This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2022.3167911, IEEE Transactions on Information Forensics and Security

Noise-based-Protection Message Dissemination Method for Insecure Opportunistic Underwater Sensor Networks

Linfeng Liu, *Member, IEEE*, Zhiyuan Xi, Jiagao Wu, *Member, IEEE*, and Jia Xu, *Senior Member, IEEE*

***Abstract*—Opportunistic Underwater Sensor Networks (OUSNs) are deployed for various underwater applications, such as underwater creature tracking and tactical surveillance. In an OUSN invaded by some eavesdroppers, the data messages disseminated by sensor nodes are probably stolen (captured and cracked) by the eavesdroppers. The data messages are disseminated through acoustic waves which could be altered by the environmental noises, i.e., the acoustic waves containing data messages could be superimposed by the environmental noises. To protect the data messages from being stolen by eavesdroppers and guarantee the required delivery ratio of data messages, we propose a Noise-based-protection Message Dissemination Method (NMDM). In NMDM, the acoustic waves containing data messages are superposed by the environmental noises and converted into some pseudo data messages. The environmental noises around source nodes are identified, encoded, and encrypted into some noise messages. Then, the pseudo data messages and noise messages are individually disseminated to the sink node. Such mechanism makes the eavesdroppers difficult to steal the data messages. Besides, the required delivery ratio of data messages is achieved by measuring the similarities between the nodes and the sink node, i.e., the pseudo data messages and noise messages are preferentially disseminated to the nodes with larger similarities to the sink node. Finally, simulation results demonstrate the superior performance of NMDM. NMDM can reduce the theft ratio of data messages and guarantee the required delivery ratio of data messages effectively.**

***Index Terms*—opportunistic underwater sensor network; environmental noise; theft ratio; eavesdropper.**

I. Introduction

Opportunistic Underwater Sensor Network (OUSN) [1], [2] technology enables various underwater applications, such as underwater creature tracking [3] and tactical surveillance [4]. In an OUSN, each sensor node is comprised of three components: a sensor, a miniature acoustic modem [5], [6], and a mobile underwater vehicle (such as a biomimetic fish, a fish, or a micro-submersible) which can carry the sensor and miniature acoustic modem.

Due to the underwater mobility of nodes, the encounters between nodes are scarce and short, and thus the data messages cannot be disseminated along stable communication paths. Besides, there is no communication infrastructure underwater, and the global scheduling of the movements of nodes for the message dissemination is impossible, i.e., the stable communication paths cannot be formed by the cooperation among nodes. Thus, the encounters between nodes should be exploited to relay the data messages to the sink node.

As shown in Fig. 1, some underwater events are monitored by the nodes, and then are encapsulated into some data messages. The data messages are typically relayed to the sink node through intermittent multi-hops, and thus the delivery ratio of data messages remains a vital issue in the message dissemination of an OUSN.

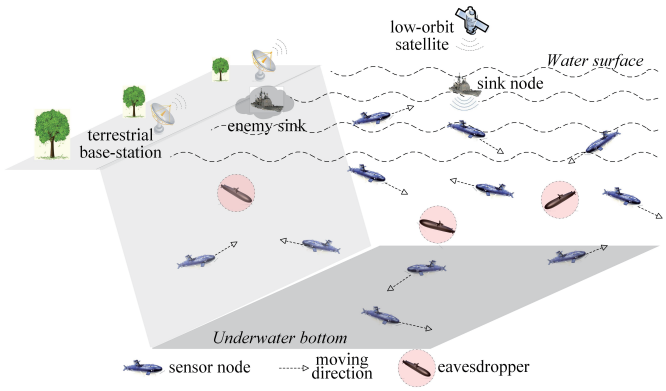


Fig. 1: An OUSN invaded by eavesdroppers.

Moreover, in an OUSN deployed for underwater military applications, some underwater spy-robots [7] termed *eavesdroppers* could be dispatched by the enemy. The eavesdroppers move around some nodes and eavesdrop on their communication channels silently, and the data messages disseminated by the nodes could be stolen (captured and cracked) by the eavesdroppers. Note that the acoustic waves containing data messages will not be altered when they are eavesdropped by the eavesdroppers, because the receivers (can be nodes or eavesdroppers) of acoustic waves can restore the data messages by identifying the wave shapes and frequencies of the received acoustic waves, without altering the acoustic waves. Especially, the locations of eavesdroppers cannot be obtained by the nodes, because each eavesdropper never actively communicates with nodes and other eavesdroppers, which makes the eavesdroppers extremely difficult to be perceived by the nodes, i.e., the nodes are blind to these eavesdroppers.

L. Liu, Z. Xi, J. Wu, and J. Xu are with the School of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China, and also with the Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing 210023, China. (Corresponding author: Jia Xu, email: xujia@njupt.edu.cn.)

Hence, the theft ratio of data messages [1] cannot be reduced through intentionally avoiding the message dissemination in the insecure areas.

Some existing methodologies can help to protect the data messages from being stolen by eavesdroppers, e.g., the data messages can be encrypted by a sophisticated encryption/decryption method before the dissemination. Besides, some identity authentication mechanisms (such as [8], [9]) can be utilized by nodes to identify the eavesdroppers and avoid the message dissemination around these eavesdroppers. However, a sophisticated encryption/decryption method or an identity authentication mechanism will give rise to a large computational cost, which is intolerable to the limited computational power of nodes in an OUSN. A light-weight (simple) encryption/decryption method can be adopted to protect the data messages. Some data messages could be captured by eavesdroppers and delivered to the enemy sink, and then the data messages encrypted by a simple encryption method are probably cracked by the enemy sink which typically has a strong computational power. To this end, this work is launched from the view point of message dissemination to protect the data messages, and the proposed message dissemination method can be technically combined with some simple encryption/decryption methods or identity authentication mechanisms to further enhance the safety of data messages.

The underwater environmental noises [10], [11] come from some artificial equipments (such as ships, wind turbines, or submarines) and some environmental events (such as water flows, rains, tides, or movements of marine mammals). The underwater environmental noises are ubiquitous especially in the shallow water. An example regarding the underwater environmental noises from an offshore wind turbine is given in [12]. [12] shows the pressure levels of the underwater environmental noises collected at a location 110 meters away from the offshore wind turbine when the offshore wind turbine is in different states. The acoustic communication [13] is taken as the typical underwater communication manner, and the environmental noises alter the acoustic waves evidently, i.e., the acoustic waves containing data messages could be superposed by some environmental noises and turn into *pseudo data messages*.

The environmental noises can be identified through Near-field Acoustic Holography (NAH) technique [14], and they can be filtered by some noise suppression techniques, such as the techniques introduced in [15] and [16]. To facilitate the sink node retrieving the original data messages, the environmental noises around source nodes are identified and encoded into several *noise messages* which should be delivered to the sink node as well. When the sink node receives the pseudo data messages and noise messages, the sink node can filter the environmental noises from pseudo data messages and retrieve the original data messages.

In this paper, we conduct a study about exploiting the environmental noises to protect the data messages from being stolen by eavesdroppers, and the message dissemination objective is to reduce the theft ratio of data messages (the proportion of data messages stolen by eavesdroppers) and guarantee the required delivery ratio of data messages (the proportion of data messages delivered to the sink node). Especially, the pseudo data messages and noise messages are individually disseminated to the sink node, and such mechanism can protect the data messages from being stolen by eavesdroppers due to the following reasons:

- The magnitudes and frequencies of environmental noises are time-varying, and the acoustic waves superposed by environmental noises seem much different over time.
- The pseudo data messages have concealed the semantic meanings of original data messages, and thereby the eavesdroppers cannot directly obtain any confidential information from the pseudo data messages.
- The eavesdroppers are hard to capture both pseudo data messages and noise messages, since the pseudo data messages and noise messages are individually disseminated.
- The eavesdroppers do not know which environmental noises are the pseudo data messages superimposed by, even though both the pseudo data messages and noise messages have been captured by them.

Besides, the link prediction technique [17] can predict the future encounters between nodes by measuring their similarities. In our proposed method, the nodes with larger similarities to the sink node are easier to become the new relay nodes, and thus the required delivery ratio of data messages can be guaranteed.

The remainder of this paper is organized as follows: Section II briefly surveys some existing related studies. Section III proposes a system model and a problem formulation. Section IV presents a Noise-based-protection Message Dissemination Method (NMDM). Section V provides some further analyses on NMDM including the method complexity, delivery ratio, and theft ratio. Simulation results for performance evaluation of NMDM are reported in Section VI. Finally, Section VII concludes this paper.

II. Related Work

A. Message Dissemination Methods in DTNs and OUSNs

Extensive studies have been conducted for the message dissemination in Delay-Tolerant Networks (DTNs) or OUSNs.

Zhang *et al.* develop a Beam width and Direction Concerned Routing protocol (BDCR) for the message dissemination problem [18]. BDCR can obtain a high delivery ratio by considering the beam width and three-dimensional direction. Reference [19] presents two cooperative forwarding schemes by leveraging the data fusion: Epidemic Forwarding (EF) with fusion and binary spray-and-wait with fusion. The data messages are considered to be spatially-temporally correlated in the dissemination process, and then the dissemination law of data messages

[1]The theft ratio of data messages is linearly related to the probability of data messages being stolen by eavesdroppers.

can be determined. In [20], a sub-gigahertz wireless technology is utilized to establish the links between mobile users in an opportunistic network, and the delivery ratio of data messages is improved. Reference [21] focuses on the information dissemination in urban scenarios with different crowd densities and renewal rates, and an epidemic protocol based on the fragmentation of data messages is proposed to increase the delivery ratio and shorten the delivery delay.

