

Review

Recent Developments in Intrusion Detection for the Internet of Underwater Things: A Systematic Review

Muhammad Haziq *

School of Engineering, Edith Cowan University, 270 Joondalup Drive, Joondalup, 6027, WA, Australia;
E-Mail: mhaziq@our.ecu.edu.au

* **Correspondence:** Muhammad Haziq; E-Mail: mhaziq@our.ecu.edu.au

Academic Editor: Yousef Farhaoui

Special Issue: [Recent Advances in Cyber Security](#)

Recent Prog Sci Eng

2026, volume 2, issue 4

doi:10.21926/rpse.2604022

Received: April 04, 2026

Accepted: October 13, 2026

Published: October 21, 2026

Abstract

The Internet of Underwater Things (IoUT) has significantly improved underwater communication. However, due to the unique and challenging characteristics of underwater acoustic channels, such as long multipaths, high propagation delay and Doppler variation, and severe energy and bandwidth limitation, IoUT networks are highly vulnerable to security threats. These conditions make conventional Internet of Things (IoT) intrusion detection systems (IDS) less effective. Improving the security of IoUT networks requires specialised intrusion detection approaches tailored to address these challenges. This paper reviews intrusion detection in IoUT. It describes the architecture of IoUT networks, and demonstrates how the channel dynamics affect the capabilities of an IDS. It also categorises threats across the physical, data link, network, transport, and application layers, together with cross-layer attacks, insider threats, and advanced persistent threats. Furthermore, the review compares anomaly-based, trust-based, and hybrid detection frameworks, with emphasis on machine learning, federated learning, blockchain-assisted security, and density-based clustering. The results are compared based on evaluation metrics, latency, energy efficiency, and computational complexity. Finally, the paper outlines open challenges and future research



© 2026 by the author. This is an open access article distributed under the conditions of the [Creative Commons by Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium or format, provided the original work is correctly cited.

directions, focusing on cross-layer detection, communication-efficient federated learning, adaptive thresholding, dataset availability, and validation at sea.

Keywords

Internet of underwater things; underwater wireless sensor networks; intrusion detection; machine learning; federated learning; underwater acoustic channel

1. Introduction

More than 70% of the Earth's surface is covered by water, and most of these water bodies are unexplored. IoUT supports the smart ocean in facilitating the exploration of these vast sub-sea environments [1, 2]. IoUT extends the Internet of Things to the underwater environment. This advanced technology has enabled the underwater environment to have intelligent network connectivity using underwater sensing and communication devices in the ocean. Sensors, autonomous underwater vehicles (AUVs), and surface gateways are interconnected to gather, analyse, and transmit data, creating an Underwater Wireless Sensor Network (UWSN) as shown in Figure 1. The field is growing in importance as UWSNs are adopted for environmental monitoring, military operations, and underwater exploration. However, as the IoUT network expands, the frequency of malicious threats increases, which disrupts underwater operations. These threats can jeopardise data integrity and manipulate underwater nodes. The importance of secure IoUT has considerably grown in recent times because of the advanced attacks directed toward underwater networks [3-5].

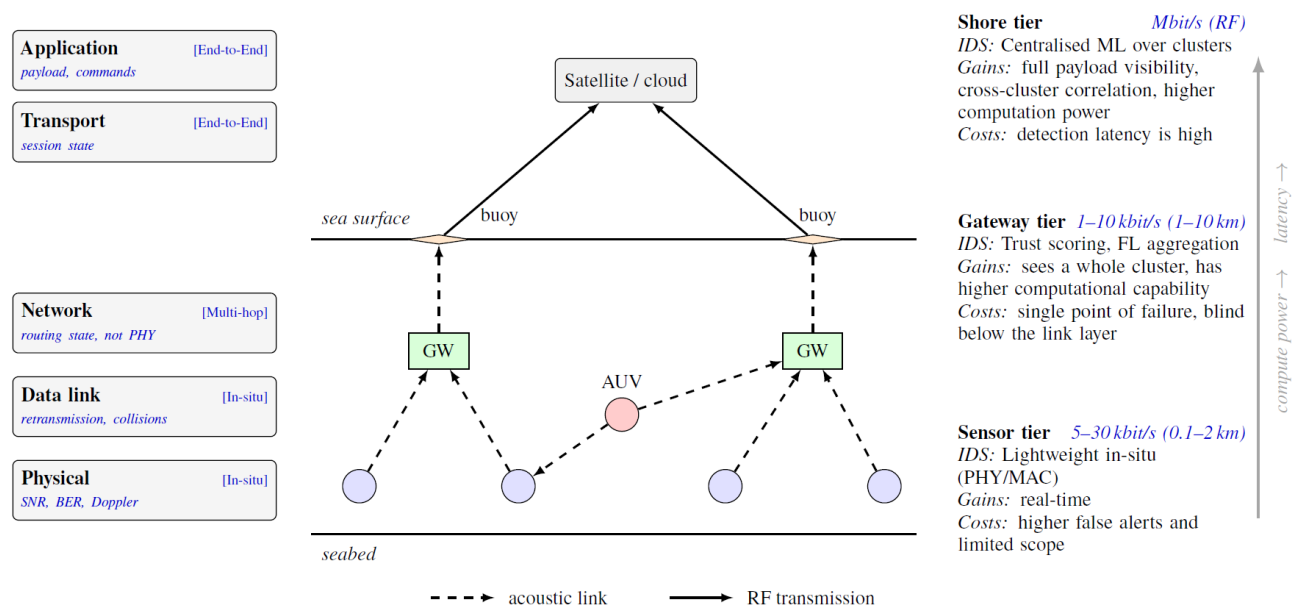


Figure 1 IoUT architecture illustrating node classes, typical link rates and ranges per tier, the protocol layers, and their impact on IDS performance. In-situ detection at the sensor tier provides real-time mitigation for PHY/MAC threats but suffers from high false-alarm rates due to harsh channel dynamics and limited node compute capabilities. Conversely, the shore tier allows complex ML-based detection over application payloads with high

accuracy, but incurs significant detection latency. Because gateways are the central aggregation points, a compromised gateway lets an adversary control what the shore tier sees of the subsea deployment.

Intrusion detection involves techniques for monitoring and safeguarding UWSNs against unauthorised access [6]. IDS are important for protecting IoUT ecosystems through detecting and mitigating cyber threats such as eavesdropping, spoofing, denial-of-service (DoS) attacks, Sybil attacks, and data tampering. Intrusion detection in IoUT is critical for network security and preventing malicious activities. However, the network and security design remain complex due to the severely challenging environment, such as limited bandwidth, power constraints, noise, Doppler effect, and high latency. IDS designed for terrestrial IoT are inadequate to adapt to the dynamic channel variations. Unlike terrestrial IoT, which uses radio frequency (RF) communication, IoUT uses an acoustic signal due to the substantial attenuation of RF signals in water. The acoustic signal is preferred due to its non-line-of-sight and long-range communication.

To support reliable data transmission, various architectures and protocols have been adopted for UWSNs to counter the varying channel dynamics. Centralised, distributed, and hybrid architectures have been developed for data collection and intrusion detection. Anomaly-based and trust-based systems and cross-layer protocols have also been adapted to improve intrusion detection. Recent work [7, 8] suggests that energy-efficient and lightweight protocols provide better efficiency and security. IoUT networks nevertheless face limitations arising from the underwater environment, which reduce the effectiveness of localisation methods and thereby increase security risks [9].

Another major challenge is resource limitation. UWSNs rely on sensor nodes deployed in remote underwater locations, where replacing or recharging batteries is difficult. As a result, these nodes often operate with limited power, memory, and processing capability, making it difficult to deploy complex security mechanisms. For this reason, existing security protocols are not well-suited for UWSNs [8]. Moreover, the dynamic nature of the underwater acoustic channels causes packet loss and interference. Hence, maintaining a stable topology is also difficult. Combined with environmental factors, node mobility can increase false positives in intrusion detection, which means security mechanisms must be able to adapt to changing network conditions [7].

These weaknesses can be exploited by attackers to disrupt communication or gain unauthorised access to sensitive data [10, 11]. Jamming attacks can interrupt acoustic signal transmission and cause data loss, while spoofing and Sybil attacks can manipulate sensor readings by introducing false identities into the network [3, 12].

To address these constraints, IoUT network designs have moved toward adaptive solutions. Recently, machine learning (ML) has been explored to improve intrusion detection in such environments [2, 13-15]. Techniques such as federated learning, hybrid AI-rule-based models, and bio-inspired algorithms have been developed to improve the detection efficiency with less computational overhead. Blockchain-based methods and cross-layer security architectures have also shown resilience against complex and multi-stage attacks [16-18].

However, despite these developments, intrusion detection for IoUT is very difficult because of the limited availability of the underwater intrusion detection datasets, detection delay, limited available power, and the dynamic nature of the underwater acoustic channel. Hence, these issues highlight the need for robust IDS that are capable of detecting anomalies in underwater

communication. Incorporating ML for anomaly detection and response offers innovative technologies for protecting UWSNs to enhance security effectiveness from advanced intrusions and attacks. Regular test and response updates are required to mitigate the impact of emerging threats and cyber-attacks in UWSNs, ensuring proactive defence against potential intrusions. Efficient resource management is vital for the sustainability of UWSN for a better network lifetime.

1.1 Relation to Existing Reviews

While existing studies examine the broader landscape of IoUT and UWSNs, intrusion detection has not been the subject of a dedicated review. Jahanbakht et al. [15] survey IoUT alongside marine data analytics but treat security only briefly. Mohsan et al. [1, 4] give broad accounts of enabling technologies, applications, and open problems. Most of these focus on the attack level: Ahmad et al. [8] classify attacks on UWSNs, Ahmed et al. [19] pair attacks with mitigation techniques, and Adam et al. [3] and Jiang et al. [5] cover security schemes and reliability across the IoUT stack. Yisa et al. [12] address security challenges in general rather than intrusion detection, and predate most of the ML and FL work discussed here.

Table 1 outlines how this review differs from the existing studies. The contributions of this paper are as follows:

- Describes how the underwater channel dynamics affect the capabilities of an IDS.
- Presents an attack taxonomy covering the physical, data link, network, transport, and application layers, as well as cross-layer attacks, insider threats, and advanced persistent threats, together with a threat model that links each attack to the layer it acts on, the vulnerability it exploits, and the countermeasures reported against it.
- Sets out the databases, publication window, and selection criteria used, so the search can be repeated.
- Compares all reviewed methods across a consistent set of criteria, covering dataset, attack type, metrics, energy cost, latency, complexity, communication overhead, network scale, test environment, and validation strategy.
- Examines the datasets used in this literature, and what class imbalance and validation practice mean for the accuracy figures reported.
- Sets out specific research questions for future work.

Table 1 Comparison of this review with existing IoUT, UWSN, and maritime security reviews. Y = covered in depth; P = partially covered; N = not covered. FL = federated learning; XAI = explainable AI.

Review	Year	Primary focus	Stated protocol	Layer-wise taxonomy	Unified metric table	Datasets discussed	FL	XAI
Jahanbakht et al. [15]	2021	IoUT architecture and marine data analytics	N	N	N	P	N	N
Yisa et al. [12]	2021	IoUT security challenges	Y	P	N	N	N	N
Ahmad et al. [8]	2021	UWSN attack taxonomy	N	Y	N	N	N	N
Mohsan et al. [4]	2022	IoUT enabling technologies	N	N	N	N	N	N
Mohsan et al. [1]	2023	IoUT advances, trends, applications	N	N	N	N	N	N
Adam et al. [3]	2024	IoUT security schemes	N	Y	N	P	P	N
Jiang et al. [5]	2024	IoUT security and reliability	N	Y	N	P	P	N
Ahmed et al. [19]	2024	UWSN attacks and mitigation	N	Y	N	N	N	N
This review	2026	Intrusion detection for IoUT	Y	Y	Y	Y	Y	Y

The remainder of this paper is organised as follows. Section 2 sets out the review methodology. Section 3 describes the background, architecture, and challenges faced by IoUT, clarifies the terminology used across the field, and differentiates IoUT from terrestrial IoT. Section 4 models the underwater acoustic channel and derives from it the requirements that any IoUT IDS must satisfy. Section 5 describes the types of intrusion reported for IoUT applications. Section 6 compares intrusion detection approaches and methods and gives their formal description. Section 7 examines the data available for training and evaluating these methods, and Section 8 covers deployment constraints, experimental infrastructure, and the standards that apply. Section 9 sets out future directions and Section 10 the lessons learned. The paper ends with a conclusion in Section 11.

2. Review Methodology

This review follows a search and screening protocol based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement. The protocol was used during the screening stage and covers the review questions, the databases searched, the search strings, the publication window, and the inclusion and exclusion criteria. The purpose is to make the selection of studies visible, so that a reader can understand and repeat the search.

2.1 Review Questions

The review is guided by four questions:

RQ1. How do the properties of the underwater acoustic channel affect the capabilities of an IDS, compared with a terrestrial IoT network?

RQ2. Which intrusions have been reported against IoUT and UWSN deployments, and which layers are affected by the attack?

RQ3. Which detection methods have been proposed, and on which technologies do they rely?

RQ4. How are these methods evaluated, and can the reported results be compared across studies?

2.2 Data Sources and Search Strings

Six databases were searched: IEEE Xplore, the ACM Digital Library, ScienceDirect, SpringerLink, MDPI, and Wiley Online Library. Scopus was not searched separately, because it largely indexes the same publishers. Google Scholar was used only to follow citations, not as a primary source. The search targeted titles, abstracts, and keywords, and combined three terms:

- underwater
- security and detection
- detection techniques

The string used to retrieve the articles was:

("underwater" OR "IoUT" OR "Internet of Underwater Things" OR "underwater wireless sensor network" OR "UWSN" OR "underwater acoustic network") AND ("intrusion detection" OR "IDS" OR "anomaly detection" OR "attack detection" OR "security" OR "trust") AND ("machine learning" OR "deep learning" OR "federated learning" OR "reinforcement learning" OR "blockchain" OR "clustering" OR "routing")

The third group restricts the results to studies that name a detection or defence technique. The publication window was January 2004 to December 2025. Studies from 2004 onward were included

because they laid the foundational work on underwater acoustic network security. Earlier work on underwater acoustics that does not address security was outside the scope.

2.3 Inclusion and Exclusion Criteria

Table 2 lists the criteria applied at the screening stage. An article had to satisfy every inclusion criterion and none of the exclusion criteria to be included.

Table 2 Inclusion and exclusion criteria applied during screening.

Type	Criterion
Inclusion	The study addresses an underwater acoustic network, a UWSN, or an IoUT deployment.
Inclusion	The study proposes, evaluates, or reviews a security, intrusion or anomaly detection system.
Inclusion	The study reports a method, an evaluation, or a structured survey.
Inclusion	The study is peer reviewed and published in a journal or in conference proceedings.
Inclusion	The study is written in English.
Exclusion	The work concerns terrestrial IoT or wireless sensor networks only, with no underwater component.
Exclusion	The work concerns underwater acoustics, localisation, or routing with no security or detection component.
Exclusion	The record is an abstract, editorial, poster, patent, preprint without peer review, or a duplicate of a record already retained.
Exclusion	The full text could not be obtained.
Exclusion	The study reports no evaluation of any kind and adds no method or classification.

2.4 Screening Procedure and Outcome

Screening was carried out in four stages, following the PRISMA guidelines of identification, screening, eligibility, and inclusion. Articles were collected from the six databases, and titles and abstracts were then screened against the criteria in Table 2. Studies passing this stage were read in full and assessed for eligibility.

The search returned 1,482 records. After 456 duplicates were removed, 1,026 records were screened on title and abstract, of which 914 were excluded. 112 full texts were assessed, and 44 were excluded, the most frequent reasons being no security component and no underwater component. A further 8 records were added through citation chasing. The final set contains 76 studies. Figure 2 shows the flow of records through the four stages.

