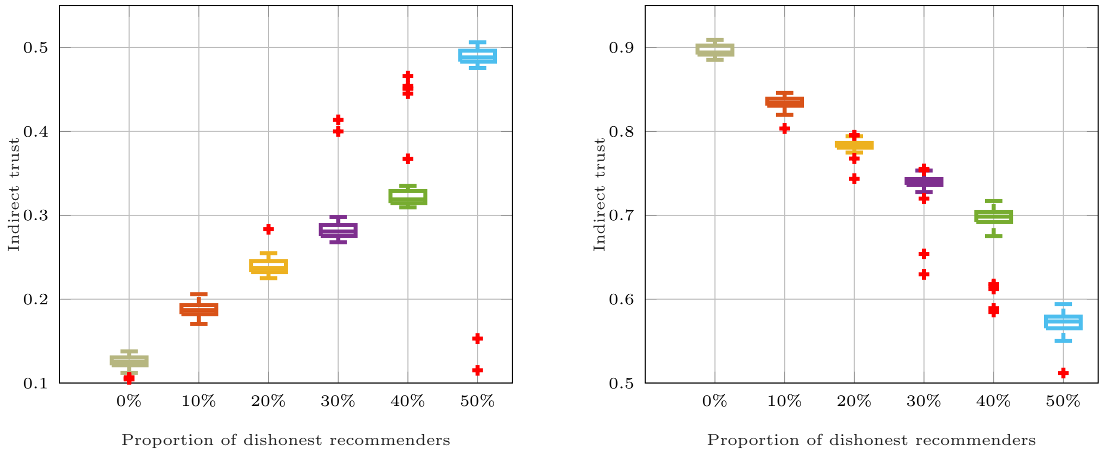


Fig. 7.5. Evaluation of $B_{ij\leftarrow k}$ over varied proportion of BM attack.

to mask their deceptive acts with seemingly normal behaviour, thereby making it more challenging for the model to accurately distinguish between malicious and benign nodes under such conditions.

7.5.2 Effectiveness Evaluation of the Indirect Trust

The indirect trust obtained toward targeted nodes by each node in the network is assessed following Equation 7.1. Figure 7.6 shows the resulting indirect trust across 20 distinct network topologies and varying attack scenarios. For each



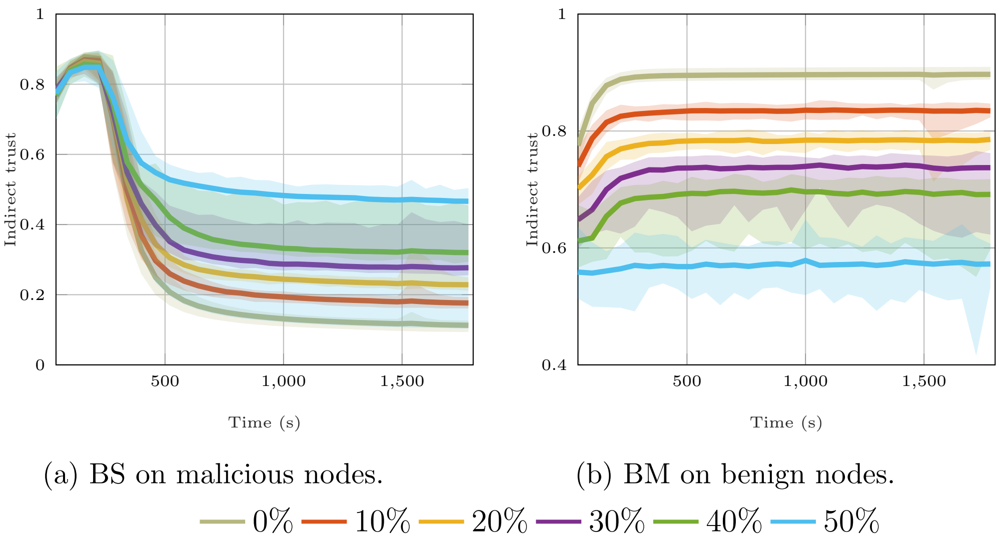
(a) BS on malicious nodes.

(b) BM on benign nodes.

Fig. 7.6. Evaluation of the obtained indirect trust over trails (for each percentage, there are 115 data points (mean values), resulting in a total of 20,550 individual indirect trust values per attack type).

attack, the indirect trust values are collected, and their distribution is assessed through quartiles (Q1, median, and Q3), highlighting the central tendency and variability. With no attacks, the proposed evaluation of recommendation methods shows a fair weighting of recommendations in both types of attacks (high trust scores for benign nodes exhibited in Figure 7.6b during BM attacks and low trust scores for malicious nodes being promoted in Figure 7.6a during BS attacks). It also shows resistance to attacks during the increase in the attack percentage. Some outliers were observed among the trials, primarily attributed to changes in the environment and simulation settings.

Figure 7.7 shows indirect trust evaluations under dishonest recommendation attacks over time. Within each graph, the line plot aggregates the average indirect trust scores from all the trials, providing a packed view of how trust in nodes is affected over time during each specific attack. The shaded area in these graphs encapsulates the spread of the scores (from min to max) across all nodes where different percentages of attacks are employed. In Figure 7.7a, the indirect trust score for malicious nodes that refuse to collaborate after consuming 10% of their energy is assessed. Initially, the indirect trust score is high, indicating appropri-



(a) BS on malicious nodes.

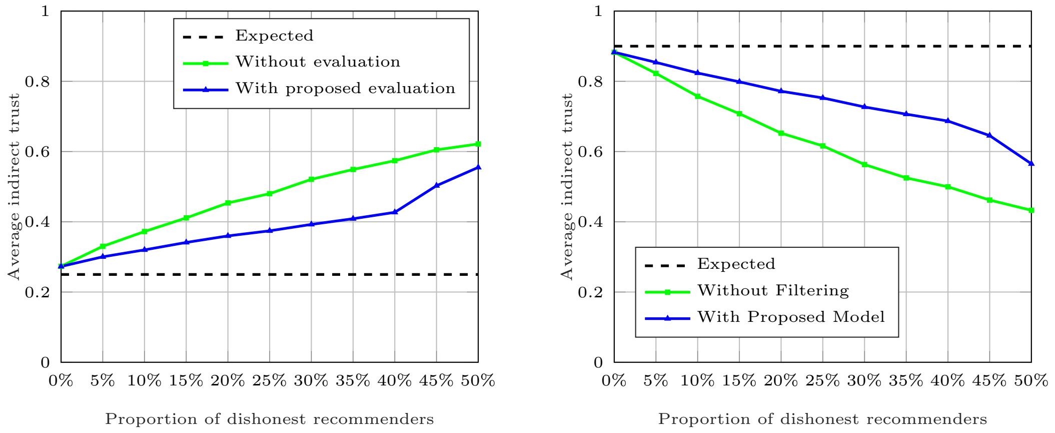
(b) BM on benign nodes.

Fig. 7.7. Evaluation of indirect trust obtained over time (solid line=average, shaded area=range from min to max across all nodes).

ate behaviour from the targeted nodes. Upon misbehaviour, in the case of 0% dishonest recommendation attacks, it is observed that there is a general decline in the indirect trust score of these malicious nodes. When dishonest recommendations are introduced, the proposed evaluation model demonstrates accuracy by further reducing the indirect trust scores and maintaining stability over time, even as the attack persists. A similar observation can be obtained in Figure 7.7b, which shows the indirect trust toward benign nodes that are facing BM attacks.

7.5.3 Attack Resistance

To evaluate the effectiveness of the proposed model, three distinct variables are examined, as shown in Figure 7.8. The first variable is the anticipated value, which signifies the recommended trust level of a node assuming no unreliable recommendations in the network. This ideal scenario assumes a high trust score (0.9) for normal nodes and a low trust score (0.25) for malicious nodes. The second variable is the computed recommended trust level using the defensive recommendation evaluation process proposed in this chapter. The third parameter



(a) BS on Malicious Nodes.

(b) BM on benign nodes.

Fig. 7.8. Evaluation of average indirect trust values obtained for nodes under attack.

is the average recommended trust level without employing any validation on the received recommendations. Specifically, within the third parameter, the indirect trust is computed using equal weights for each recommendation received.

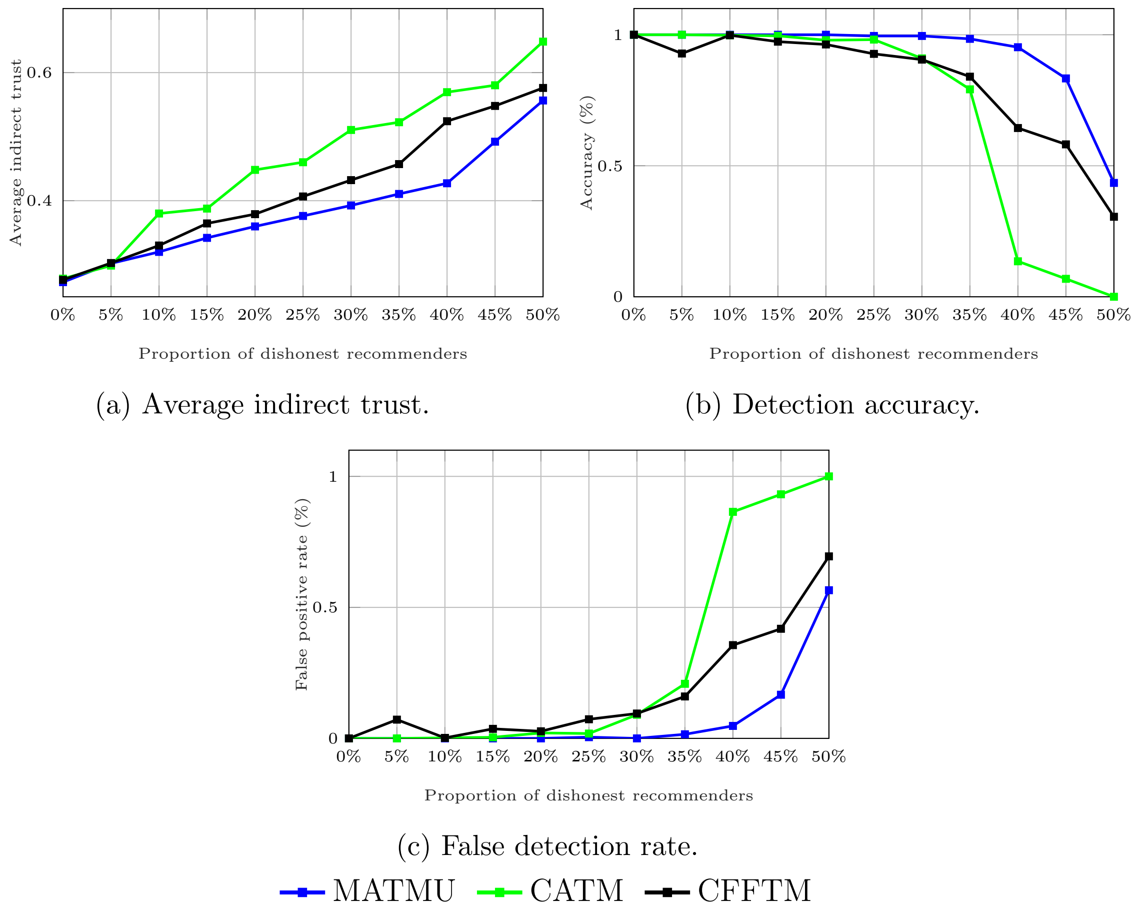
As depicted in Figure 7.8b, during BM attacks, the trustworthiness of normal nodes declines as the number of malicious nodes increases. Similarly, in BS attacks, shown in Figure 7.8a, the trust value of malicious nodes increases with their numbers, but the changes within the proposed method remain manageable. This is because, without a filtering mechanism, each recommendation node is assigned the same weight, leading evaluating nodes to accept all propagated recommendation values indiscriminately. Consequently, as the proportion of malicious nodes increases, the dissemination of false information in the recommendation sequence escalates, significantly impacting the final trust value. In contrast, the recommendation evaluation method discussed in this chapter adjusts recommendation trust values based on various attributes after filtering the recommendation received.

7.5.4 Comparison with Similar Works

To assess the effectiveness of the proposed model against comparable models in the field, CATM and CFFTM were selected, both defined as TMM with recommendation defence in underwater networks. To the best of the author's knowledge, these two models are the most recent and thus provide a valid basis for comparison. The results of this analysis are shown in Figure 7.9 and Figure 7.10, representing the evaluation under BS and BM attacks, respectively. The testing metrics are defined as follows:

1. Accuracy rate: the number of malicious nodes detected as a percentage of the total number of malicious nodes.
2. False detection rate: the number of misidentified normal nodes as a percentage of the total number of normal nodes.

First, the average indirect trust of nodes under dishonest recommendation attacks is evaluated. The aim is to examine the average recommendation trust under varying proportions of dishonest recommenders. All models exhibit some resistance to dishonest recommendations when the proportion of malicious recommendations is small. However, as the proportion of malicious nodes increases, the performance of each model fluctuates in terms of resistance to dishonest recommendations. Specifically, in Figure 7.9a, under BS attacks, CATM shows less resistance as the percentage of dishonest recommenders in the network increases. This is because CATM's validation process focuses on the performance of recommenders without accounting for the deception of recommender nodes. Conversely, CFFTM demonstrates high resistance to the attack due to its filtering technique, as explained earlier. The proposed model provides more accurate detection of attacks, even as the percentage of malicious nodes increases. Figure 7.10a illustrates the resulting average indirect trust under BM attacks in the same setting,



(a) Average indirect trust.