An opportunistic routing framework considering the characteristics of underwater networks, such as network density, traffic load, underwater environment, and underwater acoustic channels is proposed in [22]. Likewise, an asymmetric-link-based reverse routing method is designed in [23] to ensure the bidirectional communications from source nodes to destination nodes. In this method, each node maintains a neighbor table where the table items represent the link states, and the routing paths are established by prioritizing the utilization of symmetric links. Ahmed $et\ al.$ propose a mobility assisted geo-opportunistic routing paradigm based on interference avoidance for OUSNs [24], and the network volume is divided into small cubes to reduce the interferences and make proper routing decisions for efficient energy consumption.

B. Security Issues in Message Dissemination

In [25], an access control scheme with authentication for message dissemination is put forward. In this scheme, the pseudonym is integrated with the identity based signature, and thus the data messages in vehicular communications can be authenticated. Reference [26] proposes a secure message dissemination scheme with policy enforcement for vehicular ad hoc networks, and it allows the vehicles to delegate most of the decryption computations to the nearest roadside units. Wu $et\ al.$ present a data forwarding mechanism for mobile social networks to detect and resist the on/off attacks [27]. Moreover, some contact-avoidance routing protocols have been presented for the security threats in message dissemination, and most of these works seek to physically avoid the message dissemination happening in the insecure areas. For example, a secure opportunistic path model integrating the delivery probability and the safety of opportunistic paths is constructed in [28]. Then, a Contact Avoidance Routing (CAR) protocol is raised to securely disseminate a data message to the destination node against the contact-based compromise attacks. In [29], an anti-eavesdropping framework is provided to obscure the unauthorized eavesdroppers and diminish their capability of inferring the information through smart beamforming over the wiretap channel.

The above works are based on the assumption that the eavesdroppers (adversaries) can be identified through the nodal action analysis or reputation assessments [30], whereas the eavesdroppers in an OUSN are difficult to be perceived by nodes.

C. Data Protection based on Noises

Several works have employed the noises to protect the confidential data, such as [31], which evaluates the privacy and utility performance of Laplace noise addition for numeric data. The results indicate that the mechanism of Laplace noise addition can elevate the promised level of privacy effectively. In [32], the semantic noises are utilized to protect the nominal data. Through exploiting the structured knowledge sources such as ontologies, it is able to distort the nominal data while preserving preferable semantics and analytical utility. Reference [33] proposes a noise-added selection algorithm, and a location privacy protection method satisfying differential privacy is proposed to prevent the data messages from privacy disclosures. Considering the effects of underwater environmental noise on trust reliability, a new anomaly and attack resilient trust model based on the isolation forest is proposed to achieve the accurate calculation of node trust [34]. Likewise, a trust update mechanism based on reinforcement learning is proposed to deal with the inevitable dynamic fluctuations in the underwater environment [35].

However, the above-mentioned noise addition mechanisms have not been applied into the message dissemination of opportunistic networks. In an opportunistic network, each data message could be superposed by multiply noises on multiply relay nodes, and the redundant noises must be filtered for the retrieves of original data messages on the sink node.

D. Motivation of Our Work

In an OUSN invaded by some eavesdroppers, the data messages should be protected from being stolen by eavesdroppers as much as possible. However, considering the aforementioned fact that the nodes are blind to the eavesdroppers, and hence the theft ratio cannot be reduced by purposely avoiding the message dissemination in the insecure areas around eavesdroppers.

The environmental noises could alter the acoustic waves containing data messages, and thus we are motivated to exploit the environmental noises to protect the data messages. With regard to each data message, it is allowed to be superimposed by the environmental noise around the source node and converted into a pseudo data message. Besides, the environmental noise is identified, encoded, and encrypted into a noise message. Then, the pseudo data message and noise message are individually disseminated to the sink node. The eavesdroppers adjacent to the relay nodes cannot learn the environmental noise around the source node. Besides, each pseudo data message could be re-superposed by the local environmental noises around relay nodes, making the pseudo forms of the same data message quite different on different relay nodes, and thus the eavesdroppers are hard to steal the original data message.

Moreover, the similarities between the nodes and the sink node are measured for selecting the new relay nodes, and the nodes with larger similarities to the sink node are easier to become the new relay nodes.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2022.3167911, IEEE Transactions on Information Forensics and Security

III. System Model and Problem Formulation

The explanations of main notations are presented in TABLE I.

TABLE I: Main Notations

Parameter	Description
$\boldsymbol{D}$	Underwater space
N	Number of sensor nodes
N_e	Number of eavesdroppers
r_{c}	Communication range of each node
$(i,j)^{(t)}$	Potential communication link between v_i and v_j at the t-th time slot
$d(i,j)^{(t)}$	Distance between v_i and v_j at the t-th time slot
$P(i,j)^{(t)}$	Existence probability of $(i,j)^{(t)}$
$T(\tau^*)$	Theft ratio of data messages (the dissemination deadline is τ^* time slots)
$D(\tau^*)$	Delivery ratio of data messages (the dissemination deadline is τ^* time slots)
$\widetilde{\mathcal{R}}$	Required delivery ratio of data messages
${a_{js}}^{(t)}$	Similarity of v_j to the sink node v_s at the t-th time slot
${S_{js}}^{(t_1,t_2)}$	Cosine similarity of v_j (at the t_1-th time slot) to v_s (at the t_2-th time slot)
$H(\boldsymbol{m}_p)$	Set of relay nodes of pseudo data message $\boldsymbol{m}_p$
$H(\boldsymbol{n}_e)$	Set of relay nodes of noise message $\boldsymbol{n}_e$

A. OUSN Model

There are N mobile sensor nodes deployed in a convex underwater space $\boldsymbol{D}$, where $\boldsymbol{D} \in \mathbb{R}^{+3}$. The time is divided into discrete time slots, and each data message needs to be disseminated from source node to sink node during a period of τ^* time slots (dissemination deadline). Note that the number of copies of a pseudo data message (or a noise message) can be restricted by the dissemination deadline.

There are N_e eavesdroppers, and they could move around some sensor nodes and eavesdrop on their communication channels silently.

The communication range of each node is denoted by r_{c}. The coordinate of a node v_i at the t-th time slot is denoted by $C(i)^{(t)}$. The distance between two nodes v_i and v_j at the t-th time slot is denoted by $d(i,j)^{(t)}$. If $d(i,j)^{(t)} \leq r_{\mathrm{c}}$, then $(i,j)^{(t)}$ is taken as a potential communication link. Due to the signal irregularity in underwater communications, the existence probability of the link $(i,j)^{(t)}$ is determined by:

$$P(i,j)^{(t)} = \begin{cases} 0, & \text{if } d(i,j)^{(t)} > r_{\mathrm{c}}, \\ c_1 \cdot \Omega(j)^{-\zeta} \cdot d(i,j)^{(t)^{-\eta}}, & \text{otherwise}, \end{cases} \tag{1}$$

where c_1 is a constant, and $\Omega(j)$ denotes the signal irregularity around v_j which is caused by various factors, such as antenna directions, antenna gains, battery status, signal-noise-ratio threshold, and obstacles [36]. The signal irregularity around each node is assumed to obey a uniform distribution $U(\Omega_{min}, \Omega_{max})$ [37]. Equation (1) implies that the existence probability of a potential communication link is decreased with the increase of link length or the increase of signal irregularity. ζ and η are two exponents reflecting the impacts of signal irregularity and link length on the existence probability, respectively.

An example is shown in Fig. 2, two eavesdroppers are located in the communication range of a node which is disseminating a data message. An eavesdropper captures the data message, whereas another eavesdropper cannot capture the data message because the communication link from the node to the eavesdropper is disconnected due to the signal irregularity. After capturing some data messages, the eavesdropper could move back to the enemy sink for the information analysis and decision making.

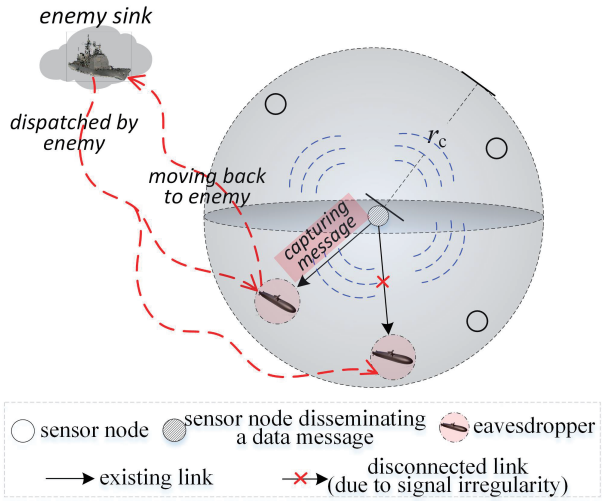


Fig. 2: Two eavesdroppers moving around a sensor node.

Besides, the underwater mobility pattern of nodes in an OUSN is comprised of autonomous movement and coordinate deviation [38].

B. Objective Function

To reduce the theft ratio of data messages and guarantee the required delivery ratio of data messages, the objective function is formulated as follows:

$$\min\ T(\tau^*), \qquad s.t.\ D(\tau^*) \geq \widetilde{\mathcal{R}}, \tag{2}$$

where $T(\tau^*)$ denotes the theft ratio which is defined as the proportion of data messages stolen by eavesdroppers during τ^* time slots. $D(\tau^*)$ denotes the delivery ratio which is defined as the proportion of data messages delivered to the sink node during τ^* time slots. $\widetilde{\mathcal{R}}$ denotes the required delivery ratio of data messages. The problem objective is that the proportion of data messages stolen by eavesdroppers is reduced as much as possible, while the proportion of data messages delivered to the sink node is larger than a threshold $\widetilde{\mathcal{R}}$.