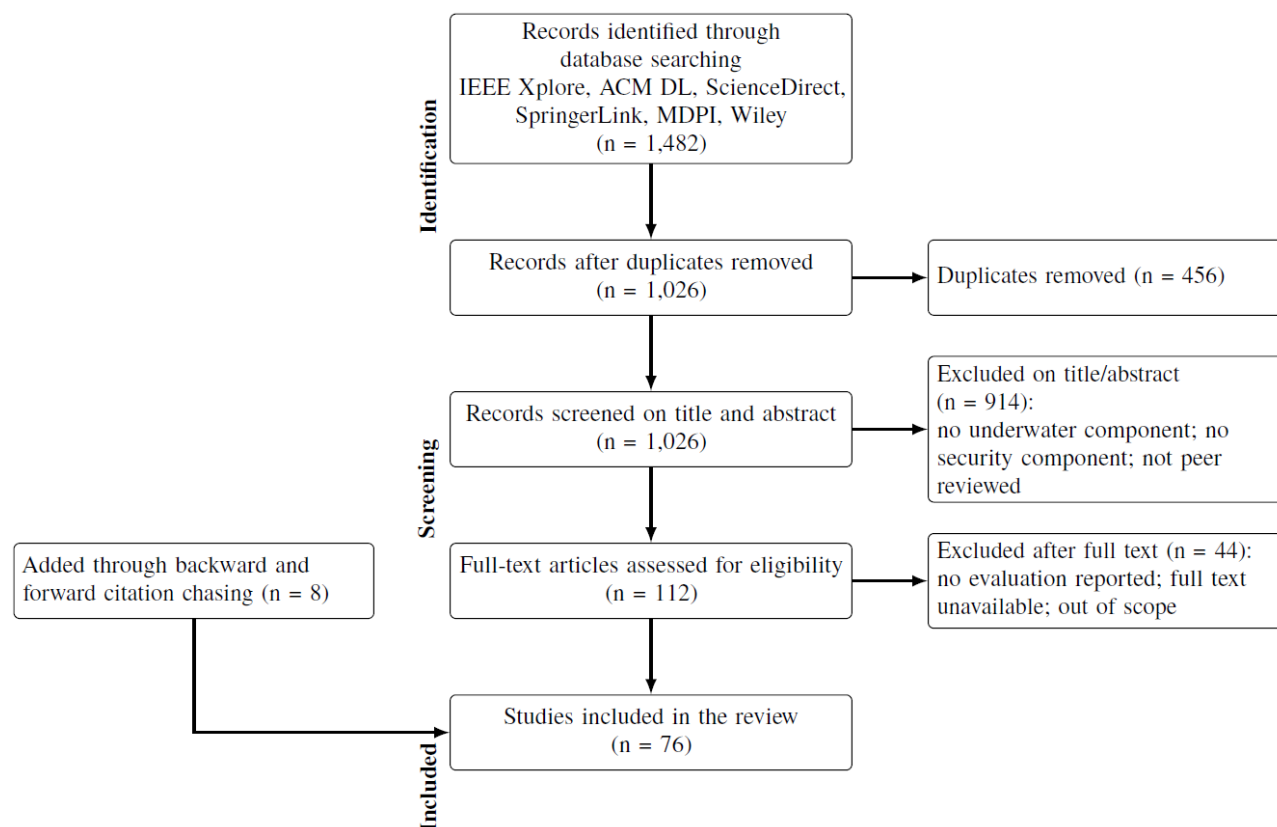


Figure 2 Flow of records through identification, screening, eligibility, and inclusion, following the PRISMA structure.

The retained set is dominated by recent work: of the 76 studies included in this review, 47 were published between 2020 and 2025 and 29 before 2020, with 59 journal articles and 17 conference papers. The older records are retained mainly for foundational material on the underwater channel and on early attack descriptions, which later work builds on rather than replaces.

2.5 Data Extraction

For each study, the following items were extracted where possible:

- the type of detection system
- the attack types addressed and the protocol layer targeted
- the dataset used, evaluation metrics, and its validation strategy
- energy cost, detection latency, and computational overhead

Items that a study does not report are recorded as not reported rather than estimated or inferred, which is presented later in Section 6.

2.6 Limitations of This Review

Three limitations should be noted. First, all screening and extraction was done by one person, so no independent check was applied. The criteria in Table 2 were used to determine the eligibility of a study, but the possibility of selection bias cannot be excluded. Second, restricting the search to English-language publications may have excluded relevant work published in other languages or released only as preprints or technical reports. Third, because the reviewed studies report different

metrics on different datasets, their results could not be aggregated into a single quantitative comparison. Section 6 compares them against a fixed set of criteria instead.

3. Background

The advancement of IoT networks has revolutionised the monitoring, data acquisition, and control of numerous systems, automated entire industries, and helped us develop smart cities. By connecting a large number of devices, operational efficiency and real-time responsiveness have been improved, which allows effective data-driven decision-making.

Extending this connectivity concept to underwater leads to IoUT [20], which aims to connect underwater nodes for improved marine monitoring, data collection, and control. Unlike IoT networks that rely on terrestrial communication networks, IoUT systems operate in underwater environments. Factors such as long multipaths, limited bandwidth, low data rate, and significant attenuation make the underwater channel one of the most complex and challenging communication environments and necessitate specialised approaches.

The integration of IoT with underwater communication nodes, such as Underwater Wireless Sensor Networks (UWSN) and underwater Autonomous Vehicles (AUVs), is very challenging, and the IoT networks developed for the terrestrial communication cannot be directly applied to underwater systems [21], whereby they require further adjustments since they do not incorporate channel propagation characteristics, such as long multipaths and high Doppler spread, in addition to the capacity of propagation channels and the availability of limited resources.

Early developments in underwater communication focused on point-to-point acoustic communication modems for short-range communication, primarily for defence. However, the surge in development of underwater sensor network has made interconnection between underwater nodes increasingly essential. With rises in offshore explorations, environmental monitoring and pollution control, underwater sensor networks, and remote broadcasting, the need for effective and reliable IoUT systems has become more critical than ever. The improvements in signal processing and power-efficient electronics have allowed for more robust data acquisition. Moreover, advances in underwater and cross-boundary communication make it possible to connect IoUT with terrestrial IoT networks, which expands coverage and reduces delay, and moves closer to the idea of the Internet of Everything (IoE) [22].

Despite the advancement of IoUT, the system faces significant hurdles. Underwater communication relies primarily on acoustic signals, which are susceptible to noise and interference caused by the channel's physicochemical properties, such as temperature gradients, salinity, and ocean currents. These challenges make UWSNs vulnerable to a range of attacks, making IoUT a distinct research area and demanding specialised security protocols, which are critical in identifying and mitigating unauthorised access and malicious activities. Moreover, due to the severe energy constraints in the deployment of underwater nodes, where devices are often battery-powered, it is difficult and costly to replace or recharge them. Therefore, energy-efficient tailored intrusion detection strategies are essential for ensuring the integrity and resilience of underwater networks [12].

This section is divided into multiple subsections. It first describes the terminology used across the field, then discusses the architecture of IoUT, and finally outlines the applications of IoUT systems and the challenges they face.

3.1 Terminology: Underwater IoT, IoUT, and UWSN

Three terms appear in this literature and are often used as if they meant the same thing. They do not, and the difference matters for what an IDS is expected to protect.

Underwater IoT, or underwater Internet of Things, is the informal term for the idea of extending IoT connectivity below the surface.

Internet of Underwater Things (IoUT) is the established name of the concept, introduced by Domingo [20]. It refers to the whole system: underwater sensor nodes, autonomous underwater vehicles, surface buoys and gateways, the acoustic, optical, and radio links between them, the surface and shore infrastructure that these connect to, and the applications and data services built on top. IoUT is therefore a network-of-networks that spans the underwater, surface, and terrestrial domains.

Underwater Wireless Sensor Network (UWSN) refers to the set of underwater nodes and the acoustic links among them, together with the routing, medium access, and localisation protocols that operate over those links. A UWSN can exist on its own, without being connected to any wider infrastructure. Every IoUT deployment contains at least one UWSN, but not every UWSN is part of an IoUT deployment.

A UWSN attack surface is confined to the underwater nodes, their acoustic links, and the protocols running over them, which is where jamming, Sybil, wormhole, sinkhole, and selective forwarding attacks operate. The IoUT attack surface additionally includes the surface gateway, the backhaul link to shore, the cloud/edge services that process the collected data, and the application interfaces exposed to users. Attacks such as application-layer abuse, credential compromise at the gateway, and tampering with the collected data once it reaches shore belong to the IoUT attack surface but not to the UWSN one.

In this paper, IoUT is used when referring to the complete system and its security, and UWSN is used when the discussion is confined to the underwater nodes and their acoustic links. Where a cited study uses one of these terms in a different sense, the usage of that study is noted. The term Underwater IoT is not used further.

3.2 Architecture

IoT interconnects devices worldwide, and IoUT extends this paradigm to underwater environments. Therefore, there are many similarities between them, such as the functionality and the architecture. However, the deployment and maintenance of an IoUT network is more complex than its counterpart. To address these gaps, we will discuss the basic architecture of IoUT and explain how it is different from IoT.

The architecture may include various underwater and surface nodes, including AUVs, sensory devices, remotely operated vehicles (ROVs), buoys, ships, and other similar nodes. In an IoUT network, nodes are interconnected with each other through the network layer, which allows information exchange.

IoUT infrastructure consists of three major communication components:

- Sensory nodes, which integrate with sensors, hydrophones, cameras, signal detectors, and storage devices.
- Data distribution nodes, including modems, relays, and repeaters.

- Sink Nodes, which are normally referred to as above-water nodes, such as offshore anchors and buoys.

In a small UWSN, a network of sensory nodes is deployed to measure underwater physical and chemical parameters. These nodes are interconnected and transmit the collected data to a local base node. However, since they are confined to a certain region and have a small transmission range, it is difficult to achieve comprehensive monitoring of the underwater environment. Therefore, multiple UWSNs are interconnected to form the IoUT, enabling real-time data acquisition and analysis on a global scale.

IoUT networks rely on acoustic signals for long-range communications; nevertheless, hybrid communication links have been proposed [23]. This hybrid approach combines magnetic induction (MI), optical and acoustic communication technologies to mitigate high signal attenuation.

Since an underwater channel is highly dispersive, these channel dynamics distort and disperse the signal, which creates major challenges for IoUT whilst relying on limited energy resources. Multiple energy-efficient IoUT network topologies have been proposed, which can be divided into three categories: aggregation methods, routing protocols, and scheduling algorithms [21]. Aggregation methods focus on data reduction to minimise the overhead. Routing protocols can be divided into tree, cluster, and chain-based routing approaches, each balancing energy efficiency, coverage, and reliability. Finally, scheduling algorithms optimise the resource allocation through priority-based scheduling and sleeping techniques. To mitigate channel interference and improve scalability, a multi-cluster-based network topology has been proposed. This multi-cluster topology was originally introduced for UWSNs [24] and later adapted for IoUT networks. In this scheme, the network is divided into multiple clusters; each cluster includes a cluster head supported by several relay nodes. The relay nodes connect sensor nodes to the cluster head, which performs routing and coordinates data transmission.

To further improve link reliability in energy-constrained underwater deployments, hop-count optimisation has also been proposed. The key idea is to increase the number of relay nodes for long-range communication, thereby reducing the hop distance between neighbouring nodes and improving link robustness. This optimisation considers the available power budget, bandwidth constraints, and transmission range to determine an appropriate number of relay nodes.

To ensure reliable operation and adapt to the dynamic channel, the IoUT network requires careful consideration. As shown in Figure 3, the main design considerations include:

- Overcoming dynamic channel variations.
- Maintaining network compatibility.
- Selecting an appropriate communication technology.
- Determining the network topology and deployment arrangement.
- Leveraging channel state information (CSI) and selecting an appropriate packet size.
- Mitigating channel dynamics (e.g., Doppler effects and time-varying multipath).
- Limited bandwidth in IoUT.
- Interfacing with existing IoT systems using predefined IoT protocols.
- Security threats and adversarial attacks.

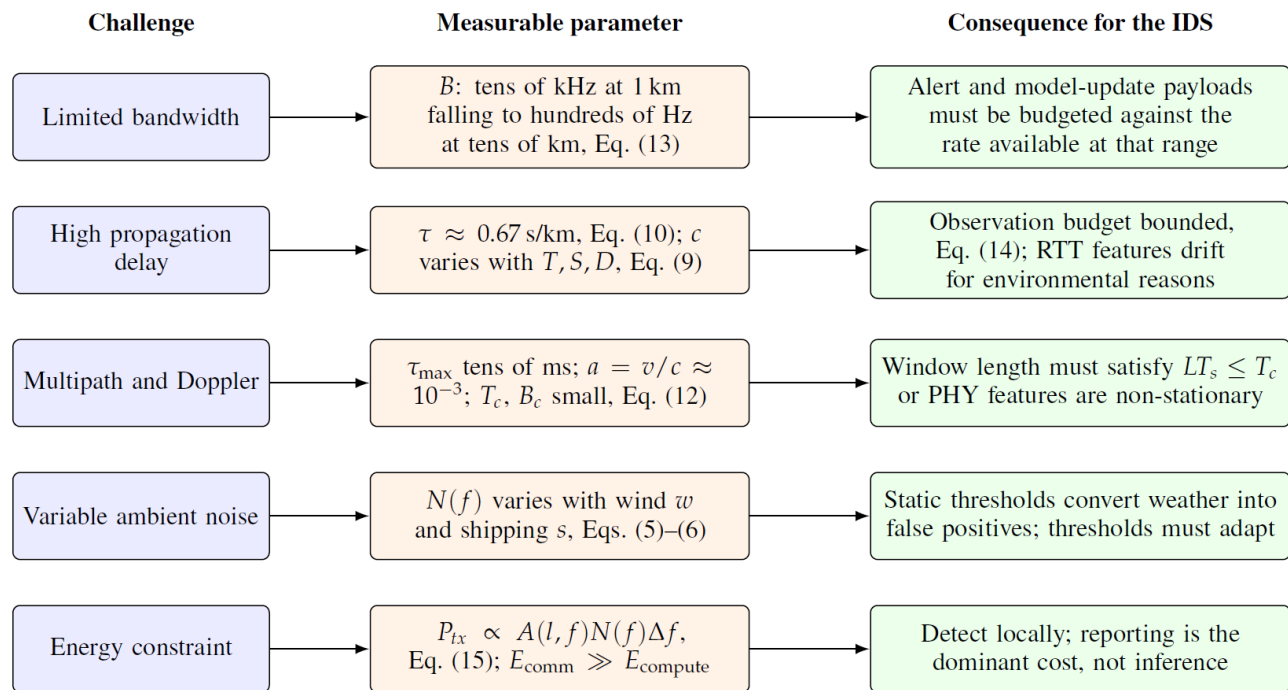


Figure 3 Challenges in IoUT design, each mapped to the measurable parameter that expresses it and to the consequence it imposes on an IDS.

3.3 Application of IoUT

IoUT networks facilitate the monitoring of the physical and chemical properties of the underwater environment, such as salinity, temperature, pressure, light conditions, pH, turbidity, water quality, pollutants, and oxygen levels. These parameters help us to understand oceanographic processes and marine ecosystems and provide data to predict the impact of environmental changes on marine life.

IoUT's development is driven by the growing recognition that underwater environments hold vast resources and critical data that can significantly impact numerous fields, including environmental science, climate change research, maritime security, and offshore energy production [25]. On a continuous and global scale, monitoring of parameters such as temperature, salinity, pollution levels, and ocean currents can provide insights into marine ecosystems and enable early detection of environmental irregularities. In addition, reliable monitoring and maintenance of pipelines, undersea cables, and oil rigs depend heavily on the robustness of the IoUT network. Figure 4 summarises the application of IoUT.

Category	Application
Environmental Monitoring	Water-quality monitoring, pollution detection, salinity and temperature monitoring.
Climate Change Research	Ocean-current analysis; monitoring of light, pH, and related environmental parameters.
Maritime Security	Surveillance and situational awareness; early detection of anomalies.
Offshore Energy & Infrastructure	Pipeline and rig maintenance; subsea cable health monitoring.

Figure 4 Applications of IoUT.

3.4 Challenges Faced in Designing IoUT Networks

The propagation characteristics of the underwater channel, and the limits they place on bandwidth, delay, and energy, are derived quantitatively in Section 4 and are not restated here. This subsection covers the remaining design challenges, which arise from the physical deployment rather than from the channel.

Underwater IoUT nodes normally operate under extremely high pressure. Moreover, the unintended growth and accumulation of algae, combined with the effect of corrosive saltwater, can degrade node performance, and deployment at remote locations further complicates routine maintenance and battery replacement [26].

Therefore, self-management is another significant challenge to the IoUT network. Self-management refers to the ability of a node to configure, heal, optimise, or protect itself from the harsh underwater environment and ensure reliable communication for long periods [20].

The energy cost of long-range communication is given by eq. (15), and dominates the total detection energy in eq. (16). Moreover, because replacing batteries underwater is difficult, IoUT applications require long node operation to avoid expensive replacement [27, 28].

The multi-cluster approach provides good performance by optimising IoUT operations, which reduces power consumption and provides network reliability and stability [24]. The design of a clustering approach must account for various underwater environmental factors, including communication distance, the positions and depths of nodes or vehicles, residual energy levels, traffic load, and signal quality. However, if the clusters are not managed properly, this strategy can increase the overall energy consumption and compromise network performance since the improper cluster head selection can deteriorate the energy efficiency of an IoUT network [29].

As IoUT deployments become increasingly widespread and complex, the need for precise node localisation and efficient resource allocation grows more critical [30]. Ensuring that underwater devices know their positions to one another not only supports reliable data routing but also optimises energy consumption—a key concern in resource-constrained underwater settings.

4. The Underwater Acoustic Channel and IDS Design Requirements

This section models the underwater acoustic channel quantitatively and derives the design requirements an IDS must satisfy.

