applied sciences

MDPI

Article

A Dynamic Trust Model for Underwater Sensor Networks Fusing Deep Reinforcement Learning and Random Forest Algorithm

Beibei Wang [1], Xiufang Yue [2], Yonglei Liu [2,*], Kun Hao [2,*], Zhisheng Li [2] and Xiaofang Zhao [2]

[1] School of Control and Mechanical Engineering, Tianjin Chengjian University, Tianjin 300384, China; wbbking@163.com
[2] School of Computer and Information Engineering, Tianjin Chengjian University, Tianjin 300384, China; xiufangyue@163.com (X.Y.); lzs@tcu.edu.cn (Z.L.); zhaoxftju@tju.edu.cn (X.Z.)
* Correspondence: yongleiliu@vip.163.com (Y.L.); kunhao@tcu.edu.cn (K.H.)

Abstract: Underwater acoustic sensor networks (UASNs) are vital for applications like marine environmental monitoring, disaster prediction, and national defense security. Due to the prolonged exposure of underwater sensor nodes in unattended and potentially hostile environments, the application of UASNs is confronted with numerous security threats. Trust models are an important means to detect anomalous nodes in UASNs and ensure security. However, when confronted with intricate underwater surroundings, the assessment of trust is prone to disruption, and current trust models lack a flexible mechanism for updating trust. Consequently, this study introduces a dynamic evaluation trust model (DRFTM) for underwater acoustic sensor networks that integrate deep reinforcement learning and the random forest algorithm. First, the DRFTM comprehensively considers indicators including communication, data, energy, and environment to provide reliable trust evidence for the next evaluation; second, under the conditions of node mobility and dynamic updating of network topology, we propose a predictive model for assessing the trust status of sensor nodes based on random forest training; last, the utilization of deep reinforcement learning is instrumental in determining the most effective trust update strategy, leading to improved detection accuracy of the trust model. The simulation results demonstrate the effectiveness of the DRFTM in detecting malicious nodes, reducing false positives, and accurately assessing trust, achieving a remarkable 99% accuracy in identifying malicious nodes.

Keywords: random forests; deep reinforcement learning; trust evaluation model; trust update mechanism; underwater acoustic sensor networks (UASNs)

Citation: Wang, B.; Yue, X.; Liu, Y.; Hao, K.; Li, Z.; Zhao, X. A Dynamic Trust Model for Underwater Sensor Networks Fusing Deep Reinforcement Learning and Random Forest Algorithm. *Appl. Sci.* **2024**, *14*, 3374. https://doi.org/10.3390/app14083374

Academic Editor: Yutaka Ishibashi

Received: 16 March 2024
Revised: 12 April 2024
Accepted: 13 April 2024
Published: 17 April 2024

1. Introduction

Underwater acoustic sensor networks (UASNs) find extensive applications in marine environmental monitoring, marine resource exploration and utilization, geological disaster prediction, as well as marine national defense and security [1–3]. UASNs consist of tens or hundreds of battery-powered underwater sensor nodes, which are typically distributed in unsupervised, or even hostile, environments and are highly vulnerable to malicious attacks and threats [4]. The unique attributes of underwater acoustic channels, including substantial transmission delays, constrained bandwidth, Doppler frequency shifts, node mobility, and pronounced multipath effects, present a formidable challenge to the security of underwater acoustic sensor networks. Ensuring the dependability and security of underwater data transmission stands as a pressing issue demanding attention [5,6].

Traditional techniques for maintaining network security, such as identity authentication and key management, are effective in resisting attacks from external malicious nodes, but they cannot effectively eliminate internal attacks and require significant computational overhead [7], these challenges render them ill-suited for resource-constrained underwater environments. Trust models have emerged as a potent solution to tackle these issues and

have increasingly become a primary approach for detecting abnormal nodes and bolstering security in underwater acoustic sensor networks (UASNs) in recent years. It is worth noting that the propagation of underwater acoustic channels is influenced by factors like ocean depth, temperature, and salinity, among others [8]. Traditional wireless sensor network security mechanisms and mature research results based on land cannot be directly applied to UASN special environments.

The environment and conditions of an underwater acoustic sensor network can change over time and with external changes, which can lead to communication disruptions between sensor nodes, data corruption, or node failures, which in turn affect the reliability and accuracy of the entire network [9]. The network must promptly detect and distinguish abnormal behaviors while also dynamically adapting its trust evaluation. Domestic and international research on trust models for UASNs has achieved certain results, and some trust update mechanisms have also been proposed. Nonetheless, a unified trust model for evaluating trust in underwater acoustic sensor networks is notably absent. Many research efforts have focused on enhancing traditional network trust models, but these adaptations often fall short of fully addressing the unique environment and requirements of underwater acoustic sensor networks. In addition, in terms of trust updating, few studies have been devoted to the problem of adaptive trust updating. Existing studies consider trust updating with a time decay factor, which only considers the time factor and ignores the dynamics of events. The results of trust updating depend heavily on the choice of the time decay factor. In proposals that involve weighting and amalgamating historical and current trust evidence, the challenge lies in determining the optimal size of the time window. This factor can significantly impact the calculation and refreshment of trust values. As a result, real-time trust updates within underwater acoustic sensor networks are imperative to adapt to environmental fluctuations, node failures, and security risks. Such real-time updates play a pivotal role in enhancing the network's reliability, precision, and security.

To address the aforementioned challenges, this paper presents a novel dynamic trust model for underwater acoustic sensor networks (DRFTM) that leverages the power of random forest and reinforcement learning. The DRFTM operates on the principles of communication trust, data trust, energy trust, and environment trust to assess trustworthiness. It employs random forest for trust determination, benefiting from its ability to mitigate the impact of noisy data generated by underwater acoustic channels and reduce the risk of overfitting by combining multiple decision trees for classification or regression. Deep reinforcement learning is harnessed to make adaptive decisions, a versatile technique with broad applications in cyber security [10]. In our proposed approach, we harness deep reinforcement learning to formulate an optimal trust update policy.

The primary contributions can be outlined as follows:

- To enhance the precision of the assessment, the influences that underwater sensor nodes are susceptible to when judging trust are extensively analyzed, and the environmental trust indicators are refined to provide reliable trust evidence for the next step of trust evaluation. Simultaneously, we account for the impact of the underwater environment on node mobility and dynamically update the network topology, rendering the simulation scenario more akin to real-world underwater conditions.
- We propose a novel dynamic trust evaluation model for underwater acoustic sensor networks (DRFTM), which fuses deep reinforcement learning and random forest algorithms to assess the trustworthiness of sensor nodes within UASNs. The DRFTM exhibits excellent trust determination capability and detection accuracy when subjected to single and hybrid attacks.
- Regarding trust updates, this paper introduces a trust update mechanism rooted in deep reinforcement learning. Leveraging the training and learning capabilities of deep reinforcement learning algorithms, this adaptive trust update strategy excels in identifying and excluding potentially malicious nodes. It ensures the timely and precise representation of trust relationships between nodes, concurrently enhancing the trust model to bolster detection accuracy.

The rest of this paper is organized as follows: In Section 2, related work on trust models is analyzed. Section 3 describes the network architecture, the attack model, and the assumptions on which this research is based. Section 4 details the design of the DRFTM model. Section 5 performs experimental simulation and performance analysis of the DRFTM. Section 6 draws conclusions and provides an outlook on future research directions.

2. Related Work

Over the recent years, numerous academics have put forward trust models designed for use in underwater acoustic sensor networks (UASNs). Broadly speaking, these trust models comprise three primary elements: the accumulation of trust evidence, the evaluation of trust, and the update of trust. This section specifically delves into the existing methods for trust evaluation and the mechanisms for trust updating.

2.1. Trust Evaluation

In [4], Han et al. introduced the Attack-Resistant Trust Model for underwater acoustic sensor networks (ARTMM). This model relies on a multidimensional trust metric to assess the trustworthiness of the network. The ARTMM specifically examines the challenges posed by unreliable underwater acoustic channels. It gathers trust-related data across various dimensions, encompassing link trust, data trust, and node trust, while also taking into account the influence of the underwater environment on trust evaluation. Nevertheless, it is worth noting that the trust evidence generation process in the ARTMM does not fully consider the potential impact of malicious attacks at each layer of the network. Furthermore, the application of fuzzy logic within the ARTMM may not adequately capture the nuances associated with uncertain trust relationships. In [11], Jiang et al. introduced a trust model tailored for underwater acoustic sensor networks, known as the trust model based on cloud theory (TMC). This model leverages the cloud theory, which is capable of accommodating factors such as randomness, fuzziness, and uncertainty inherent in trust evaluation. The TMC employs a systematic approach by conducting layer-by-layer packet loss analysis. This method helps mitigate the influence of unreliable acoustic channels and the dynamic nature of network topology on trust evaluation, ultimately enhancing the precision of detecting malicious nodes. However, it is worth mentioning that the TMC does not account for the impact of multiple malicious attacks on the trust evaluation process. In [12], Du et al. introduced an anomaly-resilient trust model centered around the isolated forest algorithm. By incorporating this algorithm, they enhanced the accuracy of detecting faulty nodes, while also introducing the concept of environmental trust, and used sub-sampling methods to reduce the probability of misclassification in trust evaluation. However, defining the metric of environmental trust as a linear model of the total noise PSD is incomplete.