(2) indicates that the main security threat in the message dissemination problem of insecure OUSNs is the theft of data messages. To this end, we exploit the environmental noises to protect the data messages, i.e., the data messages are superposed by environmental noises and converted into some pseudo data messages. The environmental noises around source nodes are identified, encoded, and encrypted into some noise messages. Then, the pseudo data messages

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2022.3167911, IEEE Transactions on Information Forensics and Security

and noise messages are individually disseminated to the sink node.

IV. Noise-based Message Dissemination Method

A. Overview of Message Dissemination Process

In our proposed Noise-based Message Dissemination Method (NMDM), the acoustic waves containing data messages are superposed by the environmental noises around source nodes and converted into pseudo data messages. Besides, the environmental noises are identified, encoded, and encrypted into some noise messages on the source nodes. The message dissemination process is illustrated in Fig. 3.

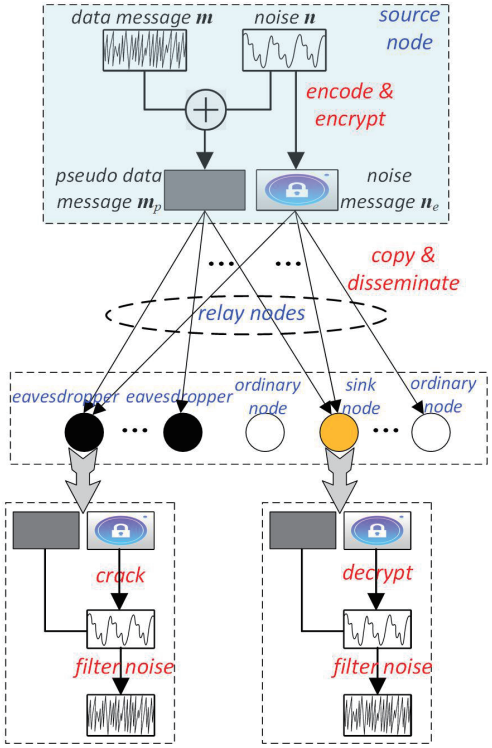


Fig. 3: Message dissemination process of NMDM.

Note that the delivery ratio of data messages will be much larger than the theft ratio of data messages, due to the following mechanisms adopted in NMDM:

(i) The noise messages are encrypted, and thus the data messages are hard to be stolen by eavesdroppers, even though the eavesdroppers have captured the pseudo data messages and noise messages.

(ii) The pseudo data messages and noise messages could be re-superposed by the local environmental noises around relay nodes, which makes these messages seem much different on different relay nodes.

(iii) The similarities between the nodes and the sink node are measured, and the messages (the pseudo data messages and noise messages) are preferentially disseminated to the nodes with larger similarities to the sink node. This mechanism can expedite the message deliveries.

B. Pseudo Data Message and Noise Message

Suppose the data message $\boldsymbol{m}$ is generated by a source node, and the environmental noise around the source node is denoted by $\boldsymbol{n}$.

$\boldsymbol{m}$ is superposed by $\boldsymbol{n}$, which produces a pseudo data message $\boldsymbol{m}_p$, i.e.,

$$\boldsymbol{m} + \boldsymbol{n} \rightarrow \boldsymbol{m}_p, \tag{3}$$

where the symbol "$+$" denotes a superposition operation on the acoustic waves of the data message $\boldsymbol{m}$ and the environmental noise $\boldsymbol{n}$. Two examples regarding the superposition of acoustic waves (the acoustic wave of a data message and the acoustic wave of an environmental noise) are given in Fig. 4.

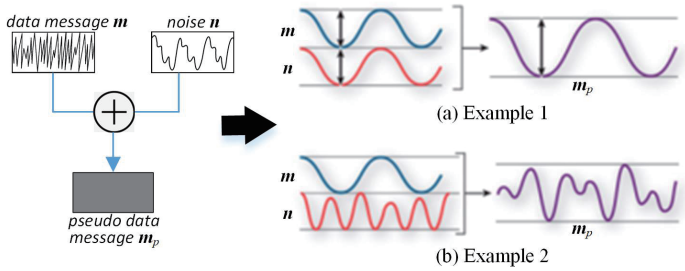


Fig. 4: Two examples regarding the superposition of acoustic waves.

The environmental noise $\boldsymbol{n}$ can be identified and localized by the NAH technique, as shown in Fig. 5, which indicates that NAH can identify and reconstruct the noise by measuring the acoustic pressure on a hologram plane.

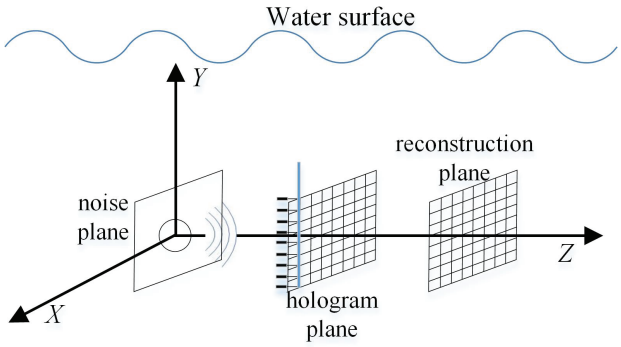


Fig. 5: Near-field acoustic holography.

Then, $\boldsymbol{n}$ is encoded and encrypted into a noise message $\boldsymbol{n}_e$ by a symmetrical encryption method, such as Advanced Encryption Standard (AES) [39]. $\boldsymbol{m}_p$ and $\boldsymbol{n}_e$ will be individually copied and disseminated to the sink node during the future τ^* time slots.

To facilitate the sink node retrieving the original data messages, when the pseudo data messages and noise messages are re-superposed by local environmental noises around relay nodes, the local environmental noises must be filtered from these messages. The noise filtering operation on a relay node is depicted in Fig. 6.

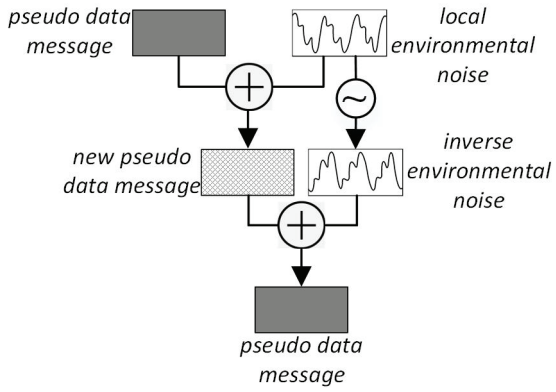


Fig. 6: Noise filtering operation on a relay node.

C. Link Prediction

Each node records the historical communication links (the encounters with other nodes), and the link record of each node is exchanged with encountered nodes to diversify and enlarge the link record.

At each time slot, the link record of each node is utilized to locally measure the similarity between itself and the sink node, and then the obtained similarity is sent to the neighbors. Besides, each node disseminates the held messages to κ neighbors (new relay nodes) which are selected based on the similarities between its neighbors and the sink node. Note that the global computations and information exchanges are not required in the link prediction process. TABLE II explains some symbols used in the description of link prediction.

TABLE II: Symbols in Link Prediction

Symbol	Definition
$msg\_list(i)^{(t)}$	List of held messages of v_i at the t-th time slot
$rec(i)^{(t)}$	Link record of v_i at the t-th time slot
$\Gamma_{in}(i)^{(t)}$	In-neighbor set of v_i at the t-th time slot
$\Gamma_{out}(i)^{(t)}$	Out-neighbor set of v_i at the t-th time slot

The details of link prediction are provided as follows:

Stage 1. *Link Record Exchange*. At the start of the t-th time slot, each node (such as v_i) sends an $inquire\_msg$, including a quintuple $(ID, t, C(i)^{(t)}, rec(i)^{(t-1)}, msg\_list(i)^{(t)})$ in the neighborhood. The in-neighbor set and out-neighbor set of v_i are first initialized by:

$$\Gamma_{in}(i)^{(t)} \leftarrow \emptyset,\ \Gamma_{out}(i)^{(t)} \leftarrow \emptyset. \tag{4}$$

The in-neighbor set of a node denotes the set of neighbors which have communication links from them to the node, and the out-neighbor set of a node denotes the set of neighbors which have communication links from the node to them. For example, in Fig. 7, the in-neighbor set of v_j is $\{v_2, v_3, v_4, v_5\}$, and the out-neighbor set of v_j is $\{v_1, v_4, v_5\}$.

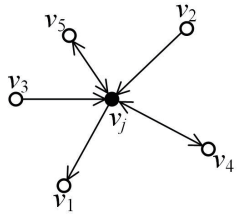


Fig. 7: In-neighbor set and out-neighbor set.