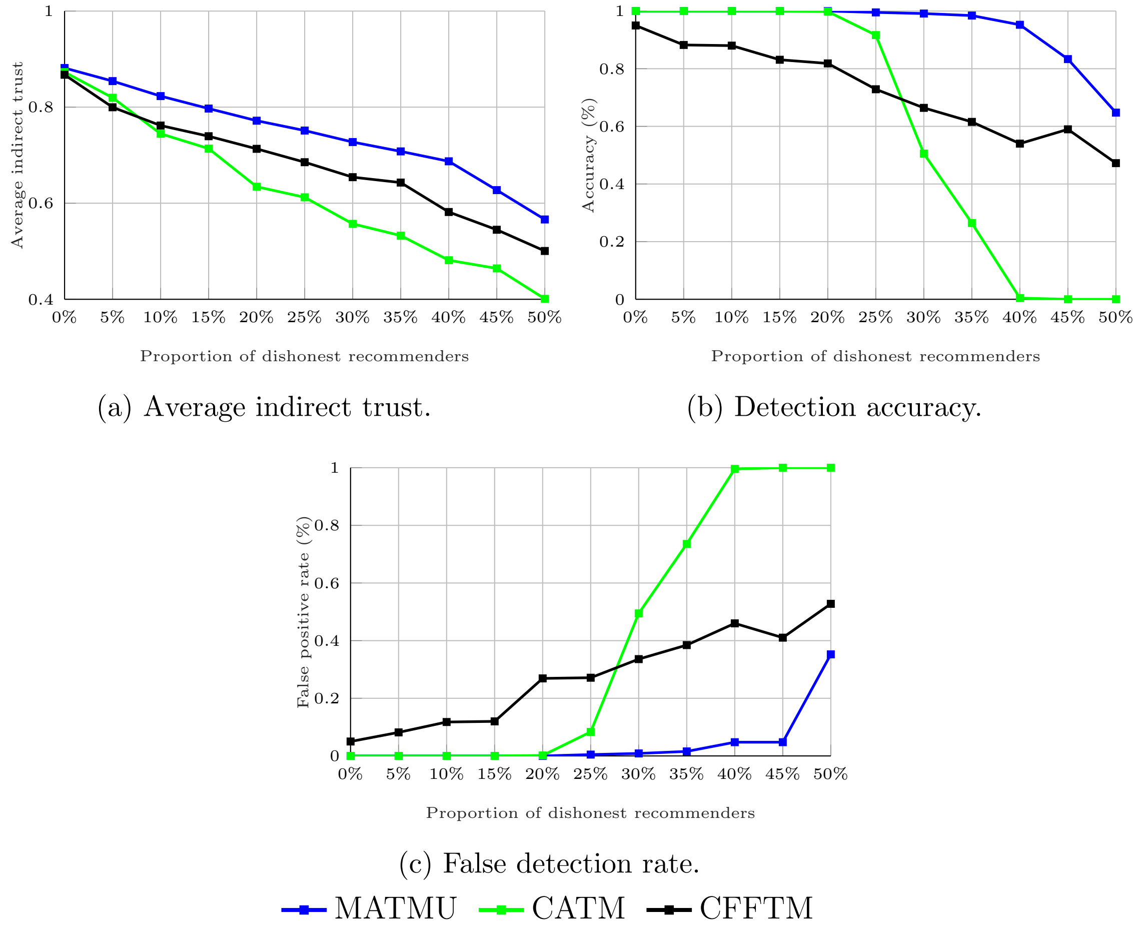
(b) Detection accuracy.

(c) False detection rate.

Fig. 7.9. Analysis of effectiveness under BS attack.

where the nodes under attack are promoted as malicious. The proposed model excels in detecting dishonest recommendations, resulting in consistently high trust toward the evaluated nodes. As expected, an increase in the number of malicious nodes challenges all models' ability to detect dishonest recommendations, yet the proposed model remains within a manageable range.

The accuracy and false detection of each model as a function of the varied proportions of malicious recommendations are evaluated. The proposed model demonstrates the highest resilience, maintaining near-perfect accuracy against up to 30% attack intensity for both BS (Figure 7.9b) and BM attacks(Figure 7.10b). Beyond this point, there is a gradual decline, but the model still outperforms others significantly. In contrast, the CFFTM model, while initially accurate, shows a significant drop in performance even at lower levels of attack intensity.



(a) Average indirect trust.

(b) Detection accuracy.

(c) False detection rate.

Fig. 7.10. Analysis of effectiveness under BM attack.

This is mainly attributed to the high false positive rate, where benign nodes are incorrectly considered malicious, as shown in Figure 7.9c and Figure 7.10c.

7.6 Conclusion

This chapter introduced a novel recommendation evaluation mechanism designed to detect and mitigate dishonest recommendations within the context of trust establishment among IoUT nodes. The proposed approach provided a method for validating received recommendations, thereby minimising the influence of fraudulent or malicious recommendations on the overall trust estimation process. The mechanism employed two key methods: an initial filtering mechanism and belief estimation. The first component of the proposed mechanism was a detection and filtering process that integrated outlier detection techniques to identify recommendations that significantly deviated from normative ones, along with deviation-based mechanisms that prioritised individual experiences. This dual-layer filtering strategy demonstrated higher accuracy in identifying and eliminating dishonest recommendations, as confirmed through primary evaluations (Section 5.4.4). The second component involved a belief-based weighting function that assigned importance to recommendations based on several social validation factors, driven by initial assumptions about what constituted an acceptable recommendation or appeared suspicious. These factors included the freshness of recommendations, which ensured relevance by prioritising recent opinions, and similarity, which favoured inputs from nodes with comparable characteristics or histories. Additionally, the trustworthiness of the recommender is involved, assigning a higher weight to recommendations from established, reliable sources, while trust decay was considered to reduce the influence of outdated trust values on the evaluation and acceptance of new opinions. Together, these mechanisms integrated filtering and weighting processes to construct a comprehensive belief profile that effectively weighed recommendations based on these factors. In other words, the importance of each received recommendation in contributing to indirect trust was determined through both mechanisms. Furthermore, a punishment method was introduced to penalise dishonest recommenders by reducing their associated trust levels.

The proposed evaluation mechanism was rigorously tested under varying proportions of dishonest recommenders who engaged in tactics such as BS and BM attacks. The results demonstrated high detection accuracy within the tested adversarial conditions, with the proposed model successfully identifying malicious nodes with high accuracy when malicious nodes comprised up to 45% of the network. The effectiveness of the proposed model was validated against benchmarked existing models through a series of comparative tests. The findings highlighted the promising performance of the recommendation evaluation mechanism, particularly in its ability to detect dishonest recommendations.

Chapter 8

Conclusions and Future Work

This chapter concludes the thesis and reflects upon the research and its contributions. It begins with a synthesis of the findings, offering a chapter-by-chapter reflection on the results in Section 8.1, then revisits the research problem statements and objectives and discusses how each has been addressed within the research in Section 8.2. Section 8.3 examines the limitations of the study and outlines potential directions for future work. Finally, Section 8.4 highlights the broader applicability of the proposed methodologies.

8.1 Synthesis of the Key Findings

Recognising the critical role of security in modern infrastructures, this thesis explored trust as a fundamental security mechanism within the distinct challenges posed by underwater networks and the IoUT. The inherently open nature of IoUT, combined with the high costs associated with regular monitoring and maintenance, renders these networks particularly susceptible to undetected hijacking and attacks. While existing research on trust-based attack detection has extensively addressed terrestrial networks such as IoT, MANET, and Ad Hoc, these solutions cannot be directly applied to IoUT due to the unique characteristics and constraints of underwater environments. The dynamic nature of underwater network topologies, influenced by water currents and motion, adds influential complexity. Additionally, limited resources, high energy demands, and challenging communication frameworks necessitate solutions that prioritise network performance. To advance the field, this research focuses on improving TMM in IoUT by analysing network structures, evaluating existing TMM, developing a mobility-aware trust model, and proposing robust mechanisms to counteract dishonest recommendations.

To establish a deeper understanding of TMM within IoUT, Chapter 2 and Chapter 3 focused on exploring IoUT and TMM, respectively. Chapter 3 provided foundational knowledge on the relatively novel concept of IoUT, highlighting its unique challenges, particularly the characteristics of underwater networks—sparsity, noisy channel, and constant motion influenced by water currents. These features affect multiple aspects of the network performance, including operability, reliability, and stability, with security standing out as a critical concern, necessitating the introduction of new security requirements, including self-stabilisation, survivability, isolation, and freshness. Building on this foundation, Chapter 3 then examined TMM by leveraging insights from various IT systems, with a particular focus on those developed for Ad Hoc networks. A taxonomy of TMM is presented, mapping current advancements in underwater networks to contextualise this thesis's contributions within the broader landscape of existing research. The extensive review of TMM in MANET, IoT, UWSN, and IoUT informs the identification of requirements essential for feasible TMM solutions. These requirements align with both underwater network security demands and the characteristics of TMM. They are classified into three main categories: platform constraints, utilised parameter engineering, and the persistence and reliability of TMM. Together, these classifications serve as a basis for guiding the development of effective TMM for underwater networks.

Chapter 4 provided essential guidelines for conducting research on underwater networks, with a particular focus on IoUT. Due to the challenges and high costs associated with real-world experiments, reliance on simulations becomes indispensable. However, this reliance often brings a recurring issue identified in the literature, where assumptions about the network and environment limit the feasibility and applicability of proposed solutions. To address this problem, the chapter focused on offering a comprehensive guide for simulating the behaviour of underwater networks and understanding the influence of network structures and water currents. It explained the key requirements for simulation from different perspectives, including the communication channel, propagation model, attenuation model, noise model, and modelling the impact of ocean current dynamics on underwater network topologies. These findings have been pivotal in enhancing existing Aqua-Sim ng, which is highlighted in Appendix C. Additionally, the chapter presented a classification of underwater network topologies based on varying application demands, along with placement strategies that help guide simulations designed to suit these topologies. When tested, the PDR and network lifetime showed alignment with the expectations of underwater networks, which supports the validity of the simulations. These insights establish a basis for designing comprehensive test cases and generating synthetic datasets. The attack models are then explored, encompassing an expanded attack space that extends beyond communication-based misuses such as DoS, SB, and SF. This broader scope includes attacks that exploit the visibility and lack of control within the network, enabling adversaries to manipulate and mislead critical network operations, including physical mobility and communication dynamics among UNs. The PM attack is introduced, which leverages these vulnerabilities, stressing its essence within the expanded attack space.

Chapter 5 proposed an in-depth evaluation of the current state-of-the-art across the pre-established network topologies and attack models introduced in the previous chapter. The chapter seeks to answer a fundamental question: whether the current models of TMM, either developed specifically for IoUT or proposed more broadly within IT systems, can be effectively adopted to meet the defined underwater network needs. This question is addressed through a combination of empirical and analytical studies, where these models are implemented and tested within simulated networks representing diverse topologies. The assessment is conducted in two primary dimensions: direct trust establishment based on multiple metrics and indirect trust mechanisms within trust management models. For the first aspect, the chapter introduced a classification of current TMM approaches of computing the trust into standalone and composite methods. Standalone approaches calculate trust independently for each metric, followed by aggregation, whereas composite approaches treat metrics holistically to derive a single trust score. These two methods provide distinct frameworks for trust computation, enabling a comprehensive evaluation of their suitability for adoption across various network topologies and attack scenarios. Under each class, a benchmark TMM utilised within IoUT is selected for analysis. The evaluation revealed that current models fail to address the diversity of attack variations across different network topology structures. The findings indicate slow convergence in certain scenarios, where the models show a lack of robustness under malicious conditions. In the domain of indirect trust, the chapter proposed a classification of techniques into node-centric and value-centric approaches. Node-centric methods focus on evaluating the reliability of individual recommendation nodes (recommenders), while value-centric methods assess the validity of recommendations received from nodes. The chapter examined methodologies for evaluating recommendations as part of indirect trust mechanisms that leverage the confidence built based on the behaviour of the recommender. When confidence measures are proposed to validate recommenders, such as those based on beta distributions or subjective logic theories, it is observed that these values alone often overlook the impact of unsuccessful communication. The analysis demonstrated that the level of confidence assigned to a node needs to incorporate the effects of communication failures to ensure reliable evaluation when it comes to the noisy communication channel of underwater networks. Additionally, under the value-centric strategies, the chapter examined two popular mechanisms in the literature: majority rule and deviation from personal experience. An analytical study evaluated the effect of combining these methods on the accuracy of detecting dishonest recommendations under various conditions. The conducted study considers sets of recommendations containing 5, 10, and 15 entries, with one-third of the recommendations being malicious. The analysis demonstrated that applying the outlier detection method followed by the deviation test yields higher accuracy, achieving an average accuracy of 92.5%, compared to other combinations. However, the analysis is certainly constrained by the proportion of dishonest recommendations injected within the network. Consequently, relying solely on a value-centric approach to identify outliers is better suited as a supporting method. This is due to challenges such as the prevalence of malicious recommendations, where the majority being dishonest greatly undermines detection efforts and necessitates the inclusion of more advanced techniques to overcome this issue of dishonest recommendations during the process of indirect trust.

Chapter 6 proposed a new TMM that leverages spatio-temporal mobility metrics for evaluating node trustworthiness, with a particular focus on mobility trust. This chapter emphasised the direct aspects of TMM, integrating five metrics to construct trust across various dimensions. The proposed model was evaluated across different topologies and attack scenarios using the same datasets from the previous chapter. It demonstrated significant improvement over benchmarked methods, achieving faster convergence toward the estimated ground truth, particularly in the presence of attacks. This performance advantage stems from the diverse metric domains incorporated into the model, enabling it to capture a wide range of behaviours and adapt effectively to various attack types. To further validate the model, its adaptability was tested against low-profile attack strategies by modifying malicious behaviours. Although these subtle attacks caused a slight reduction in convergence due to their nature being less detectable, the model still outperformed state-of-the-art approaches.