4.1 Absorption and Path Loss

The acoustic channel has geometric spreading and chemical absorption, which lead to energy loss. The absorption coefficient $\alpha(f)$, in dB/km for frequency f in kHz, is given by Thorp's empirical formula [31]:

$$\alpha(f) = \frac{0.11f^2}{1 + f^2} + \frac{44f^2}{4100 + f^2} + 2.75 \times 10^{-4}f^2 + 0.003. \quad (1)$$

The first two terms account for the relaxation of boric acid and magnesium sulphate respectively, the third for viscous absorption, and the fourth is a low-frequency residual. In eq. (1) α increases monotonically with frequency.

Total path loss over a distance l follows the Urlick model [32]:

$$A(l, f) = l^k a(f)^l, \quad a(f) = 10^{\alpha(f)/10}, \quad (2)$$

or in decibels,

$$10 \log A(l, f) = k \cdot 10 \log l + l \cdot \alpha(f), \quad (3)$$

where k is the spreading factor, taken as $k = 1$ for cylindrical spreading, $k = 2$ for spherical spreading, and $k = 1.5$ as a practical value for shallow-water deployment. The second term in eq. (3) has no terrestrial equivalent: the absorption loss in decibels grows linearly with distance, and its rate $\alpha(f)$ rises steeply with frequency.

4.2 Ambient Noise

Ambient noise is modelled as the sum of four independent noise sources [33]:

- turbulence
- distant shipping
- surface agitation from wind and waves
- thermal noise

With power spectral densities in dB re μPa per Hz and f in kHz,

$$10 \log N_t(f) = 17 - 30 \log f, \quad (4)$$

$$10 \log N_s(f) = 40 + 20(s - 0.5) + 26 \log f - 60 \log(f + 0.03), \quad (5)$$

$$10 \log N_w(f) = 50 + 7.5 w^{1/2} + 20 \log f - 40 \log(f + 0.4), \quad (6)$$

$$10 \log N_{th}(f) = -15 + 20 \log f, \quad (7)$$

where $s \in [0, 1]$ is a shipping activity factor and w is wind speed in m/s. The total is

$$N(f) = N_t(f) + N_s(f) + N_w(f) + N_{th}(f). \quad (8)$$

Across the 100 Hz to 100 kHz range, the total exhibits an approximate 18 dB/decade roll-off. Both s and w depend on the environment and change over time, so a detector that assumes a fixed noise floor will report false alarms.

4.3 Propagation Delay, Sound Speed, and Doppler

Sound speed in seawater depends on temperature T in $^\circ\text{C}$, salinity S in ppt, and depth D in metres. The Mackenzie equation [34] gives

$$\begin{aligned} c = & 1448.96 + 4.591 T - 5.304 \times 10^{-2} T^2 + 2.374 \times 10^{-4} T^3 \\ & + 1.340 (S - 35) + 1.630 \times 10^{-2} D + 1.675 \times 10^{-7} D^2 \\ & - 1.025 \times 10^{-2} T(S - 35) - 7.139 \times 10^{-13} T D^3. \end{aligned} \quad (9)$$

This yields $c \approx 1500$ m/s, so the one-way propagation delay

$$\tau = \frac{l}{c} \quad (10)$$

is about 0.67 s per kilometre, against 3.3 μ s per kilometre for radio. Underwater propagation delay is therefore about five orders of magnitude larger, and is not negligible against protocol timing. It also varies with temperature: by eq. (9), a change of a few degrees shifts c by several metres per second. Several detectors use time-of-flight or round-trip-time as their anomaly signal, and these values drift with the environment alone, which can cause false detections.

The low speed of sound makes the Doppler variation large. For a relative node with velocity v , the Doppler scaling factor is

$$a = \frac{v}{c}. \quad (11)$$

A drift of 1.5 m/s gives $a \approx 10^{-3}$, compared with $a \approx 5 \times 10^{-9}$ for a terrestrial radio link. In an underwater channel, Doppler variation does not only produce a frequency offset; it also compresses or dilates the waveform itself, which requires resampling at the receiver.

With a maximum multipath delay spread τ_{max} and Doppler spread B_d , the coherence bandwidth and coherence time are

$$B_c \approx \frac{1}{\tau_{max}}, \quad T_c \approx \frac{1}{B_d}. \quad (12)$$

τ_{max} typically reaches tens of milliseconds and B_d reaches several hertz, so both B_c and T_c are small. The channel is spread in both time and frequency, which is what is meant by describing it as doubly spread [35].

4.4 Signal-to-Noise Ratio and the Range-Dependent Usable Band

For a transmitted power spectral density $S(f)$ and receiver noise bandwidth Δf , the narrowband signal-to-noise ratio at distance l is

$$\text{SNR}(l, f) = \frac{S(f)}{A(l, f) N(f) \Delta f}. \quad (13)$$

Because $A(l, f)$ increases with frequency while $N(f)$ decreases, the factor $1/(A(l, f)N(f))$ has a maximum at a frequency that depends on range [36]. The usable bandwidth is therefore a function of range, and it narrows as range increases.

4.5 From Channel Properties to Detector Requirements

Table 3 maps each of the quantities above to the design constraint it imposes on an IDS. The following three have the largest effect on how a detector must be designed.

Table 3 Channel properties, their governing relations, and the requirements they place on an IoUT IDS.

Property	Relation	Measurable effect	Requirement placed on the IDS
Frequency-based absorption	eqs. (1), (2)	Absorption loss grows with range, and faster at higher frequency	Feature extraction must not assume a constant absorption factor
Range-dependent usable band	eq. (13)	Available bandwidth falls from tens of kHz to hundreds of Hz over tens of km	Payloads must be budgeted against the link rate
Environmental ambient noise	eqs. (5)-(8)	Noise floor varies with shipping and wind, uncontrolled	Decision thresholds must adapt
Propagation delay	eqs. (10), (14)	≈ 0.67 s/km, five orders above radio	Observation budget is bounded by range and deadline
Sound speed variation	eq. (9)	c shifts with temperature, salinity, and depth	Time-of-flight and RTT anomaly signals must be compensated for environment
Doppler scaling	eqs. (11), (12)	$\alpha \approx 10^{-3}$; waveform compression, short coherence time	Physical-layer variations are unstable across coherence intervals
Energy asymmetry	eqs. (15), (16)	$E_{comm} \gg E_{compute}$ at realistic ranges	Detection should be local

First, detection is delay-bounded. If a decision must be reached within a deadline T_d , and observations arrive at interval T_s after a propagation delay τ , the number of observations available to the detector is

$$N_{obs} \leq \left\lfloor \frac{T_d - \tau}{T_s} \right\rfloor. \quad (14)$$

Detection performance improves with N_{obs} , so eqs. (10) and (14) together place a hard ceiling on what any detector can achieve at a given range and deadline.

Second, reporting is more expensive than detecting. The transmit power required to sustain a target SNR_{req} over distance l follows from eq. (13):

$$P_{tx}(l, f) \geq SNR_{req} A(l, f) N(f) \Delta f. \quad (15)$$

Because $A(l, f)$ grows with l and f as in eq. (2), the energy cost of sending an alert to a gateway grows quickly with range, while the energy cost of evaluating a model locally does not. The total detection energy

$$E = E_{sense} + E_{compute} + E_{comm} \quad (16)$$

is therefore dominated by E_{comm} in almost any realistic deployment. This is the quantitative basis for preferring local or distributed detection over centralised detection underwater, and it applies to federated learning, as shown in Section 6.4.

Third, the environment can generate its own anomalies. The parameters s , w , T , S , and D in eqs. (5), (6), and (9) are dynamic and are not under network control. Any detector whose decision boundary is fixed at training time will drift.

4.6 Detection Performance under an Underwater Prior

Intrusion detection is a binary hypothesis test. Let H_0 denote normal operation and H_1 the presence of an intrusion, and let $\mathbf{x}$ be the observation vector. The Neyman-Pearson detector compares the likelihood ratio against a threshold γ [37]:

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x} | H_1)}{p(\mathbf{x} | H_0)} \underset{H_0}{\overset{H_1}{\geq}} \gamma, \quad (17)$$

with detection and false alarm probabilities

$$P_D = \Pr\{\Lambda > \gamma | H_1\}, \quad P_{FA} = \Pr\{\Lambda > \gamma | H_0\}. \quad (18)$$

For the canonical case of a known signal in additive Gaussian noise with deflection coefficient d^2 ,

$$P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{d^2}\right), \quad (19)$$

where $Q(\cdot)$ is the Gaussian tail function and d^2 grows with N_{obs} . eq. (14) therefore enters eq. (19) directly: range and decision deadline bound d^2 , and so bound the achievable operating point on the ROC curve.

The second consequence concerns how results are reported. Writing $\pi = \Pr\{H_1\}$ for the prior probability of intrusion, the precision of the detector is

$$Precision = \frac{\pi P_D}{\pi P_D + (1 - \pi) P_{FA}}, \quad (20)$$

and its accuracy is

$$Accuracy = \pi P_D + (1 - \pi)(1 - P_{FA}). \quad (21)$$

In an operational underwater network, intrusion events are rare, so π is small. Suppose $\pi = 10^{-3}$, $P_D = 0.95$, and $P_{FA} = 0.01$. eq. (21) returns an accuracy of 0.990, which would be reported as a strong result. eq. (20) returns a precision of 0.087: fewer than one alert in eleven is genuine. eq. (21) also shows that a detector which simply declares H_0 always achieves an accuracy of $1 - \pi = 0.999$ while detecting nothing.

This matters for reading the literature. Reported accuracies above 95% are common in the studies reviewed here, but the proportion of intrusions in the evaluation data is rarely stated. On a balanced or near-balanced dataset, $\pi \approx 0.5$ and eq. (21) is well behaved, but such figures do not transfer to a deployment where π is three or four orders of magnitude smaller. The point is not that the reported results are wrong, but that accuracy on a balanced dataset does not predict field

performance, and that the operating π is almost never stated. Section 6 returns to this when comparing the reviewed methods.

5. Types of Intrusions for IoUT

The IoUT interconnects physical devices equipped with embedded sensors and onboard software to support data collection, analysis, and decision-making. Recently, IoUT has received significant research attention. However, the underwater channel imposes constraints that are very different from the terrestrial IoT. In particular, radio and optical signals, which are widely used in terrestrial networks, are generally unsuitable for underwater because of severe attenuation and absorption. Hence, most UWSNs use acoustic communication. Although acoustic links support longer transmission ranges, they come with the constraints quantified in Section 4.

Therefore, these limitations, combined with the limited power, memory, and processing capability of underwater nodes, make IoUT networks more exposed to security threats. Protecting data confidentiality and integrity in such conditions requires sophisticated intrusion detection schemes that are specifically designed for underwater systems. Recent progress in high-performance computing and machine learning has also strengthened physical-layer security by using channel characteristics to improve resistance to eavesdropping and tampering.

Accordingly, intrusions in IoUT can be categorised based on the protocol layer, as summarised in Figure 5. Adam et al. [3] identified the types of intrusions at each network level, and these are described in Sections 5.1 to 5.4.

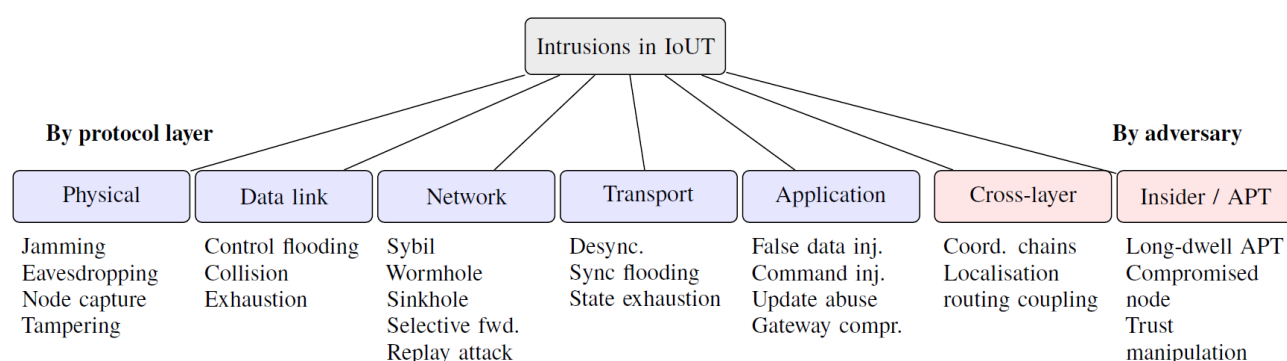


Figure 5 Taxonomy of intrusions in IoUT. The five protocol layers are shown on the left, the categories defined by adversary rather than by layer are shown on the right.

Three further categories are needed to describe the threats faced by a complete IoUT system. The first is the application layer, which a UWSN does not have but which is part of the IoUT attack surface as defined in Section 3.1. The second is cross-layer attacks, whose defining property is that no single-layer detector observes an anomaly large enough to trigger. The third covers threats that are defined by the adversary rather than by the protocol: insider threats from nodes that hold valid credentials, and advanced persistent threats that exploit the long maintenance intervals of underwater deployment. Sections 5.5 to 5.8 address these.

Section 5.9 then draws the categories together into a threat model that links each attack to the layer it acts on, the underwater property that makes it viable, the signal it produces that a detector could observe, and the countermeasures reported against it. Section 5.10 compares the attacks on severity, likelihood, scope, and attacker cost.

Table 4 summarises the impact of cyber-attacks across the protocol layers.

Table 4 Layer-wise intrusions in IoUT.

Layer	Attacks	Primary impact	Common mitigation directions
Physical/Device	Jamming, eavesdropping, physical tampering	Throughput collapse, energy drain, data leakage	FHSS/SS, duty-cycling, physical-layer security, tamper-aware design
Data link (MAC)	Control flooding, collisions, exhaustion	Contention increase, retransmissions, battery drain	Rate limiting, MAC authentication, adaptive backoff, anomaly monitoring
Network	Sybil, wormhole, sinkhole, selective forwarding, MitM	Routing disruption, packet loss, localisation errors	Trust models, secure routing, consistency checks, multipath routing
Transport	Desync, sync flooding, replay affecting timing	Connection failure, state exhaustion, degraded reliability	SYN cookies, stateless design, replay protection, robust sync
Application	False data injection, command injection, firmware and update abuse, credential compromise at the gateway	Corrupted measurements, unauthorised actuation, loss of control over nodes and vehicles	Message authentication, signed updates, command authorisation, plausibility checks against physical models
Cross-layer	Coordinated low-rate jamming, induced route flapping, energy depletion chains	Battery drain, persistent instability, evasion of single-layer detectors	Joint statistics across layers, correlated anomaly scoring, shared state between detectors

5.1 Physical and Device-Level Intrusions

At the physical layer, attackers can exploit the doubly spread underwater acoustic channel to enhance the attacks. A primary threat is jamming, in which disruptive signals are injected into the channel to block communication [38-41]. Moreover, it interrupts data delivery and significantly shortens the lifetime of the nodes by forcing them to retransmit [41].

Eavesdropping is another physical layer threat, where an attacker intercepts transmitted signals to obtain sensitive information. In underwater acoustic nodes, both attacks and defences are strongly influenced by channel impairments. Analytical models have therefore been used to characterise eavesdropping and intrusion probability under realistic channel conditions [42]. To increase robustness against interference, techniques such as spread spectrum, frequency-hopping spread spectrum, and direct-sequence spread spectrum have been widely explored [43]. Additional mitigation strategies include adaptive duty-cycling, which switches nodes to sleep mode during an attack and resumes operation afterwards to limit energy loss. More recent approaches integrate

machine learning to adjust protection policies as channel conditions and attack patterns evolve. Covert communication techniques are also used to hide the transmissions.

5.2 Data-Level Intrusions

At the data link layer, MAC protocols coordinate access to the shared acoustic channel under severe bandwidth constraints. Compared to physical-layer jamming, MAC-layer jamming can be more strategic and protocol-aware. For example, an attacker may repeatedly trigger control exchanges (e.g., RTS/CTS behaviour) to disrupt coordination among nodes and reduce overall throughput. *Collision attacks* occur when packets are transmitted in a way that intentionally causes collisions, corrupting frames and increasing retransmissions. An *exhaustion attack* keeps the channel busy and forces a victim to expend energy on repeated reception, processing, and retransmission. This may be achieved by flooding the network with excessive requests, which increases channel contention and drains node batteries.

5.3 Network-Level Intrusions

Network level threats in IoUT mainly target routing and packet forwarding. A replay attack occurs when an attacker retransmits valid messages to mislead the receiver or interfere with normal protocol operation. In selective forwarding, a compromised node appears to behave normally but deliberately drops packets, which lowers the overall throughput while making the attack difficult to detect [44]. *Man-in-the-middle (MitM)* attacks are also a serious concern because an attacker can intercept and modify communication between legitimate nodes.

Other common network layer attacks include Sybil, wormhole, and sinkhole attacks. In a *Sybil attack*, one adversary assumes multiple identities to disrupt trust relationships and influence routing decisions [45-47]. In a *wormhole attack*, messages are tunnelled between two distant points in the network to create the false impression of a shorter route. This can seriously affect routing, localisation, and data integrity [45, 48-50]. Wormhole attacks may also support denial-of-service behaviour by enabling packet dropping, traffic analysis, or other coordinated attacks. In a *sinkhole attack*, a malicious node advertises an attractive route to draw in surrounding traffic and then drops or manipulates the packets it receives.