In [13], Shaikh et al. introduced a trust model known as the Group-based Trust Management System (GTMS). This model involves the clustering of sensor nodes into groups for trust evaluation. Within the GTMS framework, trust evaluation occurs at multiple levels: node level, cluster head level, and base station level, achieved through the observation of communication behaviors among neighboring nodes. Node trust values are computed individually by each cluster member, with cluster heads responsible for gathering and maintaining the trust status of their respective cluster members, thus forming a global trust value for nodes within the group. However, it is important to note that the increased overhead associated with packet transmission and data storage tends to reduce the cluster head's operational lifespan compared to that of a typical node. Additionally, the potential delay arising from the transmission of node-to-node trust values was not addressed in the GTMS model. In [14], He et al. introduced the SVM-based Collaborative Trust Model (STMS) designed for application in underwater acoustic sensor networks. This model incorporates a clustering network structure and introduces a dual cluster head mechanism. Notably, the slave cluster head's role is to oversee and monitor the master cluster head, enhancing the overall system's security. The STMS effectively tackles the challenge of limited evidence in sparsely deployed underwater environments. Nonetheless,

it is worth noting that the STMS model does not take into account the potential influence of complex underwater environmental factors on the precision of trust prediction. In [15], Su et al. proposed a trust management mechanism for underwater acoustic sensor networks with redeemable SVM-DS fusion. The SDFTM takes into account the challenges posed by intricate underwater environments on node assessment. It addresses the issue of misclassifying node distrust by considering both historical performance and environmental impact factors. This approach helps prevent the erroneous classification of normal nodes as malicious ones and enhances the precision of malicious node detection.

2.2. Trust Update Mechanism

The trust update process involves fine-tuning and revising the trust evaluations of nodes to align with the evolving network environment and node behavior. In [16], Jiang et al. put forth an efficient distributed trust model tailored for wireless sensor networks. This model employs a sliding time window concept to facilitate trust value updates. This time window is composed of multiple time slots, each representing a cycle time. Additionally, aging factors β: $\beta = e^{t_i - t_{i+1}}$ are introduced with weight values $w_i = \beta$, $w_{i+1} = 1 - \beta$ to account for the decay in trust values. Trust update methodologies founded on the sliding time window mechanism find widespread application in the literature [4,10,12]. These approaches hinge on both real-time and historical windows to execute the update process.

From a sociological point of view, trust is a generalization and summary of historical experience, and from a statistical point of view, trust has the property of decaying over time. In [17], Peng et al. introduced a dynamic trust renewal model characterized by multiple constraints. This model incorporates a time aging factor to signify the gradual decay of trust over time. Additionally, it employs a reward and punishment factor to distinguish between the consequences of successful interactions and failed interactions. Several factors are introduced to ensure that trust develops gradually while declining swiftly, mirroring the trust-building process observed in human societies.

In [18], Jiang et al. introduced a trust evaluation update mechanism known as TEUC, which relies on the C4.5 decision tree, that pointed out the problems in the sliding time window mechanism and time forgetting factor for trust updating, designed two trust updating mechanisms based on event triggering and time triggering and defined a reward and punishment factor, activated when a specific quantity of trust evidence shifts from its predefined fuzzy trust level, prompting a trust update. Furthermore, if the time window exceeds its limit, trust updating is triggered, irrespective of whether there is a significant change in node trust evidence. Nonetheless, it is important to clarify that TEUC does not consolidate trust evidence, and it is limited to executing updates for individual trust evidence instances, not accommodating multiple sources simultaneously.

In [19], He et al. introduced the Trust Update Mechanism based on Reinforcement Learning (TUMRL), designed specifically for UASNs. They employ the concept of criticality to assess node importance, enabling customized trust update strategies. By integrating reinforcement learning into the environmental model along with criticality, trust updates are made more efficient and adaptive to changing attack patterns. However, when using reinforcement learning, the weights of the states in the trust update process are limited, and the weights cannot be adjusted adaptively.

Building upon the diverse trust updating schemes discussed earlier, this paper suggests the development of a deep reinforcement learning-based trust updating model capable of dynamically selecting appropriate trust update strategies.

3. System Architecture and Assumptions

In this section, we provide an overview of the network architecture, attack model, and reasonable assumptions employed in the DRFTM methodology under study.

3.1. Network Architecture

As shown in Figure 1, this paper considers UASNs that are constructed based on a cluster-based network. The sensor nodes in the network are assigned unique IDs. Furthermore, all underwater nodes possess identical initial energy levels, and their communication, computational, and storage capabilities are uniform and finite. It is assumed that the positional information of sensor nodes can be obtained through established localization algorithms [20]. Where the Sink node is deployed on the surface of the water, sensor nodes are randomly dispersed underwater, malicious nodes are hidden in the network, and the AUV cruises through the waters following a predefined path. The sensor nodes are divided into clusters based on a spectral clustering algorithm [21], and each cluster consists of a primary cluster head (CH), a supervisory cluster head (SCH), and several ordinary sensor nodes (SNs). The sensor nodes (SNs) are responsible for sensing environmental data in the marine surroundings and transmitting the data to the cluster head (CH) through either single-hop or multi-hop communication. The cluster head (CH) then consolidates the gathered data and transmits them to the nearby AUV, which transmits them to the Sink nodes, and data transmission between the Sink node and the Base Station is facilitated through satellite communication. In this setup, the supervisory cluster head (SCH) assumes the role of overseeing the actions of the cluster head (CH), with its conduct subject to monitoring by the ordinary sensor nodes (SNs).

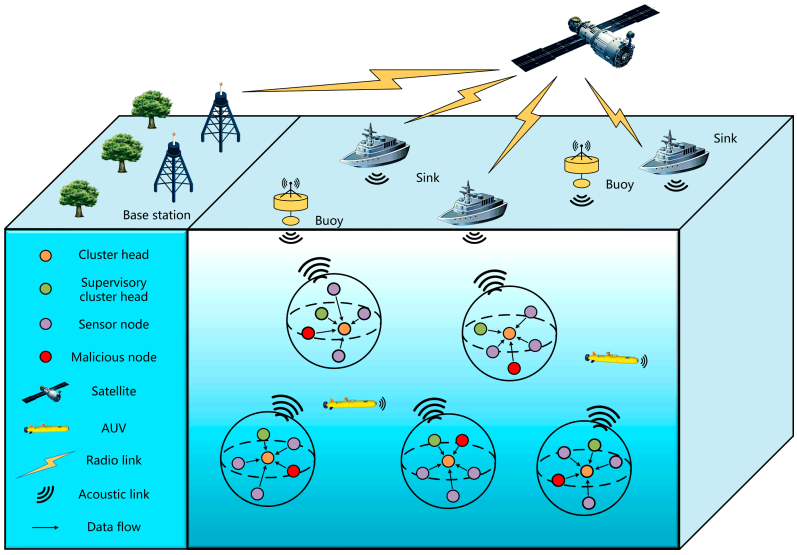


Figure 1. Network architecture.

In the actual underwater environment, the position of sensor nodes is not static but moves with the currents. Changes in node locations affect the network links and the network topology, so it is necessary to introduce an ocean current model to simulate the realistic underwater environment. In this paper, the meandering current flow (MCM) model is introduced as an ocean current model [22]. In a shallow marine environment, the currents move mainly horizontally, and the vertical motion is negligible. The trajectory of a sensor node under complex ocean current motion is represented by the flow function ϕ. When the sensor node's initial position is (x, y), the node's distance traveled at moment t is calculated as follows [23]:

$$x = -\frac{\partial \phi(x, y, t)}{\partial y}, \; y = \frac{\partial \phi(x, y, t)}{\partial x}, \tag{1}$$

where x, y denote the latitudinal and radial vectors of the moment t velocity field, with the unit of km.

3.2. Attack Model

UASNs are susceptible to a variety of malicious threat attacks. The capabilities of the malicious nodes discussed in this paper align with the Dolev–Yao model [24] and can be modeled as the following attack models (malicious nodes that only perform eavesdropping attacks are not considered in this paper, as the attacker performs eavesdropping attacks without active attack behavior and cannot be easily monitored) [25]:

- DDoS attack: Distributed denial of service attack. During operation, the malicious node continuously sends meaningless data traffic to neighboring nodes, preventing them from properly functioning. Malicious node b keeps sending virtual packets to neighboring nodes, as shown in Figure 2a.
- Selective forwarding attack: The malicious node selectively forward packets that pass through it, dropping some of them. As shown in Figure 2b, malicious node b drops packets received from neighboring node c.
- Data pollution attack: The malicious node can carry out data pollution attacks by means such as tampering with data packets, injecting false information, or corrupting data. As shown in Figure 2c, malicious node c transmits altered fake packets to node b, which transmits the packets to cluster head a as it should, and cluster head a receives the altered fake packets from node b.
- On–off attack: The attackers continuously toggle between normal and malicious behaviors to obfuscate and evade detection. The malicious node takes measures to conceal its malicious activities to avoid detection by other nodes within the network.

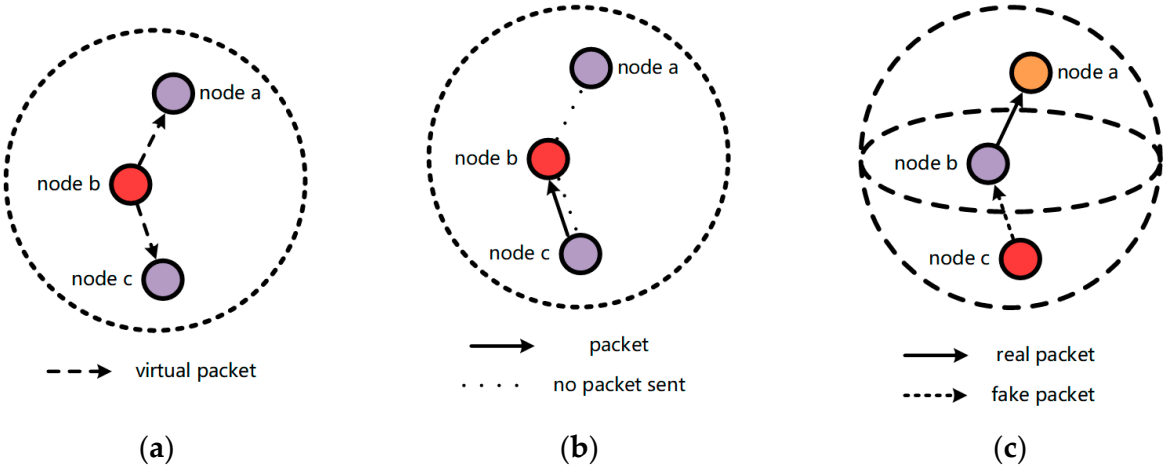


Figure 2. Attack model: (**a**) DDoS attack; (**b**) selective forwarding attack; (**c**) data pollution attack.