Chapter 7 introduced a novel method for evaluating recommendations as part of the indirect trust framework within the proposed TMM. This method addressed the persistent issue of dishonest recommendations, a challenge that arises when malicious UNs manipulate trust-seeking behaviours within the network. These nodes spread deceptive trust scores, either boosting the reputation of harmful nodes (BS attacks) or unfairly damaging the reputation of trustworthy ones (BM attacks), thereby compromising the integrity of the recommendation system and the overall trustworthiness of the network. Both BS and BM attacks were examined by varying the proportion of malicious recommenders. At lower levels of dishonesty, typically below 20% to 25%, traditional recommendation models relying on majority-rule mechanisms can mitigate the effects of dishonest recommendation attacks favouring the honest majority. However, as the proportion of malicious recommenders exceeds this threshold, attackers begin to dominate, rapidly distorting trust scores and achieving their objectives. These effects are more pronounced over time, as cumulative attacks exacerbate the disruption. To overcome these challenges, a two-phase recommendation evaluation method was developed to assess both the trustworthiness of recommendations and the long-term behaviour of recommenders. This approach demonstrated strong resilience, even in scenarios with high levels of malicious activity. When tested against existing recommendation-based methods that were proposed for addressing dishonest recommendations in TMM for UWSN and IoUT, the proposed model consistently outperformed them both analytically and through extensive simulations. The proposed model showed particular robustness in scenarios with high proportions of malicious recommenders. During BS attacks, it maintained an accuracy of 95.24% with 40% dishonest recommenders, decreasing to 43.45% at 50%. In comparison, existing models performed significantly worse, achieving only 13.54% accuracy at 40% and collapsing entirely to 0% at 50%. Similarly, for BM attacks, the proposed model achieved 64.76% accuracy at 50% dishonest recommenders, while competing models began failing at 30%, dropping to 50.56%

accuracy or worse and eventually collapsing beyond 40%. These results highlight the effectiveness and robustness of the proposed method in combating dishonest recommendations within TMM for UWSN and IoUT.

Revisiting the requirements established in Section 3.5, and through its fast *convergence*, the TMM proposed in Chapter 6 and Chapter 7 generate *accurate* trust values quickly, even under consistent or malicious node behaviour. Incorporating multiple metrics as trust parameters ensures *robustness*, enabling accurate and *context-specific* trust evaluation based on the network's operational environment. Its dynamic weighting mechanism provides *adaptability*, adjusting trust dimensions in real time to address changes in network topology and mobility constraints. The model arguably *lightweight* in computational design, which minimises computational overhead and ensures efficient trust calculation with low latency, making it suitable for resource-constrained devices in IoUT. Mechanisms for *maintaining historical trust* allow nodes to store and leverage past interactions within limited storage, supporting a *decentralised* architecture. Trust values are continuously *updated* to maintain accuracy and reflect nodes' actual behaviour reliably. Its *resilience* against malicious attacks, demonstrated through test cases, ensures accurate evaluations even under adversarial actions. Furthermore, the model aligns with the requirements of *scalability*, *decentralisation*, and *availability*, supporting independent evaluations and functioning reliably under limited evidence or attacks. Combined, these properties establish the model as a practical and effective solution for IoUT environments.

8.2 Thesis Contributions

The broader security aspects, including the TMM, within the context of IoUT remain a relatively immature area of study. Each core chapter of this thesis makes a contribution to advancing knowledge in this domain. Some contributions stem from comprehensive and critical reviews of existing literature (Chapter 2, Chapter 3, Chapter 4), while others are the result of extensive research conducted through both analytical and empirical methods (Chapter 5, Chapter 6, Chapter 7). Revisiting the research problem statements established on Section 1.2 and guided by the research objectives, the key contributions of this thesis are summarised in the following sections.

- ***PS.1*: The lack of attack-based underwater datasets and the shortcomings of existing simulation tools in addressing the diverse potential applications of IoUT**

  Motivated by the lack of a comprehensive understanding of the current and potential structure of underwater networks, this thesis provides an in-depth analysis of key aspects of such networks. Specifically, it investigates how UAC are simulated, including the propagation model, noise model, and attenuation model, offering guidelines to enable researchers to simulate underwater networks more realistically. The thesis proposes a classification basis for different network topologies, derived from advancements in UNs and application demands. These topologies include static, floating, anchored, and mobile configurations. Considering the unique characteristics of each topology, the thesis explores the requirements for deploying these networks in two-dimensional and three-dimensional spaces, along with appropriate deployment strategies. This work establishes a clear guidlines for developing realistic simulation case studies that accurately represent the characteristics and mobility of each network topology. These contributions facilitate broader testing and validation of security models, directly addressing **Objective 1**.

- ***PS.2*: Examining the readiness of current TMM with physical misbehaviour**

  This contribution proposes an investigation into how current TMM handle diverse underwater network topologies and the attack scenarios. It expands the scope beyond communication-related attacks to include those arising from the physical domain, with PM being a notable example of malicious attacks. A significant limitation observed in existing models is their inability to achieve fast convergence when detecting attacks. This highlights the need to incorporate metrics derived from the physical domain, as PM attacks often go undetected in the current TMM. Furthermore, the indirect trust mechanisms within existing TMM show limitations in meeting trust management requirements, particularly in mitigating the impact of dishonest recommendations. Different methodologies were examined within the context of underwater networks, and potential improvements were identified. Consequently, this contribution addresses **Objective 2** by identifying key limitations in current models and highlighting opportunities for enhancement.

- ***PS.3*: The needs for lightweight and decentralised multi-metric TMM for IoUT**

  This thesis introduces a novel multi-metric decentralised TMM, referred to as the Mobility-Aware Trust Model (MATMU). The proposed model integrates metrics from both physical and communication domains, utilising spatial and temporal metrics derived from the velocity of UN to establish the mobility trust dimension. Additionally, it incorporates metrics related to communication reliability, delay, and energy consumption. Trust is determined by evaluating how effectively a trustee communicates with minimal delay, maintains stable energy levels, and exhibits movement patterns consistent with expected behaviour. These factors collectively shape the TMM, with a single trust value constructed from these dimensions through a dynamic weighting mechanism that adapts to changes in the contributing factors. Simulation tests demonstrated that the proposed TMM effectively detects misbehaviour across diverse network topologies and attack scenarios, thereby addressing **Objective 3**.

- ***PS.4*: The problem of dishonest recommendations**

  This thesis proposed a novel recommendation evaluation mechanism that demonstrates robust performance as the proportion of malicious and dishonest recommendations increases, as shown in Chapter 7. The mechanism combines an initial filtering phase, using outlier detection followed by deviation from personal experience $O \rightarrow D$, with a belief-based evaluation phase that incorporates social validation factors. These factors include recommender similarity, long-term trustworthiness, recommendation freshness, and trust score age. This approach acknowledges the contextual nature of trust, moving beyond absolute trust to evaluate recommender behaviour over time. The proposed method has shown strong resilience in mitigating dishonest recommendations, effectively addressing **Objective 4**

8.3 Limitations and Future Work

While this thesis has offered valuable insights to advancing TMM in the IoUT domain, it is important to recognise some limitations in the proposed solutions, which pave the way for future research opportunities. The key areas are identified below.

- **The scalability demands beyond the TMM**

  While the adoption of decentralised TMM can arguably enhance scalability— allowing each individual UN to independently manage the TMM without reliance on central authorities— it is reasonable to assume that the efficiency of this process is inherently tied to the capabilities of the UNs themselves. Given the low and variable underlying resources available, concerns arise that, even with lightweight TMM, the ability to monitor and store the necessary data might gradually introduce additional overhead to the UNs' management processes, especially when there are many newcomers due to the dynamic nature of the network. The broader need for an efficient, scalable solution extends beyond the security and TMM requirements to encompass the overall functional demands of the IoUT. One potential approach to address this challenge is by improving the network design to reduce the overhead imposed on UNs. In our earlier work (Almutairi et al. 2022), we proposed a model integrating TMM with a novel network design, wherein additional components within the system architecture were introduced to evaluate and maintain the TMM process (see Appendix B for more information). Augmented scalability in design to support a variety of network-related demands within activities, such as traffic routing and some specific functions, e.g., load balancing, thereby opening new avenues for research to further enhance and optimise the network.

- **The trade-off between enhancing accuracy of TMM and the resource constraint IoUT**

  In the proposed work, a dynamic weighting mechanism was employed as an effective and adaptable approach to balancing trust evaluation in resource-constrained environments. While this method provides flexibility, it inherently relies on predetermined thresholds. These thresholds, calibrated to average values derived in this study, meet the requirements of typical deployment scenarios. However, like most threshold-based solutions, there is a potential risk of exploitation by sophisticated malicious entities capable of working around these thresholds. To further enhance robustness, future work could explore complementary approaches such as fuzzy logic or lightweight machine learning models like support vector machines for establishing trust across multiple metrics. While promising, these methods would need careful evaluation, particularly regarding their energy consumption, to ensure compatibility with the lightweight design principles essential for underwater networks. The current solution strikes a balance between security, efficiency, and network longevity while leaving room for future enhancements to address emerging challenges.

- **Advancing TMM in complex and heterogeneous networks**

  The work in this thesis focuses on TMM within single-hop topologies, which offer notable advantages for underwater sensing applications. As noted in prior research, single-hop topologies align well with the constraints of underwater networks due to their simplicity, which helps mitigate the challenges of energy consumption faced by battery-powered sensor nodes (Gorma 2019). While this focus on single-hop networks provides a clear and controlled context for evaluating TMM, future research could explore extending these models to multi-hop topologies. These networks, though more complex, are essential for larger-scale underwater deployments where nodes must relay information across multiple intermediaries to reach the destination. Investigating how TMM can be adapted to maintain trust across multiple hops while addressing challenges such as increased latency, energy consumption, and the compounded risk of malicious behaviour is a promising direction for future work.

  Another relevant avenue for future research is examining TMM in heterogeneous network environments. These networks involve diverse node types, such as surface buoys, underwater vehicles, and stationary sensors, each with varying capabilities and roles. Exploring how TMM can account for this heterogeneity while maintaining trust evaluation and efficient resource utilisation would further validate its applicability to real-world underwater applications.

- **Expanding evaluation against sophisticated attack strategies**

  This work extended the attack model to include threats originating from both communication and physical domains, providing a comprehensive foundation for understanding malicious behaviours in underwater networks. However, it is reasonable to anticipate that attackers will evolve, employing more sophisticated and collaborative strategies beyond those explored in this thesis. The attack domain within IoUT remains an area that requires further investigation to fully understand the dynamics of malicious entities, particularly as they interact with complex and interconnected systems extending beyond the underwater environment. Future research could focus on developing and evaluating TMMs that are robust against coordinated and advanced attack strategies, such as collusion-based attacks, Sybil attacks, or hybrid strategies combining multiple threat vectors. This line of inquiry would not only enhance the resilience of IoUT networks but also provide insights into designing adaptive and scalable TMM capable of addressing emerging security challenges.

- **Integration with communication protocols**

  While this work focuses on enhancing the TMM, the applicability of the proposed method across different contexts remains unexplored. Future research may also leverage the proposed TMM to enhance the security of routing protocols, particularly in decision-making processes such as route selection. By incorporating trust as a key factor, routing decisions could account for the trustworthiness of UNs, ensuring greater resilience against malicious behaviour. For example, protocols like the Vector-Based Forwarding Protocol (VPF), which primarily relies on distance as the decision criterion for selecting relay nodes, could be augmented with the proposed TMM. This integration would allow security considerations, derived from trust evaluations to influence the relay node selection process, thereby strengthening the overall robustness of the network. Expanding the research to explore such integrations would not only validate the versatility of the TMM but also pave the way for more secure and adaptive communication protocols in underwater networks.

8.4 On Enhancing TMM in IoUT

The IoUT has emerged as a transformative technology, enabling critical advancements in environmental monitoring, marine research, and maritime operations. As IoUT systems grow in complexity and scale, the demand for robust security and TMM becomes increasingly crucial. While IoUT shares characteristics with other emerging technologies like IoT and MANET, such as resource constraints and dynamic topologies, it introduces distinct challenges specific to its underwater context, including the reliance on underwater UAC, limited physical accessibility, and environmental factors affecting network stability. These challenges define its unique requirements, making conventional TMM insufficient. Ensuring secure and reliable communication in IoUT networks is not just a technical necessity but also a cornerstone for the success and longevity of IoUT applications. Addressing these demands requires innovative solutions to tackle threats arising from both communication and physical domains.

The work presented in this thesis addresses these challenges by enhancing TMM solutions to account for the unique complexities of IoUT. Expanding metric space to include physical-domain metrics with conventional communication metrics creates new opportunities for advancing TMM and detecting deceptive attacks. Additionally, analysing the behaviour of recommenders to identify the roots of dishonest practices—incorporating social and logical perspectives—further strengthens the ability of TMM to counteract exploitation by malicious entities. These contributions underscore the importance of adopting interdisciplinary approaches to trust and security, bridging the gap between theoretical models and practical applications. Moreover, the methodologies developed in this research provide valuable tools to support emerging technologies facing similar constraints, broadening the impact of these contributions.

While the methodologies were developed for the IoUT domain, the approaches presented in this thesis are largely domain-independent and can be adapted, with minimal modifications, to other domains. The integration of trust evaluation across communication and physical domains as well as the proposed recommendation evaluation mechanism, has demonstrated its utility in IoUT, but the same principles could extend to applications such as anomaly detection in IoT systems, assessing fraud in financial networks, optimising autonomous vehicle interactions, and even trust evaluation in human-machine collaboration.

References

Aaqib, Muhammad et al., 2023. "IoT trust and reputation: a survey and taxonomy". In: *Journal of Cloud Computing* 12.1, p. 42.

Adam, Nadir et al., 2024. "State-of-the-art security schemes for the Internet of Underwater Things: A holistic survey". In: *IEEE Open Journal of the Communications Society*.

Adewuyi, Anuoluwapo A. et al., 2019. "CTRUST: A dynamic trust model for collaborative applications in the Internet of Things". In: *IEEE Internet of Things Journal* 6.3, pp. 5432–5445. ISSN: 2327-4662.

Adewuyi, Anuoluwapo Amarachukwu, 2021. *Trust Modelling and Management for Collaborative and Composite Applications in the Internet of Things*. Liverpool John Moores University (United Kingdom).

Akyildiz, Ian F et al., 2016. "SoftWater: Software-defined networking for next-generation underwater communication systems". In: *Ad Hoc Networks* 46, pp. 1–11. ISSN: 1570-8705.

Alam, SM Nazrul et al., 2008. "Coverage and connectivity in three-dimensional underwater sensor networks". In: *Wireless Communications and Mobile Computing* 8.8, pp. 995–1009.