Network security mechanisms are intended to maintain communication performance under attack by reducing packet loss and preserving throughput [51, 52]. IDS play an important role by monitoring network behaviour and identifying suspicious activity [53]. In practice, IDS is often combined with encryption, network segmentation, and anomaly detection to improve resilience. Effective network-layer defence is therefore essential for reliable routing and secure data exchange in IoUT deployments.

5.4 Transport-Layer Intrusions

At the transport layer, IoUT is vulnerable to DoS-style attacks such as *desynchronisation* and *synchronisation flooding*. A desynchronisation attack disrupts an established connection by injecting forged packets with incorrect sequence information, causing the nodes to lose alignment and trigger repeated retransmissions. Since underwater networks often rely on GPS-free synchronisation mechanisms [54], manipulation of timing and control information can have an impact on the

network. In addition, Sybil and replay attacks can generate false measurements that degrade time synchronisation.

A synchronisation flooding attack targets protocols that maintain per-connection state by repeatedly sending connection requests until resource limits are reached. The resulting traffic surge overloads the network and makes it difficult to distinguish legitimate sessions from malicious activity. One possible mitigation is the use of connectionless transport protocols; however, this can reduce transport-layer functionality required by applications. A common defence is the use of SYN cookies [55], which avoids allocating server-side state until a client proves reachability. While effective against certain flooding behaviours, this approach can increase computation and introduce additional overhead.

5.5 Application-Layer Intrusions

As defined in Section 3.1, a UWSN has no application layer of its own. An IoUT system does: the surface gateway, the backhaul to shore, and the data services built on the collected measurements are all part of the system, and all are reachable by an adversary.

False data injection targets the integrity of the measurements. An adversary that controls a node can make it report false data, and because the node still holds valid credentials the readings pass every check the network applies, so downstream decisions are corrupted. This is difficult to detect by traffic analysis, because packet rates and routing behaviour remain normal. *Command injection and unauthorised actuation* target the downlink, and could give an adversary control of AUVs and actuated underwater nodes. *Firmware and update abuse* exploits a constraint specific to this setting. Updating a deployed node over an acoustic link is slow, so update mechanisms are designed for minimum payload, and signature verification is sometimes skipped to save bandwidth and energy. An unauthenticated update path gives an adversary control of a node that cannot be physically reached until the next maintenance cycle. *Gateway and service compromise* targets the gateways, which connect the underwater nodes to the surface. An adversary at the gateway sees all the traffic and can inject malicious data in either direction.

5.6 Cross-Layer Attacks

Sections 5.1 to 5.5 treat each layer as an independent target. A cross-layer attack affects several layers, but its effect at each individual layer stays below that layer's detection threshold.

Consider an adversary that uses low-rate jamming at the physical layer, at a low duty cycle, so that the resulting increase in bit error rate stays within the variation that wind noise already produces eq. (6). The MAC layer retransmits the corrupted frames, which is normal behaviour, and the rate of retransmissions stays low. However, the retransmissions increase the energy cost of the node. The routing layer may then select an alternative path, and the adversary repeats the pattern on that path. This results in route flapping and battery depletion.

Let T_i be the test statistic of a detector operating at layer i with threshold γ_i . A layer-wise detection scheme raises an alarm when

$$\exists i: T_i > \gamma_i, \quad (22)$$

and a cross-layer attack is one for which $T_i \leq \gamma_i$ holds for every i while a joint statistic $T_{joint}(T_1, \dots, T_n)$ exceeds its own threshold. Lowering the individual γ_i to catch such an attack raises the per-layer false alarm rate, and by eq. (20) the operating precision degrades sharply. Detection therefore requires the joint statistic rather than per-layer thresholds.

Localisation-routing coupling exploits the dependence of underwater routing on position estimates. An attack on the ranging or timing used for localisation propagates into routing decisions without ever appearing as a routing anomaly.

5.7 Insider Threats and Compromised Nodes

An insider node is a trusted node inside a network that holds valid credentials but is controlled by an adversary. It is difficult to identify, because the node still authenticates correctly. Insider threats arise in IoUT through node capture, where an adversary physically recovers a deployed node, and through firmware compromise, as described in Section 5.5.

To identify a compromised node, trust and reputation models calculate node scores from observed behaviour rather than from credentials. These models are complex to deploy in an underwater network, and their design is outside the scope of this study.

A malicious node and a failing node produce similar behavioural signatures. A node suffering biofouling, sensor drift, corrosion, or battery decline will drop packets, report implausible values, and respond inconsistently, which is also the profile of selective forwarding or false data injection.

The consequence is that a trust model tuned to isolate malicious nodes will also isolate degraded nodes, reducing coverage in a network that was already sparse, while a trust model tuned to tolerate degradation could allow an insider node to operate. None of the trust-based schemes reviewed in Section 6 reports how it separates these two cases, and we identify this as an open problem in Section 9.

5.8 Advanced Persistent Threats

An advanced persistent threat is characterised by an adversary pursuing a long-term objective while avoiding detection. IoUT networks are exposed for two reasons. First, dwell time is long: underwater deployments are serviced after long intervals, so a node compromised in a deployment cycle may remain compromised. Second, reconnaissance is passive and therefore undetectable: acoustic eavesdropping requires only a hydrophone within range and emits nothing, so it leaves no trace on the network.

A typical example of this attack begins with passive acoustic data collection to map the topology and traffic patterns, followed by compromise of a small number of nodes through capture or update abuse, firmware modification, and manipulation at a rate low enough to remain within normal variation. Detecting this pattern is extremely difficult and requires sophisticated resources. However, an underwater node has neither the storage to retain a long observation history nor the energy budget. Detection must therefore be performed where the history can be held, which is at the gateway or on shore. None of the detection methods reviewed in Section 6 operates on a timescale longer than a single evaluation run.

5.9 Threat Model

Table 5 sets out the threat model. Each attack is linked to the layer or layers it acts on, the underwater property that makes it viable, the observable signal a detector could use, the countermeasures reported in the reviewed literature, and the limitation of those countermeasures. The third column distinguishes this from a general IoT threat model: in most cases the attack is not new, but the property that makes it hard to detect is specific to the acoustic channel.

Table 5 Threat model for IoUT, linking each attack to the layer it acts on, the underwater property that makes it viable, the signal available to a detector, and the reported countermeasures with their limitations.

Attack	Layer	Underwater property exploited	Observable behaviour	Reported countermeasures	Limitation
Jamming	Physical	Narrow usable band eq. (13)	SNR drop and raised BER	Spread spectrum, frequency hopping, duty-cycling	Ineffective against low-rate jamming below threshold
Eavesdropping	Physical	Passive reception and long acoustic range	None, as attack listens for signals	Physical-layer security, covert communication, encryption	Undetectable by design; only prevention is possible
Node capture	Physical/ device	Nodes recoverable by vessel or ROV	Absence, then return with valid credentials	Tamper-resistant hardware, key rotation	Rotation limited by link budget; capture may not be observable at all
Control flooding	Data link	Long propagation delay inflates handshake windows	Elevated control-to-data ratio	Rate limiting, MAC authentication	Legitimate retransmission under poor channel resembles flooding
Collision/ exhaustion	Data link	Shared low-rate channel; energy cost of reception	Rising retransmission and collision counts	Adaptive backoff, anomaly monitoring	Indistinguishable from congestion caused by environmental noise eq. (6)
Sybil	Network	Identity binding is weak without reliable localisation	Inconsistent position and timing claims	Trust models, position verification	Position verification degraded by sound speed variation eq. (9)
Wormhole	Network	High propagation delay makes a shortcut highly attractive to routing	Route latency inconsistent with geometric distance	Distance bounding, consistency checks	Distance bounding requires accurate c , which varies eq. (9)

Sinkhole	Network	Sparse topology; few alternative routes	Abnormal route advertisement and traffic concentration	Secure routing, multipath routing	Multipath routing costs energy the deployment does not have eq. (15)
Selective forwarding	Network	High baseline packet loss masks deliberate dropping	Delivery ratio below neighbourhood baseline	Trust models, watchdog monitoring	Cannot be separated from environmental loss or node degradation
Replay	Network/ transport	Long delay widens the acceptance window for stale packets	Duplicate sequence numbers; timing inconsistency	Nonces, timestamps, sequence checks	Timestamping requires synchronisation, itself an attack target
Desynchronisation	Transport	GPS-free synchronisation; delay variability	Repeated retransmission; state loss	Robust synchronisation, replay protection	Synchronisation error and attack produce the same symptom
False data injection	Application	No independent ground truth underwater	Values inconsistent with physical models	Plausibility checks against environmental models	Requires a validated channel and environment model; rarely implemented
Command injection	Application	Downlink authentication weakened to save bandwidth	Commands outside mission envelope	Command authorisation, envelope checks	Consequences are physical and may be irreversible before detection
Update abuse	Application	Update payloads minimised for a low-rate link	Unexpected firmware version or behaviour change	Signed updates, staged rollout	Signature verification cost traded against link budget
Gateway compromise	Application	Gateway is the only high- bandwidth point in the system	Anomalies in shore- side traffic and access patterns	Standard IT controls, segmentation	Outside the scope of every IoUT IDS reviewed here

Cross-layer chain	Multiple	Per-layer tolerances widened to accommodate a poor channel	Only visible in joint statistics eq. (22)	Cross-layer detection, correlated scoring	No reviewed IoUT IDS implements a joint statistic
Insider node	Multiple	Long service intervals; degradation mimics malice	Behavioural deviation from neighbourhood baseline	Trust and reputation models	Cannot separate malicious from failing nodes
Advanced persistent threat	Multiple	Months-long dwell time; passive reconnaissance; no patch bandwidth	Weak evidence correlated over months	None reported for IoUT	Node storage and energy cannot support long-horizon correlation

5.10 Comparative Risk Assessment

Table 6 compares these attacks on four dimensions. No standardised scoring system exists for underwater networks, and applying a terrestrial framework such as CVSS directly would misstate the likelihood since it does not account for acoustic range, deployment cost, or maintenance interval.

Table 6 Comparative risk assessment of IoUT attacks. Ratings are qualitative and assigned against the definitions given in Section 5.10.

Attack	Severity	Likelihood	Scope	Attacker cost	Basis for rating
Jamming	High	High	Cluster	Low	Commodity projector within acoustic range disables communication; no credentials needed; widely reported
Eavesdropping	Medium	High	Network	Low	Passive hydrophone suffices and emits nothing; severity limited to confidentiality where payloads are encrypted
Node capture	High	Low	Network	High	Yields valid key material and persistent access, but requires vessel or ROV time and knowledge of Node position
Control flooding	Medium	Medium	Cluster	Low	Degrades throughput and drains energy; requires protocol access but no credentials
Collision/exhaustion	Medium	Medium	Node	Low	Effect is local and recoverable, but energy loss is not; easily confused with congestion
Sybil	High	Medium	Network	Medium	Corrupts routing and trust simultaneously; requires the ability to originate multiple identities
Wormhole	High	Medium	Network	Medium	Distorts routing and localisation together; requires two coordinated positions and an out-of-band link
Sinkhole	High	Medium	Cluster	Medium	Concentrates traffic for dropping or inspection; sparse topology makes it effective with one node
Selective forwarding	Medium	Medium	Cluster	Medium	Individually small loss but sustained and very hard to attribute against a High loss baseline
Replay	Medium	Medium	Cluster	Low	Requires only capture and retransmission; wide acceptance windows make it practical
Desynchronisation	Medium	Low	Node	Medium	Disrupts an established connection but requires accurate sequence state
False data injection	High	Medium	Network	Medium	Corrupts the output of the deployment while leaving all Network metrics normal

Command injection	High	Low	Network	High	Physical and potentially irreversible consequences, but requires downlink access and mission knowledge
Update abuse	High	Low	Network	High	Grants persistence for the remainder of the deployment cycle; requires access to the update path
Gateway compromise	High	Medium	Network	Medium	Exposes the entire deployment and needs no underwater capability; terrestrial attack surface
Cross-layer chain	High	Low	Network	High	Evades every reviewed detector by construction, but requires detailed knowledge of the deployed thresholds
Insider Node	High	Low	Network	High	Bypasses authentication entirely; requires prior compromise by capture, update abuse, or supply chain
Advanced persistent threat	High	Low	Network	High	Combines several of the above with long dwell time; assumes a state-level or comparably resourced adversary

The dimensions are defined as follows. *Severity* is the consequence if the attack succeeds: high denotes loss of network function or of data integrity across the deployment, medium denotes degraded performance or localised loss, and low denotes a limited or recoverable effect. *Likelihood* reflects how easily the attack can succeed given the channel, and how often it is reported in the studies. High likelihood denotes an attack requiring only off-the-shelf acoustic hardware, medium denotes one requiring network access or knowledge of the protocol, and low denotes one requiring node compromise or physical access. *Scope* is the extent of the effect. *Attacker cost* covers equipment, vessel time, positioning, and prior knowledge.

Two observations follow from Table 6 and are relevant to the future directions in Section 9. The attacks with the lowest attacker cost, jamming and eavesdropping, are also among the most likely, and neither is well addressed by the detection methods reviewed in Section 6, which concentrate on network-layer attacks. The attacks with the highest severity are concentrated at the application layer and in the multi-layer categories, which go beyond the four network layers usually used to classify attacks.

6. Intrusion Detection Approaches and Methods

IoUT-based IDS have been developed to counter security threats and improve overall network security. Early works between 2015 and 2020 focus mainly on basic security mechanisms, before shifting towards intrusion detection. The growing risk of wormhole, black hole, and Sybil attacks led researchers to develop adaptive anomaly detection methods. For example, the Virtual-Lattice Monitoring (VLM) algorithm [56] used both mobile and static nodes to detect intrusions in sparse underwater networks, while hybrid optical-acoustic clustering methods [57] improved the detection of wormhole attacks.

From 2021, ML and optimisation-based IDS models were presented. Moreover, energy-efficient and cluster-based frameworks were introduced to improve detection accuracy while reducing the impact on network lifetime. A recent study highlighted the need for advanced detection methods [19, 58], while techniques such as SMOTE-based data balancing and random forest classifiers [59] enhanced the detection of rare and multitype attacks, ensuring greater resilience against sophisticated threats.

Federated learning (FL), blockchain-based security, and hybrid deep learning models were introduced to further improve the performance. IDS frameworks based on FL [18, 60] enabled distributed anomaly detection in UWSNs. Moreover, CNN-Bi-LSTM architectures optimised with NADAM and AdamW techniques emerged as powerful tools for spatial-temporal attack detection, to improve detection efficiency and adaptability in dynamic underwater environments.

Currently, IoUT networks combine multiple ML-based IDS approaches, which include RL, FL, and clustering-based detection [18, 59, 61, 62]. Despite these advancements, practical validation of federated and blockchain-based IDS remains limited. Figure 6 shows a generic IDS workflow used in IoUT systems.

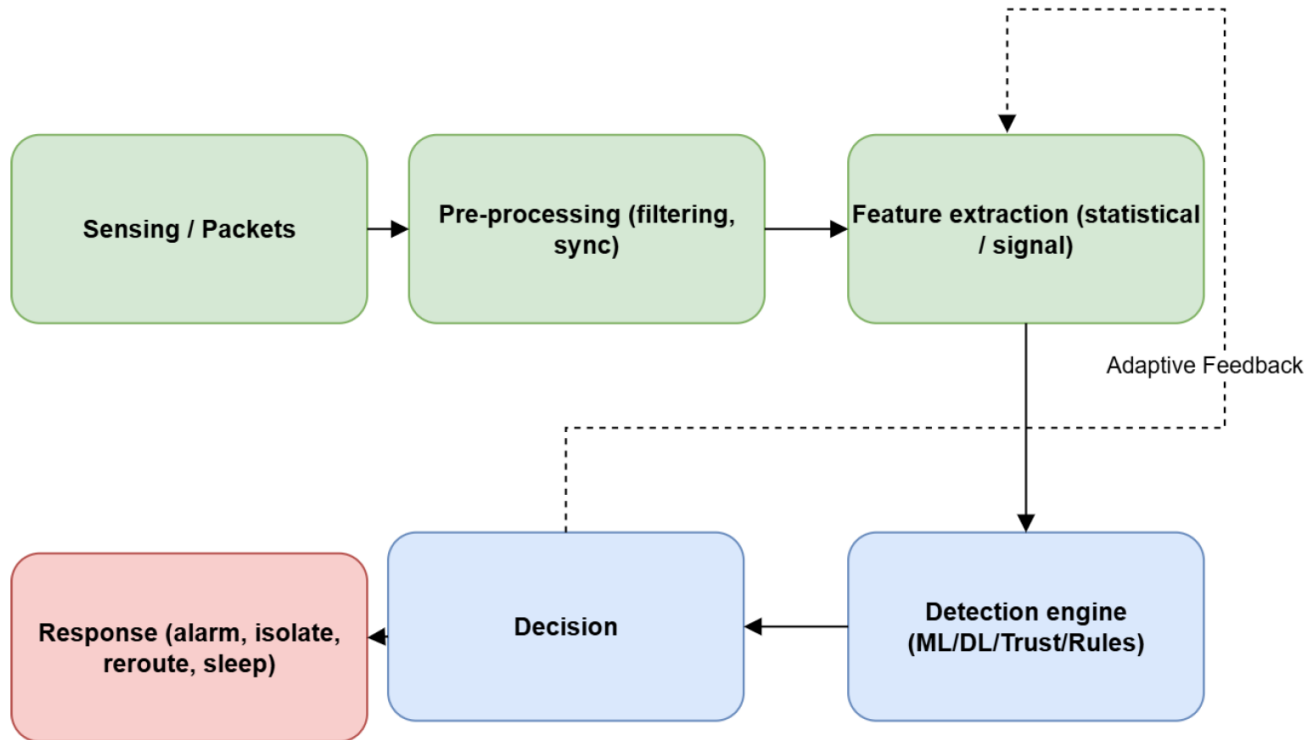


Figure 6 Intrusion detection workflow for IoUT.