Stage 2. *Neighbor Identification*. If v_j has received the $inquire\_msg$ from v_i, then $\Gamma_{in}(j)^{(t)}$ and $rec(j)^{(t)}$ are updated by:

$$\begin{cases} \Gamma_{in}(j)^{(t)} \leftarrow \Gamma_{in}(j)^{(t)} \bigcup v_i, \\ rec(j)^{(t)} \leftarrow rec(j)^{(t-1)} \bigcup rec(i)^{(t-1)} \bigcup (i,j)^{(t)}. \end{cases} \tag{5}$$

After receiving the $inquire_msg$s from neighbors, v_j sends a $reply\_msg$ which is expressed as a triple $(ID, t, \Gamma_{in}(j)^{(t)})$. If v_i $(v_i \in \Gamma_{in}(j)^{(t)})$ has received the $reply\_msg$ from v_j, then $\Gamma_{out}(i)^{(t)}$ and $rec(i)^{(t)}$ are updated by:

$$\begin{cases} \Gamma_{out}(i)^{(t)} \leftarrow \Gamma_{out}(i)^{(t)} \bigcup v_j, \\ rec(i)^{(t)} \leftarrow rec(i)^{(t)} \bigcup (j,i)^{(t)}. \end{cases} \tag{6}$$

Stage 3. *Similarity Calculation*. ${a_{js}}^{(t+1)}$ denotes the similarity between v_j and the sink node v_s at the $(t+1)$-th time slot, and it can measure the existence probability of the potential communication link $(j,s)^{(t+1)}$. ${a_{js}}^{(t+1)}$ is expressed as:

$${a_{js}}^{(t+1)} = \sum_{t_1=1}^{t} \sum_{t_2>t_1}^{t} w^{(t_1,t_2)} \cdot {S_{js}}^{(t_1,t_2)}, \tag{7}$$

where ${S_{js}}^{(t_1,t_2)}$ denotes the cosine similarity between v_j (at the t_1-th time slot) and v_s (at the t_2-th time slot), where t_1 and t_2 are two time slots not later than t. (7) indicates that the similarity is calculated as a weighted cosine similarity over previous time slots.

Specifically, the cosine similarity ${S_{js}}^{(t_1,t_2)}$ is obtained by:

$${S_{js}}^{(t_1,t_2)} = \begin{cases} 1, & \text{if } t_1 = t_2 \text{ and } (j,s)^{(t_1)} \text{ exists}, \\ \dfrac{\left|\Gamma_{out}(j)^{(t_1)} \bigcap \Gamma_{in}(s)^{(t_2)}\right|}{\sqrt{\left|\Gamma_{out}(j)^{(t_1)}\right| \cdot \left|\Gamma_{in}(s)^{(t_2)}\right|}}, & \text{otherwise}, \end{cases} \tag{8}$$

where $\Gamma_{out}(j)^{(t_1)}$ denotes the out-neighbor set of v_j at the t_1-th time slot, $\Gamma_{in}(s)^{(t_2)}$ denotes the in-neighbor set of v_s at the t_2-th time slot, and $|\cdot|$ denotes the cardinality of a set.

(8) implies that more common neighbors in the out-neighbor set of v_j and the in-neighbor set of v_s give rise to a larger cosine similarity between v_j and v_s.

$w^{(t_1,t_2)}$ denotes a weight of the cosine similarity to reflect the impacts of t_1 and t_2, and $w^{(t_1,t_2)}$ is given by:

$$w^{(t_1,t_2)} = \frac{\frac{(t_1 \cdot t_2)^{\alpha}}{(|t_1 - t_2| + 1)^{\beta}}}{\sum_{t_1=1}^{t} \sum_{t_2>t_1}^{t} \left[\frac{(t_1 \cdot t_2)^{\alpha}}{(|t_1 - t_2| + 1)^{\beta}}\right]}, \tag{9}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2022.3167911, IEEE Transactions on Information Forensics and Security

which indicates that a smaller $w^{(t_1,t_2)}$ is obtained when the temporal difference (the difference between t_1 and t_2) becomes larger, or the numerical values of t_1 and t_2 become smaller.

The implications of (9) are given as follows: (*i*) The communication links generated more recently provide more important reference for predicting the future communication links. (*ii*) The common neighbors exist during a shorter time interval are more valuable for predicting the future communication links.

After that, v_j sends the obtained ${a_{js}}^{(t+1)}$ to v_i.

Fig. 8 provides an example to illustrate the similarity expression. The communication links are asymmetry due to the different signal irregularities around different nodes, and v_2 is taken as a common neighbor of v_j and v_s according to the communication links $(j,2)^{(t)}$ and $(2,s)^{(t+n)}$, i.e., v_j can disseminate a message to v_2 at the t-th time slot, and afterwards v_2 disseminates the message to v_s at the $(t+n)$-th time slot.

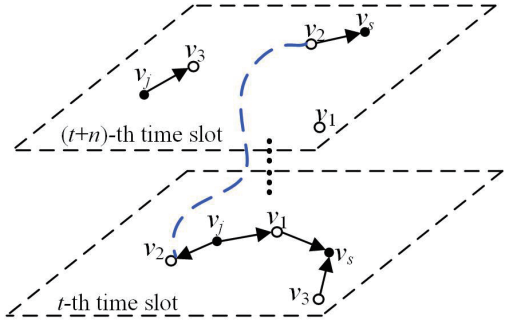


Fig. 8: Similarity between v_j and v_s.

D. Relay Node Candidates

Suppose there is a pseudo data message $\boldsymbol{m}_p$ and an encrypted noise message $\boldsymbol{n}_e$. Each relay node of $\boldsymbol{m}_p$ (or $\boldsymbol{n}_e$) produces and disseminates κ new copies of $\boldsymbol{m}_p$ (or $\boldsymbol{n}_e$) at each time slot. Suppose v_i is a relay node of $\boldsymbol{m}_p$ (or $\boldsymbol{n}_e$) at the t-th time slot. With regard to another node v_j ($v_j \in \Gamma_{out}(i)^{(t)} \bigcap \Gamma_{in}(i)^{(t)}$), the similarity ${a_{js}}^{(t+1)}$ is calculated by v_j and then sent to v_i.

The similarities received from neighbors (within the communication range of r_c) are sorted in a descending order by v_i, and the new relay nodes of $\boldsymbol{m}_p$ (or $\boldsymbol{n}_e$) are selected according to the following cases:

- *Case* 1. If $v_j \in H(\boldsymbol{m}_p)$ and $v_j \in H(\boldsymbol{n}_e)$, where $H(\boldsymbol{m}_p)$ and $H(\boldsymbol{n}_e)$ denote the sets of relay nodes of $\boldsymbol{m}_p$ and $\boldsymbol{n}_e$, respectively. Then, v_j is not taken as a relay node candidate.
- *Case* 2. If $v_j \in H(\boldsymbol{m}_p)$, $v_j \notin H(\boldsymbol{n}_e)$ and ${a_{js}}^{(t+1)} \geq \widetilde{a}$ ($\widetilde{a}$ is a similarity threshold), then v_j is taken as a relay node candidate of $\boldsymbol{n}_e$.
- *Case* 3. If $v_j \notin H(\boldsymbol{m}_p)$, $v_j \in H(\boldsymbol{n}_e)$ and ${a_{js}}^{(t+1)} \geq \widetilde{a}$, then v_j is taken as a relay node candidate of $\boldsymbol{m}_p$.
- *Case* 4. If $v_j \in H(\boldsymbol{m}_p)$, $v_j \notin H(\boldsymbol{n}_e)$ and ${a_{js}}^{(t+1)} < \widetilde{a}$, then v_j is not taken as a relay node candidate.
- *Case* 5. If $v_j \notin H(\boldsymbol{m}_p)$, $v_j \in H(\boldsymbol{n}_e)$ and ${a_{js}}^{(t+1)} < \widetilde{a}$, then v_j is not taken as a relay node candidate.
- *Case* 6. If $v_j \notin H(\boldsymbol{m}_p)$, $v_j \notin H(\boldsymbol{n}_e)$ and ${a_{js}}^{(t+1)} \geq \widetilde{a}$, then v_j is taken as a relay node candidate of both $\boldsymbol{m}_p$ and $\boldsymbol{n}_e$.
- *Case* 7. If $v_j \notin H(\boldsymbol{m}_p)$, $v_j \notin H(\boldsymbol{n}_e)$ and ${a_{js}}^{(t+1)} < \widetilde{a}$, then v_j is taken as a relay node candidate of either $\boldsymbol{m}_p$ or $\boldsymbol{n}_e$.

Especially, Case 7 indicates that when the probability of v_j encountering the sink node v_s is very small, v_j cannot become the relay node candidate of $\boldsymbol{m}_p$ and $\boldsymbol{n}_e$ simultaneously, and thus the probability of $\boldsymbol{m}_p$ and $\boldsymbol{n}_e$ being captured by the same eavesdropper can be significantly reduced.

An example is given in Fig. 9 to illustrate the selection of relay node candidates. There are five neighbors v_1, v_2, v_3, v_4, v_5, and some of them are selected to become the relay node candidates of $\boldsymbol{m}_p$ and/or $\boldsymbol{n}_e$ according to their current statuses and the similarities to v_s.

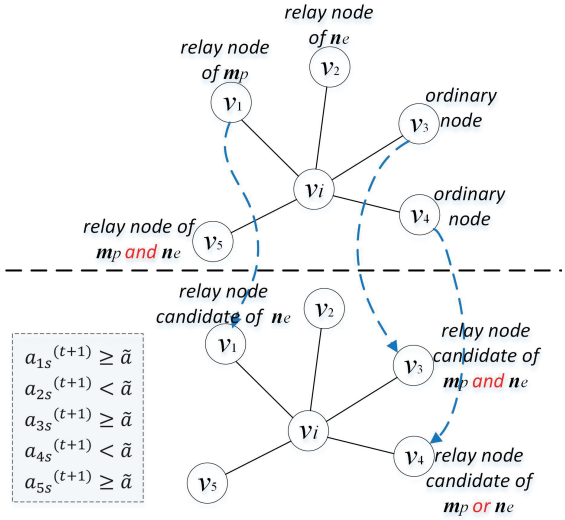


Fig. 9: Relay node candidates.