Alamu, Olumide et al., 2023. "Energy harvesting techniques for sustainable underwater wireless communication networks: A review". In: *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, p. 100265.

Almutairi, Abeer et al., 2022. "A Multi-Level Trust Framework for the Internet of Underwater Things". In: *2022 IEEE International Conference on Cyber Security and Resilience (CSR)*, pp. 370–375. DOI: 10.1109/CSR54599.2022.9850334.

Arifeen, Md Murshedul et al., 2019. "Anfis based trust management model to enhance location privacy in underwater wireless sensor networks". In: *International Conference on Electrical, Computer and Communication Engineering (ECCE)*. IEEE, pp. 1–6. ISBN: 1538691116.

Bello, Oladayo et al., 2022. "Internet of underwater things communication: Architecture, technologies, research challenges and future opportunities". In: *Ad Hoc Networks* 135, p. 102933.

Bhattacharjya, K. et al., 2021. "IoUT: Modelling and simulation of Edge-Drone-based Software-Defined smart Internet of Underwater Things". In: *Simulation Modelling Practice and Theory* 109, p. 102304. ISSN: 1569-190x. DOI: `ARTN10230410.1016/j.simpat.2021.102304`. URL: `%3CGo%20to%20ISI%3E://WOS:000638003500007`.

Blinken, Max, 2019. "The Future of Maritime Remote Systems". In: *Defense.info*. Accessed: 2022-04-11. URL: `https://defense.info/maritime-dynamics/2019/11/the-future-of-maritime-remote-systems/`.

Bolster, Andrew, 2017. *An investigation into trust and reputation frameworks for autonomous underwater vehicles*. The University of Liverpool (United Kingdom).

Boufares, Nadia et al., 2015. "Three dimensional mobile wireless sensor networks redeployment based on virtual forces". In: *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*. IEEE, pp. 563–568.

Buchegger, Sonja et al., 2002. "Performance analysis of the CONFIDANT protocol". In: *Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking computing*, pp. 226–236.

Caiti, Andrea et al., 2013. "Mobile underwater sensor networks for protection and security: field experience at the UAN11 experiment". In: *Journal of Field Robotics* 30.2, pp. 237–253.

Cambridge University Press, n.d. *Trust.* https://dictionary.cambridge.org/dictionary/english/trust. Accessed: 2025-01-01.

Carminati, Barbara et al., 2022. *Security and trust in online social networks.* Springer Nature.

Caruso, Antonio et al., 2008. "The meandering current mobility model and its impact on underwater mobile sensor networks". In: *IEEE INFOCOM 2008-The 27th Conference on Computer Communications.* IEEE, pp. 221–225.

Chaudhary, Monika et al., 2022. "Underwater wireless sensor networks: Enabling technologies for node deployment and data collection challenges". In: *IEEE Internet of Things Journal* 10.4, pp. 3500–3524.

Chen, Shenlong et al., 2012. "Dealing with dishonest recommendation: The trials in reputation management court". In: *Ad Hoc Networks* 10.8, pp. 1603–1618. ISSN: 1570-8705.

Cho, A-ra et al., 2022. "Survey of Acoustic Frequency Use for Underwater Acoustic Cognitive Technology". In: *Journal of Ocean Engineering and Technology* 36.1, pp. 61–81.

Cho, Jin-Hee et al., 2010. "A survey on trust management for mobile ad hoc networks". In: *IEEE communications surveys tutorials* 13.4, pp. 562–583. ISSN: 1553-877X.

Cho, Jin-Hee et al., 2015. "A survey on trust modeling". In: *ACM Computing Surveys (CSUR)* 48.2, pp. 1–40.

Choudhary, Monika et al., 2021. "Node deployment strategies in underwater wireless sensor network". In: *International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE).* IEEE, pp. 773–779. ISBN: 1728177413.

Del Grosso, Vincent A., 1974. "New equation for the speed of sound in natural waters (with comparisons to other equations)". In: *The Journal of the Acoustical Society of America* 56.4, pp. 1084–1091. ISSN: 0001-4966.

Dini, Gianluca et al., 2012. "A secure communication suite for underwater acoustic sensor networks". In: *Sensors* 12.11, pp. 15133–15158.

Domingo, Mari Carmen, 2012. "An overview of the internet of underwater things". In: *Journal of Network and Computer Applications* 35.6, pp. 1879–1890. ISSN: 1084-8045.

Domingues, Rémi et al., 2018. "A comparative evaluation of outlier detection algorithms: Experiments and analyses". In: *Pattern recognition* 74, pp. 406–421. ISSN: 0031-3203.

Du, Jiaxin et al., 2020. "ITrust: an anomaly-resilient trust model based on isolation forest for underwater acoustic sensor networks". In: *IEEE Transactions on Mobile Computing*. ISSN: 1536-1233.

Du, Jiaxin et al., 2022. "LTrust: an adaptive trust model based on LSTM for underwater acoustic sensor networks". In: *IEEE Transactions on Wireless Communications* 21.9, pp. 7314–7328. ISSN: 1536-1276.

Duong, Thuy Van T. et al., 2012. "A weighted combination similarity measure for mobility patterns in wireless networks". In: *arXiv preprint arXiv:1206.1418*.

Dynamics, General, 2023. *Bluefin-21 Unmanned Underwater Vehicle (UUV)*. Available online, accessed on 20 March 2024. General Dynamics. URL: `https://gdmissionsystems.com/products/underwater-vehicles/bluefin-21-autonomous-underwater-vehicle`.

Garratt, J. R., 1977. "Review of drag coefficients over oceans and continents". In: *Monthly weather review* 105.7, pp. 915–929. ISSN: 1520-0493.

Gong, Zaiwu et al., 2020. "Measuring trust in social networks based on linear uncertainty theory". In: *Information Sciences* 508, pp. 154–172.

Gorma, Wael, 2019. "Effective Medium Access Control for Underwater Acoustic Sensor Networks". PhD thesis. University of York.

Govindan, Kannan et al., 2011. "Trust computations and trust dynamics in mobile adhoc networks: A survey". In: *IEEE Communications Surveys Tutorials* 14.2, pp. 279–298. ISSN: 1553-877X.

Goyal, Nitin et al., 2022. “An anchor-based localization in underwater wireless sensor networks for industrial oil pipeline monitoring”. In: *IEEE Canadian Journal of Electrical and Computer Engineering* 45.4, pp. 466–474.

Guo, Ji et al., 2011. “A new trust management framework for detecting malicious and selfish behaviour for mobile ad hoc networks”. In: *IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE, pp. 142–149. ISBN: 145772135X.

Han, Guangjie et al., 2015. “An attack-resistant trust model based on multidimensional trust metrics in underwater acoustic sensor network”. In: *IEEE Transactions on Mobile Computing* 14.12, pp. 2447–2459. ISSN: 1536-1233.

Han, Guangjie et al., 2019. “A synergetic trust model based on SVM in underwater acoustic sensor networks”. In: *IEEE transactions on vehicular technology* 68.11, pp. 11239–11247. ISSN: 0018-9545.

Holmes, Jason D et al., 2010. “An overview of unmanned underwater vehicle noise in the low to mid frequencies bands”. In: *Proceedings of Meetings on Acoustics*. Vol. 9. AIP Publishing.

Hou, Xiangwang et al., 2021. “Machine-learning-aided mission-critical Internet of Underwater Things”. In: *IEEE Network* 35.4, pp. 160–166. ISSN: 0890-8044.

Hua, Shanshan et al., 2021. “A lightweight trust management mechanism based on conflict adjudication in underwater acoustic sensor networks”. In: *Computing, Communications and IoT Applications (ComComAp)*. IEEE, pp. 258–262. ISBN: 1665427981.

Huang, Xumin et al., 2017. “Distributed reputation management for secure and efficient vehicular edge computing and networks”. In: *IEEE Access* 5, pp. 25408–25420. ISSN: 2169-3536.

Iltaf, Naima et al., 2013. “A mechanism for detecting dishonest recommendation in indirect trust computation”. In: *EURASIP Journal on Wireless Communications and Networking* 2013, pp. 1–13.

Information, National Centers for Environmental, 2020. *World Ocean Database*. Web Page. URL: `https://www.ncei.noaa.gov/products/world-ocean-database`.

Islam, Kazi Yasin et al., 2022. "A survey on energy efficiency in underwater wireless communications". In: *Journal of Network and Computer Applications* 198, p. 103295. ISSN: 1084-8045.

Jahanbakht, Mohammad et al., 2021. "Internet of underwater things and big marine data analytics—a comprehensive survey". In: *IEEE Communications Surveys Tutorials*. ISSN: 1553-877X.

James Jr, Harvey S, 2002. "The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness". In: *Journal of Economic Behavior & Organization* 47.3, pp. 291–307.

Jiang, Bin et al., 2024. "Hybrid trust model for identifying malicious attacks in underwater acoustic sensor network". In: *IEEE Sensors Journal*.

Jiang, Jinfang et al., 2017. "A trust model based on cloud theory in underwater acoustic sensor networks". In: *IEEE Transactions on Industrial Informatics* 13.1, pp. 342–350. ISSN: 1551-3203.

Jiang, Jinfang et al., 2020. "A dynamic trust evaluation and update mechanism based on C4. 5 decision tree in underwater wireless sensor networks". In: *IEEE Transactions on Vehicular Technology* 69.8, pp. 9031–9040. ISSN: 0018-9545.

Jiang, Jinfang et al., 2022. "Controversy-adjudication-based trust management mechanism in the internet of underwater things". In: *IEEE Internet of Things Journal* 10.3, pp. 2603–2614. ISSN: 2327-4662.

Jiang, Shengming, 2017. "State-of-the-art medium access control (MAC) protocols for underwater acoustic networks: A survey based on a MAC reference model". In: *IEEE Communications Surveys Tutorials* 20.1, pp. 96–131. ISSN: 1553-877X.

Jiang, Shengming, 2018. “On securing underwater acoustic networks: A survey”. In: *IEEE Communications Surveys Tutorials* 21.1, pp. 729–752. ISSN: 1553-877X.

Jøsang, Audun, 2001. “A logic for uncertain probabilities”. In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 9.03, pp. 279–311. ISSN: 0218-4885.

Josang, Audun et al., 2002. “The beta reputation system”. In: *Proceedings of the 15th bled electronic commerce conference*. Vol. 5. Citeseer, pp. 2502–2511.

Jouhari, Mohammed et al., 2019. “Underwater wireless sensor networks: A survey on enabling technologies, localization protocols, and internet of underwater things”. In: *IEEE Access* 7, pp. 96879–96899.

Kao, Chien-Chi et al., 2017. “A comprehensive study on the internet of underwater things: applications, challenges, and channel models”. In: *Sensors* 17.7, p. 1477.

Kesari Mary, Delphin Raj et al., 2022. “Energy Optimization Techniques in Underwater Internet of Things: Issues, State-of-the-Art, and Future Directions”. In: *Water* 14.20, p. 3240. ISSN: 2073-4441.

Khalil, Ruhul et al., 2020. “Towards the Internet of underwater things: Recent developments and future challenges”. In: *IEEE Consumer Electronics Magazine*. ISSN: 2162-2248.

Khalil, Ruhul Amin et al., 2021. “Bayesian multidimensional scaling for location awareness in hybrid-Internet of Underwater Things”. In: *IEEE/CAA Journal of Automatica Sinica* 9.3, pp. 496–509.

Khan, Muhammad Toaha Raza et al., 2019. “REMEDY: Receiver-initiated MAC based on energy-efficient duty-cycling in the IoUT”. In: *IEEE Access* 7, pp. 105202–105211.

Khedim, Farah et al., 2015. “Dishonest recommendation attacks in wireless sensor networks: A survey”. In: *12th International Symposium on Programming and Systems (ISPS)*. IEEE, pp. 1–10. ISBN: 1479976997.

Lagerspetz, Olli, 2013. *Trust: The tacit demand.* Vol. 1. Springer Science & Business Media.

Lenard, Teri et al., 2023. “Exploring Trust Modelling and Management Techniques in the Context of Distributed Wireless Networks: A Literature Review”. In: *IEEE Access.*

Li, Deyi et al., 2017. *Artificial intelligence with uncertainty.* CRC press. ISBN: 1315366959.

Li, Hong et al., 2015. “Security and privacy in localization for underwater sensor networks”. In: *IEEE Communications Magazine* 53.11, pp. 56–62.

Li, Huaizhi et al., 2007. “Trust management in distributed systems”. In: *Computer* 40.2, pp. 45–53.

Li, Ruidong et al., 2009. “A novel hybrid trust management framework for MANETs”. In: *29th IEEE International Conference on Distributed Computing Systems Workshops.* IEEE, pp. 251–256. ISBN: 0769536603.

Liang, Wei et al., 2019. “TBRS: A trust based recommendation scheme for vehicular CPS network”. In: *Future Generation Computer Systems* 92, pp. 383–398. ISSN: 0167-739X.

Lowney, M. Phil et al., 2018. “Trust Management in Underwater Acoustic MANETs based on Cloud Theory using Multi-Parameter Metrics”. In: *2018 International Carnahan Conference on Security Technology (ICCST)*, pp. 1–5. DOI: 10.1109/CCST.2018.8585499.

Luhmann, Niklas, 2018. *Trust and power.* John Wiley & Sons.

Luo, H. J. et al., 2018. “Software-Defined Architectures and Technologies for Underwater Wireless Sensor Networks: A Survey”. In: *Ieee Communications Surveys and Tutorials* 20.4, pp. 2855–2888. ISSN: 1553-877X. DOI: 10.1109/Comst.2018.2842060. URL: %3CGo%20to%20ISI%3E://WOS:000451262800011.

Luo, Junhai et al., 2009. “A trust model based on fuzzy recommendation for mobile ad-hoc networks”. In: *Computer networks* 53.14, pp. 2396–2407.

Luo, Junhai et al., 2021. "Localization algorithm for underwater sensor network: A review". In: *IEEE Internet of Things Journal* 8.17, pp. 13126–13144.