4. Design of the DRFTM

The DRFTM trust model comprises three primary components: (1) trust evidence acquisition, (2) trust evaluation based on random forests, and (3) dynamic trust update based on deep reinforcement learning. The system architecture of the model is depicted in Figure 3. Trust evidence acquisition involves collecting relevant information from various nodes and sources. Random forests are utilized for node trust level evaluation, while deep reinforcement learning is employed for the dynamic update of node trust values to adapt to node mobility and network fluctuations. This comprehensive model contributes to enhancing network security and reliability, ensuring efficient trust management among nodes.

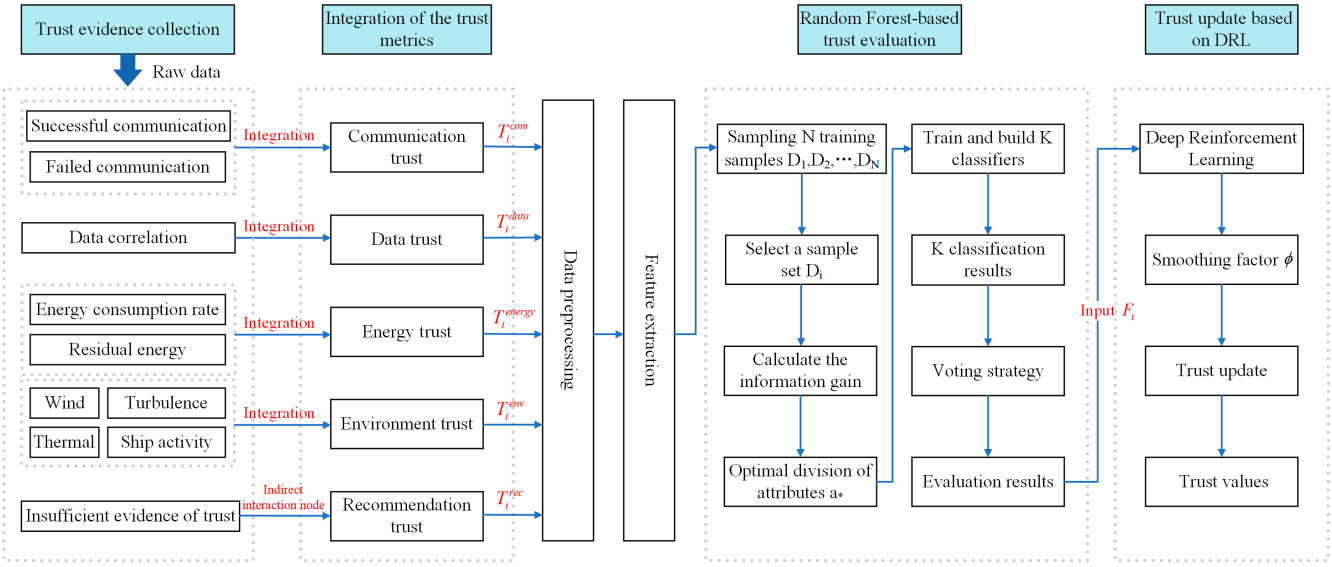


Figure 3. DRFTM architecture diagram.

4.1. Access to Evidence of Trust

4.1.1. Communication Trust

To a certain extent, the communication success rate between adjacent nodes reflects the credibility of their neighbor nodes to some extent. Therefore, the communication success rate is usually an important indicator to evaluate the communication trust degree between adjacent nodes. Assuming node n_i and its m neighboring nodes $(n_1, n_2, \ldots, n_j, \ldots, n_m)$, this paper incorporates the beta distribution to simulate the behavior of an individual node [26]. The beta distribution is a density function that acts as a conjugate prior distribution to the Bernoulli or binomial distribution and is suitable for modeling trust distributions and is expressed in terms of the function Γ as [27]:

$$P(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \forall 0 \le x \le 1, \tag{2}$$

where $a \ge 0$, $b \ge 0$. For communication trust, a represents normal behavior, i.e., the ability to successfully transmit between nodes n_i and n_j, b represents abnormal behavior, i.e., the failure of transmission between nodes n_i and n_j, and the total interaction count between nodes n_i and n_j is equal to the sum of the normal transmission count and the abnormal transmission count, which is represented as $(a+b)$.

When forecasting the behavior of node n_i, node n_i's conduct is denoted by σ and follows a uniform distribution: $P(\sigma) = Beta(1,1)$ in the absence of a priori knowledge. The posterior probabilities obey a Beta distribution with the following parameters: $P(\sigma) = Beta(a+1, b+1)$. Based on this beta distribution, node n_j computes the communication trust value for node n_i as

$$T_i^{com} = E(Beta(a+1, b+1)) = \frac{a+1}{a+b+2}. \tag{3}$$

4.1.2. Data Trust

Data trust is an assessment of the credibility and reliability of data, typically encompassing considerations of data fault tolerance and consistency. In a certain period, data detected by adjacent nodes have temporal and spatial correlations, and the trustworthiness of node n_i is judged by the difference between the data transmitted by node n_i and the data perceived by neighboring node n_j. Suppose data collected by various sensor nodes over a defined time period are denoted as $\{d_1, d_2, \ldots, d_m, \ldots, d_n\}$, we introduce a normal

distribution to model the probability density of data items detected by adjacent nodes as [28,29]:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{4}$$

where μ represents the mean of this set of data and σ represents the variance of this set of data.

The mean value of data obtained from a normal distribution is primarily influenced by the majority of data points, making it considered the most representative value that effectively reflects the similarity of data values. Consequently, it is regarded as having the highest trustworthiness. Trustworthiness evidence based on the received data values is calculated as

$$T_i^{data} = \begin{cases} 2\left(0.5 - \int_{\mu}^{d_m} f(x)dx\right) = 2\int_{d_m}^{\infty} f(x)dx, & d_m \geq \mu \\ 2\left(0.5 - \int_{d_m}^{\mu} f(x)dx\right) = 2\int_{-\infty}^{d_m} f(x)dx, & d_m < \mu \end{cases}. \tag{5}$$

4.1.3. Energy Trust

Due to the limited energy resources of sensor nodes in underwater environments, when the energy levels of these nodes drop too low, it can disrupt the normal operation of the entire sensor network. Malicious nodes execute distributed denial of service attacks or selective forwarding attacks, and the node energy consumption rate will be abnormal. Under stable conditions, the rate of energy consumption by nodes should remain relatively constant. Given that underwater sensor nodes rely on batteries, the amount of remaining energy serves as an indicator of the node's capacity to sustain its operations. The computation of trust evidence based on energy levels is performed in the following manner:

$$T_i^{energy} = \begin{cases} \frac{E^{res}}{E^{init}}, & E^{res} \geq \theta \\ 0, & otherwise' \end{cases} \tag{6}$$

where E^{res} represents the remaining energy of the node, E^{init} represents the node's initial energy, and θ signifies the minimum energy required for the node to accomplish its tasks.

4.1.4. Environment Trust

It is inaccurate to judge the trustworthiness of nodes by considering only the interaction behavior between them, and communication between two nodes in an underwater environment is also affected by unreliable underwater acoustic channels. Marine ambient noise is a major factor affecting underwater acoustic propagation, and turbulence, shipping, waves, and thermal noise are the most common sources of marine ambient noise. The empirical equation below gives the continuous power spectral density of the four noise components [30], with a unit of dB re 1 $\mu Pa^2/Hz$, and it varies with the frequency f of the sound waves:

$$10\log N_t(f) = 17 - 30\log f, \tag{7}$$

$$10\log N_s(f) = 40 + 20(s - 0.5) + 26\log f - 60\log(f + 0.03), \tag{8}$$

$$10\log N_w(f) = 50 + 7.5\sqrt{w} + 20\log f - 40\log(f + 0.4), \tag{9}$$

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

$$N(f) = N_t(f) + N_s(f) + N_w(f) + N_{th}(f), \tag{11}$$

where N_t stands for turbulence noise, N_s represents shipping noise, N_w refers to sea surface noise, N_{th} is thermal noise, and $N(f)$ represents the summation of various environmental noises mentioned above. Considering the application scenario, the transmission frequency of UASNs is significantly influenced by shipping noise and sea surface noise.

The propagation loss in underwater acoustic channels is typically influenced by the distance due to the attenuation of sound waves when propagating through water. In general, the propagation loss of sound waves can be described using the following formula [31]:

$$A(d, f) = d^k a(f)^d, \tag{12}$$

whereas the variable d represents the Euclidean distance, and k is the propagation geometric shape expansion factor. Different values of k reflect the geometric characteristics of sound waves when propagating in water, specifically whether they exhibit a cylindrical or spherical shape. $a(f)$ denotes the rate at which energy is lost due to absorption during the propagation of sound waves in seawater, expressed as

$$10 \log a(f) = \frac{0.11 f^2}{1 + f^2} + \frac{44 f^2}{4100 + f^2} + \frac{2.75 f^2}{10^4} + 0.003. \tag{13}$$

By utilizing the ocean environment noise and accounting for underwater acoustic propagation losses, one can derive the signal-to-noise ratio (SNR) at a distance d of the receiving node:

$$SNR(d, f) = \frac{E_b}{N(f) A(d, f)}, \tag{14}$$

where E_b represents the amount of energy used to transmit data per unit bit.