6.1 On the Comparability of Reported Results

Table 7 summarises the reviewed IoUT intrusion detection methods and their evaluation setting, and Table 8 gives the quantitative results they report. We recorded what each study reports, and presenting the results this way shows that the methods cannot be ranked against one another, because they lack a common benchmark and reporting method. None of the studies reports detection latency, communication overhead, or computational complexity in a form that allows comparison. The evaluations use different data sources, different network scales, and different attack sets. One study evaluated its underwater method only on KDD-99 and its revision NSL-KDD, which are terrestrial datasets, with no underwater network simulation.

Table 7 Reviewed IoUT intrusion detection methods: approach, scope, and evaluation setting. nr indicates the study does not report the item.

Study	Approach and technique	Attacks addressed	Data source	Network scale	Implementation environment	Validation strategy
Das et al. [2]	Hybrid signature and anomaly; Adaptive Random Forest, One-Class SVM, concept drift detection	Blackhole, grayhole, flooding	Author-generated UW-ASN traces, publicly released	16-64 nodes	NS-2 with Aqua-Sim; detector on Raspberry Pi 3B+	Prequential (test-then-train) on the data stream
Samson and Nagarajan [61]	Hybrid ML and RL; CNN-BiLSTM with reinforcement learning	Listening, forging, flooding	NR	NR	NR	NR
Samson and Nagarajan [18]	Federated and deep learning; FL, CNN-BiLSTM, blockchain	Not specified	NR	NR	NR	NR
Zhang et al. [62]	Anomaly-based; DBSCAN clustering over opportunistic routing	Sinkhole, Sybil, hybrid, on-off variants	Simulated network traces	100 nodes, 5500 ³ m ³ , 20 kbit/s, 2000 m range	NR (no simulator named)	Unsupervised; single simulation campaign
Arivumani and Nagarajan [63]	Hybrid deep learning; CNN-LSTM with NADAM optimiser	Data tampering	NR	NR	NR	NR
Popli et al. [59]	Federated learning with blockchain	Zero-day threats	NR	NR	NR	NR
Huang et al. [64]	Anomaly-based; rough set feature reduction, Random Forest, SMOTE	DoS, Probe, R2L, U2R (terrestrial taxonomy)	NSL-KDD (20% subset) and KDD-99; no underwater data	Not applicable; no network simulation	NR	Train and test split on the NSL-KDD subset

Yan et al. [56]	Barrier coverage with virtual-lattice monitoring and actuator mobility	Physical intruder traversal; not cyber attacks	Simulated deployment geometry	NR	NR	Simulation; detection probability rather than classification
Murgod and Sundaram [57]	Anomaly-based; cluster detection with hybrid optical-acoustic routing	Wormhole	NR	NR	NR	NR
Jiang et al. [65]	Trust-based; hybrid trust model with Random Forest over trust evidence	Multiple malicious behaviours with attack-type identification	Author-constructed trust dataset from direct and recommendation evidence	Clustered UASN	MATLAB R2019a	Out-of-bag error; no external test set

Table 8 Reported quantitative results for the reviewed methods. nr indicates the study does not report the quantity.

Study	Accuracy	Prec.	Rec.	F1	Energy	Latency	Computational complexity	Comm. overhead
Das et al. [2]	0.977 (signature stage)	1	0.978	nr (AUC 0.989)	0.72 J/s detection overhead, measured on hardware	NR	CPU and memory usage measured; no asymptotic analysis	Qualitative only
Samson and Nagarajan [61]	NR	NR	NR	NR	NR	NR	NR	NR
Samson and Nagarajan [18]	NR	NR	NR	NR	NR	NR	NR	NR
Zhang et al. [62]	Relative gain of 3-15% over baselines; no absolute value	NR	NR	NR	Energy tax 9.45-9.87 J after response, against 10.07-43.98 J before	NR	NR	NR
Arivumani and Nagarajan [63]	96.70%	NR	NR	NR	NR	NR	NR	NR
Popli et al. [59]	NR	NR	NR	NR	NR	NR	NR	NR
Huang et al. [64]	Stated as above 99%	NR	NR	NR	NR	NR	NR	NR
Yan et al. [56]	Not applicable; reports average intruder detection probability	NR	NR	NR	Communication energy model; comparative results only	NR	NR	NR
Murgod and Sundaram [57]	NR	NR	NR	NR	NR	NR	NR	NR
Jiang et al. [65]	Above 94%, rising above 99% with more trust evidence	NR	NR	NR	NR	Running time linear in tree count; no absolute figure	NR	NR

Where the validation strategy is stated, it also differs between the studies (Table 7): one uses prequential evaluation on a data stream, one a single train and test split, one only out-of-bag error, and two simulation runs with no separate test set, so overfitting to the evaluation data cannot be ruled out. None of the ten studies reports confidence intervals, repeated runs with different random seeds, or a significance test on its detection results, so a difference of a few percentage points between two methods cannot be shown to exceed run-to-run variation.

6.2 Machine Learning-Based Intrusion Detection Approaches

As mentioned earlier, protocols designed for terrestrial networks often perform poorly in underwater channels, which highlights the need for approaches that adapt to varying channel dynamics. Recent studies have explored methods such as hybrid deep learning models, FL for decentralised security, and adaptive techniques that aim to improve detection accuracy without increasing the computational complexity of resource-limited networks.

Several studies have investigated hybrid deep learning-based models that improve IDS performance by combining spatial and temporal feature extraction. In 2024, authors [63] proposed an IDS using a convolutional LSTM architecture optimised with NADAM for detecting intrusions in IoUT UWSNs, achieving 96.7% accuracy. Later, this idea was extended by integrating RL-based adaptive attack detection. For instance, authors in another study [18] developed a FL-based hierarchical IDS that uses CNNs for spatial analysis and Bidirectional Long Short-Term Memory (Bi-LSTM) networks for capturing temporal behaviours. Similarly, another study [61] proposed a multi-tier intrusion detection framework that combines CNNs, Bi-LSTMs, RL, and optimisation with AdamW to enhance adaptability and reduce false positives in UIoTWSNs.

Decentralised IDS solution based on FL has gained attraction for resource-constrained underwater environments. Unlike centralised models, FL supports distributed training across UWSNs while preserving data privacy and reducing communication overhead. Popli et al. [59] proposed a FL framework with a focus on privacy preservation and collaborative anomaly detection. Furthermore, another study proposed a FL-based IDS integrated with blockchain to secure model aggregation in hierarchical architectures [18] and its application in the broader IoUT through the Fed-IoUT framework proposed by Pei et al. [66]. Another study [67] has extensively reviewed the role of FL in IoUT, highlighting its advantages in handling heterogeneous underwater data sources and constrained environments.

In addition to DL and FL, adaptive ML models have been proposed to address the energy limitations and changing attack patterns in IoUT networks. Das et al. [2] introduced an adaptive intrusion detection and prevention system based on Adaptive Random Forest (ARF) and concept drift detection methods such as ADWIN, allowing the model to update as network conditions change. Classical ML-IDS methods such as Support Vector Machines (SVMs) also remain useful, as shown by a study on malicious attack detection in UWSNs [60].

Clustering-based IDS focus on local anomaly detection, hence minimising communication overhead, and efficiently managing energy resources. Several studies have explored density-based clustering for IDS. For example, Zhang et al. [62] proposed the DOIDS model, which uses DBSCAN clustering to detect routing attacks such as sinkhole attacks in opportunistic routing environments. In another study [57], authors developed a cluster-based method for detecting wormhole attacks through round-trip time analysis, reporting better energy efficiency and network throughput. Huang

et al. [64] combined neighbourhood rough sets for dimensionality reduction with a random forest classifier to improve the detection of multiple attack types in clustered UWSNs.

Moreover, trust-based IDS methods provide another way to identify malicious nodes. Jiang et al. [65] proposed a cloud theory-based trust model, which was extended by [59]. They proposed hybrid trust frameworks that incorporate random forest classifiers to improve detection accuracy. These methods are intended to handle uncertainty in node behaviour while reducing misclassification in dynamic and noisy underwater environments.

Recent studies [3, 68, 69] have examined the ML-based IDSs in both IoT and IoUT networks. These studies highlight key challenges such as dataset scarcity, the need for real-world validation, and the trade-offs between detection accuracy and computational efficiency in underwater systems.

Most of the reviewed literature lacks real-world validation. One example is the use of ML models for SONAR-based mine detection [67], which shows how such techniques can be applied in a practical underwater defence network. Additionally, emerging techniques like Graph Neural Networks (GNNs) and Digital Twins remain under-explored despite their potential for modelling complex underwater attacks [70].

Overall, current research suggests that ML-based IDS is a promising direction for IoUT UWSNs. Future work should focus on addressing these limitations through approaches such as transfer learning, synthetic-to-real adaptation, and newer methods including GNNs and Digital Twins.

6.3 Learning-Based Detectors

This subsection formally states the detectors reviewed in this paper, so that their assumptions and costs can be compared.

6.3.1 Problem Statement

Let $\mathbf{x}_t \in \mathbf{R}^d$ be the feature vector extracted from a node at time step t , typically comprising link-layer counters, routing statistics, energy readings, and physical-layer measurements such as received level and round-trip time. Detection operates on a window of L consecutive steps,

$$\mathbf{X}_t = [\mathbf{x}_{t-L+1}, \dots, \mathbf{x}_t] \in \mathbf{R}^{d \times L}, \quad (23)$$

and produces $\hat{y}_t = f_{\theta}(\mathbf{X}_t) \in [0, 1]$, an estimate of $Pr\{H_1\}$. By eq. (12), L cannot be chosen freely: a window longer than the channel coherence time T_c mixes samples drawn from different channel realisations, so physical-layer features are no longer stationary across the window.

6.3.2 Convolutional Feature Extraction

The convolutional stage produces

$$\mathbf{h}^{(l)} = \sigma(\mathbf{W}^{(l)} * \mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}), \quad \mathbf{h}^{(0)} = \mathbf{X}_t, \quad (24)$$

where $*$ denotes discrete convolution, $\mathbf{W}^{(l)}$ the kernels of layer l , and $\sigma(\cdot)$ a pointwise nonlinearity, in this literature almost always $\sigma(z) = \max(0, z)$. The convolution is applied along the time axis, so the stage extracts patterns that are local in time and shared across positions. Its assumption is translation invariance: that an attack signature has the same shape wherever it occurs in the

window. This holds for jamming and flooding, whose signatures are short and repeated, and holds less well for selective forwarding, whose signature is a change in a long-run rate.

6.3.3 Bidirectional Recurrent Stage

The recurrent stage models dependence across the window. For a long short-term memory cell [71], with $\odot$ denoting the elementwise product,

$$\mathbf{i}_t = \sigma_g(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i), \quad (25)$$

$$\mathbf{f}_t = \sigma_g(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f), \quad (26)$$

$$\mathbf{o}_t = \sigma_g(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o), \quad (27)$$

$$\mathbf{g}_t = \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c), \quad (28)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \mathbf{g}_t, \quad (29)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t), \quad (30)$$

where σ_g is the logistic function. The bidirectional form [72] runs one such recurrence forward and a second backward over the same window and concatenates their states,

$$\mathbf{h}_t^{bi} = [\vec{\mathbf{h}}_t; \overleftarrow{\mathbf{h}}_t]. \quad (31)$$

The backward pass is what makes the stage bidirectional, and it carries a consequence that the reviewed papers do not discuss: computing $\overleftarrow{\mathbf{h}}_t$ requires the whole window to be present, so a bidirectional detector cannot emit a decision until the window closes. It adds at least LT_s to the decision latency, which enters the budget in eq. (14) alongside the propagation delay. A bidirectional model is therefore a poorer fit to a tight deadline than its unidirectional equivalent, whatever its offline accuracy.

6.3.4 Objective Functions and Class Imbalance

Training minimises a loss over the labelled set. The usual choice is binary cross-entropy,

$$L_{\text{BCE}}(\boldsymbol{\theta}) = -\frac{1}{N} \sum_{i=1}^N [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)]. \quad (32)$$

Under the small intrusion prior discussed in Section 4.6, eq. (32) is minimised close to the trivial predictor. Two standard corrections apply. Class weighting rescales the two terms,

$$L_{\text{W}}(\boldsymbol{\theta}) = -\frac{1}{N} \sum_{i=1}^N [w_1 y_i \log \hat{y}_i + w_0 (1 - y_i) \log(1 - \hat{y}_i)], \quad (33)$$

with w_1/w_0 typically set to the inverse class ratio. Focal loss [73] instead down-weights examples the model already classifies confidently,

$$L_F(\boldsymbol{\theta}) = -\frac{1}{N} \sum_{i=1}^N \alpha_t (1 - p_{t,i})^\gamma \log p_{t,i}, \quad (34)$$

where $p_{t,i} = \hat{y}_i$ if $y_i = 1$ and $1 - \hat{y}_i$ otherwise, and $\gamma > 0$ controls the down-weighting. Neither correction changes the operating π ; both change where the decision boundary sits relative to it. Rebalancing the dataset itself, commonly with SMOTE, alters π during training, and the deployed prior is then left unstated.

6.3.5 Optimisers

Several of the reviewed detectors are distinguished chiefly by their optimiser, so the update rules are given here for completeness. Adam [74] maintains bias-corrected first and second moment estimates $\hat{\mathbf{m}}_t$ and $\hat{\mathbf{v}}_t$ of the gradient $\mathbf{g}_t$ and applies

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t} + \epsilon}. \quad (35)$$

NADAM [75] substitutes a Nesterov-accelerated momentum term for $\hat{\mathbf{m}}_t$,

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \frac{\eta}{\sqrt{\hat{\mathbf{v}}_t} + \epsilon} \left(\beta_1 \hat{\mathbf{m}}_t + \frac{(1 - \beta_1) \mathbf{g}_t}{1 - \beta_1^t} \right), \quad (36)$$

and AdamW [76] decouples weight decay from the adaptive term,

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \left(\frac{\hat{\mathbf{m}}_t}{\sqrt{\hat{\mathbf{v}}_t} + \epsilon} + \lambda \boldsymbol{\theta}_t \right). \quad (37)$$

These differ in convergence behaviour during training. None of them affects inference cost, energy consumption, or detection latency on a deployed node, which is worth stating because the reviewed work sometimes presents the choice of optimiser as though it were a property of the deployed detector.

6.3.6 Reinforcement Learning Formulation

Where reinforcement learning is used for adaptive detection, the problem is a Markov decision process (S, A, P, R, γ) . The state $\mathbf{s}_t \in S$ aggregates recent detector outputs and network observations; the action $a_t \in A$ selects a detection threshold, a sampling rate, or a response; and the objective is

$$\max_{\pi} \mathbf{E} \left[\sum_{t=0}^{\infty} \gamma^t r(\mathbf{s}_t, a_t) \right], \quad (38)$$

with the action-value function satisfying

$$Q^*(\mathbf{s}, a) = \mathbf{E} \left[r + \gamma \max_{a'} Q^*(\mathbf{s}', a') \mid \mathbf{s}, a \right]. \quad (39)$$

The reward is normally a weighted combination of detection benefit and false alarm and energy cost, for example $r = \beta_1 \mathbf{1}[\text{true detection}] - \beta_2 \mathbf{1}[\text{false alarm}] - \beta_3 E_t$. Two assumptions embedded in eq. (39) deserve scrutiny underwater. It assumes the state is observed without delay, whereas eq. (10) says observations arrive at least τ late; and it assumes stationary transition dynamics P , whereas the environmental variation described in Section 4 makes P time-varying. Neither assumption is addressed in the reviewed work.

6.3.7 Detection Pipeline

Algorithm 1 states the general pipeline shared by the learning-based detectors reviewed here.