E. Message Dissemination

With regard to a data message $\boldsymbol{m}$, the pseudo data message $\boldsymbol{m}_p$ and the encrypted noise message $\boldsymbol{n}_e$ are individually disseminated to the sink node. The relay nodes of $\boldsymbol{m}_p$ or $\boldsymbol{n}_e$ must filter the local environmental noises before the further dissemination. Each relay node sends the copies of $\boldsymbol{m}_p$ or $\boldsymbol{n}_e$ to κ new relay nodes which are with the larger similarities to the sink node. The flowchart of disseminating a data message is shown in Fig. 10.

After receiving the copies of $\boldsymbol{m}_p$ and $\boldsymbol{n}_e$, the sink node extracts $\boldsymbol{m}$ through filtering $\boldsymbol{n}_e$ from $\boldsymbol{m}_p$, indicating that $\boldsymbol{m}$ is delivered to the sink node. If $\boldsymbol{m}$ has been delivered to the sink node before the dissemination deadline, an announcement originated from the sink node will be broadcasted to all the relay nodes of $\boldsymbol{m}_p$ and $\boldsymbol{n}_e$ to stop the further dissemination. Note that the size of an announcement is very small and can be piggybacked with other messages. Thus,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2022.3167911, IEEE Transactions on Information Forensics and Security

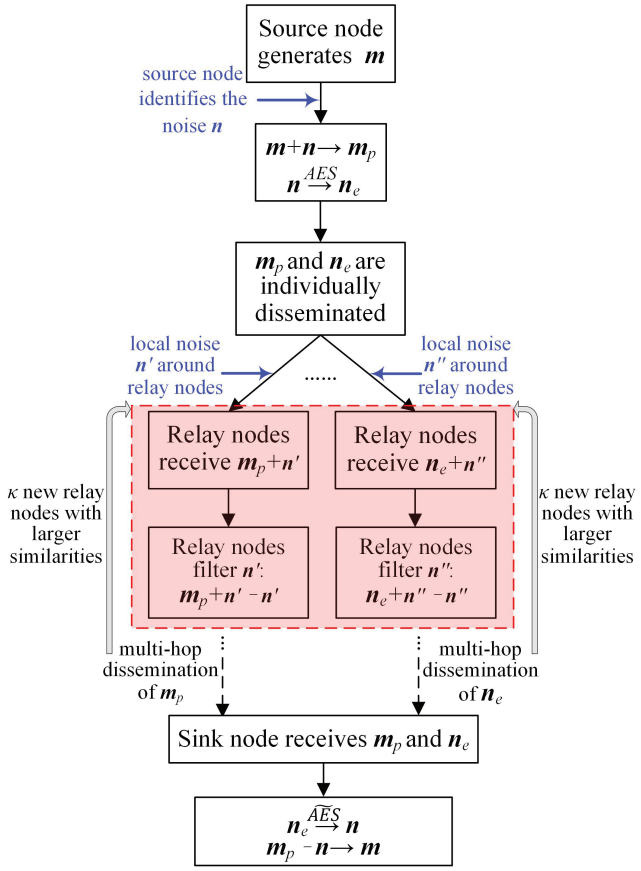


Fig. 10: Flowchart of disseminating a data message.

the effect of broadcast is negligible. If the dissemination deadline has been expired, then $\boldsymbol{m}_p$ and $\boldsymbol{n}_e$ are discarded by the relay nodes.

If both $\boldsymbol{m}_p$ and $\boldsymbol{n}_e$ are captured by an eavesdropper, and $\boldsymbol{n}_e$ is successfully cracked by the eavesdropper. Then, $\boldsymbol{m}$ is considered to be stolen by the eavesdropper.

Fig. 11 provides an example of message dissemination. In Fig. 11, a data message is decomposed into a pseudo data message and a noise message. Two eavesdroppers capture the pseudo data message and the noise message respectively, while another eavesdropper captures both the pseudo data message and noise message.

V. Method Analysis

A. Method Complexity

In NMDM, each data message is disseminated through disseminating a pseudo data message and an encrypted noise message, and thus the number of disseminated messages is related to the number of data messages. However, some information exchanges are required for the selection of new relay nodes. When the number of neighbors of each node is related to N, the communication complexity of NMDM is up to $\mathrm{O}(N^2)$.

There are three parts contributed to the computational complexity of NMDM: (*i*) The encryptions of the environmental noises around source nodes, and the computational complexity is $\mathrm{O}(N)$. (*ii*) The filtering operations of local

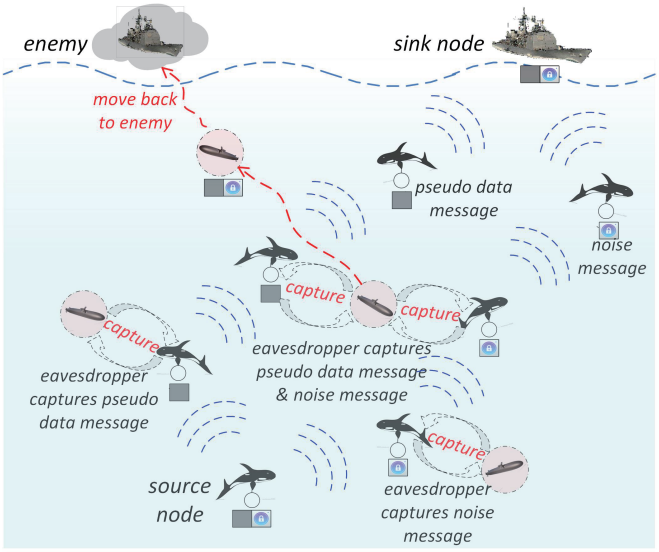


Fig. 11: An example of message dissemination.

environmental noises around relay nodes, and the computational complexity is $\mathrm{O}(N)$ as well. (*iii*) The sorts of the similarities of nodes, and the computational complexity is $\mathrm{O}(N \cdot \log_2 N)$ when *Quick Sort* method is adopted. Therefore, the computational complexity of NMDM is expressed as $\mathrm{O}(N \cdot \log_2 N)$.

B. Expected Delivery Ratio

The number of in-neighbors of a node is referred to as the in-degree of the node, and the number of out-neighbors of a node is referred to as the out-degree of the node. We first prove that the in-degrees of nodes follow a power law distribution.

Lemma 1: The in-degrees of nodes follow a power law distribution.

Proof: The in-degree of a node v_i is given by:

$$\begin{aligned} deg_i^{(t)} &= \sum_{l=1}^{N-1} P(l,i)^{(t)} \approx c_1 \cdot \Omega(i)^{-\zeta} \cdot \frac{\frac{4}{3}\pi \cdot {r_c}^3}{|\boldsymbol{D}|} \cdot \left(\frac{3r_c}{4}\right)^{-\eta} \\ &= c_2 \cdot \Omega(i)^{-\zeta} \cdot {r_c}^{3-\eta}, \end{aligned} \tag{10}$$

where $c_2 = \left(\frac{4}{3}\right)^{\eta+1} \cdot \frac{c_1 \cdot \pi}{|\boldsymbol{D}|}$. ζ and η are two exponents reflecting the impacts of signal irregularity and link length on the existence probability, as defined in (1).

According to the expression of $deg_i^{(t)}$, we have that $deg_i^{(t)} > deg_j^{(t)}$ when $i < j$. Then, the cumulative in-degree distribution of v_i is expressed as:

$$P_{\mathrm{cd}}(deg_i^{(t)}) = \frac{\Omega(i) - \Omega_{min}}{\Omega_{max} - \Omega_{min}}. \tag{11}$$

Let $deg_i^{(t)} = k$, and hence there is $k = c_2 \cdot \Omega(i)^{-\zeta} \cdot {r_c}^{3-\eta}$. We obtain that $\Omega(i)^{-\zeta} = \frac{k}{c_2} \cdot {r_c}^{\eta-3}$, and then $P_{\mathrm{cd}}(k)$ can be expressed as:

$$P_{\mathrm{cd}}(k) = \frac{k^{-\frac{1}{\zeta}}}{\left(\frac{c_2}{{r_c}^{\eta-3}}\right)^{-\frac{1}{\zeta}} \cdot (\Omega_{max} - \Omega_{min})} - \frac{\Omega_{min}}{\Omega_{max} - \Omega_{min}}. \tag{12}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2022.3167911, IEEE Transactions on Information Forensics and Security

Therefore, the in-degree distribution of nodes satisfies that $Pr(k) \propto k^{-\left(\frac{1}{\zeta}+1\right)}$ approximatively. □

Without loss of generality, let $Pr(\Gamma_{in}) = \xi \cdot \Gamma_{in}^{-\gamma}$ (where $2 < \gamma \leq 3$, $\xi = \frac{1}{\sum_{\Gamma_{in}=1}^{N-1} \Gamma_{in}^{-\gamma}}$). The nodes with different signal irregularities are assumed to be uniformly distributed, and hence the out-degree of each node Γ_{out} is equal to $\frac{4\overline{p}\cdot\pi\cdot r_c^3\cdot N}{3|\boldsymbol{D}|}$, where $\overline{p}$ denotes the expected existence probability of a potential communication link. $\overline{p}$ is expressed as:

$$\begin{aligned}\overline{p} =& \frac{c_1 \cdot \int_{\Omega_{min}}^{\Omega_{max}} \omega^{-\zeta} d\omega}{\Omega_{max} - \Omega_{min}} \cdot \frac{\sum_{k=1}^{\frac{r_c}{r_0}} \left[k^3 - (k-1)^3\right] \cdot (k \cdot r_0)^{-\eta}}{(\frac{r_c}{r_0})^3} \\ =& \frac{c_1 \cdot ({\Omega_{max}}^{1-\zeta} - {\Omega_{min}}^{1-\zeta})}{(1-\zeta)\cdot(\Omega_{max} - \Omega_{min})} \cdot \frac{\sum_{k=1}^{\frac{r_c}{r_0}} (3k^2 - 3k + 1) \cdot (k \cdot r_0)^{-\eta}}{(\frac{r_c}{r_0})^3},\end{aligned} \tag{13}$$

where r_0 denotes the minimum distance between two neighbors.