Lurton, Xavier, 2002. *An introduction to underwater acoustics: principles and applications*. Vol. 2. Springer.

Mahmood, Adnan et al., 2023. "Toward a distributed trust management system for misbehavior detection in the internet of vehicles". In: *ACM Transactions on Cyber-Physical Systems* 7.3, pp. 1–25. ISSN: 2378-962X.

Marche, Claudio et al., 2020. "Trust-related attacks and their detection: A trust management model for the social IoT". In: *IEEE Transactions on Network and Service Management* 18.3, pp. 3297–3308. ISSN: 1932-4537.

Markets, R., 2022. *4.3 Billion Autonomous Underwater Vehicle (AUV) Markets - Global Forecast to 2026*. Accessed: 2025-02-24. URL: https://www.globenewswire.com/news-release/2021/11/05/2328501/28124/en/4-3-Billion-Autonomous-Underwater-Vehicle-AUV-Markets-Global-Forecast-to-2026.html.

Martin, Robert et al., 2017. "Aqua-Sim Next generation: An NS-3 based underwater sensor network simulator". In: *Proceedings of the 12th International Conference on Underwater Networks Systems*, pp. 1–8.

Mazinani, Sayyed Majid et al., 2018. "A vector-based routing protocol in underwater wireless sensor networks". In: *Wireless Personal Communications* 100.4, pp. 1569–1583.

McKenzie, Grant et al., 2021. "Measuring urban regional similarity through mobility signatures". In: *Computers, Environment and Urban Systems* 89, p. 101684. ISSN: 0198-9715.

Michiardi, Pietro et al., 2002. "Core: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks". In: *Advanced Communications and Multimedia Security: IFIP TC6/TC11 Sixth Joint Working Conference on Communications and Multimedia Security September 26–27, 2002, Portorož, Slovenia*. Springer, pp. 107–121.

Mohsan, Syed Agha Hassnain et al., 2023. “Recent advances, future trends, applications and challenges of internet of underwater things (iout): a comprehensive review”. In: *Journal of Marine Science and Engineering* 11.1, p. 124. ISSN: 2077-1312.

Morozs, Nils et al., 2020. “Channel modeling for underwater acoustic network simulation”. In: *IEEE Access* 8, pp. 136151–136175. ISSN: 2169-3536.

Nain, Mamta et al., 2024. “A survey on node localization technologies in UWSNs: Potential solutions, recent advancements, and future directions”. In: *International Journal of Communication Systems*, e5915.

Nanthakumar, Sathish et al., 2024. “A comparative study of range based and range free algorithms for node localization in underwater”. In: *e-Prime-Advances in Electrical Engineering, Electronics and Energy* 9, p. 100727.

Nkenyereye, Lewis et al., 2024. “Internet of underwater things: A survey on simulation tools and 5G-based underwater networks”. In: *Electronics* 13.3, p. 474.

north.io, 2024. “Underwater Infrastructure Big Data Project Secures €2.4 Million Grant”. In: *Marine Technology News*. URL: `https://www.marinetechnologynews.com/news/underwater-infrastructure-project-secures-638517`.

Ocean, Sofar, 2024. *Sofar Ocean: Connecting the World’s Oceans*. `https://www.sofarocean.com/`. Accessed: 2024-10-21.

Pompili, Dario et al., 2009. “Three-dimensional and two-dimensional deployment analysis for underwater acoustic sensor networks”. In: *Ad Hoc Networks* 7.4, pp. 778–790. ISSN: 1570-8705.

Porter, Michael B, 2011. “The bellhop manual and user’s guide: Preliminary draft”. In: *Heat, Light, and Sound Research, Inc., La Jolla, CA, USA, Tech. Rep* 260.

Pourghebleh, Behrouz et al., 2019. “A comprehensive study on the trust management techniques in the Internet of Things”. In: *IEEE Internet of Things Journal* 6.6, pp. 9326–9337. ISSN: 2327-4662.

Pourkabirian, Azadeh et al., 2023. “An accurate RSS/AoA-based localization method for internet of underwater things”. In: *Ad Hoc Networks* 145, p. 103177.

Pranitha, B et al., 2020. “Analysis of underwater acoustic communication system using equalization technique for ISI reduction”. In: *Procedia Computer Science* 167, pp. 1128–1138.

Price, James F. et al., 1987. “Wind-driven ocean currents and Ekman transport”. In: *Science* 238.4833, pp. 1534–1538. ISSN: 0036-8075.

Qiu, Tie et al., 2019. “Underwater Internet of Things in smart ocean: System architecture and open issues”. In: *IEEE transactions on industrial informatics* 16.7, pp. 4297–4307. ISSN: 1551-3203.

Rasmusson, Lars et al., 1996. “Simulated social control for secure internet commerce”. In: *Proceedings of the 1996 workshop on New security paradigms*, pp. 18–25.

Rotter, Julian B, 1980. “Interpersonal trust, trustworthiness, and gullibility.” In: *American psychologist* 35.1, p. 1.

Scientific, Sea-Bird, 2024. *Navis Biogeochemical (BGC) Autonomous Profiling Float*. Accessed: 2024-10-22. URL: `https://www.seabird.com/navis-biogeochemical-bgc-autonomous-profiling-float/product?id=54627925752`.

Shabut, Antesar M. et al., 2014. “Recommendation based trust model with an effective defence scheme for MANETs”. In: *IEEE Transactions on mobile computing* 14.10, pp. 2101–2115. ISSN: 1536-1233.

Shabut, Antesar M. et al., 2018. “A multidimensional trust evaluation model for MANETs”. In: *Journal of Network and Computer Applications* 123, pp. 32–41. ISSN: 1084-8045.

Sharma, Avani et al., 2020. “Towards trustworthy Internet of Things: A survey on Trust Management applications and schemes”. In: *Computer Communications* 160, pp. 475–493. ISSN: 0140-3664.

Singh, Ashish et al., 2024. “Trust management in online computing environment: a complete review”. In: *Journal of Ambient Intelligence and Humanized Computing* 15.1, pp. 491–545.

Song, Aijun et al., 2019. “Editorial underwater acoustic communications: Where we stand and what is next?” In: *IEEE Journal of Oceanic Engineering* 44.1.

Spiesberger, John L et al., 1991. “New estimates of sound speed in water”. In: *The Journal of the Acoustical Society of America* 89.4, pp. 1697–1700.

Srinivasan, Avinash et al., 2006. “DRBTS: distributed reputation-based beacon trust system”. In: *2nd IEEE international symposium on dependable, autonomic and secure computing*. IEEE, pp. 277–283. ISBN: 0769525393.

Stull, Ronald B., 2015. *Practical meteorology: an algebra-based survey of atmospheric science*. University of British Columbia.

Su, Xin et al., 2020. “A review of underwater localization techniques, algorithms, and challenges”. In: *Journal of Sensors* 2020. ISSN: 1687-725X.

Su, Yishan et al., 2021. “A redeemable SVM-DS fusion-based trust management mechanism for underwater acoustic sensor networks”. In: *IEEE sensors journal* 21.22, pp. 26161–26174. ISSN: 1530-437X.

Sun, Yan Lindsay et al., 2006. “Information theoretic framework of trust modeling and evaluation for ad hoc networks”. In: *IEEE Journal on Selected Areas in Communications* 24.2, pp. 305–317. ISSN: 0733-8716.

Tuna, Gurkan et al., 2017. “A survey on deployment techniques, localization algorithms, and research challenges for underwater acoustic sensor networks”. In: *International Journal of Communication Systems* 30.17, e3350. ISSN: 1074-5351.

Valeport, 2023. *MIDAS WLR Water Level Recorder*. Available online, accessed on 20 March 2023. Teledyne Valeport Ltd: Totnes, UK. URL: `https://www.valeport.co.uk/products/midas-wlr-water-level-recorder/`.

Vegni, Anna Maria et al., 2021. “A vlc-based footprinting localization algorithm for internet of underwater things in 6g networks”. In: *2021 17th International symposium on wireless communication systems (ISWCS)*. IEEE, pp. 1–6.

Wang, Biao et al., 2023. “Adaptive Power-Controlled Depth-Based Routing Protocol for Underwater Wireless Sensor Networks”. In: *Journal of Marine Science and Engineering* 11.8, p. 1567.

Wang, Hongzhi et al., 2019. “Progress in outlier detection techniques: A survey”. In: *IEEE Access* 7, pp. 107964–108000. ISSN: 2169-3536.

Wei, Lijun et al., 2022. “Trust management for Internet of Things: A comprehensive study”. In: *IEEE Internet of Things Journal* 9.10, pp. 7664–7679.

Xu, Bo et al., 2022. “A node location optimization algorithm based on mobility prediction for underwater wireless sensor networks”. In: *2022 34th Chinese Control and Decision Conference (CCDC)*. IEEE, pp. 2176–2182.

Yang, Guang et al., 2019. “Challenges and security issues in underwater wireless sensor networks”. In: *Procedia Computer Science* 147, pp. 210–216. ISSN: 1877-0509.

Yang, Yue et al., 2021. “A survey of autonomous underwater vehicle formation: Performance, formation control, and communication capability”. In: *IEEE Communications Surveys Tutorials* 23.2, pp. 815–841. ISSN: 1553-877X.

Yang, Zhe et al., 2022. “TADR-EAODV: A trust-aware dynamic routing algorithm based on extended AODV protocol for secure communications in wireless sensor networks”. In: *Internet of Things* 20, p. 100627.

Yisa, Aliyu Gana et al., 2021. “Security challenges of internet of underwater things: A systematic literature review”. In: *Transactions on Emerging Telecommunications Technologies* 32.3, e4203. ISSN: 2161-3915.

Zhang, Hongwei et al., 2021. “Subsea pipeline leak inspection by autonomous underwater vehicle”. In: *Applied Ocean Research* 107, p. 102321. ISSN: 0141-1187.

Zhang, Mengjie et al., 2023. “A recommendation management defense mechanism based on trust model in underwater acoustic sensor networks”. In: *Future Generation Computer Systems* 145, pp. 466–477. ISSN: 0167-739X.

Zhang, Yongguang et al., 2003. “Intrusion detection techniques for mobile wireless networks”. In: *Wireless Networks* 9, pp. 545–556. ISSN: 1022-0038.

Zhao, Zhenyi et al., 2022. “A Cooperative Hunting Method for Multi-AUV Swarm in Underwater Weak Information Environment with Obstacles”. In: *Journal of Marine Science and Engineering* 10.9, p. 1266. ISSN: 2077-1312.

Zhu, Rongxin et al., 2024. “Design Guidelines on Trust Management for Underwater Wireless Sensor Networks”. In: *IEEE Communications Surveys & Tutorials.*

Zhu, Zhengliang et al., 2023. “Internet of underwater things infrastructure: A shared underwater acoustic communication layer scheme for real-world underwater acoustic experiments”. In: *IEEE Transactions on Aerospace and Electronic Systems* 59.5, pp. 6991–7003.

Zouridaki, Charikleia et al., 2005. “A quantitative trust establishment framework for reliable data packet delivery in MANETs”. In: *Proceedings of the 3rd ACM workshop on Security of ad hoc and sensor networks*, pp. 1–10.

Zouridaki, Charikleia et al., 2009. “E-Hermes: A robust cooperative trust establishment scheme for mobile ad hoc networks”. In: *Ad Hoc Networks* 7.6, pp. 1156–1168.

Appendices

Appendix A

List of Equations

Eq. No.	Equation Explanation
	Chapter 4
Equation 4.1	$c = \sqrt{\frac{E}{\rho}} = \sqrt{\frac{1}{\chi\rho}}$. — Acoustic propagation speed.
Equation 4.2	$\begin{aligned} c = 1448.96 + 4.591t - 5.304 \times 10^{-2}t^2 + 2.374 \times 10^{-4}t^3 \\ +1.340(s-35) + 1.630 \times 10^{-2}d + 1.675 \times 10^{-7}d^2 \\ -1.025 \times 10^{-2}t(s-35) - 7.139 \times 10^{13}td^3. \end{aligned}$ — Del Grosso equation of computing the propagation speed.
Equation 4.3	$\frac{I_2}{I_1} = \left(\frac{4\pi R_1}{4\pi R_2}\right)^2 = \left(\frac{R_1}{R_2}\right)^2$. — Geometrical spreading loss.
Equation 4.4	$L_{\text{spr}}(R) = k \times 10 \log_{10}\left(\frac{R}{R_{1m}}\right)$. — The spreading transmission loss, considered from the reference unit ($R_{1m} = 1m$).
Equation 4.5	$L_{\text{abs}}(f) = A_1P_1\frac{f_1f^2}{f_1^2+f^2} + A_2P_2\frac{f_2f^2}{f_2^2+f^2} + A_3P_3f^2$. —The absorption loss.
Equation 4.6	$\begin{aligned} L_{\text{abs}}(f) = 0.11\frac{f^2}{1+f^2} + \frac{44f^2}{4100+f^2} \\ +3 \times 10^{-4}f^2 + 33 \times 10^{-3}. \end{aligned}$ — The absorption loss based on the Thorp formula.