In the physical layer of acoustic modems, the widely employed modulation technique is binary phase shift keying (BPSK) [32]. Under specific conditions characterized by a distance of d and a frequency of f, the error probability of BPSK modulation can be expressed as

$$p_e(d, f) = \frac{1}{2} \left(1 - \sqrt{\frac{SNR(d, f)}{1 + SNR(d, f)}} \right), \tag{15}$$

The ambient trust (transmission success rate of a packet of received length l) based on the effect of noise is expressed as

$$T_i^{env} = p(d, f) = (1 - p_e(d, f))^l. \tag{16}$$

4.1.5. Recommendation Trust

In UASNs, due to the sparse deployment and constant movement of nodes, the trust evidence obtained through direct interactions between nodes may not be sufficient to support the judgment of node trust, which requires the introduction of recommendation trust. In [33], Zhang et al. made a careful study of recommendation trust and distinguished between unreliable suggestions and dishonest nodes in the network. Here, we adopt the same computational method and do not elaborate on the specific process, and the recommendation trust is expressed as T_i^{rec}.

4.2. Trust Evaluation Based on Random Forest

In the current context of trust evaluation, different techniques such as Bayesian distribution, subjective logic, fuzzy logic, cloud theory, SVM classification, and various machine learning methods are used [17]. However, this paper proposes a new approach named DRFTM, which utilizes random forest to predict and assess the trustworthiness of sensor nodes. Random forest is a technique introduced by Leo Breiman that relies on aggregating multiple decision trees [34]. Compared with a single decision tree, random forest can better process high-dimensional features and large-scale data sets and has strong anti-noise ability. For complex underwater environments where decision trees do not work well, random forest can reduce the impact of noise data generated by underwater acoustic channels on trust evaluation and limit overfitting without significantly reducing prediction accuracy. At the same time, the random forest can automatically select important features and can

effectively process missing data. In addition, random forest can also assess the importance of features to provide a more comprehensive reference for trust determination.

Trust evaluation using the random forest algorithm involves three key steps: (1) preprocessing collected trust evidence data, (2) selecting an appropriate decision tree generation method, and (3) using the random forest algorithm to generate the optimal classification result through a majority vote of multiple decision trees, thereby determining the trust status of each node.

4.2.1. Data Preprocessing

The trust evidence collected may be affected by noise in the underwater acoustic channel, and the calculations for different trust evidence have varying standards. Therefore, it is not possible to directly input these trust evidence data into the model. To address this issue, all collected trust evidence needs to be normalized, meaning it is scaled to fall within the range of [0, 1].

4.2.2. Generation of Decision Trees

A decision tree is a basic method for classification and regression tasks. It forms a tree-like structure, where data points in a specified feature space are progressively allocated from the root node to child nodes and ultimately reach a leaf node, which represents the data point's classification. Depending on the choice of the best attribute selection method, decision trees are generally categorized into three main methods, including ID3, C4.5, and CART [17]. In this paper, C4.5 is chosen as the decision tree generation algorithm. For the five types of trust evidence obtained through the method in Section 4.1, the trust evidence obtained within a time window can be expressed as sample D, i.e., $\left\{\left(T_1^c, T_1^d, T_1^{en}, T_1^{env}, T_1^{rec}\right), \left(T_2^c, T_2^d, T_2^{en}, T_2^{env}, T_2^{rec}\right), \ldots, \left(T_n^c, T_n^d, T_n^{en}, T_n^{env}, T_n^{rec}\right)\right\}$. The primary focus in decision tree learning is discovering the best attribute for grouping data. The C4.5 algorithm selects attributes by calculating the information gain ratio and prioritizes the attributes with higher information gain: it is assumed that there are $V(1, 2, \ldots v, \ldots, V)$ possible values of $\{a_1, a_2, \ldots, a_v, \ldots, a_V\}$ for discrete attribute a, where all the samples of attribute a in sample D with value a_v are denoted as D_v. k denotes the classification result of the sample, and D_{vk} denotes the sample classified as k. The corresponding information entropy is as follows:

$$H(D_v) = -\sum_{k \in N} p(k|v) \log p(k|v) = -\sum_{k \in N} \frac{|D_{vk}|}{|D_v|} \log \frac{|D_{vk}|}{|D_v|}, \tag{17}$$

Each value contains a different number of samples, and we assign weights $|D_v|/|D|$ to the likelihood of each value. Consequently, we can calculate the information gain obtained when attribute a is used to partition the sample set D:

$$Gain(D, a) = H(D) - \sum_{v=1}^{v} \frac{|D_v|}{|D|} H(|D_v|) = H(D) - H(D, a) = I(D, a). \tag{18}$$

The information gain criterion tends to favor attributes with a larger number of values. To mitigate this preference, the C4.5 decision tree algorithm uses the information gain ratio to select the optimal splitting attribute. The information gain rate is defined as follows:

$$Gain_{ratio}(D, a) = \frac{Gain(D, a)}{IV(a)}, \tag{19}$$

where $IV(a) = -\sum_{v=1}^{V} \frac{|D_v|}{|D|} \log \frac{|D_v|}{|D|} = H(a)$.

4.2.3. Trust Status Determination of Nodes

This study utilizes the random forest algorithm to categorize trust levels for trust evaluation. The determination of a node's trust status directly corresponds to the classification outcome produced by the random forest algorithm. The trust level of nodes is represented by $\{C_1, C_2, C_3\}$, which corresponds to {high trust, ordinary trust, low trust}, respectively. Every decision tree is a classifier. For the trust evidence set obtained in Section 4.1, it is assumed that it contains K decision trees $\{H_1, H_2, \ldots, H_K\}$, and K trees will have K kinds of trust judgment results. The random forest algorithm combines all trust evaluation results and determines the final trust status of a node by majority voting, selecting the result with the highest number of votes. The implementation steps of the random forest algorithm are presented as follows:

Input: given training sample data and test sample data.

1. The original training set contains N samples. Each time a tree is constructed, n samples are randomly chosen from it to form the training set for that tree. This process is repeated K times, resulting in the creation of K training sample sets. Finally, these sample sets are used to build K classification trees.
2. The feature dimension of the samples is M, and we need to choose a constant $m < M$, and m features are randomly selected from the M features.
3. Each tree is prepruned according to pruning strategies during growth.
4. Multiple classification trees are assembled into a random forest. This random forest classifier is then used to evaluate and classify new data. The final classification result is determined by the voting consensus of the individual decision tree classifiers.

Voting strategy: The decision tree H_i will predict a final result from the set of category labels $\{C_1, C_2, C_3\}$. The predicted output of H_i on the training set x is expressed as a three-dimensional vector $\left(H_i^1(x), H_i^2(x), H_i^3(x)\right)$, where $H_i^j(x)$ is the output of H_i on the category marker C_j. The absolute majority voting method is used to determine the final classification result, meaning that if a label receives more than half of the votes, it is predicted as that label; otherwise, it is labeled as ordinary trust:

$$H(x) = \begin{cases} C_j, & \text{if } \sum_{i=1}^{K} H_i^j(x) > 0.5 \sum_{n=1}^{N} \sum_{i=1}^{K} H_i^n(x) \\ \text{Ordinary trust}, & \text{otherwise} \end{cases}. \tag{20}$$

4.3. Trust Update Based on Deep Reinforcement Learning

A timely update of the trust status helps the node know the current status of the neighbor so that it can perform the next action. In the face of changing environment and noise conditions, nodes need to constantly improve their trust update mode according to new data and experience. Hence, the adoption of an adaptive trust update strategy enables nodes to dynamically modify the update frequency and weight allocation in response to evolving conditions. The incorporation of deep reinforcement learning represents an innovative approach within the trust update process, allowing for adaptive learning based on environmental cues and feedback. The deep reinforcement learning model determines the optimal strategy based on the evaluated node's state and acquired trust evidence. It selects an action A_t to maximize rewards R_t according to the current state S_t, and as a result of action A_t execution, the environment undergoes changes, accompanied by feedback in the form of rewards R_t. This iterative process continues until the optimal strategy is identified, leading to the achievement of optimal trust updates.

4.3.1. Update Rule

In this paper, the exponential smoothing function is used for node trust update [35]:

$$Trust_t = \phi Trust_{t-1} + (1 - \phi) F_t, \tag{21}$$

where $Trust_{t-1}$ is the trust value at time $t-1$ and $\phi \in (0,1)$ represents the smoothing factor. When a node exhibits high trust, F_t is set to 1, and for low trust, F_t is set to 0. The selection of ϕ parameter values involves finding a trade-off between low and high values, where a lower ϕ value indicates that the current behavior carries a higher weight in trust updates, and the trust value of a node will rise or fall rapidly, which is a powerful trust update strategy. A strong trust update strategy will help nodes rapidly increase or decrease their trust value. Conversely, when ϕ has a higher value, it indicates that past trust values hold greater influence in trust updates, and the trust value of the node will slowly rise or fall. Therefore, the accurate trust value can be obtained only by determining the appropriate parameter ϕ. This paper utilizes deep reinforcement learning to find the best strategy for updating trust.