Because the movement of each node follows the same underwater mobility model, each node can be approximatively considered to encounter another node with an equivalent probability. When the link records of nodes have been adequately exchanged for a long period, the similarities obtained by nodes tally with the pattern of the underwater mobility model. Thus, the numbers of nodes in $\Gamma_{out}(j)^{(t_1)}$, $\Gamma_{in}(s)^{(t_2)}$, and $\Gamma_{out}(j)^{(t_1)} \bigcap \Gamma_{in}(s)^{(t_2)}$ are approximatively equivalent at different time slots, which yields that ${S_{js}}^{(t_1,t_2)} = {S_{js}}^{(t_1',t_2')}$. Hence, we have that ${a_{js}}^{(t+1)} = {S_{js}}^{(t_1,t_2)}$.

Then, the complementary cumulative distribution function can be expressed as:

$$\begin{aligned}F(\widetilde{a}) &= P\left(\frac{\frac{\Gamma_{out}\cdot\Gamma_{in}}{N-2}}{\sqrt{\Gamma_{out}\cdot\Gamma_{in}}} \geq \widetilde{a}\right) = P\left(\Gamma_{in} \geq \frac{\widetilde{a}^2\cdot(N-2)^2}{\Gamma_{out}}\right) \\ &= \sum_{\Gamma_{in}=\frac{\widetilde{a}^2\cdot(N-2)^2}{\Gamma_{out}}}^{N-1} \xi\cdot{\Gamma_{in}}^{-\gamma} \approx \xi\cdot\int_{\frac{\widetilde{a}^2\cdot(N-2)^2}{\Gamma_{out}}}^{N-1} {\Gamma_{in}}^{-\gamma} d\Gamma_{in} \\ &= \frac{\xi}{\gamma-1}\cdot\left\{\left[\frac{\widetilde{a}^2\cdot(N-2)^2}{\Gamma_{out}}\right]^{1-\gamma} - (N-1)^{1-\gamma}\right\}.\end{aligned} \tag{14}$$

Thus, when a data message $\boldsymbol{m}$ has been generated for $(\tau-1)$ time slots, the number of relay nodes holding both $\boldsymbol{m}_p$ and $\boldsymbol{n}_e$ is written as $P\left(\frac{|\Gamma_{out}\bigcap\Gamma_{in}|}{\sqrt{\Gamma_{out}\cdot\Gamma_{in}}} \geq \widetilde{a}\right)\cdot(1+\kappa)^{\tau-1}$, where κ denotes the copy number of a message disseminated by a relay node at each time slot.

Besides, the number of relay nodes holding either $\boldsymbol{m}_p$ or $\boldsymbol{n}_e$ is written as $\left\{1 - P\left(\frac{|\Gamma_{out}\bigcap\Gamma_{in}|}{\sqrt{\Gamma_{out}\cdot\Gamma_{in}}} \geq \widetilde{a}\right)\right\}\cdot(1+\kappa)^{\tau-1}$. Therefore, the expected delivery ratio of data messages is given by:

$$\begin{aligned}\mathbb{E}(D(\tau)) = \overline{p}\cdot&\left[1 - \left(1 - \frac{2\pi\cdot {r_c}^3}{3\,|\boldsymbol{D}|}\right)^{F(\widetilde{a})\cdot\frac{(1+\kappa)^\tau-1}{\kappa}}\right] \\ &+ \left\{\overline{p}\cdot\left[1 - \left(1 - \frac{2\pi\cdot {r_c}^3}{3\,|\boldsymbol{D}|}\right)^{[1-F(\widetilde{a})]\cdot\frac{(1+\kappa)^\tau-1}{\kappa}}\right]\right\}^2.\end{aligned} \tag{15}$$

The sink node is floating on the water surface. In (15), $\frac{2}{3}\pi\cdot {r_c}^3$ denotes the volume of communication region of the sink node, as shown in Fig. 12. In the hemispheric region, the nodes can communicate with the sink node directly. The first part of (15) denotes the probability of $\boldsymbol{m}_p$ and $\boldsymbol{n}_e$ being delivered to v_s from the same relay node, and the second part of (15) denotes the probability of $\boldsymbol{m}_p$ and $\boldsymbol{n}_e$ being delivered to v_s from two different relay nodes.

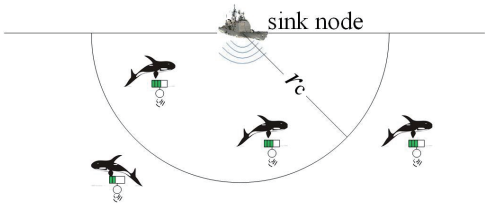


Fig. 12: Communication region of sink node.

C. Expected Theft Ratio

The expected theft ratio of data messages is expressed as:

$$\mathbb{E}(T(\tau)) = N_e\cdot p_c\cdot\left\{\begin{array}{c}\overline{p}\cdot\left[1 - \left(1 - \frac{4\pi\cdot {r_c}^3}{3|\boldsymbol{D}|}\right)^{F(\widetilde{a})\cdot\frac{(1+\kappa)^\tau-1}{\kappa}}\right] + \\ \left\{\overline{p}\cdot\left[1 - \left(1 - \frac{4\pi\cdot {r_c}^3}{3|\boldsymbol{D}|}\right)^{[1-F(\widetilde{a})]\cdot\frac{(1+\kappa)^\tau-1}{\kappa}}\right]\right\}^2\end{array}\right\}, \tag{16}$$

where p_c denotes the probability of an encrypted noise message being successfully cracked by an eavesdropper.

Note that the increase of $\mathbb{E}(D(\tau))$ leads to the increase of $\mathbb{E}(T(\tau))$, and the optimal setting of κ should be obtained from the following formula:

$$\min\ \mathbb{E}(T(\tau^*)),\ s.t.\ \mathbb{E}(D(\tau^*)) = \widetilde{\mathcal{R}}. \tag{17}$$

The value of κ obtained from (17) will be applied in our simulations.

VI. Simulations

In this section, NMDM is evaluated by observing the performance variations with respect to different parameters and by comparing with other algorithms (EF, BDCR, and CAR). We develop a simulator using Python language to evaluate the performance of NMDM. At each time slot, each node generates a new data message by the probability ρ, and thus $N\cdot\rho$ new data messages are generated by nodes.

The underwater vessel noise database in [40] is taken as the environmental noises in our simulations. The main parameter settings are given in TABLE III. Note that the nodes are sparsely distributed, and the encounters between nodes are scarce, according to the size of deployment space, the number of nodes, and the communication range of each node, as shown in TABLE III.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2022.3167911, IEEE Transactions on Information Forensics and Security

TABLE III: Simulation Parameters

Parameter	Description	Value
N	Number of sensor nodes	1,600
N_e	Number of eavesdroppers	3
$\|\boldsymbol{D}\|$	Size of underwater deployment space	$400\times150\times100$ m^3
τ^*	Dissemination deadline (number of time slots)	8
r_c	Communication range of each node	16 m
r_0	Minimum distance between nodes	2 m
ρ	Probability of generating a data message at each time slot	0.05
c_1	Coefficient in signal irregularity formula	0.679
ζ	Exponent in signal irregularity formula	0.77
η	Exponent in signal irregularity formula	2
γ	Exponent in in-degree distribution of nodes	2.3
$\widetilde{a}$	Similarity threshold	5×10^{-3}
Ω_{min}	Minimum signal irregularity	0.1
Ω_{max}	Maximum signal irregularity	0.9
κ	Copy number of a message disseminated by a relay node at each time slot	2
p_c	Probability of an encrypted noise message being cracked by an eavesdropper	0.2
$\widetilde{\mathcal{R}}$	Required delivery ratio	0.65
$\mathcal{V}_m$	Maximum autonomous speed of nodes	1.5 m/s
τ_s	Duration of a time slot	6.02 s

A. Setting of κ

The setting of κ is obtained from (17), and the numerical results are illustrated in Fig. 13.

Fig. 13 indicates that a larger κ is obtained when a smaller τ^* or a larger $\widetilde{\mathcal{R}}$ is given. This is because more message copies are disseminated to guarantee a larger required delivery ratio during a shorter dissemination deadline. Specially, the largest value of κ reaches 3.2, when $\tau^*=6$ and $\widetilde{\mathcal{R}}=0.8$.