Require: window length L , coherence time T_c , sample interval T_s , deadline T_d , propagation delay τ , model f_θ , threshold γ_t

Ensure: decision $\hat{y}_t \in \{H_0, H_1\}$

```

assert  $LT_s \leq T_c \triangleright 2$ : assert  $LT_s + \tau \leq T_d$            else features are non-stationary across the window, see eq. (12)
for each time step  $t$  do                                      $\triangleright$  observation budget, see eq. (14)
    collect  $\mathbf{x}_t$ : link-layer counters, routing statistics, energy, RSS, RTT
    compensate RTT and time-of-flight features for  $c(T, S, D)$  using (9)
    normalise  $\mathbf{x}_t$  per link using the current estimate of  $A(l, f)$  from (2)
    form window  $\mathbf{X}_t$  as in (23)
     $\hat{y}_t \leftarrow f_\theta(\mathbf{X}_t)$ 
    update threshold  $\gamma_t$  from the current noise estimate  $N(f)$  in (8)
    if  $\hat{y}_t > \gamma_t$  then
        raise  $H_1$  locally
        transmit alert only if benefit exceeds  $E_{\text{comm}}$  from (15)
    end if
    if drift detected in the distribution of  $\mathbf{x}_t$  then
        distinguish environmental drift from adversarial drift before retraining
    end if
end for

```

Algorithm 1 Learning-based IoUT intrusion detection pipeline.

Lines 1, 2, 5, 6, 9 and 12 of Algorithm 1 are the underwater-specific steps. None of the detectors reviewed in Table 7 implements all of them, and most implement none, which is a more precise statement of the gap than saying that terrestrial methods transfer poorly.

6.4 Federated Learning: Formulation, Aggregation, and Threat Model

Federated learning has recently been proposed for IoUT. This subsection states the formulation, the robust aggregation rules, and the threat model, and then evaluates them against the underwater constraints.

6.4.1 Formulation

Let K nodes hold local datasets $D_1, \dots, D_K$ with $n_k = |D_k|$ and $n = \sum_k n_k$. Federated averaging [77] minimises

$$F(\mathbf{w}) = \sum_{k=1}^K \frac{n_k}{n} F_k(\mathbf{w}), \quad F_k(\mathbf{w}) = \frac{1}{n_k} \sum_{i \in D_k} l(\mathbf{w}; \mathbf{x}_i, y_i), \quad (40)$$

by alternating local optimisation with aggregation. In round t each selected node runs E local epochs from the global model $\mathbf{w}_t$ to obtain $\mathbf{w}_{t+1}^k$, and the server forms

$$\mathbf{w}_{t+1} = \sum_{k \in S_t} \frac{n_k}{n_{S_t}} \mathbf{w}_{t+1}^k. \quad (41)$$

6.4.2 Non-IID Data and Client Drift

The convergence guarantees for eq. (41) assume that local datasets are drawn from a common distribution, but for underwater nodes they are not. Nodes at different depths and locations observe different traffic, different noise conditions, and different attack exposure. Let δ_k be the gradient divergence of node k ,

$$\delta_k = \|\nabla F_k(\mathbf{w}) - \nabla F(\mathbf{w})\|, \quad (42)$$

the degradation of eq. (41) relative to centralised training grows with $\max_k \delta_k$ and with the number of local epochs E [78]. Increasing E is the standard way to reduce communication rounds, so underwater deployments usually use a higher E . Proximal correction terms [79] bound local divergence at the cost of slower progress per round.

6.4.3 Communication Cost

For a FL model with $|\mathbf{w}|$ parameters at b bits and R rounds, the payload per node is

$$C = 2R |\mathbf{w}| b \quad \text{bits}, \quad (43)$$

and over a link of rate B the airtime is C/B , to which eq. (10) adds $2R\tau$ of propagation. A modest CNN-BiLSTM detector with $|\mathbf{w}| = 10^6$ parameters at $b = 32$ bits transmits 3.2×10^7 bits per direction per round. Over an acoustic link at $B = 5$ kbit/s, one direction of one round occupies approximately 6.4×10^3 s, or about 1.8 hours. Fifty rounds of training would occupy roughly 178 hours of channel time per node, on a shared channel, at a transmit power set by eq. (15). Quantising the weights to reduce b can degrade detection performance.

6.4.4 Threat Model

Assume an adversary that controls a subset $B \subset \{1, \dots, K\}$ of nodes with $|B| = f$, has full knowledge of the model architecture and can read and control the nodes. Under this model three attacks are relevant.

In an *untargeted poisoning* attack, a node in B submits $\mathbf{w}_{t+1}^k = \mathbf{w}_{t+1}^{\text{honest}} + \epsilon$ with ϵ chosen to move the aggregate away from the optimum and degrade overall accuracy. In a *targeted or backdoor* attack, the adversary trains on data carrying a trigger pattern δ and submits a scaled update so that the aggregate satisfies $f_{\mathbf{w}}(\mathbf{x} + \delta) = y_{\text{target}}$ while behaving normally on clean inputs; in an intrusion detection setting the natural target is to have a specific attack signature classified as benign. In a *gradient inversion* attack [80], an honest-but-curious server reconstructs a node's training data from its submitted update by solving

$$\min_{\mathbf{x}'} \|\nabla l(\mathbf{w}; \mathbf{x}', y') - \nabla l(\mathbf{w}; \mathbf{x}, y)\|^2 + \lambda R(\mathbf{x}'), \quad (44)$$

where R is a prior on the input. This defeats the privacy argument usually given for adopting federated learning in the first place.

6.4.5 Robust Aggregation Rules

Robust aggregation replaces the weighted mean in eq. (41) with a rule that resists a bounded number of Byzantine updates. Krum [81] scores each update by its distance to its nearest neighbours. Writing Δ_i for the update of node i and N_i for the set of $K-f-2$ updates closest to Δ_i in Euclidean distance, the score is

$$s(i) = \sum_{j \in N_i} \|\Delta_i - \Delta_j\|^2, \quad (45)$$

and the rule selects the single update with the lowest score,

$$\text{Krum}(\Delta_1, \dots, \Delta_K) = \Delta_{i^*}, \quad i^* = \arg \min_i s(i). \quad (46)$$

Multi-Krum selects the m updates with the lowest scores and averages them, trading some robustness for the variance reduction of averaging. Byzantine resilience of eq. (46) requires

$$K \geq 2f + 3. \quad (47)$$

Alternative rules include coordinate-wise median and trimmed mean, and geometric-median approaches such as RFA [82]. Differential privacy [83] addresses eq. (44) rather than poisoning, by adding calibrated noise to each update; for the Gaussian mechanism, (ϵ, δ) -differential privacy requires a noise scale

$$\sigma \geq \frac{\Delta_2 \sqrt{2 \ln(1.25/\delta)}}{\epsilon}, \quad (48)$$

where Δ_2 is the l_2 sensitivity of the update.

Condition eq. (47) is hard to satisfy in underwater clusters, because they are small. A cluster of $K = 8$ nodes tolerates $f \leq 2$ by eq. (47), and a cluster of $K = 5$ tolerates none. Krum also requires all K updates to be present before scoring, so a single node whose acoustic link is degraded stalls the round. Both properties conflict with sparse, intermittently connected underwater topologies. The federated IoUT studies reviewed here cite robust aggregation as a mitigation without reporting K, f , or whether eq. (47) is satisfied in their configuration. Establishing which robust aggregation rules remain meaningful at the cluster sizes and connectivity typical of underwater deployment is an open problem, and it is stated as such in Section 9.

6.5 Blockchain and Clustering: Assumptions Against the Channel

A blockchain relies on consensus among participants, here the nodes, on the state of a shared ledger. However, consensus among the nodes requires communication. For example, a system in which K nodes must agree on each block exchange messages on the order of K per round for a leader-based protocol and K^2 for the Byzantine agreement protocols, usually invoked when an adversarial participant is assumed. According to eq. (43) and the link rates implied by eq. (13), this

is the same budget problem that constrains federated learning in Section 6.4, because consensus is required per transaction rather than per training round.

Moreover, consensus protocols assume a participant set that is known and reachable, whereas underwater links fail intermittently. None of the blockchain-assisted IoUT schemes reviewed here reports its message complexity, its block interval, or its storage requirement. In their current form, distributed ledgers are not applicable to underwater networks, because the consensus mechanism needs to be investigated for this setting. Convergence under asynchronous and delayed communication has been studied in the distributed consensus literature [84, 85], and those results can be applied to underwater networks.

Cluster-based detection assumes that a cluster head can observe its members, and that members are comparable to one another. The first assumption fails in sparse deployments, where a cluster head may have few members and each member few neighbours, so the neighbourhood baseline used for anomaly scoring is estimated from a handful of samples. The second assumes that the channel is the same for all members, but, as shown in Section 4, nodes at different depths experience different noise and channel variations, so deviation from the cluster mean is not evidence of misbehaviour.

Cluster maintenance in UWSNs includes energy-efficient aggregation within clusters [86] and the interaction between data gathering, node mobility, and topology change [87]. The aggregation strategy determines what a cluster head can observe, and therefore what it can detect. Position estimates underwater are uncertain and are obtained by optimisation over noisy range measurements [88, 89], so a detector that scores nodes on position consistency, as several trust and Sybil detection schemes do, inherits that uncertainty as false positives. Separating localisation error from adversarial position falsification is an open problem.

6.6 Explainability of Learning-Based Detection (XAI)

None of the detectors reviewed in Table 7 reports any explanation mechanism. For a deployed IoUT system this is a practical problem. In a terrestrial network, an alert can be checked by examining the host and the traffic, and confirmed or dismissed at low cost. This is not possible for nodes deployed underwater. The usual response is to isolate the node, which removes its coverage from a network that is already sparse and may be exactly what an adversary intends. Section 5.7 showed that a failing node and a malicious node produce similar signatures, so an alert with no reason attached is hard to act on, and is likely to be ignored.

Post-hoc attribution methods explain an individual decision of a trained model. SHAP [90] assigns each input feature a contribution to the prediction using Shapley values from cooperative game theory, and LIME [91] fits a locally faithful linear surrogate around the instance being explained. Both are model-agnostic and both are expensive: SHAP requires many model evaluations per explanation, which is not ideal for a node operating under limited energy. The practical arrangement underwater is therefore to detect locally and analyse at the gateway, where the compute budget is higher. This requires that enough of the feature vector is transmitted to support the explanation.

Two requirements follow that are not addressed in the reviewed work. First, an explanation must distinguish adversarial cause from environmental cause. Second, the explanation must be cheap enough to transmit. By eq. (15) the cost of reporting grows with range, so an explanation that doubles the alert payload doubles the energy cost of alerting.

7. Datasets for IoUT Intrusion Detection

Section 6.1 showed that the reviewed methods cannot be compared partly because they are not evaluated on common data. This section examines the data available, since it underlies most of the evaluation problems identified above.

7.1 What is Available

The datasets used in this literature fall into three groups, and none of them is an IoUT intrusion dataset in the sense that CICIDS2017 and UNSW-NB15 are terrestrial intrusion datasets. KDD-99 and its revised version NSL-KDD [92] are commonly used, followed by UNSW-NB15 [93] and CICIDS2017 [94]. WSN-DS is also reported in a study [95] as it is generated from a wireless sensor network, but it is designed for a terrestrial LEACH-based network with radio links. Using these datasets for underwater environment means evaluating against a threat model that does not match the acoustic channel modelled in Section 4.

Most of the remaining studies generate their own data from a network simulator, typically NS-2 with Aqua-Sim, NS-3, or MATLAB. This does reproduce underwater propagation characteristics, but each study generates its own data, with its own topology, node count and mobility model, traffic profile, and attack implementation. The exception among the studies examined here is the dataset released by [2], which is publicly available.

Measured underwater acoustic data exists, but it was collected to characterise the communication channel rather than security. The Watermark benchmark [96], for instance, provides recorded at-sea channel responses for the replay-based evaluation of modulation schemes. Such data contains no intrusions and no labels, so it cannot be used directly to train or evaluate a detector, though it is a realistic basis for generating a channel against which synthetic attacks could be injected.

7.2 Why the Gap Persists

Generating a labelled underwater intrusion dataset is much harder than a terrestrial one. Deploying and recovering a test network is complex and expensive. Moreover, it is difficult to establish that a node behaved maliciously. Real attacks against a deployed network in open water raise legal and safety questions. Furthermore, the resulting data would be specific to one site, one season, and one topology, so its generality would be limited by exactly the environmental variability described by eqs. (5), (6), and (9).

7.3 Class Imbalance and Labelling

IoUT intrusion datasets are imbalanced, so a model trained on them without correction is biased towards the majority class. Common responses are resampling, for example SMOTE [97], and loss reweighting as in eqs. (33), and (34). Resampling changes the class ratio the model is trained on, and therefore the calibration of its output, whereas reweighting changes the decision boundary. For this reason the class ratio after resampling should be reported.

Labels are a second problem. In a simulation the label is exact, because the attack was scripted. In a real deployment it is not, since a failing node and a malicious node can behave alike (Section 5.7), so the label is itself an estimate.

7.4 Synthetic Data Generation

Given the data collection complexity, synthetic data is generated to simulate the IoUT network. *Simulator-generated traces* are the most common source, but their fidelity is bounded by the channel model in the simulator, which is typically the Urick and Thorp model eq. (2), and eq. (1) with a statistical multipath model. *Replay-based generation* uses measured channel responses such as those in [96] and injects synthetic attack traffic through them. This retains real channel statistics, including the variability that Section 4 identifies as a source of false alarms. *Generative models* learn the distribution of an existing dataset and sample from it. Generative adversarial networks [98] and their tabular variants [99] have been proposed for augmenting minority attack classes. However, a generative model trained on simulator output reproduces the simulator's assumptions, including its channel model and the scripted attack signatures. Augmentation increases the quantity of training data, but not its information content.

7.5 What A Usable Dataset Would Require

A useful IoUT intrusion dataset would need to specify the channel conditions under which it was collected or generated, including sea state and shipping activity; the topology, node count, and ranges, since these set the propagation delay and hence the observation budget in eq. (14); and the labelling procedure and its uncertainty. For a real deployment it should also state the hardware on which the detector ran.

8. Deployment, Testbeds, and Standards

The detectors reviewed in Section 6 were designed and evaluated based on simulation or on recorded data. This section covers the deployment constraints that simulation does not capture, the testbeds that exist, and the standards that apply.

8.1 Deployment Constraints

Deploying UWSNs at scale is constrained by several factors that simulation rarely captures.

- **Battery life.** A deployed node runs on a fixed battery that cannot be replaced without recovering the node.
- **Limited hardware.** Nodes use low-power embedded processors, which limits the size of the model that can run on them.
- **Biofouling and corrosion.** Marine growth and corrosion degrade sensors and transducers over the course of a deployment.
- **Infrequent servicing.** Nodes are recovered every few months to years, so a fault or compromise can go unaddressed for a long time.
- **No ground truth in the field.** A deployed detector cannot be checked against a known label, so its false alarm rate cannot be measured directly.

8.2 Experimental Infrastructure

Infrastructure for at-sea experimentation exists and is more mature than the reviewed security literature suggests. The Littoral Ocean Observatory Network operated by the NATO STO Centre for

Maritime Research and Experimentation in the Gulf of La Spezia is a cabled, persistent, remotely operable testbed for underwater communication and networking experiments [100]. It was extended and connected, together with facilities at the University of Porto and others, into the federated set of testbeds assembled by the SUNRISE project, accessible through a common interface [101]. On the software side, SUNSET [102] and DESERT Underwater [103] allow the same protocol implementation to be run in simulation, in emulation with real modems, and in a field trial, which removes the usual reimplementation step between a simulated result and a trial.

8.3 Digital Twins and Attack Emulation

A digital twin is a synchronised model of a physical system, updated from live telemetry and used for prediction and monitoring. For IoUT networks, it allows a candidate detector to be tested against attack scenarios that cannot be mounted against a real deployment. Moreover, it offers a reference for expected behaviour against which model performance can be compared. System-level digital twin modelling for IoUT networks has been proposed, combining a twin network with federated learning and a security prediction component [70].

8.4 Security Standards and Regulatory Frameworks

The first internationally adopted digital underwater acoustic communication standard is JANUS, developed at CMRE and promulgated as NATO STANAG 4748 in March 2017 [104, 105]. It is open and free to use, and lets otherwise incompatible modems communicate with each other. Its design favoured interoperability over security, and the baseline packet has no authentication field [106]. Anyone within acoustic range can transmit using this protocol, including an adversary, so the replay, injection, and spoofing attacks in Table 5 are easier to mount. Work on secure and interoperable underwater acoustic communication is under way in the NATO IST-174 research task group [107], and authentication protocols built on JANUS have been proposed [106].

IEC 62443 [108] addresses industrial automation and control systems and provides a zone and conduit model, which maps reasonably onto the IoUT deployment. NIST SP 800-82 [109] covers operational technology security and starts from similar assumptions: devices last a long time, are hard to patch, and are physically exposed. All of these hold for underwater networks. However, neither standard accounts for the acoustic channel.