4.3.2. Deep Reinforcement Learning

Q-learning is a classical reinforcement learning algorithm employed to acquire the optimal action-value function. However, in the MDP process, When the state and action spaces are extensive, and it demands a substantial number of samples to explore the entire state space, Q-learning-based schemes may encounter difficulties. This limitation is mainly related to Q-table storage and updating. When the state space is large, you need to use a huge Q-table to store the value corresponding to each state-action, which takes up a lot of memory. Also, the speed of updating the Q-table is affected because each iteration step needs to traverse the entire Q-table. In the scenario of underwater acoustic sensor networks, deep reinforcement learning, particularly through the use of Deep Q-Networks (DQNs), offers a strong solution to this problem.

4.3.3. Process of Deep Reinforcement Learning

In this paper, the sliding time window is used to store the trust characteristics of sensor nodes, and the trust state of nodes is determined by random forest. Additionally, the deep reinforcement learning model can record the update strategy of previous trust updates, determine the best trust update strategy according to the update strategy of previous trust updates, the trust evidence score of sensor nodes, and the trust state of nodes, and choose action A_t to maximize the reward R_t according to the optimal strategy. The following details the deep reinforcement learning parameters used in the DRFTM:

- State: the deep reinforcement learning model predicts the state of the environment at the time point based on the random forest algorithm. States are defined as

$$S_t = \left\{ \phi_{t-1}; T^c, T^d, T^{en}, T^{env}, T^{rec}; C_P \right\}, \tag{22}$$

  where ϕ_{t-1} is the smoothing factor of the previous trust update strategy, $\left\{ T^c, T^d, T^{en}, T^{env}, T^{rec} \right\}$ is the trust evidence score of the sensor node, and C_P is the node trust state determined by the random forest according to the result of the last round of trust update combined with the current trust evidence.
- Action: the action of deep reinforcement learning is to select the smoothing factor ϕ, that is, the trust update strategy, and it is expressed as

$$A_t = (\phi_1, \phi_2, \ldots, \phi_m), \tag{23}$$

  where $(\phi_1, \phi_2, \ldots, \phi_m)$ is the m smoothing factor of trust renewal strategy.
- Reward: reward is defined as the detection accuracy of the trust state of the node determined by the trust model at a given time, which is defined as

$$R_t = \frac{m_t}{M_t}, \tag{24}$$

  where m_t is the number of malicious nodes detected by the trust model and M_t is the actual total number of malicious nodes.

When the Sink node collects a new round of trust evidence, that is, a new state S_t appears in the network, a trust update strategy, represented as an action A_t, is chosen based on the new trust evidence and the trust state of the node determined by the trust evidence. Given the state, the action A_t is determined by the optimal trust update policy π. This action A_t is then executed, resulting in a reward R_t obtained from the environment. The environment operates according to a Markov decision process (MDP) and transitions to a new state S_{t+1}, in accordance with the MDP rules. The goal of reinforcement learning is to maximize the expected long-term discounted rewards by identifying the best policy. Q-values can be utilized to find the optimal policy and can be updated using the following formula:

$$Q_{t+1}(s_t, a_t) = (1 - \alpha)Q_t(s_t, a_t) + \alpha[R + \gamma \max Q_t(s_{t+1}, a_{t+1})], \tag{25}$$

where α is the learning rate. The best strategy can be expressed as

$$\pi^* = \underset{a \in A}{\operatorname{argmax}} Q^*(s, a). \tag{26}$$

DQN uses neural network $Q(s, a; w)$ to approximate $Q^*(s, a)$, takes state as input, and outputs Q value estimation for each action. Relating inputs are approximated to relating outputs using weight or coefficients. The objective is to minimize the error by adjusting the network's weights through gradient descent. First, the weights of the deep Q network coefficients are randomly initialized. Over time, the deep Q-network updates its weights based on the difference between the expected reward and the true reward. In each iteration of the deep Q-network, the objective is to minimize the loss function as follows:

$$L(w) = \sum_{(s_t, a_t)} (y_t - Q(s_t,\ a_t;\ w))^2, \tag{27}$$

where $Q(s_t,\ a_t;\ w)$ is the predicted value and y_t is the estimated value of the actual value, which can be expressed as

$$y_t = R_t + \gamma \underset{a_t \in A}{\max} Q(s_t,\ a_t;\ w), \tag{28}$$

where R_t is the corresponding reward and γ is the discount factor.

5. Simulation Results and Analysis

5.1. Simulation Settings

This paper presents a comparative analysis of the DRFTM model and four existing models, namely the ARTMM [4], STMS [12], TEUC [17], and TUMRL [18]. The ARTMM is renowned as the pioneer trust model implemented in underwater environments. On the other hand, the STMS is a notable trust model specifically designed for layer-clustered networks. The TEUC stands out by utilizing a decision tree algorithm for trust evaluation, incorporating both event and time-based updating mechanisms. Lastly, the TUMRL is constructed on reinforcement learning principles for trust update. To evaluate the performance of these trust models, MATLAB 2021 simulation is employed. Specifically, the simulation incorporates sensor nodes that are uniformly distributed within a network space of dimensions 500 m × 500 m × 500 m. The default simulation parameters can be found in Table 1.

The performance evaluation indicators are described as follows:

- Detection accuracy (%): reflects the overall predictive accuracy of the model for all samples.
- Error rate (%): measures the model's ability to incorrectly predict negative samples as positive.
- Trust value: a visual reflection of the evaluation result of node trust.

Table 1. Simulation parameters.

Parameters	Default Value
Network size	500 m × 500 m × 500 m
Number of nodes	100
Node placement	Randomly deployed
Maximum communication radius	200 m
Node initial energy	65 J
Packet length	2048 bits
Proportion of malicious nodes	30%
Phase velocity c	0.12
Average width of ocean currents	1.2

5.2. Validation Effectiveness of the DRFTM

5.2.1. Discussion of Tree Number Parameters

The initial performance of a single decision tree in a random forest model tends to be relatively poor due to the risk of overfitting. The introduction of attribute perturbation in random forests can further increase model variance, resulting in decreased performance. Additionally, using numerous decision trees in random forests can make the algorithm slow, which requires finding a balance between efficiency and detection capability in trust evaluation models. A commonly used metric for evaluating classification models is area under the curve (AUC). AUC reflects the area under the ROC curve and measures the classification performance, specifically the ability to detect malicious nodes in a trust evaluation model. It provides a numerical value that helps determine the better classifier.

In Figure 4, we observe the interplay between the number of decision trees, AUC, and runtime. As illustrated in Figure 4a, there is a rapid increase in AUC as the number of decision trees rises. Beyond a certain threshold, the AUC stabilizes at a relatively high value, showing only marginal improvement with further increases in the number of trees. Figure 4b demonstrates a linear correlation between runtime and the number of decision trees. While, theoretically, a higher number of decision trees may lead to more accurate trust evaluation, our experiments indicate that an optimal detection accuracy is achieved at around 25 decision trees. Given the consideration of runtime impact, we set the number of decision trees to 25 in this study.

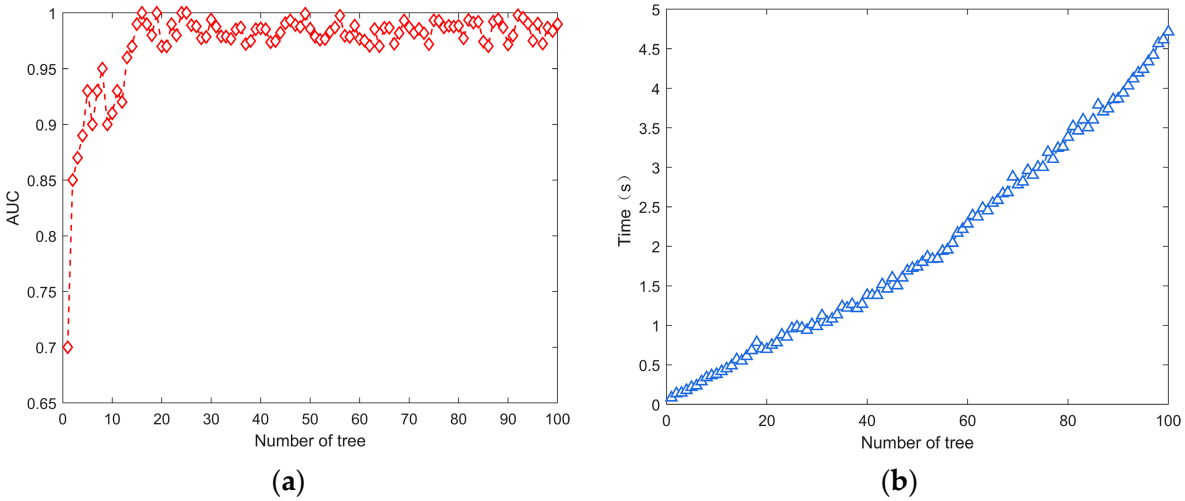
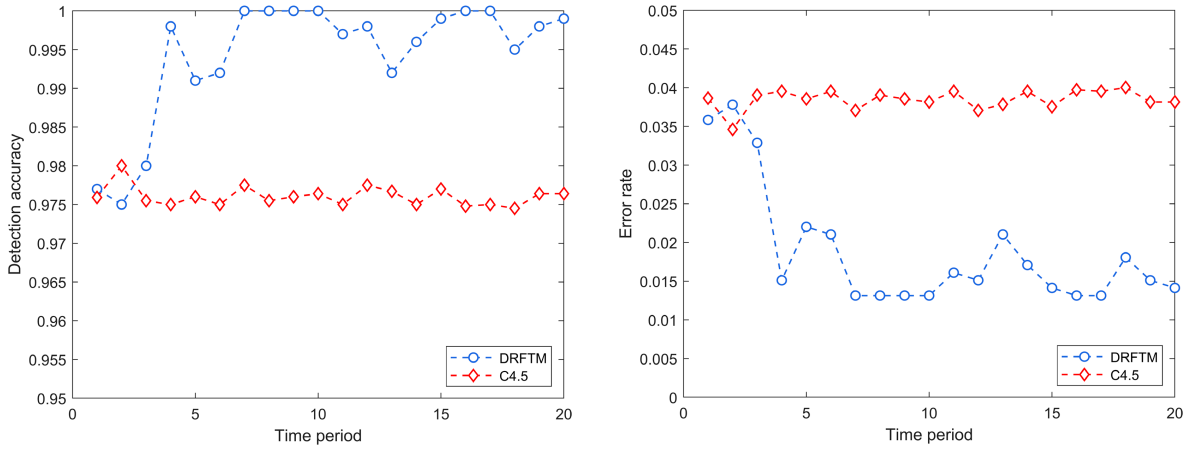


Figure 4. Impact of the number of trees on the DRFTM performance: (**a**) AUC vs. tree number; (**b**) running time vs. tree number.