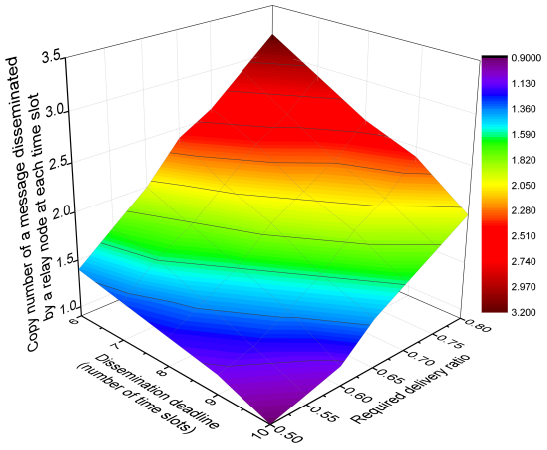
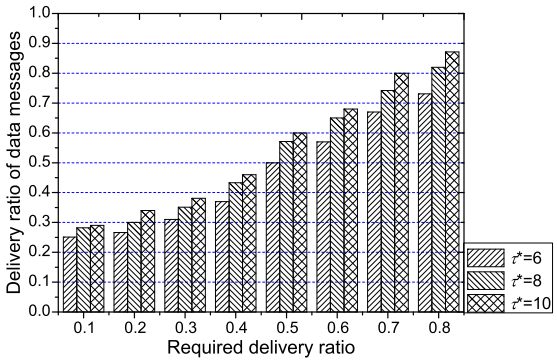


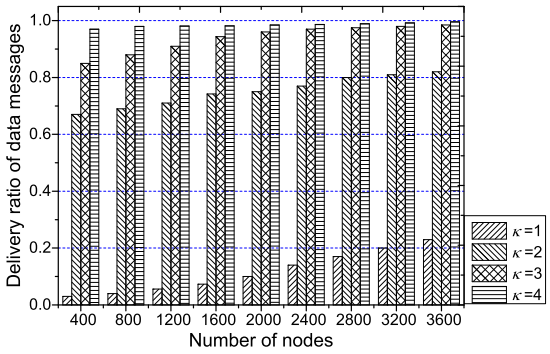
Fig. 13: Setting of κ vs. $\widetilde{\mathcal{R}}$ and τ^*.

B. Delivery Ratio

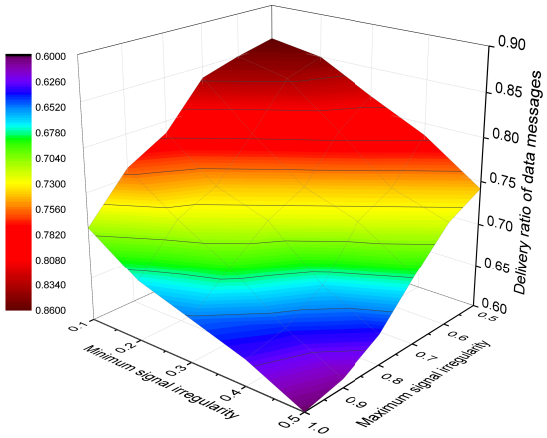
The delivery ratio results are provided in Fig. 14, and some observations are obtained as follows:



(a) Delivery ratio vs. $\widetilde{\mathcal{R}}$ and τ^*



(b) Delivery ratio vs. N and κ



(c) Delivery ratio vs. Ω_{max} and Ω_{min}

Fig. 14: Delivery ratio of data messages.

In Fig. 14(a), the delivery ratio of data messages increases with the continuous increase of $\widetilde{\mathcal{R}}$. The reason is that more message copies will be disseminated when a larger $\widetilde{\mathcal{R}}$ should be guaranteed, and then more data messages can be delivered to the sink node during the dissemination deadline. Besides, Fig. 14(a) also illustrates that the bar

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2022.3167911, IEEE Transactions on Information Forensics and Security

with a larger τ^* is higher than that with a smaller τ^*, since the data messages are more difficult to be delivered to the sink node during a shorter dissemination deadline, although a larger κ is set by (17) under a smaller τ^*.

The delivery ratio bars rise up with the increase of κ or N, as depicted in Fig. 14(b), which is attributed to the fact that a larger κ makes the pseudo data messages and noise messages propagated much more quickly, and the delivery ratio of data messages can be improved when more nodes are deployed in the OUSN.

The delivery ratio of data messages is reduced with the increase of Ω_{min} or Ω_{max}, and this is because a larger signal irregularity reduces the existence probabilities of potential communication links and restricts the message dissemination more seriously.

C. Theft Ratio

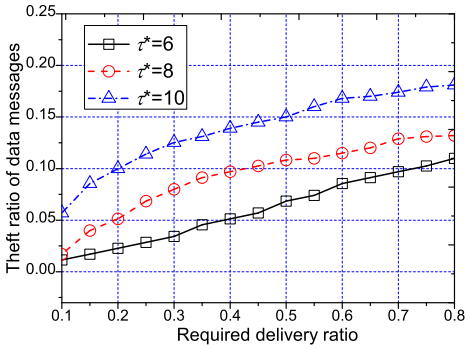
The theft ratio of data messages is calculated as the proportion of the data messages stolen by eavesdroppers to all data messages generated by nodes. Similar to the phenomena in Fig. 14, in Fig. 15 the theft ratio of data messages is increased with the increase of $\widetilde{\mathcal{R}}$ or N (Fig. 15(a)), and the theft ratio of data messages is reduced with the increase of Ω_{min} or Ω_{max} (Fig. 15(c)). Besides, more data messages are stolen by the eavesdroppers when more eavesdroppers invade the OUSN, as shown in Fig. 15(b).

Note that the theft ratio of data messages is much smaller than the delivery ratio of data message, due to the following mechanisms in NMDM: (*i*) The acoustic waves of data messages are superimposed by environmental noises. (*ii*) The noise messages are encrypted before the dissemination. (*iii*) The pseudo data messages and noise messages are individually disseminated by different relay nodes.

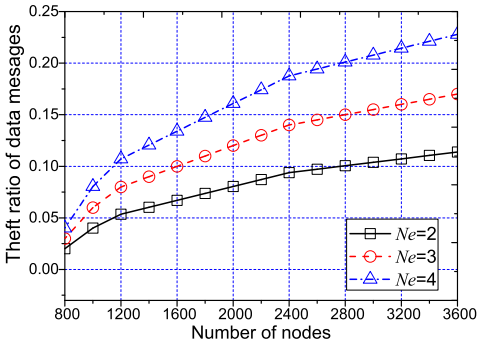
D. Algorithm Comparisons

To further analyze the merits of NMDM, we compare NMDM with other algorithms (EF, BDCR, and CAR) under the underwater mobility model introduced in [38]. EF adopts the epidemic dissemination manner, and the data messages held by each node will be disseminated to all the encountered nodes, which indicates that EF achieves the largest delivery ratio and consumes the largest communication cost. BDCR disseminates the data messages by exploiting the beam width and three-dimensional direction, and it can obtain a preferable delivery ratio. In CAR, each node disseminating a data message tries to avoid a contact with adversaries, and the secure opportunistic paths are determined by integrating the delivery probability and the safety measurement. The computational complexity of these algorithms is first compared, as shown in TABLE IV.

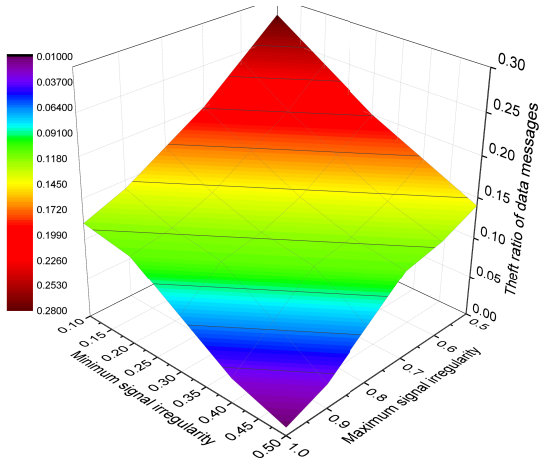
Then, these algorithms are compared in terms of delivery ratio and theft ratio, and the simulation results are presented in Fig. 16. Recall that CAR is a secure routing protocol against the contact-based attacks, as introduced in Section II.B. Although the eavesdroppers are extremely difficult to be perceived by the nodes in an OUSN, in this simulation



(a) Theft ratio vs. $\widetilde{\mathcal{R}}$ and τ^*



(b) Theft ratio vs. N and N_e



(c) Theft ratio vs. Ω_{max} and Ω_{min}

Fig. 15: Theft ratio of data messages.

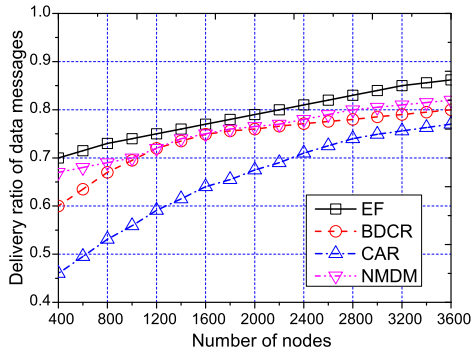
we assume that two-thirds of eavesdroppers can be identified and labeled in CAR. However, some data messages are still stolen due to the unlabeled eavesdroppers (one-thirds of eavesdroppers).

We observe that NMDM achieve the minimum theft ratio among these algorithms, while the delivery ratio of NMDM can be guaranteed larger than the required delivery ratio $\widetilde{\mathcal{R}}$. These phenomena imply that NMDM makes a preferable tradeoff between the theft ratio and the delivery ratio through protecting the data messages with environmental

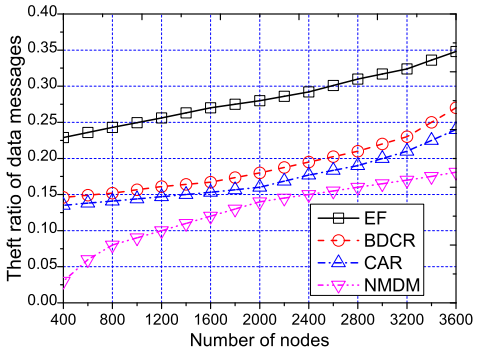
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2022.3167911, IEEE Transactions on Information Forensics and Security

TABLE IV: Computational Complexity

Algorithm	Computational complexity
EF	$\mathrm{O}(N)$
BDCR	$\mathrm{O}(N)$
CAR	$\mathrm{O}(N^2)$
NMDM	$\mathrm{O}(N \cdot \log_2 N)$



(a) Delivery ratio



(b) Theft ratio

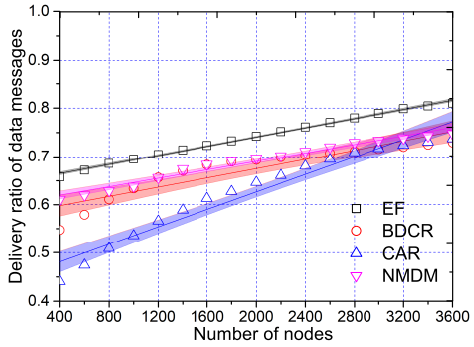
Fig. 16: Algorithm comparisons under underwater mobility model.

noises. EF achieves the largest delivery ratio by adopting an epidemic dissemination manner, whereas EF is not suitable for most of the practical applications due to the extremely high communication complexity.