Table continued on next page

Eq. No.	**Equation Explanation**
Equation 4.7	$P(d, f) = L_{\text{spr}}(d) + d_{\text{km}} L_{\text{abs}}(f)$. —The attenuation of sound underwater.
Equation 4.8	$N_s(f) = 40 + 20(s_h - 0.5) + 26\log_{10}(f) - 60\log_{10}(f + 0.03)$. — Noise estimated from shipping source.
Equation 4.9	$N_w(f) = 50 + 7.5\sqrt{w} + 20\log_{10}(f) - 40\log_{10}(f + 0.4)$. — Noise estimated from wind source.
Equation 4.10	$N_{\text{th}}(f) = -15 + 20\log_{10}(f)$. — Noise estimated from thermal source.
Equation 4.11	$N_t(f) = 17 - 30\log_{10}(f)$. — Noise estimated from turbulence source.
Equation 4.12	$N(f) = N_s(f) + N_w(f) + N_{th}(f) + N_t(f)$. — The net noise computation.
Equation 4.13	$$v_l = N \times \frac{4}{3}\pi R_s^3,$$ $$v_u = v_l + 2\left(\frac{h \times l}{\pi R_s^2} + \frac{h \times w}{\pi R_s^2} + \frac{l \times w}{\pi R_s^2}\right) + 4\left(\frac{h}{R_s} + \frac{l}{R_s} + \frac{w}{R_s}\right).$$ — The nodes placement distribution equation.
Equation 4.14	$u = -\frac{\partial\psi}{\partial y}; \quad v = \frac{\partial\psi}{\partial x}$. — Water current velocity based on Meandering current.
Equation 4.15	$D = \sqrt{\frac{2K}{f_c}}$ — The Ekman layer depth.
Equation 4.16	$$\overline{u} = \left[\frac{w_{fv}}{\sqrt{Kf_c}}\right]\left[e^{\frac{z}{D}}\cos(\frac{z}{D} - \frac{\pi}{4})\right]$$ $$\overline{v} = \left[\frac{w_{fv}}{\sqrt{Kf_c}}\right]\left[e^{\frac{z}{D}}\sin(\frac{z}{D} - \frac{\pi}{4})\right]$$ — Horizontal ocean-current components based on Ekman theory.

Table continued on next page

Eq. No.	**Equation Explanation**
Equation 4.17	$w_{fv} = \frac{\rho_{air}}{\rho_{water}}\alpha_{fv}$. — The friction of the velocity of water based on Ekman theory.
Equation 4.18	$t_{ij} = \frac{\sum_{i=1}^{n}\left(S_{i(t)}-D_{i(t)}\right)}{N}$. — The average end-to-end delay.
Chapter 5	
Equation 5.1	$$\phi_{j,m}^{t} = \frac{\min_j \lvert\alpha_{j,m}^{t} - \mathrm{g}_m^t\rvert + \rho \max_j \lvert\alpha_{j,m}^{t} - \mathrm{g}_m^t\rvert}{\lvert\alpha_{j,m}^{t} - \mathrm{g}_m^t\rvert + \rho \max_j \lvert\alpha_{j,m}^{t} - \mathrm{g}_m^t\rvert},$$ $$\varphi_{j,m}^{t} = \frac{\min_j \lvert\alpha_{j,m}^{t} - \mathrm{b}_m^t\rvert + \rho \max_j \lvert\alpha_{j,m}^{t} - \mathrm{b}_m^t\rvert}{\lvert\alpha_{j,m}^{t} - \mathrm{b}_m^t\rvert + \rho \max_j \lvert\alpha_{j,m}^{t} - \mathrm{b}_m^t\rvert}.$$ — Grey Relational Coefficient (GRC) proposed by MTACMM.
Equation 5.2	$[\phi_j^t, \varphi_j^t] = \left[\sum_{m=0}^{M} h_m \phi_{j,m}^t, \sum_{m=0}^{M} h_m \varphi_{j,m}^t\right]$ — Trust component proposed by MTACMM.
Equation 5.3	$T_{ij} = \frac{1}{1+\left(\frac{\phi_j^t}{\varphi_j^t}\right)^2}$. — Trust computation proposed by MTACMM.
Equation 5.4	$T_{ij} = w_1 T_{\mathrm{c}} + w_2 T_{\mathrm{d}} + w_3 T_{\mathrm{e}}$. — Trust computation proposed by CATM.
Equation 5.5	$c_{ik} = 1 - \sqrt{\frac{12\alpha_{ik}\beta_{ik}}{(\alpha_{ik}+\beta_{ik})^2(\alpha_{ik}+\beta_{ik}+1)}}$ — Confidence estimated in beta trust model.
Equation 5.6	$c_{ik} = 1 - u_{ik}$. — Confidence estimated in subjective logic trust model.
Chapter 6	
Equation 6.1	$M_{\mathrm{ss}} = \frac{1}{2}\left(\frac{\vec{v_i}(t)\cdot\vec{v_j}(t)}{\Vert\vec{v_i}(t)\Vert\Vert\vec{v_j}(t)\Vert}\frac{\min(\Vert\vec{v_i}(t)\Vert,\Vert\vec{v_j}(t)\Vert)}{\max(\Vert\vec{v_i}(t)\Vert,\Vert\vec{v_j}(t)\Vert)}+1\right)$. —Spatial mobility similarity between two nodes.
Equation 6.2	$M_{\mathrm{ts}} = \frac{1}{2}\left(\frac{\vec{v_j}(t)\cdot\vec{v_j}(t')}{\Vert\vec{v_j}(t)\Vert\Vert\vec{v_j}(t')\Vert}\frac{\min(\Vert\vec{v_j}(t)\Vert,\Vert\vec{v_j}(t')\Vert)}{\max(\Vert\vec{v_j}(t)\Vert,\Vert\vec{v_j}(t')\Vert)}+1\right)$. — Temporal mobility similarity between two nodes.
Equation 6.3	$T_{\mathrm{m}} = wM_{ss}+(1-w)M_{ts}$. — Trust based on mobility factor.

Table continued on next page

Eq. No.	**Equation Explanation**
Equation 6.4	$T_c = b + ur.$ — Trust based on communication factor.
Equation 6.5	$b_{ij} = \frac{\sum_{ta}^{t_b} s_{ij}}{\sum_{ta}^{t_b} s_{ij} + \sum_{ta}^{t_b} f_{ij} + 2}, u_{ij} = \frac{2}{\sum_{ta}^{t_b} s_{ij} + \sum_{ta}^{t_b} f_{ij} + 2}.$ — Estimation of trust on communication based on subjective logic and beta theory.
Equation 6.6	$T_{\text{td}} = \left(\frac{1}{p} \sum_{i=1}^{p} \frac{d_{ij}/c}{\text{Ed}_{ij}} \right).$ — Trust based on delay factor.
Equation 6.8	$T_e = 1 - \frac{EC_i^m}{E_i}.$ — Trust based on energy factor.
Equation 6.9	$T_{ij}^d(t) = w_m T_m + w_d T_{\text{td}} + w_c T_c + w_e Te.$ — Direct trust estimation.
Equation 6.10	$w_i = \frac{e^{(\theta_i - T_i)/\nu}}{\sum_{m=1}^{M} e^{(\theta_i - T_i)/\nu}}.$ — Dynamic weighting mechanism of trust factors.
Equation 6.11	$\Lambda_{slot}^t = \frac{e^{t/M}}{\sum_{m=1}^{M} e^{m/M}}.$ — Trust update process.
Chapter 7	
Equation 7.1	$T_{ij}^r(t) = \sum_{k=1}^{K} B_{ij \leftarrow k} T_{kj}, j \neq k.$ — Indirect trust computation.
Equation 7.2	$\lvert T_{ij(t-\epsilon)} - T_{kj(t)}^r \rvert (1 - d_{T(ij)}) \leq \delta_h.$ — Trust filtering based on deviation method.
Equation 7.3	$c_{ik} = \frac{\alpha_{ik}}{(\alpha_{ik} + \beta_{ik})}.$ — Established confidence based on beta model.
Equation 7.4	$c_{ik} = b_{ik} + u_{ik} r.$ — Established confidence based on subjective logic model.
Equation 7.5	$f_{ik} = \frac{S_i^k}{S_i^K} \tau^{\frac{1}{S_i^k}}.$ — Established familiarity factor.
Equation 7.6	$s_{ik(t)} = w f_{ik} + (1 - w) c_{ik}.$ — Established similarity factor.
Equation 7.7	$\Gamma_{kj} = \eta_{sc} (t - t_{T_{kj}})^{-\varepsilon}.$ — Established timeliness factor.
Equation 7.8	$B_{ij \leftarrow k} = \frac{s_{ik(t)} \Gamma_{kj} (1 - d_{T(ij)}) T_{ki(t)}}{\sum_{k'}^{K} \left(s_{ik'(t)} \Gamma_{k'j} (1 - d_{T(ij)}) T_{k'i(t)} \right)}, k \neq k'.$ — Belief weighting model.
Equation 7.9	$T_{ki} = T_{ki}(t) \left(1 - \frac{1}{\exp^{\lambda}} \right).$ — Established penalty factor.

Appendix B

Conceptual Multi-level Trust model design for scalable IoUT

B.1 Challenges in the current work

In general, all the current developments in underwater network trust (both centralised and distributed approaches) come with several constraints that will take additional effort to overcome. Existing studies of trust management in IoUT are based mainly on the premise of a trustworthy environment at the outset. This is not realistic for underwater networks where nodes are usually left unattained for a long time, making them easily exposed to adversaries and vulnerable to demolishing attacks. Furthermore, despite having some strength factors of the existing trust policies, they also involve various drawbacks, such as memory expense, communication overhead, and work under assumptions such as platform reliance, which makes them unrealistic for large-scale IoUT. Moreover, the existing work neglects the essential properties of IoUT architecture, such as heterogeneity of the network, node mobility and the scalability requirement. According to Li et al. (2017), any feasible trust mechanism for an IoT system must involve scalability and heterogeneity requirements since a typical IoT system contains both resource constraints and a significant and continuously increasing number of connected devices. Since IoUT inherits these requirements from IoT, any trusted system for IoUT should also address these requirements. Some of the critical issues that are worthwhile to highlight are discussed below.

1. While the fully distributed trust management system attains the IoUT scalability requirement, having an entirely distributed system is not practicable due to the performance constraints of devices underwater and the limited memory and computing capabilities of each underwater node. In a distributed approach, each node must preserve an up-to-date record of the whole network's trust values in some form of data structure. This structure's size is directly proportional to the network's size. A single sensor node cannot store and compute the entire network's trust values using node resources, which will eventually degrade network performance.

2. The same can be said for the centralised trust management strategy used in underwater networks; it is ineffective since it wastes energy owing to additional routing overhead. When the sink node is far away from the underwater nodes, the overall routing cost for exchanging trust values between the nodes and the sink node is highly energy-intensive in large-scale networks. Furthermore, this approach is reliant on the central entity's availability as well as constant and stable connectivity between nodes and this entity, which is not optimal given the current state of underwater communication. When nodes experience problems in communicating with the central entity, the trust value cannot be computed, and the system as a whole becomes exposed to a single point of failure. Furthermore, such a system will not scale well due to network congestion caused by a high volume of nodes sending requests to the same entity.

3. Unlike many other systems that obtain a substantial indication of misbehaviour from a single metric, the accuracy of trust formation for underwater networks depends on several metrics because of various factors that influence the stability of underwater networks, such as the dynamic environment, node mobility, frequent signal loss, energy and resource constraints. Relying on multi-metric trust methods raises the complexity of the entire trust es-

tablishment process, from trust composition, computation logarithms, the size of packets transmitted, and the continuous need to update and assess the trust values on a regular basis. This complexity requires a further evaluation of the impact of the trust established on the performance of the network.

4. Due to the dynamic nature of the underwater network, trust evidence may be uncertain and incomplete. This uncertainty will affect the judgment and could lead to paralysing the whole network by either trusting a malicious node or causing a high volume of false alarms and preventing benign nodes from functioning properly.

Existing studies have attempted to address specific challenges related to trust mechanisms in underwater networks; nevertheless, a general conclusion from these contributions is the lack of a comprehensive investigation that adequately addressed the problems mentioned above in conjunction with the unique requirements of an IoUT network.

B.2 Proposed Framework

In order to address the aforementioned limitations, this study proposes the concept of a Multi-level Trust Framework for IoUT, as shown in Fig. B.1. The framework employs a hybrid approach that takes advantage of the IoUT network's fundamental components to simplify the process of computing trust between nodes, distribute the heavy computation between underwater nodes and above-water sink nodes, and maintain a reliable trust system throughout the network. The suggested framework incorporates three levels of trust management:

- a simple low-level trust mechanism between undersea nodes (level 0);

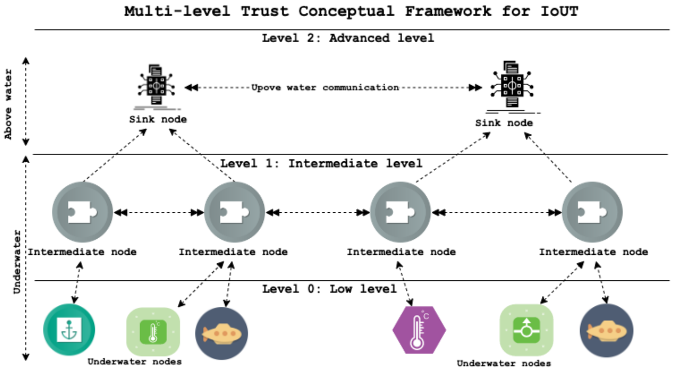


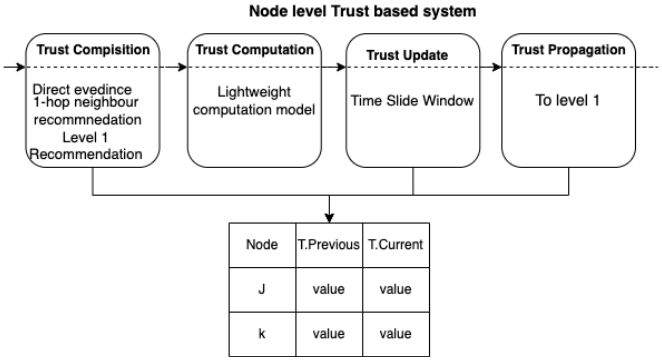
Fig. B.1. Conceptual framework of multi-level trust management for IoUT.

- an intermediate layer to aggregate trust values and serve as a link between underwater and above-water sink nodes (level 1).
- a more sophisticated assessment method at the sink node (level 2);

To make accurate trust judgments of nodes with minimal overheads, the framework utilises both distributed and centralised techniques. All of the layers work collaboratively to maintain the functionality of a trust-based system. The following subsections describe the fundamental functionality of each level in the proposed framework.