5.2.2. Comparison of Random Forest and C4.5 Decision Trees

This paper employs the random forest algorithm and conducts a comparative analysis with another classification method, namely the C4.5 decision tree algorithm, as depicted in

Figure 5. Random forest exhibits better detection accuracy and a lower error rate than C4.5 for different malicious node rates. This is due to the fact that in the random forest algorithm, multiple decision trees are used for classification. Finally, the final classification result is arrived at by voting or averaging, etc. In comparison, the C4.5 decision tree algorithm uses only a single decision tree for classification. Therefore, the random forest algorithm is able to classify more accurately, whereas the C4.5 algorithm is relatively weaker.



(**a**) (**b**)

Figure 5. Comparison between random forest and C4.5: (**a**) detection accuracy over time; (**b**) error rate over time.

5.2.3. Validation of the Trust Update Mechanism

In this context, node trustworthiness is represented by a trust value on a scale from 0 to 1, where 0 signifies a complete lack of trustworthiness and 1 represents absolute trustworthiness. Initially, during the network's inception, all nodes are assumed to possess a trust value of 1. To evaluate the effectiveness of the DRFTM in discerning between normal and malicious nodes, we monitor the evolution of trust values over time. For this experimental configuration, we introduce an initial 30% of malicious nodes to initiate a hybrid attack (Section 3.2). As depicted in Figure 6, normal nodes consistently maintain a high trust value, whereas malicious nodes undergo a swift decrease in trust, eventually stabilizing at a low trust value of 0.1. This finding highlights the DRFTM's proficiency in effectively distinguishing between normal and malicious nodes.

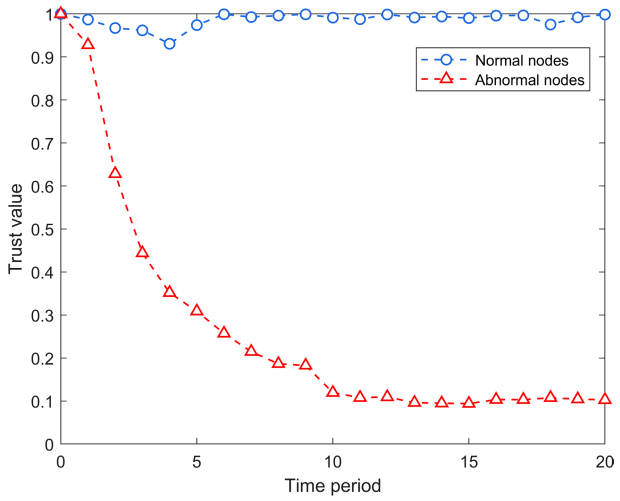


Figure 6. Change in trust value between normal and malicious nodes.

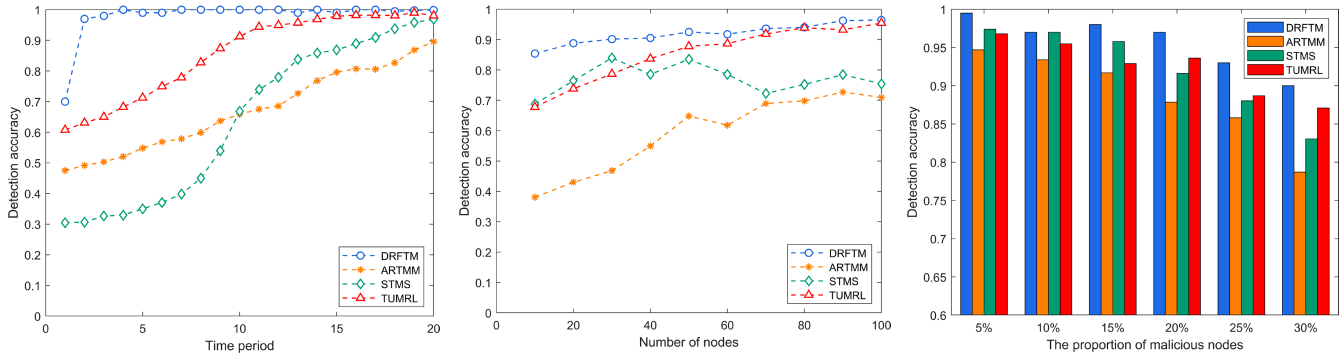
5.3. Comparison of the DRFTM Performance in Different Attack Modes

To validate the performance of the DRFTM model under single-mode and hybrid attacks, we compare the DRFTM with three established trust models (ARTMM, STMS, and TUMRL). In a single-mode attack, only one mode of attack occurs at the same time, i.e., DDoS attack, selective forwarding attack, data contamination attack, and on–off attack. A hybrid attack occurs when all four attack modes occur at the same time.

5.3.1. Detection Accuracy under Hybrid Attack

This paper evaluates the detection accuracy of four models with respect to time, different percentages of malicious nodes, and different numbers of nodes and calculates the detection accuracy of several models over a period of 20 cycles. Among them, the DRFTM model proposed in this paper fuses deep reinforcement learning with a random forest algorithm and is compared with the weighted fusion computation method of the ARTMM, the SVM classification algorithm of the STMS model, and the reinforcement learning method of the TUMRL model.

First, the detection accuracy of different models is tested with time variation. The malicious nodes initiate hybrid attacks right from the start of network operation, with the malicious node percentage set at 30%. As illustrated in Figure 7a, the detection accuracy of the DRFTM exhibits a swift initial increase, eventually stabilizing at a level exceeding 98% after several iteration cycles. At the beginning of the simulation, the detection accuracies of the ARTMM, STMS, and TUMRL are relatively low, and as more and more trust evidence is obtained from the neighboring nodes, the detection accuracies are continuously improved. The STMS requires the training of a trust evidence set for predicting sensor node trust. Initially, during the simulation's outset, there is a shortage of trust evidence, resulting in low detection accuracies. However, by the 10th training cycle, the STMS achieves higher detection accuracies compared to the ARTMM. The detection accuracy of the TUMRL is better than the ARTMM and STMS but always lower than the detection accuracy of the DRFTM. The DRFTM relies on the random forest algorithm to assess the trust status of nodes. It employs a voting mechanism based on decision tree classifications to derive optimal classification results. Additionally, it integrates deep reinforcement learning for adaptive strategy selection during updates. This synergy results in an accelerated increase in the accuracy of the DRFTM. It can be seen that the DRFTM model outperforms the ARTMM, STMS, and TUMRL in terms of detection accuracy.



(**a**) (**b**) (**c**)

Figure 7. Detection accuracy: (**a**) detection accuracy over time; (**b**) number of nodes vs. detection accuracy; (**c**) the ratio of malicious nodes vs. accuracy.

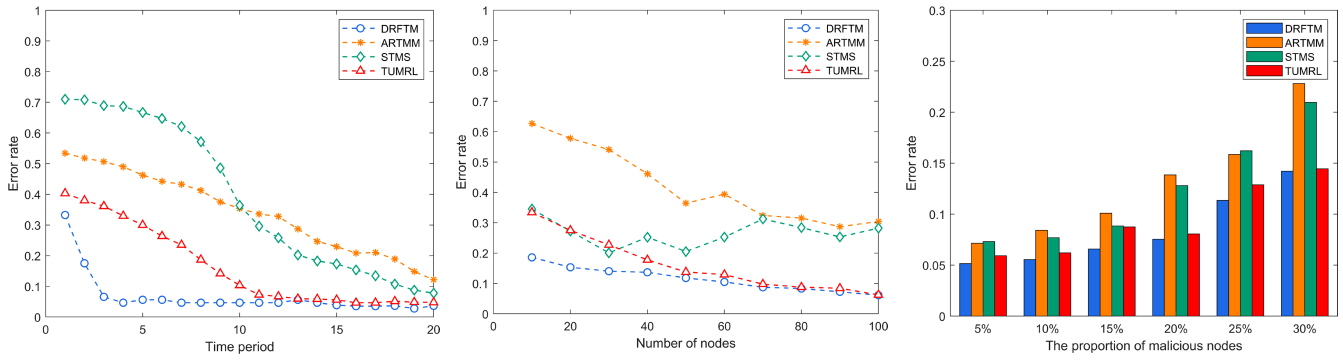
At a fixed network size, the fluctuation in node count indicates the level of node deployment sparsity (the communication range is appropriately adjusted with the sparsity). As shown in Figure 7b, the impact of node count variation on detection accuracy is observed at the 10th cycle of simulation, with a 30% proportion of malicious nodes. The

experiments demonstrate that the DRFTM exhibits strong performance even in underwater environments characterized by sparse sensor node deployment.