On the regulatory side, IMO Resolution MSC.428(98) requires cyber risk to be addressed in a ship's safety management system [110], and the guidelines in MSC-FAL.1/Circ.3 follow the functions of the NIST Cybersecurity Framework [111]. Both are written for ships and the ship-to-port interface, so they do not obviously cover a seabed sensor network connected to a shore facility. We did not find a security standard written specifically for underwater networks.

Across the attacks summarised in Table 5, two needs remain unaddressed: authenticating acoustic traffic within tight link budgets, and incident response for nodes that cannot be reached physically. Section 9 states both as open problems.

9. Future Direction

We frame the challenges below as concrete research questions rather than broad topics. Each subsection examines the key obstacle that currently blocks progress, and Figure 7 summarises the overall landscape.

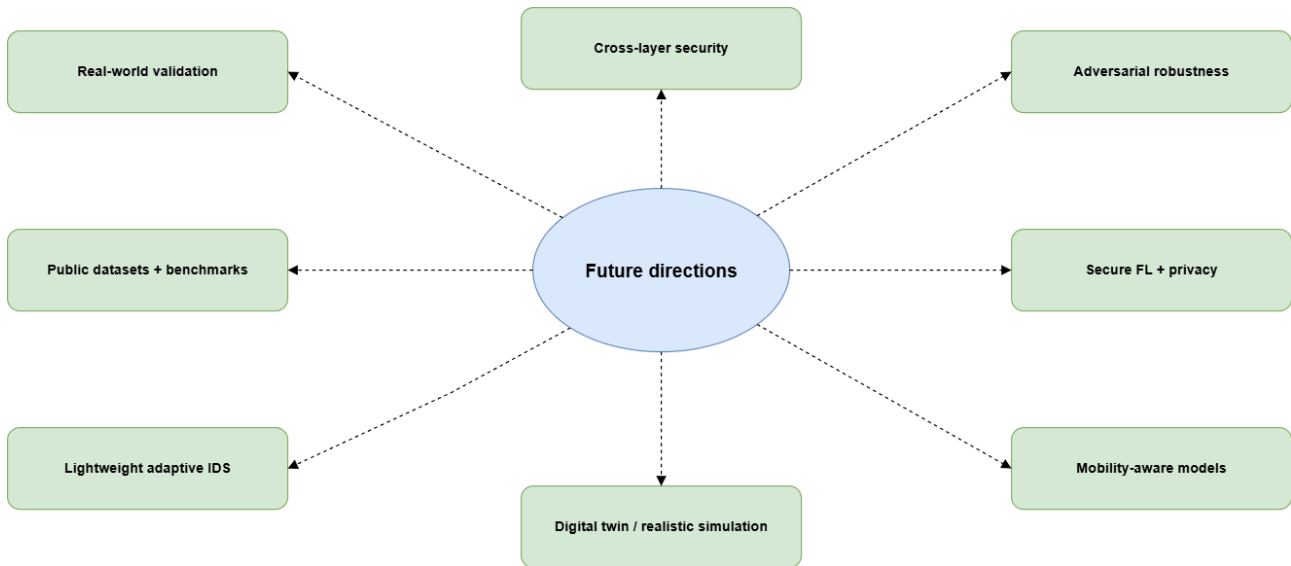


Figure 7 Possible Future Research Direction for IoUT.

9.1 Cross-Layer Detection Under a Joint Statistic

Section 5.6 showed that an attack constructed with $T_i \leq \gamma_i$ at every layer evades any layer-wise detection scheme, and that lowering the individual thresholds degrades precision. The open problem is to construct a joint statistic T_{joint} over per-layer observations that detects such attacks without requiring each layer to raise its own alarm. This problem breaks down into three key questions:

- which per-layer features must be shared, given the transmission energy costs?
- what baseline dependencies exist across layers during benign operation, ensuring that deviations reliably indicate an attack?
- can nodes compute the test statistic incrementally within the memory limits?

9.2 Communication-Efficient Federated Detection

Section 6.4 showed that federated learning at realistic model sizes does not fit an acoustic link. The open problem is to specify a federated scheme whose total payload under eq. (43) fits the link budget available at the deployment range, while retaining the accuracy of centralised training. This requires joint optimisation of quantisation, sparsification, and round reduction, since reducing rounds increases local epochs and therefore divergence.

9.3 Robust Aggregation at Underwater Cluster Sizes

Condition eq. (47) requires $K \geq 2f + 3$, which underwater cluster sizes do not satisfy: eight nodes tolerate two adversaries and five tolerate none. Krum additionally blocks until all updates arrive,

which intermittent links prevent. The open problem is to identify aggregation rules that remain meaningful at $K \leq 10$ and tolerate missing updates, or alternatively to determine the cluster organisation under which existing rules become applicable.

9.4 Adaptive Thresholding Against Environmental Drift

Sections 4 and 5.7 identified two drift sources that are not adversarial: environmental variation, and monotonic sensor degradation from biofouling. A detector with a fixed decision boundary drifts with both. The open problem is a thresholding rule that tracks the environment without tracking the adversary, so that a slowly increasing attack is not absorbed into the baseline. This is harder than drift adaptation because the adversary can adapt to the drift.

9.5 Separating Malicious Nodes from Failing Ones

Section 5.7 identified this as the central weakness of trust-based detection underwater. A node suffering from biofouling, corrosion, or low battery tends to produce the behavioural signature of selective forwarding or false data injection. The open problem is to identify observable quantities that differ between the two cases.

9.6 Explainability Within the Alert Budget

Section 6.6 argued that an alert that cannot be justified will be ignored, because the response to an alert is expensive and slow. The open problem is an explanation that distinguishes adversarial from environmental cause. This is a joint constraint on explanation method and payload format, and it is not addressed by applying SHAP or LIME unmodified, since both assume the explanation is cheap to compute and free to deliver.

9.7 Long-Horizon Detection for Persistent Threats

Section 5.8 showed that IoUT is unusually exposed to slow attacks, because service intervals are long, reconnaissance is passive, and communication is slow. Detecting these requires correlating weak evidence over a long period of time, which conflicts with storage and energy limitation. The open problem is to determine what part of a node's history must be transmitted for long-horizon detection at the gateway, and how a summary of it can be maintained on the node within its memory and energy budget.

9.8 A Common Reporting Template and a Shared Dataset

Section 6.1 showed that the reviewed methods cannot be ranked, and Section 7 showed that no common dataset exists. The open problem is an agreement on a reporting template and on at least one shared evaluation dataset.

9.9 Validation at Sea

No detector reviewed here reports a sea trial, yet Section 8 showed that the facilities exist and that frameworks such as SUNSET and DESERT Underwater allow the implementation to move from

simulation to emulation with real modems. Running an IoUT intrusion detection method on real hardware over a real acoustic link, and reporting what changes, is complex and expensive.

9.10 Authentication and Response Within the Standards Gap

Section 8.4 identified two needs that no existing study addresses: authentication of acoustic traffic at a cost compatible with the link budget, and a defined response procedure for a compromise that cannot be physically reached. The first is a research problem, since the payload and computation budget for authentication under eqs. (13) and (16) is severe and JANUS has no authentication [106]. The second is mainly a standardisation problem.

10. Lessons Learned

This survey highlights unique challenges, types, and detection strategies required to secure the underwater network. IoUT operates in a harsh environment where the conventional IoT model is ineffective. The delayed propagation and long-range properties of acoustic communication make the IoUT network highly vulnerable to jamming. Variations of underwater characteristics complicate the security implementation. Moreover, nodes are battery-powered, and it is difficult to replace them. Energy-efficient algorithms are necessary to sustain operation in such an environment while maintaining detection accuracy. Machine learning, federated learning, and blockchain-based IDS have shown promise but require further optimisation for real-world underwater deployment. The limited availability of datasets is one of the biggest hurdles in designing secure UWSNs. Current studies rely mainly on synthetic data, which does not capture the variability of real underwater conditions, and in some cases terrestrial datasets that do not model the underwater channel.

Relying on a single detection layer is not enough for IoUT networks. Because attacks can target several layers at once, and can be constructed so that no single layer registers an anomaly, cross-layer protocols are required. Hybrid IDS models that combine signature-based, anomaly-based, and trust-based techniques may offer better adaptability and improved detection accuracy against threats. Federated learning also provides a decentralised way to strengthen security while preserving data privacy, whereas DOIDS can help detect anomalies with lower overhead. Since underwater conditions change over time, adaptive systems and machine learning-based IDS may become more practical for responding to new and emerging attacks.

Despite the challenges of deploying UWSNs, recent advances in hybrid IDS, machine learning, federated learning, and blockchain-based security show strong potential. However, practical deployment needs to focus on energy efficiency, and adaptive security mechanisms that can support resilient and scalable underwater networks.

11. Conclusion

IoUT is reshaping the monitoring and management of underwater environments and infrastructure, and effective intrusion detection is essential for maintaining network security, protecting data integrity, and ensuring reliable communication in these systems. The properties of the acoustic channel limit what a detector can achieve, since delay, bandwidth, and noise restrict the observations available to it. The attack surface of a complete IoUT deployment extends beyond the protocol layers, to application-layer, cross-layer, insider, and persistent threats, and the highest-

severity attacks fall in precisely those categories. The reviewed detection methods cannot be compared with each other, because they do not evaluate their models against a unified set of criteria. As underwater applications continue to grow, there is a need for robust, adaptive, and lightweight intrusion detection frameworks.

Author Contributions

The author solely wrote the manuscript.

Funding

This research received no external funding.

Competing Interests

The author declares no competing interests.

AI-Assisted Technologies Statement

ChatGPT was used during manuscript revision to improve grammar and overall language. The scientific content, literature interpretation, analysis, and conclusions were developed by the author. All scientific content, data interpretation, and conclusions were developed independently by the author. The author has thoroughly reviewed and edited the AI-assisted text to ensure its accuracy and accept full responsibility for the content of the manuscript.

References

1. Mohsan SAH, Li Y, Sadiq M, Liang J, Khan MA. Recent advances, future trends, applications and challenges of internet of underwater things (IoUT): A comprehensive review. *J Mar Sci Eng*. 2023; 11: 124.
2. Das S, Pasikhani AM, Gope P, Clark JA, Patel C, Sikdar B. AIDPS: Adaptive intrusion detection and prevention system for underwater acoustic sensor networks. *IEEE ACM Trans Netw*. 2023; 32: 1080-1095.
3. Adam N, Ali M, Naeem F, Ghazy AS, Kaddoum G. State-of-the-art security schemes for the internet of underwater things: A holistic survey. *IEEE Open J Commun Soc*. 2024; 5: 6561-6592.
4. Mohsan SA, Mazinani A, Othman NQ, Amjad H. Towards the internet of underwater things: A comprehensive survey. *Earth Sci Inf*. 2022; 15: 735-764.
5. Jiang B, Feng J, Cui X, Wang J, Liu Y, Song H. Security and reliability of internet of underwater things: Architecture, challenges, and opportunities. *ACM Comput Surv*. 2024; 57: 1-37.
6. Khraisat A, Alazab A. A critical review of intrusion detection systems in the internet of things: Techniques, deployment strategy, validation strategy, attacks, public datasets and challenges. *Cybersecurity*. 2021; 4: 18.
7. Shah S, Munir A, Waheed A, Alabrah A, Mukred M, Amin F, et al. Enhancing security and efficiency in underwater wireless sensor networks: A lightweight key management framework. *Symmetry*. 2023; 15: 1484.

8. Ahmad I, Rahman T, Zeb A, Khan I, Ullah I, Hamam H, et al. Analysis of security attacks and taxonomy in underwater wireless sensor networks. *Wirel Commun Mob Comput*. 2021; 2021: 1444024.
9. Goyal N, Dave M, Verma AK. SAPDA: Secure authentication with protected data aggregation scheme for improving QoS in scalable and survivable UWSNs. *Wirel Pers Commun*. 2020; 113: 1-15.
10. Yang G, Dai L, Wei Z. Challenges, threats, security issues and new trends of underwater wireless sensor networks. *Sensors*. 2018; 18: 3907.
11. Khan JR, Khan SM, Siddiqui FA. Investigative analysis of vulnerabilities and attacks on underwater wireless sensor networks. *Ad Hoc Sens Wirel Netw*. 2025; 60. doi: 10.32908/ahsw.n.v60.10299.
12. Yisa AG, Dargahi T, Belguith S, Hammoudeh M. Security challenges of internet of underwater things: A systematic literature review. *Trans Emerg Telecommun Technol*. 2021; 32: e4203.
13. Gyamfi E, Jurcut A. Intrusion detection in internet of things systems: A review on design approaches leveraging multi-access edge computing, machine learning, and datasets. *Sensors*. 2022; 22: 3744.
14. Lin H, Xue Q, Feng J, Bai D. Internet of things intrusion detection model and algorithm based on cloud computing and multi-feature extraction extreme learning machine. *Digit Commun Netw*. 2023; 9: 111-124.
15. Jahanbakht M, Xiang W, Hanzo L, Azghadi MR. Internet of underwater things and big marine data analytics-A comprehensive survey. *IEEE Commun Surv Tutor*. 2021; 23: 904-956.
16. Ahmad SZ, Qamar F. A hybrid AI based framework for enhancing security in satellite based IoT networks using high performance computing architecture. *Sci Rep*. 2024; 14: 30695.
17. Mounnan O, Manad O, Boubchir L, El Mouatasim A, Daachi B. A review on deep anomaly detection in blockchain. *Blockchain Res Appl*. 2024; 5: 100227.
18. Samson SA, Nagarajan M. A federated learning-based hierarchical intrusion detection system for underwater IoT wireless sensor networks. *Proceedings of the 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*; 2024 December 12-13; Gobichettipalayam, India. New York, NY: IEEE.
19. Ahmed KM, Shams R, Khan FH, Luque-Nieto MÁ. Securing underwater wireless sensor networks: A review of attacks and mitigation techniques. *IEEE Access*. 2024; 12: 161096-161133.
20. Domingo MC. An overview of the internet of underwater things. *J Netw Comput Appl*. 2012; 35: 1879-1890.
21. Ali ES, Saeed RA, Eltahir IK, Khalifa OO. A systematic review on energy efficiency in the internet of underwater things (IoUT): Recent approaches and research gaps. *J Netw Comput Appl*. 2023; 213: 103594.
22. Luo H, Wang J, Bu F, Ruby R, Wu K, Guo Z. Recent progress of air/water cross-boundary communications for underwater sensor networks: A review. *IEEE Sens J*. 2022; 22: 8360-8382.
23. Wang J, Shi W, Xu L, Zhou L, Niu Q. Design of optical-acoustic hybrid underwater wireless sensor network. *J Netw Comput Appl*. 2017; 92: 59-67.
24. Rani S, Ahmed SH, Malhotra J, Talwar R. Energy efficient chain based routing protocol for underwater wireless sensor networks. *J Netw Comput Appl*. 2017; 92: 42-50.

25. Al-Habob AA, Dobre OA, Poor HV. Role assignment for spatially-correlated data aggregation using multi-sink internet of underwater things. *IEEE Trans Green Commun Netw.* 2021; 5: 1570-1579.
26. Hasan WK, Ahmad I, Habibi D, Phung QV, Al-Fawa'reh M, Islam KY, et al. A survey on energy efficient medium access control for acoustic wireless communication networks in underwater environments. *J Netw Comput Appl.* 2025; 235: 104079.
27. Haziq M, Phung QV, Lachowicz S, Habibi D, Ahmad I. Modulation techniques for underwater acoustic communication: A comprehensive survey. *IEEE Access.* 2025; 13: 150715-150755.
28. Subba B, Phung QV, Lachowicz S, Hasan W, Haziq M, Habibi D, et al. Orthogonal time frequency space modulation for underwater acoustic communication systems: A review. *Signal Process.* 2025; 240: 110359.
29. Liou EC, Kao CC, Chang CH, Lin YS, Huang CJ. Internet of underwater things: Challenges and routing protocols. *Proceedings of the 2018 IEEE international conference on applied system invention (ICASI)*; 2018 April 13-17; Chiba, Japan. New York, NY: IEEE.
30. Pan X, Shen Y, Zhang J. IoUT based underwater target localization in the presence of time synchronization attacks. *IEEE Trans Wirel Commun.* 2021; 20: 3958-3973.
31. Thorp WH. Analytic description of the low-frequency attenuation coefficient. *J Acoust Soc Am.* 1967; 42: 270.
32. Urick RJ. *Principles of underwater sound*. 3rd ed. Westport, CT: Peninsula Publishing; 1983.
33. Coates RFW. *Underwater acoustic systems*. London, UK: Macmillan Education; 1990.
34. Mackenzie KV. Nine-term equation for sound speed in the oceans. *J Acoust Soc Am.* 1981; 70: 807-812.
35. Dhanoa JS, Ormondroyd RF, Hughes EJ. A robust digital communication system for doubly-spread underwater acoustic channels. *Proceedings of the Oceans '04 MTS/IEEE Techno-Ocean '04 (IEEE Cat. No.04CH37600)*; 2004 November 09-12; Kobe, Japan. New York, NY: IEEE.
36. Stojanovic M. On the relationship between capacity and distance in an underwater acoustic communication channel. *Mob Comput Commun Rev.* 2007; 11: 34-43.
37. Kay SM. *Fundamentals of statistical processing, Volume 2: Detection theory*. Upper Saddle River, NJ: Prentice Hall; 1998.
38. Zuba M, Shi Z, Peng Z, Cui JH. Launching denial-of-service jamming attacks in underwater sensor networks. *Proceedings of the WUWNet'11: Sixth ACM International Workshop on Underwater Networks*; 2011 December 01-02; Seattle, WA, USA. New York, NY: Association for Computing Machinery.
39. Misra S, Dash S, Khatua M, Vasilakos AV, Obaidat MS. Jamming in underwater sensor networks: Detection and mitigation. *IET Commun.* 2012; 6: 2178-2188.
40. Bagali S, Sundaraguru R. Efficient channel access model for detecting reactive jamming for underwater wireless sensor network. *Proceedings of the 2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*; 2019 March 21-23; Chennai, India. New York, NY: IEEE.
41. Xiao L, Li Q, Chen T, Cheng E, Dai H. Jamming games in underwater sensor networks with reinforcement learning. *Proceedings of the 2015 IEEE Global Communications Conference (GLOBECOM)*; 2015 December 06-10; San Diego, CA, USA. New York, NY: IEEE.
42. Wang Q, Dai HN, Li X, Wang H, Xiao H. On modeling eavesdropping attacks in underwater acoustic sensor networks. *Sensors.* 2016; 16: 721.