B.2.1 Level 0: Low-level trust sub-model

The lowest level focuses on establishing trust between aquatic nodes while taking into account the constraints of underwater communication and node capacity. Lightweight and acceptable trust mechanisms between these nodes are to be considered. The trade-off between the robustness and the complexity of trust calculation for underwater nodes is also considered with a tendency towards preserving node performance. The primary purpose of this level is to speed up the node-to-node trust establishment process and alleviate the demand for high precision in each computed trust value while saving the node's limited resources. Each node will be responsible for the trust establishment process, as shown in Fig. B.2, by



Node	T.Previous	T.Current
J	value	value
k	value	value

Fig. B.2. Representation of trust process in level 0.

considering a lighter version for each stage of the trust establishment process. This means that each process of trust establishment: trust composition, computation, propagation, and update will address the limitation of the underwater nodes and will strive to provide a reasonable trust with minimum overhead.

The complexity of establishing trust between nodes and the minimum level of confidence required at this level are the main factors that affect which metrics to consider during this phase. Both direct and indirect evidence are required. A subset of the trust metrics (successful/unsuccessful communication, node-centric evidence) is considered for the direct evidence. The reason behind that choice is to reduce the complexity of trust computation and rely on the most valuable evidence. For the indirect evidence and due to the possibility of missing recommendations from nearby nodes, this model uses several procedures to request recommendations: either from 1-hop neighbours who have a relation with the trustee node or by requesting recommendations from level 1 nodes-which will be further explained in the next subsection. While the considered evidence can be seen as a preliminary for computing a robust trust system, it could guarantee an acceptable trust computation that managed to handle the limitations of both computation and communication overhead and, in the same way, detect any blatant misbehaviour. The resulting trust value propagates to any level 1 node nearby.

Table B.1: Specifications of Trust Process for Level 0.

Trust Process	Description	Merits	Shortcoming
Trust composition - Direct evidence	Successful communication, node-centric evidence	Essential in measuring the trustworthiness of a network.	High chance of false positives.
Trust composition - Indirect evidence	From nodes in level 1 and from 1-hop neighbours.	To reduce the communication overhead	Insufficient recommendations.
Trust Computation	A lightweight computation logarithm.	Reduces the computation overhead.	Limited accuracy.
Trust Update	Time slide window mechanism with two-time slots (previous - current).	Reduces the storage requirement.	Depends on the predetermined time interval.
Trust Propagation	Transmitted to (level 1) and the 1-hop neighbour upon request.	Reduces the heavy transmission cost on the limited bandwidth.	Misleading recommendation.

There are several factors to consider when computing the trust value, such as reducing memory and transmission overhead. One way is by adopting a light logarithm to compute the trust with minimal overhead. For instance, weighting mechanisms, subjective logic methods, or statistical approach algorithms could be utilised to compute the trust without sabotaging the performance of underwater nodes Kao et al. (2017). Another way is by reducing the trust value representation method to save space, which plays a significant role in exchanging trust values among nodes in less transmission and power overhead. Table B.1 determined the merits and shortcomings of each step in the process of establishing trust at this level. The following two levels in this framework are designed to overcome the shortcomings of the trust establishment process at this low level while maintaining the advantages.

B.2.2 Level 1: Intermediate level trust sub-model

The suggested framework uses a number of intermediate nodes that operate as a connection unit between the underwater nodes and the above-water sink nodes. They are positioned in different locations to maintain the connectivity required. As shown in Fig. B.3, these intermediary nodes are linked to one another, establishing a chain of connected nodes, and are responsible for a variety of functions that strengthen the overall trust process's robustness. Each node in the set of

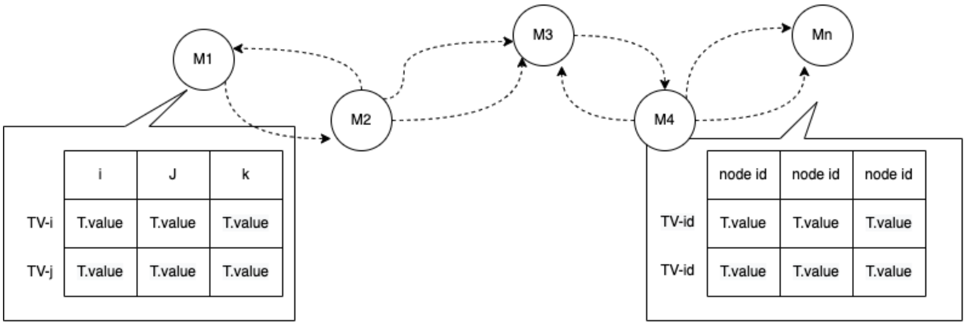


Fig. B.3. Representation of level 1 intermediate nodes structure.

intermediate nodes collects the computed trust value for each undersea node in its range and combines it into an aggregated trust, and sends it to the sink node to examine at the above level. In order to alleviate the scarcity of insufficient recommendations, these nodes will be responsible for supporting the underwater nodes by providing them with recommendations upon request, as mentioned in Level 0.

B.2.3 Level 2: Advanced level trust sub-model

An accurate trust computation requires a reliable and sophisticated component to handle the complexity of computation. Since underwater nodes are incapable of dealing with this level of complexity, this framework adopts a collective of boosted agents (in sink nodes) to assess the trust values generated by underwater nodes as depicted by Fig. B.4. At this level, the sink nodes are put in use to examine the entire trust system due to their high-performance capability compared to the underwater nodes. The sink nodes handle a variety of collaborative intelligent activities to excessively assess the trustworthiness of nodes and compare it to the computed trust in Level 0. Sink nodes will proceed with the assessment process upon receiving the aggregated trust from Level 1 nodes. Evaluating each node's trust begins with the extraction and familiarisation of various complicated metrics, such as the projected environmental change in the current situation, the data collected by underwater nodes and obtained by sink nodes, and the

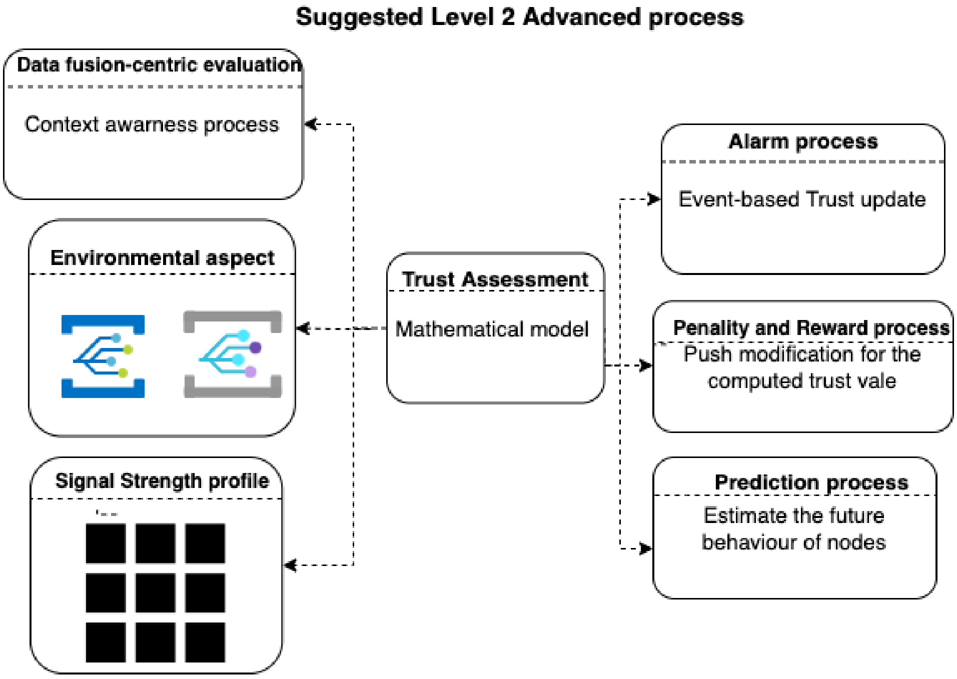

Suggested Level 2 Advanced process
Data fusion-centric evaluation
Context awarness process
Environmental aspect
Signal Strength profile
Trust Assessment
Mathematical model
Alarm process
Event-based Trust update
Penality and Reward process
Push modification for the computed trust vale
Prediction process
Estimate the future behaviour of nodes

Fig. B.4. Representation of level 2 sub-model.

predicted signal changes profile of a specific area. By utilising the effect of each of these factors and extracting a context-related observation, an examination of the computed trust value against these factors should indicate the accuracy of the computed trust. At this stage, two possible situations are expected: the computed trust value by nodes underwater is accurate or inaccurate. The insufficiency of the computed trust could result from misbehaviour (malicious or selfish behaviour) or a false positive case due to any changes in the environment. Therefore, this model also assesses the reason behind the variation in the computed trust. In case of an attack, an urgent alert will be triggered to notify the underwater nodes of a penalty for the misbehaved node. The notification can be achieved by using an event-based trust update mechanism generated from the sink node and propagated to the lower level in the framework. In case of environmental changes, the sink node will request an update to the trust value of this node.

B.2.4 Case study

Consider the simple scenario in Fig. B.5, which describes the trust establishment process in the underwater network by following the proposed framework. In the diagram, node Y attempts to communicate with node X for the first time, while another node Z is maliciously compromised and presented as a threat in the network. Node X will start establishing trust with Y during this phase, according to level 0. In this situation, trust will be formed using the lightweight and simple version of trust, which is based on X's communication and node proof, as well as Z and M1's recommendations (the nearest level 1 node to node X). The accuracy of the computed trust value is subject to the manipulation of the compromised node Z. At this stage, if there is no attack, the trust establishment will occur successfully. However, due to the presence of attacking node Z, the accuracy will be low. According to the framework, the computed trust will be transmitted to the sink node at the surface of the ocean (level 2) with the help of M1. The sink node will thoroughly examine the accuracy of the computed trust. In case node Z performs a bad-mouthing attack to decay the trust of node Y or a Ballot-stuffing attack to fabricate a positive and fake recommendation of node Y, the sink nodes will be able to detect the inaccuracy of the computed trust value and generate an alarm to punish node Z and trigger an event-based update to propagate the accurate trust value of Y.

B.3 Discussion

The trust management in IoUT highly influences the security and reliability of not only underwater nodes and networks but also the collected data. The proposed framework aims to design a robust trust system while preserving the whole network performance and involves the requirement of IoUT. The following sub-

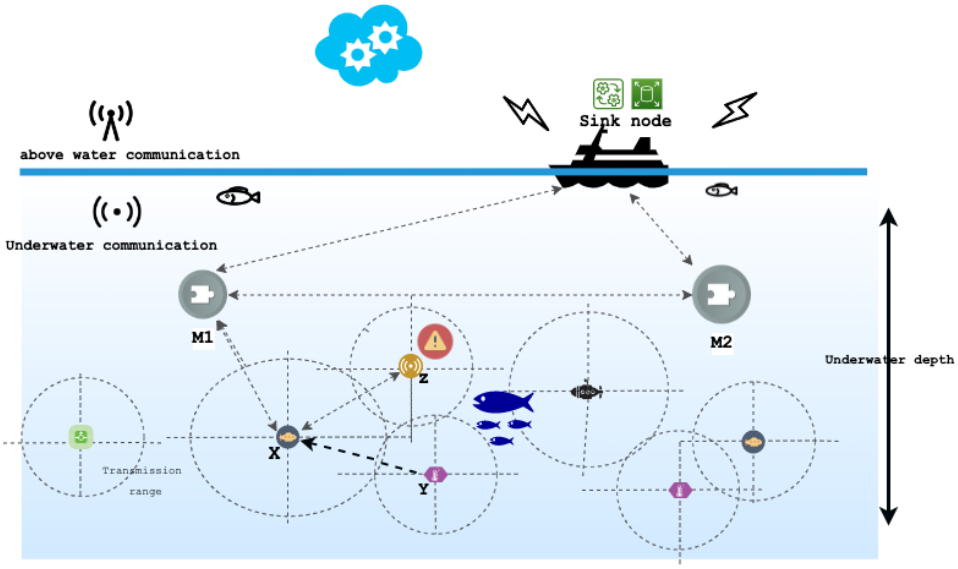


Fig. B.5. Scenario of trust communication in the proposed framework.

sections show an attempt to examine the feasibility of the proposed framework.

Addressing IoUT requirement

As stated in section III, there is a lack of involvement in the requirement of IoUT during the design of the trust system. The suggested framework succeeds in capturing these needs and performing appropriately by assessing nodes' scalability, heterogeneity, and mobility. For instance, this approach supports the scalability requirement as nodes can be added to the network freely without decreasing the network's performance. The reason behind that is the utilisation of a local trust agent in each node and the global trust agent at the sink nodes. Moreover, while almost all of the proposed frameworks assume that underwater nodes are stationary as a prerequisite for trust management to have an accurate result, the proposed framework addresses nodes' free mobility to reflect a more realistic application of IoUT. As with the scalabilities, and since the trust process is distributed between nodes under the direction of the sink nodes, no matter where the node's location is in the given area, the trust management framework will be able to compute trust values.

Overall performance

Due to the restricted resources, high energy consumption, and difficult-to-maintain communication paradigm, network performance is crucial for any solution in underwater networks. It is worth noting that the majority of existing studies are more concerned with measuring the correctness of their proposed methodologies than with the overall network performance. Similarly, the current development lacks the adoption of a realistic environment to mimic the dynamicity and complexity of the underwater networks. Both communication and computation overhead, as well as energy usage, are required indicators of the trust management solution's feasibility. By dividing computing complexity across several layers, the proposed approach aims to reduce the complexity of the entire trust formation process during the design phase. In comparison to the existing fully distributed and centralised approaches to address the limited communication bandwidth and high probability of signal loss, the proposed framework reduces the communication overhead when exchanging trust recommendations and messages between nodes to establish trust. Energy can be harvested from a variety of sources for IoUT, including communication, processing, and mobility. By utilising lightweight algorithms for computing trust locally and dividing the process throughout subsea components, the proposed framework is intended to decrease the overhead of the trust process.