As depicted in Figure 7c, it illustrates the detection accuracy of malicious nodes at the 10th simulation cycle for various proportions of malicious nodes. With an increase in the proportion of malicious nodes, all four models experience a decline in detection accuracy. The ARTMM exhibits the most rapid decrease, followed by the STMS. In contrast, the DRFTM demonstrates the smallest decrease and maintains a robust detection accuracy of 90% even when the proportion of malicious nodes reaches 30%. This resilience can be credited to the robustness of the random forest algorithm, which adeptly manages noise while integrating environmental trust factors to ensure steady detection of normal nodes. In contrast, other algorithms overlook the dynamics of the underwater environment. The DRFTM, through learning from historical interactions and trust evidence, discerns the characteristics of malicious nodes, thus sustaining high detection accuracy even in scenarios with limited trust evidence.

5.3.2. Error Rate under Hybrid Attack

To further assess the performance of the DRFTM trust model, we compare the false detection rates of the four trust models under various parameters, maintaining the same experimental setup as detailed in Section 5.3.1. As illustrated in Figure 8a, the DRFTM consistently maintains an error rate below 0.1 after stabilization. In contrast, the ARTMM and STMS exhibit initially high error rates that gradually decrease over time. Figure 8b demonstrates that a lower node count corresponds to higher error rates. Reduced node numbers result in limited trust evidence, which can negatively impact the model's trust determination. With an increasing node count, the DRFTM's error rate steadily decreases, establishing a stable pattern of low error rates. Regarding different ratios of malicious nodes, Figure 8c reveals that when the ratio remains below 15%, all four trust models maintain an error rate below 0.1. However, as the proportion of malicious nodes increases, the ARTMM and STMS experience a rapid escalation in error rates. This is attributed to the growing ratio of malicious nodes introducing false information into the trust evidence. The ARTMM, employing a weighted fusion method for trust calculation, exhibits limited ability to discern false trust, resulting in high error rates. In contrast, the STMS employs an unsupervised method for label assignment in the first stage, and errors in this initial phase have cascading negative effects on the subsequent anomaly identification process in the second stage. The DRFTM and TUMRL perform better with a slower increasing trend in the error rate, which stabilizes at an error rate of less than 0.15.



(**a**) (**b**) (**c**)

Figure 8. Error rate: (**a**) error rate over time; (**b**) number of nodes vs. error rate; (**c**) the ratio of malicious nodes vs. error rate.

5.3.3. Performance of the DRFTM under Single Mode Attack

To provide a conclusive comparison, we assessed the performance of the four models exclusively in single-mode attack scenarios. In these experiments, we initiated each test

with an initial percentage of 30% anomalous nodes, focusing on a single type of attack. No attack occurs in the first five iterations, and after five iterations, 30% of the nodes switch to anomalous behavior state (experimental verification confirms that altering anomalous behavior after a specific number of iterations results in a consistent trend in trust value changes. The figure illustrates this trust value variation of malicious nodes when switching anomalous behavior within five iterations as an illustrative example), launching a DDoS attack, a selective forwarding attack, a data contamination attack, and an on–off attack, respectively (Section 3.2). We compared the average trust levels of malicious nodes across the four trust models.

As shown in Figure 9, during the initial five iterations when the malicious node refrains from launching an attack, all four models consistently estimate high average trust values for the nodes. However, when the malicious node initiates an attack, the DRFTM rapidly detects and reflects this change by observing a sharp drop in the average trust value assigned to the malicious node. While trust values exhibit slight fluctuations under various attack scenarios, the overall trend remains consistent. This is because the DRFTM model uses a deep reinforcement learning adaptive trust update strategy, which is able to react quickly to outliers in the trust evidence and determine the node as low trust during trust evaluation, assigning it a very low trust value. The other three models also respond to the actions of the malicious node, albeit with a slower reaction time, resulting in a gradual decline in the trust value assigned to the malicious node. Figure 9d shows that while the malicious node launches a DDoS attack, selective forwarding attack, and data contamination attack, the malicious node performs an on–off attack. Because the switch attack is launched intermittently, the trust values estimated for the malicious node by the four models exhibit varying degrees of fluctuation, but the fluctuation magnitude of the DRFTM is small and the trust state is stable. This is because the DRFTM model not only considers the current trust state but also combines the node's historical behavior when updating the trust. The malicious node executes the on–off attack, and even when the node occasionally performs well, the node's trust value will not be restored to the normal node's trust state, and the overall trend shows a fast decline and a slow rise. Taken together, the above results show that malicious nodes can be quickly identified by their trust values, which indicates the reliability of the DRFTM model to detect malicious nodes and their resistance to face multiple attacks.

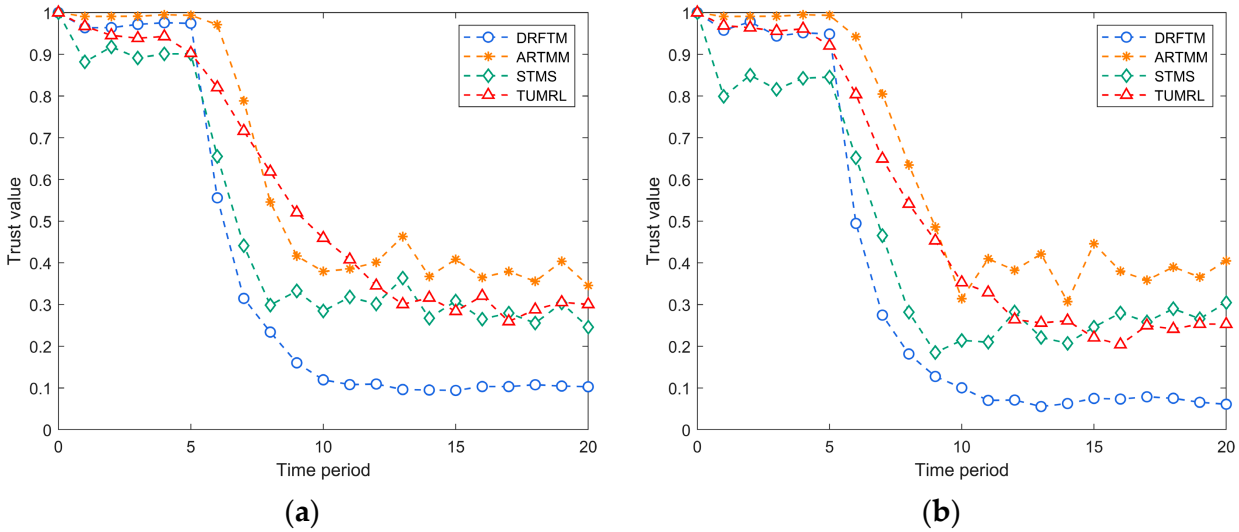


Figure 9. *Cont.*

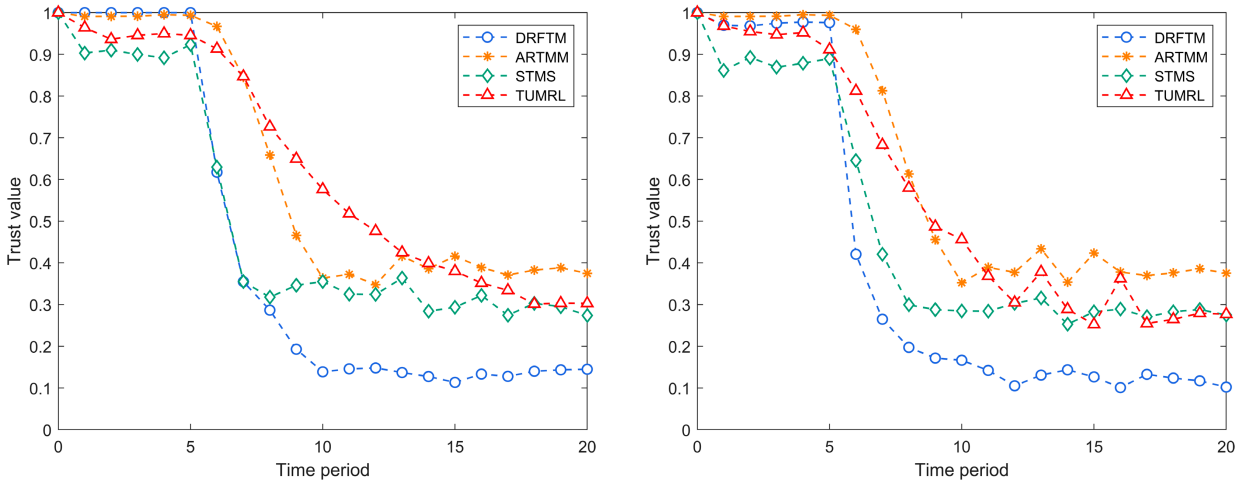


Figure 9. Comparison of performance in single-mode attacks: (**a**) DDoS attack; (**b**) selective forwarding attack; (**c**) data pollution attack; (**d**) on–off attack.

6. Conclusions

This paper addresses the challenge of detecting malicious nodes within complex underwater acoustic sensor networks, accommodating various attack behaviors of malicious nodes. It introduces a dynamic trust evaluation model that combines deep reinforcement learning and the random forest algorithm. In this framework, behavioral characteristics of malicious nodes are leveraged to select three types of trust evidence: communication trust, data trust, and energy trust. The model also accounts for the influence of intricate underwater environments on node trust determination, enhancing trust indicators based on environmental factors. A trust evaluation model, based on the random forest algorithm, is trained to assess the trust status of sensor nodes. Real-world scenarios, where ocean currents induce dynamic movement of underwater sensor nodes, are considered. This paper analyzes and simulates node mobility, accounting for the dynamic evolution of network topology. Nodes are adaptively updated using a trust update mechanism rooted in deep reinforcement learning. This mechanism facilitates timely and precise reflection of trust relationships between nodes, thereby enhancing the trust model's detection accuracy. Experimental analyses demonstrate that DRFTM excels in promptly detecting malicious behaviors, outperforming existing approaches, and maintaining robust performance even in sparsely deployed underwater environments.