43. Lanbo L, Shengli Z, Jun-Hong C. Prospects and problems of wireless communication for underwater sensor networks. *Wirel Commun Mob Comput*. 2008; 8: 977-994.
44. Pandarinath P. Secure localization with defense against selective forwarding attacks in wireless sensor networks. *Proceedings of the 2011 3rd International Conference on Electronics Computer Technology*; 2011 April 08-10; Kanyakumari, India. New York, NY: IEEE.
45. Shahapur SS, Khanai R. Localization, routing and its security in UWSN-A survey. *Proceedings of the 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*; 2016 March 03-05; Chennai, India. New York, NY: IEEE.
46. Domingo MC. Securing underwater wireless communication networks. *IEEE Wirel Commun*. 2011; 18: 22-28.
47. Mian SM, Kumar R. Security analysis and issues in underwater wireless sensor auditory and multipath network. *Int J Anal Exp Modal Anal*. 2019; 11: 2269-2271.
48. Dargahi T, Javadi HH, Shafiei H. Securing underwater sensor networks against routing attacks. *Wirel Pers Commun*. 2017; 96: 2585-2602.
49. Xiao L, Zhu Y. Modeling the wormhole attack in underwater sensor network. *Proceedings of the 2012 8th International Conference on Wireless Communications, Networking and Mobile Computing*; 2012 September 21-23; Shanghai, China. New York, NY: IEEE.
50. Wang W, Kong J, Bhargava B, Gerla M. Visualisation of wormholes in underwater sensor networks: A distributed approach. *Int J Secur Netw*. 2008; 3: 10-23.
51. Furdek M, Wosinska L, Goścień R, Manousakis K, Aibin M, Walkowiak K, et al. An overview of security challenges in communication networks. *Proceedings of the 2016 8th International Workshop on Resilient Networks Design and Modeling (RNDM)*; 2016 September 13-15; Halmstad, Sweden. New York, NY: IEEE.
52. Korhonen J, Wang Y. Effect of packet size on loss rate and delay in wireless links. *Proceedings of the IEEE Wireless Communications and Networking Conference*; 2005 March 13-17; New Orleans, LA, USA. New York, NY: IEEE.
53. Baykara M, Das R. A novel honeypot based security approach for real-time intrusion detection and prevention systems. *J Inf Secur Appl*. 2018; 41: 103-116.
54. Akyildiz IF, Pompili D, Melodia T. Underwater acoustic sensor networks: Research challenges. *Ad Hoc Netw*. 2005; 3: 257-279.
55. Shah D, Kumar V. TCP SYN cookie vulnerability [Internet]. New York, NY: arXiv; 2018. Available from: <https://arxiv.org/abs/1807.08026>.
56. Yan J, Li X, Luo X, Guan X. Virtual-lattice based intrusion detection algorithm over actuator-assisted underwater wireless sensor networks. *Sensors*. 2017; 17: 1168.
57. Murgod TR, Sundaram SM. Cluster based detection and reduction techniques to identify wormhole attacks in underwater wireless sensor networks. *Int J Adv Comput Sci Appl*. 2020; 11: 58-63.
58. Mary DRK, Ko E, Kim SG, Yum SH, Shin SY, Park SH. A systematic review on recent trends, challenges, privacy and security issues of underwater internet of things. *Sensors*. 2021; 21: 8262.
59. Popli MS, Singh RP, Popli NK, Mamun M. A federated learning framework for enhanced data security and cyber intrusion detection in distributed network of underwater drones. *IEEE Access*. 2025; 13: 12634-12646.
60. Ahmed MR, Tahsien SM, Aseeri M, Kaiser MS. Malicious attack detection in underwater wireless sensor network. *Proceedings of the 2015 IEEE International Conference on*

- Telecommunications and Photonics (ICTP); 2015 December 26-28; Dhaka, Bangladesh. New York, NY: IEEE.
61. Samson SA, Nagarajan M. Multi-tier intrusion detection system for underwater IoT wireless sensor networks using hybrid machine learning and reinforcement learning approaches. Proceedings of the 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS); 2024 October 23-25; Erode, India. New York, NY: IEEE.
 62. Zhang R, Zhang J, Wang Q, Zhang H. DOIDS: An intrusion detection scheme based on DBSCAN for opportunistic routing in underwater wireless sensor networks. *Sensors*. 2023; 23: 2096.
 63. Arivumani SS, Nagarajan M. Adaptive convolutional-LSTM neural network with NADAM optimization for intrusion detection in underwater IoT wireless sensor networks. *Eng Res Express*. 2024; 6: 035243.
 64. Huang H, Liu N, Chen D, Yang Q, Huang X. Research on the intrusion detection model of underwater sensor networks. *J Sens*. 2022; 2022: 2323747.
 65. Jiang B, Zhou R, Luo F, Cui X, Liu Y, Song H. Hybrid trust model for identifying malicious attacks in underwater acoustic sensor network. *IEEE Sens J*. 2024; 24: 26743-26754.
 66. Pei J, Liu W, Wang L, Liu C, Bashir AK, Wang Y. Fed-IoUT: Opportunities and challenges of federated learning in the Internet of Underwater Things. *IEEE Internet Things Mag*. 2023; 6: 108-112.
 67. Reddy BN, Vaithilingam CA, Kamalakkannan S. SONAR based under water mine detection using machine learning algorithms. Proceedings of the 2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM); 2024 February 21-23; Noida, India. New York, NY: IEEE.
 68. Saini J, Kait R. Exploring machine learning strategies for intrusion detection in wireless sensor networks. Proceedings of the 2024 IEEE 9th International Conference for Convergence in Technology (I2CT); 2024 April 05-07; Pune, India. New York, NY: IEEE.
 69. Samson SA, Nagarajan M. Comparative analysis of deep learning models for intrusion detection system in UIoT. Proceedings of the 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS); 2023 October 18-20; Erode, India. New York, NY: IEEE.
 70. Wang L, Yan L, Li X, Han S. System-level digital twin modeling for underwater wireless IoT networks. *J Mar Sci Eng*. 2024; 13: 32.
 71. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997; 9: 1735-1780.
 72. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. *IEEE Trans Signal Process*. 1997; 45: 2673-2681.
 73. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2017 October 22-29; Venice, Italy. New York, NY: IEEE.
 74. Kingma DP, Ba J. Adam: A method for stochastic optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015); 2015 May 07-09; San Diego, CA, USA. Available from: <https://www.scirp.org/reference/referencespapers?referenceid=4351247>.
 75. Dozat T. Incorporating Nesterov momentum into Adam. Proceedings of the 4th International Conference on Learning Representations (ICLR 2016); 2016 May 02-04; San Juan, PR, USA. Available from: <https://www.scirp.org/reference/referencespapers?referenceid=3121839>.

76. Loshchilov I, Hutter F. Decoupled weight decay regularization. Proceedings of the 7th International Conference on Learning Representations (ICLR 2019); 2019 May 06-09; New Orleans, LA, USA. Available from: <https://www.scirp.org/reference/referencespapers?referenceid=2885414>.
77. McMahan B, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (PMLR 2017); 2017 April 20-22; Fort Lauderdale, FL, USA. Available from: <https://www.scirp.org/reference/referencespapers?referenceid=3429224>.
78. Zhao Y, Li M, Lai L, Suda N, Civin D, Chandra V. Federated learning with non-IID data [Internet]. New York, NY: arXiv; 2018. Available from: <https://arxiv.org/abs/1806.00582>.
79. Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. Proc Mach Learn Syst. 2020; 2: 429-450.
80. Geiping J, Bauermeister H, Dröge H, Moeller M. Inverting gradients-how easy is it to break privacy in federated learning? Proceedings of the Advances in Neural Information Processing Systems 33 (NeurIPS 2020); 2020 December 06-12; Virtual. Available from: https://proceedings.neurips.cc/paper_files/paper/2020/hash/c4ede56bbd98819ae6112b20ac6bf145-Abstract.html.
81. Blanchard P, El Mhamdi EM, Guerraoui R, Stainer J. Machine learning with adversaries: Byzantine tolerant gradient descent. Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017); 2017 December 04-09; Long Beach, CA, USA. Available from: https://proceedings.neurips.cc/paper_files/paper/2017/hash/f4b9ec30ad9f68f89b29639786cb62ef-Abstract.html.
82. Pillutla K, Kakade SM, Harchaoui Z. Robust aggregation for federated learning. IEEE Trans Signal Process. 2022; 70: 1142-1154.
83. Papernot N, Abadi M, Erlingsson U, Goodfellow I, Talwar K. Semi-supervised knowledge transfer for deep learning from private training data [Internet]. New York, NY: arXiv; 2016. Available from: <https://arxiv.org/abs/1610.05755>.
84. Li W, Yan S, Shi L, Yue J, Shi M, Lin B, et al. Multiagent consensus tracking control over asynchronous cooperation-competition networks. IEEE Trans Cybern. 2025; 55: 4347-4360.
85. Shi L, Yan S, Li W. Consensus and products of substochastic matrices: Convergence rate with communication delays. IEEE Trans Syst Man Cybern Syst. 2025; 55: 4752-4761.
86. Goyal N, Dave M, Verma AK. Improved data aggregation for cluster based underwater wireless sensor networks. Proc Natl Acad Sci India Sect A Phys Sci. 2017; 87: 235-245.
87. Gupta O, Goyal N. Retracted article: The evolution of data gathering static and mobility models in underwater wireless sensor networks: A survey. J Ambient Intell Human Comput. 2021; 12: 9757-9773.
88. Kumar M, Goyal N, Qaisi RM, Najim M, Gupta SK. Game theory based hybrid localization technique for underwater wireless sensor networks. Trans Emerg Telecommun Technol. 2022; 33: e4572.
89. Nain M, Goyal N, Rani S, Popli R, Kansal I, Kaur P. Hybrid optimization for fault-tolerant and accurate localization in mobility assisted underwater wireless sensor networks. Int J Commun Syst. 2022; 35: e5320.
90. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017); 2017 December 04-09;

Long Beach, CA, USA. Available from:

https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

91. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 August 13-17; San Francisco, CA, USA. New York, NY: Association for Computing Machinery.
92. Tavallae M, Bagheri E, Lu W, Ghorbani AA. A detailed analysis of the KDD CUP 99 data set. Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications; 2009 July 08-10; Ottawa, ON, Canada. New York, NY: IEEE.
93. Moustafa N, Slay J. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). Proceedings of the 2015 Military Communications and Information Systems Conference (MilCIS); 2015 November 10-12; Canberra, ACT, Australia. New York, NY: IEEE.
94. Sharafaldin I, Lashkari AH, Ghorbani AA. Toward generating a new intrusion detection dataset and intrusion traffic characterization. Proceedings of the 4th International Conference on Information Systems Security and Privacy; 2018 January 22-24; Funchal, Madeira, Portugal. Setúbal, Portugal: Science and Technology Publications.
95. Almomani I, Al-Kasasbeh B, Al-Akhras M. WSN-DS: A dataset for intrusion detection systems in wireless sensor networks. J Sens. 2016; 2016: 4731953.
96. van Walree PA, Socheleau FX, Otnes R, Jenserud T. The watermark benchmark for underwater acoustic modulation schemes. IEEE J Oceanic Eng. 2017; 42: 1007-1018.
97. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res. 2002; 16: 321-357.
98. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. Proceedings of the Advances in Neural Information Processing Systems 27 (NIPS 2014); 2014 December 08-13; Montreal, Quebec, Canada. Available from: https://proceedings.neurips.cc/paper_files/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html.
99. Xu L, Skoularidou M, Cuesta-Infante A, Veeramachaneni K. Modeling tabular data using conditional GAN. Proceedings of the Advances in Neural Information Processing Systems 32 (NeurIPS 2019); 2019 December 08-14; Vancouver, BC, Canada. Available from: https://proceedings.neurips.cc/paper_files/paper/2019/hash/254ed7d2de3b23ab10936522dd547b78-Abstract.html.
100. Alves J, Potter J, Guerrini P, Zappa G, LePage K. The LOON in 2014: Test bed description. Proceedings of the 2014 Underwater Communications and Networking (UComms); 2014 September 03-05; Sestri Levante, Italy. New York, NY: IEEE.
101. Petrioli C, Petroccia R, Spaccini D, Vitaletti A, Arzilli T, Lamanna D, et al. The SUNRISE GATE: Accessing the SUNRISE federation of facilities to test solutions for the internet of underwater things. Proceedings of the 2014 Underwater Communications and Networking (UComms); 2014 September 03-05; Sestri Levante, Italy. New York, NY: IEEE.
102. Petrioli C, Petroccia R, Potter JR, Spaccini D. The SUNSET framework for simulation, emulation and at-sea testing of underwater wireless sensor networks. Ad Hoc Netw. 2015; 34: 224-238.

103. Masiero R, Azad S, Favaro F, Petrani M, Toso G, Guerra F, et al. DESERT underwater: An NS-Miracle-based framework to design, simulate, emulate and realize test-beds for underwater network protocols. Proceedings of the 2012 Oceans - Yeosu; 2012 May 21-24; Yeosu, Korea. New York, NY: IEEE.
104. Potter J, Alves J, Green D, Zappa G, Nissen I, McCoy K. The JANUS underwater communications standard. Proceedings of the 2014 Underwater Communications and Networking (UComms); 2014 September 03-05; Sestri Levante, Italy. New York, NY: IEEE.
105. North Atlantic Treaty Organization. NATO - STANAG 4748: Digital Underwater Signalling Standard for Network Node Discovery & Interoperability [Internet]. Brussels, Belgium: NATO; 2017. Available from: <https://standards.globalspec.com/std/10202042/stanag-4748>.
106. Téglásy BZ, Wengle E, Potter JR, Katsikas S. Authentication of underwater assets. *Comput Netw*. 2024; 241: 110191.
107. Hamilton A, Barnett J, Hobbs AM, Pelekanakis K, Petroccia R, Nissen I, et al. Towards secure and interoperable underwater acoustic communications: Current activities in NATO IST-174 research task group. *Procedia Comput Sci*. 2022; 205: 167-178.
108. International Electrotechnical Commission. IEC 62443: Security for industrial automation and control systems. Geneva, Switzerland: IEC; 2018.
109. Stouffer K, Pease M, Tang C, Zimmerman T, Pillitteri V, Lightman S, et al. Guide to operational technology (OT) security. Gaithersburg, MD: NIST; 2023; NIST SP 800-82r3. Available from: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-82r3.pdf>.
110. International Maritime Organization. Resolution MSC.428(98): Maritime cyber risk management in safety management systems [Internet]. London, UK: IMO; 2017. Available from: [https://wwwcdn.imo.org/localresources/en/OurWork/Security/Documents/Resolution%20MSC.428\(98\).pdf](https://wwwcdn.imo.org/localresources/en/OurWork/Security/Documents/Resolution%20MSC.428(98).pdf).
111. International Maritime Organization. MSC-FAL.1/Circ.3/Rev.2: Guidelines on maritime cyber risk management [Internet]. London, UK: IMO; 2022. Available from: [https://wwwcdn.imo.org/localresources/en/OurWork/Security/Documents/MSC-FAL.1-Circ.3-Rev.2-%20Guidelines%20On%20Maritime%20Cyber%20Risk%20Management%20\(Secretariat\)%20\(1\).pdf](https://wwwcdn.imo.org/localresources/en/OurWork/Security/Documents/MSC-FAL.1-Circ.3-Rev.2-%20Guidelines%20On%20Maritime%20Cyber%20Risk%20Management%20(Secretariat)%20(1).pdf).