Appendix C

Enhancement on Simulation Model

The Aqua-Sim ng has been utilised due to its layer-wise protocol design, which enables the integration of trust within the communication stack. Figure C.1 illustrates the modifications made to the current simulation environment. More specifically, the green classes in the diagram highlight the enhancements to the simulation model. The `AquaSimTrust` class provides an abstraction for inspecting packets, as well as analysing physical metrics and readings, which are utilised by various trust models. Several existing models from the literature, along with the proposed trust model, have been developed based on this abstract class. The `AquaSimAttack` class has been modified to include attacks from the physical domain. Additionally, a new mobility model, based on water currents, has been introduced under `ns3:MobilityHelper`. A snapshot of the code structure, including the main class constructors, and the workflow of trust within the communication stack, is shown below.

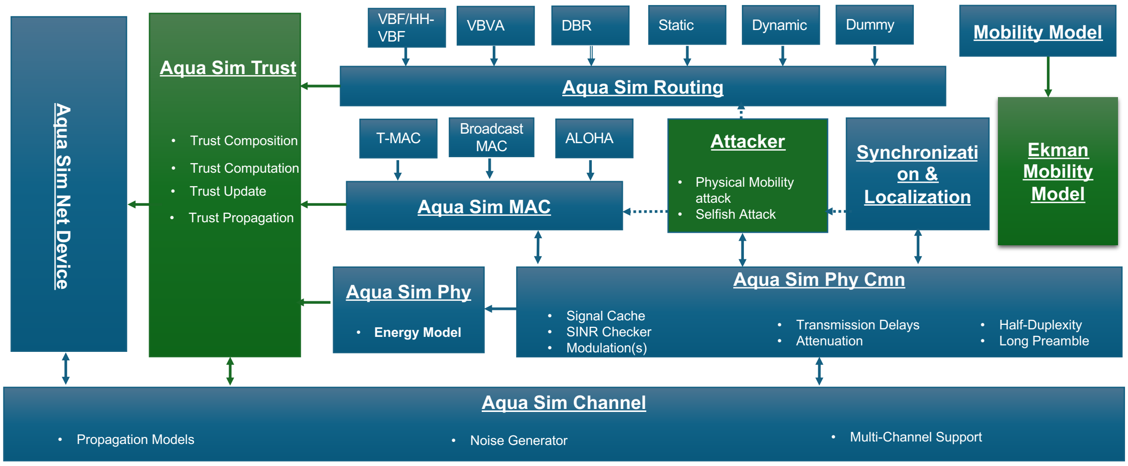


Fig. C.1. Modified Aqua-Sim ng architecture.

C.1 The Aqua Sim Trust Interface

The workflow of trust establishment begins in the `AquaSimChannel` class, where network traffic is monitored and subsequently passed to `AquaSimTrust` ?? C.1. Within the simulation scripts, nodes are assigned the trust model in a manner similar to other functionalities as illustrated in ??. This integration ensures that all nodes within the simulation can utilise `AquaSimTrust` seamlessly, just as they would with other core components of AquaSim.

Listing C.1: Trust Monitor on the `AquaSimChannel`

```
bool
AquaSimChannel::Recv(Ptr<Packet> p, Ptr<AquaSimPhy> phy)
{NS_LOG_FUNCTION(this);
  NS_LOG_DEBUG("Packet:" << p << " Phy:" << tifp << " Channel:" << this);
  Ptr<AquaSimNetDevice> sender = Ptr<AquaSimNetDevice>(tifp->GetNetDevice());
  Ptr<AquaSimNetDevice> recver;
  Ptr<AquaSimPhy> rifp;
  Time pDelay;
  std::vector<PktRecvUnit> * recvUnits = m_prop->ReceivedCopies(sender, p,
      m_deviceList);
  if (recver->IsTrustOn()){
        double s=AquaSimAddress::ConvertFrom (sender->GetAddress ()).GetAsInt ()
            ;
        double r=AquaSimAddress::ConvertFrom (recver->GetAddress ()).GetAsInt ()
            ;
        Vector ph_data ;
        ph_data.x=pstamp.GetFreq();
        ph_data.y=pstamp.GetPt();
        ph_data.z=pstamp.GetPr();
        recver->GetTrustModel()->RecvPhysical(p,ph_data);    Simulator::Schedule
            (pDelay, &AquaSimPhy::Recv, rifp, p->Copy());}
```

Listing C.2: The Constructor of AquaSimTrustBase

```
AquaSimTrustBase::AquaSimTrustBase ():
    m_totalPktRecv (0),
    m_totalPktDelay (0),
    m_seq (0),
    m_pr (0),
    m_pt (0),
    m_freq (0),
    m_node (0),
    m_device(nullptr) ,
    m_trace_time (40),
    m_slot (6),
    m_numOfRecommendation_Request (0),
    m_numOfRecommendation_Response (0),
    sentCount (0),
    isInitiated (false),
    m_num_nodes (0),
    m_lostpkt(0),
    m_trustPeriod (0),
      responsesReceived ( false),
      timeoutOccurred (false),
{ NS_LOG_FUNCTION(this);
  m_trustGloabalRecord = std::list<std::list<trust_table_base>> ();}}
```

Listing C.3: Initiating Trust Process in the Simulation Script

```
for (NodeContainer::Iterator i = NodesCon.Begin(); i != NodesCon.End();
                            i++) {
                        Ptr<AquaSimNetDevice> newDevice = CreateObject<
                            AquaSimNetDevice>();
                        devices.Add(asHelper.Create(*i, newDevice));
                        newDevice->GetPhy()->SetTransRange(range);
                        Ptr<AquaSimEnergyModel> energy = newDevice->
                            GetEnergyModel();
                             energy->SetInitialEnergy(70000);  //in watt
                        //------------------trust -----------------------
                        Ptr<AquaSimTrustMATMU> trust_model = CreateObject<
                            AquaSimTrust>();
                        trust_model->SetNumNodesToInitiate(num_nodes);
                        trust_model->SetDevice(newDevice);
                        trust_model->InitiateTrustTable(AquaSimAddress::
                            ConvertFrom(newDevice->GetAddress()).GetAsInt());
                        trust_model->ScheduleTrustProcess(60,simStop);
                        trust_model->RunTrust();
                        newDevice->SetTrustModel(trust_model);
                    }}
```

C.2 The Ekman Mobility Model

The mobility of water current mimicked by the Ekman theory explained in Section 4.3.3.1 is added under `ns-3 EkmanDriftingMobilityModel`.

Listing C.4: Ekman Model Constructor

```
TypeId
EkmanDriftingMobilityModel::GetTypeId(void)
{ Ptr<NormalRandomVariable> rand = CreateObject<NormalRandomVariable> ();
  static TypeId tid = TypeId("ns3::EkmanDriftingMobilityModel")
     .SetParent<MobilityModel> ()
        .SetGroupName ("Mobility")
        .AddConstructor<EkmanDriftingMobilityModel> ()
    .AddAttribute ("EddyVisscosity", "eddy viscosity factor",
      DoubleValue(0.00168),
      MakeDoubleAccessor(&EkmanDriftingMobilityModel::m_eddy_viscosity),
      MakeDoubleChecker<double>())
    .AddAttribute ("SurfaceVelocity", "Water current surface velocity",
      DoubleValue(rand->GetValue(10, 200)),
      MakeDoubleAccessor(&EkmanDriftingMobilityModel::m_surf_wv),
      MakeDoubleChecker<double>())
      .AddAttribute ("SurfaceAngle", "Water current surface angle movement",
          DoubleValue(rand->GetValue(0, 90)),
          MakeDoubleAccessor(&EkmanDriftingMobilityModel::m_surf_wa),
          MakeDoubleChecker<double>())
      .AddAttribute ("Latitude", "Latitude angle",
          DoubleValue(rand->GetValue(-90, 90)),
          MakeDoubleAccessor(&EkmanDriftingMobilityModel::m_latitude),
          MakeDoubleChecker<double>())
      .AddAttribute ("TimeStep",
                        "Change current direction and speed after moving for
                            this time.",
                        TimeValue (Seconds (1.0)),
                        MakeTimeAccessor (&EkmanDriftingMobilityModel::
                            m_timeStep),
                        MakeTimeChecker ()) ;
  return tid;}
```

Appendix D

Existing Trust Models

D.1 CATM Process

Controversy-adjudication-based Trust Management (CATM) is a recent trust management mechanism designed for the IoUT. It integrates three main processes: trust calculation, trust recommendation, and trust evaluation, which collectively represent overall trust based on either direct trust (T_d) or recommendation trust (T_{ind}) obtained from one-hop neighbours. Direct trust is evaluated using three key metrics: trust based on packet delivery ratio (T_c), trust based on transmission delay (T_d), and trust based on energy consumption (T_e). These dimensions of trust are combined to represent the overall trust that node n_i holds toward node n_j as:

$$T_{ij} = w_1 T_c + w_2 T_d + w_3 T_e, \tag{D.1}$$

where $w_1 + w_2 + w_3 = 1$ and w_1, w_2, and w_3 represent the corresponding weights of each metric in the overall trust.

Trust based on packet delivery ratio (T_c) is determined by monitoring the successful to unsuccessful communication ratio. A successful attack scenario could result in packet drops, and thus, the packet delivery ratio is a common metric for indicating misbehaviour. However, in underwater environments, packet drops can also occur due to non-malicious factors, making it challenging to differentiate between malicious and non-malicious causes of packet loss. To address this, a penalty factor η is introduced to refine the trust estimation. The trust related to communication is then represented by:

$$\begin{aligned} T_c &= \frac{\alpha}{\alpha + \eta\beta} \\ \eta &= 1 + \log_2\left(1 + \left(\frac{u}{us + s}\right)\right) \end{aligned} \tag{D.2}$$

where s and us represent the successful and unsuccessful communication events, respectively, and $\alpha = s + 1$, $\beta = us + 1$. The penalty factor η is expressed as:

$$\eta = 1 + \log_2\left(1 + \left(\frac{u}{us + s}\right)\right) \tag{D.3}$$

Transmission delay (T_d) is another essential metric for evaluating the trustworthiness of a connection. A shorter delay generally indicates a more dependable connection. Incorporating delay as an additional metric in the trust model aids in detecting attacks that cause significant communication delays. In underwater communication, the propagation delay can be highly variable due to environmental factors like depth and temperature. The proposed trust model accounts for these environmental impacts on the delay metric to ensure a more accurate evaluation. Trust based on delay is computed using:

$$T_d = \left(\frac{1}{pkt}\sum_{i=1}^{pkt}\frac{d_{ij}/p_s}{dt_{ij}}\right), \tag{D.4}$$

where d_{ij} is the distance between the sender node n_i and the receiver node n_j, and dt_{ij} is the expected time delay in successfully delivering packets (pkt). The underwater propagation speed p_s.

Energy consumption (T_e) is the third metric used to evaluate trust. It represents the trustworthiness of a node based on its energy usage behaviour. Trust based on energy consumption is calculated as:

$$T_e = 1 - \frac{E_j^c}{E}, \tag{D.5}$$

where E_j^c is the energy consumed by the trustee node, and E represents the initial registered energy.

At any point, a node n_i can send a message to its one-hop neighbours requesting a recommendation regarding a specific node n_j. Neighbours who have a trust record for that node are required to provide a recommendation response, including their computed trust score based on their own experience. Upon receiving these recommendations, the evaluation process begins by assessing the degree of certainty for each recommender. CATM utilises both link stability and node reliability to evaluate the received recommendations.

Link stability is determined by analysing the signal-to-noise ratio (SNR), received signal strength (RSS), and the distance between the recommender and the trustor. The SNR is represented as the ratio of the signal power to the noise power spectral density and is given by:

$$SNR(d, f) = \frac{E/A(d, f)}{N(f)B} \tag{D.6}$$

where E is the transmission power in watts, B is the bandwidth in hertz, and $N(f)$ represents the noise power spectral density. Additionally, the received signal strength indicator (RSSI) is used to measure the strength of the acoustic signal, accounting for environmental factors like absorption, scattering, and ambient noise. To evaluate the link quality, grey relational analysis (GRA) is applied to link indicators such as SNR_r, $RSSI_r$, and d_{ir}. The grey correlation coefficient is computed as:

$$\begin{gathered}\phi_{r,i}^t = \frac{\min_r |\alpha_{r,i}^t - \text{good}_i^t| + \rho \max_r |\alpha_{r,i}^t - \text{good}_i^t|}{|\alpha_{r,i}^t - \text{good}_i^t| + \rho \max_r |\alpha_{r,i}^t - \text{good}_i^t|} \\ \varphi_{r,i}^t = \frac{\min_r |\alpha_{r,i}^t - \text{bad}_i^t| + \rho \max_r |\alpha_{r,i}^t - \text{bad}_i^t|}{|\alpha_{r,i}^t - \text{bad}_i^t| + \rho \max_r |\alpha_{r,i}^t - \text{bad}_i^t|}\end{gathered} \tag{D.7}$$

where $\alpha_{r,i}^t$ represents the evaluated metrics, ρ is a chosen distinguishing coefficient, and "good" and "bad" refer to reference metric sequences. The link quality score is then calculated as:

$$LQ_r = \frac{1}{1 + \left(\frac{\phi_{r,i}^t}{\varphi_{r,i}^t}\right)^2} \tag{D.8}$$

Node reliability is evaluated by the entropy of the ratio of recommendations sent over time to the number of packets transmitted. If a node consistently responds to recommendation requests, it is considered reliable. Node reliability is computed as:

$$\begin{aligned} R_r &= \sum_{n=1}^{N} \frac{Pkt(n)}{Pkt_{max}(n)}.E(p)) \\ E(p) &= -plog(p) - (1-p)log(1-p) \end{aligned} \tag{D.9}$$

where $E(p) = -p\log(p) - (1-p)\log(1-p)$, and p is the ratio of recommendations to packets sent.

The overall recommendation weight from node n_r is then determined by combining the link quality and node reliability scores:

$$W_r = \frac{1}{1 + (\alpha.R_r + \beta.LQ_r)} \tag{D.10}$$

where $\alpha + \beta = 1$ and $0 < \alpha, \beta < 1$. The trust score based on recommendations is finally expressed as:

$$T_{ind} = \sum_{r=1}^{N} \frac{W_r}{sum(W_r)}.Trust_{(direct)}^{rj} \tag{D.11}$$