However, this research also has its limitations. The attack model simulated in this study excludes scenarios where malicious nodes engage solely in eavesdropping attacks. Nevertheless, eavesdropping attacks pose a significant threat to the confidentiality, integrity, and overall security of UASNs. These attacks can jeopardize confidentiality, manipulate data, and operate covertly, presenting challenges in detection and mitigation. In future studies, we will focus on enhancing security through secure routing protocols, developing strategies to resist eavesdropping attacks, and exploring more practical trust models for UASNs.

Author Contributions: Conceptualization, B.W. and X.Y.; methodology, B.W. and Y.L.; software, X.Y. and Y.L.; validation, K.H., Y.L. and X.Y.; formal analysis, Y.L.; investigation, X.Y.; resources, X.Z.; data curation, X.Z.; writing—original draft preparation, X.Y.; writing—review and editing, K.H. and B.W.; visualization, Z.L.; supervision, Z.L.; project administration, K.H.; funding acquisition, K.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Chunhui Cooperation Program of the Ministry of Education, HZKY20220590-202200265; National Natural Science Foundation of China under Grant 61902273.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Han, G.; Long, X.; Zhu, C.; Guizani, M.; Zhang, W. A High-Availability Data Collection Scheme based on Multi-AUVs for Underwater Sensor Networks. *IEEE Trans. Mob. Comput.* **2020**, *19*, 1010–1022. [CrossRef]
2. Han, G.; Shen, S.; Wang, H.; Jiang, J.; Guizani, M. Prediction-Based Delay Optimization Data Collection Algorithm for Underwater Acoustic Sensor Networks. *IEEE Trans. Veh. Technol.* **2019**, *68*, 6926–6936. [CrossRef]
3. Wang, H.; Han, G.; Liu, Y.; Li, A.; Jiang, J. AUV-Assisted Stratified Source Location Privacy Protection Scheme based on Network Coding in UASNs. *IEEE Internet Things J.* **2023**, *10*, 10636–10648. [CrossRef]
4. Han, G.; Jiang, J.; Shu, L.; Guizani, M. An Attack-Resistant Trust Model Based on Multidimensional Trust Metrics in Underwater Acoustic Sensor Network. *IEEE Trans. Mob. Comput.* **2015**, *14*, 2447–2459. [CrossRef]
5. Han, G.; Shen, S.; Song, H.; Yang, T.; Zhang, W. A Stratification-Based Data Collection Scheme in Underwater Acoustic Sensor Networks. *IEEE Trans. Veh. Technol.* **2018**, *67*, 10671–10682. [CrossRef]
6. Liu, X.; Du, X.; Zhang, S.; Han, D. Cooperative Computing Offloading Scheme via Artificial Neural Networks for Underwater Sensor Networks. *Appl. Sci.* **2023**, *13*, 11886. [CrossRef]
7. Jiang, J.; Han, G.; Zhu, C.; Chan, S.; Rodrigues, J.J.P.C. A Trust Cloud Model for Underwater Wireless Sensor Networks. *IEEE Commun. Mag.* **2017**, *55*, 110–116. [CrossRef]
8. Jiang, S. State-of-the-Art Medium Access Control (MAC) Protocols for Underwater Acoustic Networks: A Survey Based on a MAC Reference Model. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 96–131. [CrossRef]
9. Xue, L.; Cao, C. Joint Channel and Power Assignment for Underwater Cognitive Acoustic Networks on Marine Mammal-Friendly. *Appl. Sci.* **2023**, *13*, 12950. [CrossRef]
10. Nguyen, T.T.; Reddi, V.J. Deep Reinforcement Learning for Cyber Security. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 3779–3795. [CrossRef]
11. Jiang, J.; Han, G.; Shu, L.; Chan, S.; Wang, K. A Trust Model Based on Cloud Theory in Underwater Acoustic Sensor Networks. *IEEE Trans. Ind. Inform.* **2017**, *13*, 342–350. [CrossRef]
12. Du, J.; Han, G.; Lin, C.; Martinez-Garcia, M. ITrust: An Anomaly-Resilient Trust Model Based on Isolation Forest for Underwater Acoustic Sensor Networks. *IEEE Trans. Mob. Comput.* **2022**, *21*, 1684–1696. [CrossRef]
13. Shaikh, R.A.; Jameel, H.; d'Auriol, B.J.; Heejo, L.; Sungyoung, L.; Young-Jae, S. Group-Based Trust Management Scheme for Clustered Wireless Sensor Networks. *IEEE Trans. Parallel Distrib. Syst.* **2009**, *20*, 1698–1712. [CrossRef]
14. Han, G.; He, Y.; Jiang, J.; Wang, N.; Guizani, M.; Ansere, J.A. A Synergetic Trust Model Based on SVM in Underwater Acoustic Sensor Networks. *IEEE Trans. Veh. Technol.* **2019**, *68*, 11239–11247. [CrossRef]
15. Su, Y.; Ma, S.; Zhang, H.; Jin, Z.; Fu, X. A Redeemable SVM-DS Fusion-Based Trust Management Mechanism for Underwater Acoustic Sensor Networks. *IEEE Sens. J.* **2021**, *21*, 26161–26174. [CrossRef]
16. Jiang, J.; Han, G.; Wang, F.; Shu, L.; Guizani, M. An Efficient Distributed Trust Model for Wireless Sensor Networks. *IEEE Trans. Parallel Distrib. Syst.* **2015**, *26*, 1228–1237. [CrossRef]
17. Peng, S.; Yang, A.; Zhong, H.; Feng, Z. A dynamic trust updating model based on multiple constraints in wireless mesh networks. In Proceedings of the 2013 IEEE Third International Conference on Information Science and Technology (ICIST), Yangzhou, China, 23–25 March 2013; pp. 815–819.
18. Jiang, J.; Zhu, X.; Han, G.; Guizani, M.; Shu, L. A Dynamic Trust Evaluation and Update Mechanism Based on C4.5 Decision Tree in Underwater Wireless Sensor Networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 9031–9040. [CrossRef]
19. He, Y.; Han, G.; Jiang, J.; Wang, H.; Martinez-Garcia, M. A Trust Update Mechanism Based on Reinforcement Learning in Underwater Acoustic Sensor Networks. *IEEE Trans. Mob. Comput.* **2022**, *21*, 811–821. [CrossRef]
20. Su, Y.; Guo, L.; Jin, Z.; Fu, X. A Mobile-Beacon-Based Iterative Localization Mechanism in Large-Scale Underwater Acoustic Sensor Networks. *IEEE Internet Things J.* **2021**, *8*, 3653–3664. [CrossRef]
21. Jia, H.; Ding, S.; Xu, X.; Nie, R. The latest research progress on spectral clustering. *Neural Comput. Appl.* **2013**, *24*, 1477–1486. [CrossRef]
22. Caruso, A.; Paparella, F.; Vieira, L.F.M.; Erol, M.; Gerla, M. The meandering current mobility model and its impact on underwater mobile sensor networks. In Proceedings of the IEEE INFOCOM 2008-The 27th Conference on Computer Communications, Phoenix, AZ, USA, 13–18 April 2008; pp. 221–225.
23. Chen, Y.; Zhu, J.; Wan, L.; Fang, X.; Tong, F.; Xu, X. Routing failure prediction and repairing for AUV-assisted underwater acoustic sensor networks in uncertain ocean environments. *Appl. Acoust.* **2022**, *186*, 108479. [CrossRef]
24. Cervesato, I. The Dolev-Yao intruder is the most powerful attacker. In Proceedings of the 16th Annual Symposium on Logic in Computer Science—LICS, Boston, MA, USA, 16–19 June 2001; pp. 1–2.

25. Yang, G.; Dai, L.; Wei, Z. Challenges, threats, security issues and new trends of underwater wireless sensor networks. *Sensors* **2018**, *18*, 3907. [CrossRef]
26. Velusamy, D.; Pugalendhi, G.; Ramasamy, K. A cross-layer trust evaluation protocol for secured routing in communication network of smart grid. *IEEE J. Sel. Areas Commun.* **2019**, *38*, 193–204. [CrossRef]
27. Wu, X.; Huang, J.; Ling, J.; Shu, L. BLTM: Beta and LQI Based Trust Model for Wireless Sensor Networks. *IEEE Access* **2019**, *7*, 43679–43690. [CrossRef]
28. Lim, H.-S.; Moon, Y.-S.; Bertino, E. Provenance-based trustworthiness assessment in sensor networks. In Proceedings of the Seventh International Workshop on Data Management for Sensor Networks, Singapore, 13 September 2010; pp. 2–7.
29. Shao, K.; Luo, F.; Mei, N.; Liu, Z. Normal distribution based dynamical recommendation trust model. *J. Softw.* **2012**, *23*, 3130–3148. [CrossRef]
30. Stojanovic, M. On the relationship between capacity and distance in an underwater acoustic communication channel. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **2007**, *11*, 34–43. [CrossRef]
31. Coates, R.F.W. *Underwater Acoustic Systems*; Springer: London, UK, 1990.
32. Coutinho, R.W.L.; Boukerche, A.; Loureiro, A.A.F. A novel opportunistic power controlled routing protocol for internet of underwater things. *Comput. Commun.* **2020**, *150*, 72–82. [CrossRef]
33. Zhang, M.; Feng, R.; Zhang, H.; Su, Y. A recommendation management defense mechanism based on trust model in underwater acoustic sensor networks. *Future Gener. Comput. Syst.* **2023**, *145*, 466–477. [CrossRef]
34. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
35. Fan, M.; Tan, Y.; Whinston, A.B. Evaluation and design of online cooperative feedback mechanisms for reputation management. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 244–254.