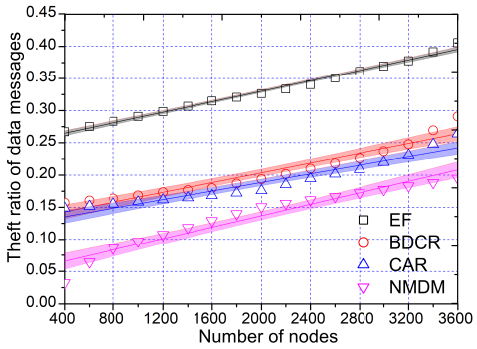
Moreover, the algorithm comparisons under random waypoint model [41] and nomadic community mobility model [42] are implemented, as shown in Fig. 17 and Fig. 18.

The delivery ratio results in Fig. 17(a) are slightly smaller than those in Fig. 16(a), while the theft ratio results in Fig. 17(b) are slightly larger than those in Fig. 16(b). These phenomena are attributed to the fact that the movements of nodes under random waypoint model are more irregular than those under underwater mobility model, which makes the similarities harder to be accurately measured.

In the nomadic community mobility model, the nodes belong to several nomadic communities (groups). In each nomadic group, there are some ordinary floating nodes and



(a) Delivery ratio



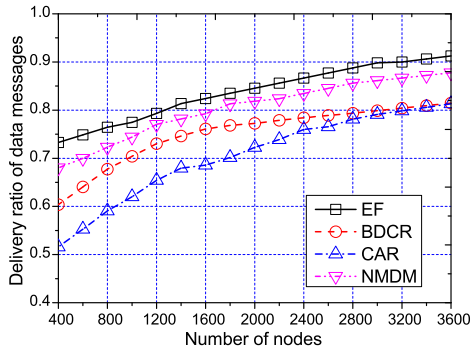
(b) Theft ratio

Fig. 17: Algorithm comparisons under random waypoint model (fitted curves with 95% confidence intervals).

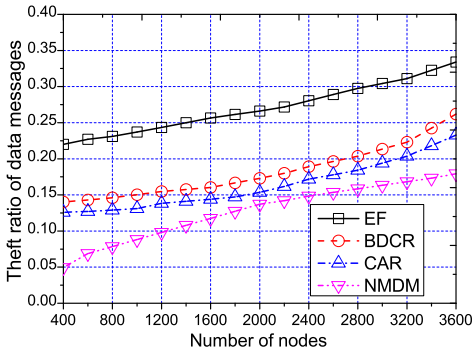
a reference floating node. Each reference floating node moves from one location to another, while the ordinary floating nodes in the same nomadic group always roam around the reference floating node. In the simulations, each group includes 31 ordinary floating nodes and 1 reference floating node, and the movement range of the reference floating node during a time slot is limited to 10 meters. Compared with the simulation results in Fig. 16, the delivery ratio is slightly increased (Fig. 18(a)), and the theft ratio is slightly decreased (Fig. 18(b)), due to the fact that the pseudo data messages and noise messages are easy to be disseminated between nomadic groups, especially when the number of nomadic groups is large.

VII. Conclusions

This study explores a noise-based-protection message dissemination method for the insecure OUSNs invaded by some eavesdroppers. To reduce the theft ratio of data messages and guarantee the required delivery ratio of data messages, the data messages are superimposed by the environmental noises around source nodes and converted into some pseudo data messages. Besides, the environmental noises are identified, encoded, and encrypted into some noise messages. Then, the pseudo data messages and noise messages are individually disseminated to the sink node.



(a) Delivery ratio



(b) Theft ratio

Fig. 18: Algorithm comparisons under nomadic community mobility model.

Thus, the data messages can be protected from being stolen by eavesdroppers. Simulation results demonstrate that NMDM can reduce the theft ratio of data messages and guarantee the required delivery ratio of data messages effectively, and NMDM is especially suitable for the insecure OUSNs surrounded by various environmental noises.

The movements of eavesdroppers may be much intelligent, and the pseudo data messages and noise messages could be purposefully stolen by eavesdroppers. For example, with regard to a data message, an eavesdropper could move around several adjacent relay nodes which disseminate the pseudo data message and noise message respectively. Thus, the data message could be stolen when the eavesdropper captures both the pseudo data message and noise message. To avoid such situation, the pseudo data message and noise message should be disseminated by the relay nodes which are far away from each other. Our future research will focus on investigating this issue.

ACKNOWLEDGMENTS

This research is supported by National Natural Science Foundation of China under Grant Nos. 61872191, 61872193, 61972210; Six Talents Peak Project of Jiangsu Province under Grant No. 2019-XYDXX-247.

REFERENCES

[1] R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, *et al*, "EnOR: Energy Balancing Routing Protocol for Underwater Sensor Networks," *IEEE International Conference on Communications (ICC)*, pp. 574–582, Paris, France, 2017.

[2] G. Han, J. Du, C. Lin, *et al*, "An Energy-Balanced Trust Cloud Migration Scheme for Underwater Acoustic Sensor Networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1636–1649, 2020.

[3] R. Davis, M. Baumgartner, A. Comeau, *et al*, "Tracking Whales on the Scotian Shelf using Passive Acoustic Monitoring on Ocean Gliders," *IEEE OCEANS'16*, Monterey, USA, 2016.

[4] B. G. Ferguson, and K. W. Lo, "Acoustic Detection, Localization, and Tracking of Tactical Autonomous Aerial and Underwater Vehicles," *Journal of the Acoustical Society of America*, vol. 140, no. 4, 2016.

[5] D. Sarriá, O. Pallarés, J. del-Río-Fernández, *et al*, "Low Cost OFDM based Transmitter for Underwater Acoustic Communications," *2013 MTS/IEEE OCEANS*, Bergen, Norway, 2013.

[6] P. Goulet, C. Guinet, R. Swift, *et al*, "A Miniature Biomimetic Sonar and Movement Tag to Study the Biotic Environment and Predator-prey Interactions in Aquatic Animals," *Deep-Sea Research*, vol. 148, pp. 1–11, 2019.

[7] G. N. A. H. Yar, A. Ahmad, and K. Khurshid, "Low Cost Assembly Design of Unmanned Underwater Vehicle (UUV)," *International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*, Islamabad, Pakistan, 2021.

[8] T. Li, J. Ma, P. Feng, *et al*, "Lightweight Security Authentication Mechanism Towards UAV Networks," *2019 International Conference on Networking and Network Applications (NaNA)*, Daegu, Korea, 2019.

[9] M. Wazid, A. Kumar Das, V. Bhat K, *et al*, "LAM-CIoT: Lightweight Authentication Mechanism in Cloud-based IoT Environment," *Journal of Network and Computer Applications*, vol. 150, 2020.

[10] P. H. Dahl, J. H. Miller, D. H. Cato, *et al*, "Underwater Ambient Noise," *Acoustics Today*, vol. 3, no. 1, pp. 23–33, 2007.

[11] R. Chen, A. Poulsen, and H. Schmidt, "Spectral, Spatial, and Temporal Characteristics of Underwater Ambient Noise in the Beaufort Sea in 1994 and 2016," *The Journal of the Acoustical Society of America*, vol. 145, no. 2, pp. 605–614, 2019.

[12] K. Betke, M. S. Glahn, and R. Matuschek, "Underwater Noise Emissions from Offshore Wind Turbines," *Proceedings of the joint congress CFA/DAGA'04*, pp. 591–592, Strasbourg, France, 2004.

[13] R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, *et al*, "Design Guidelines for Opportunistic Routing in Underwater Networks," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 40–48, 2016.

[14] H. Wu, D. Li, L. Yu, *et al*, "A Boundary Element Method based Near Field Acoustic Holography in Noisy Environments," *The Journal of the Acoustical Society of America*, vol. 147, no. 5, 2020.

[15] H. Jiao, and V. Kursun, "Ground Bouncing Noise Suppression Techniques for Data Preserving Sequential MTCMOS Circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 19, no. 5, pp. 763–773, 2011.

[16] E. Ahmed, and A. M. Eltawil, "On Phase Noise Suppression in Full-Duplex Systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1237–1251, 2015.

[17] M. Naravani, D. G. Narayan, S. Shinde, *et al*, "A Cross-Layer Routing Metric with Link Prediction in Wireless Mesh Networks," *Procedia Computer Science*, vol. 171, pp. 2215–2224, 2020.

[18] S. Zhang, and D. Li, "A Beam width and Direction Concerned Routing for Underwater Acoustic Sensor Networks," *IEEE 9th International Conference on Mobile Ad-hoc and Sensor Networks*, pp. 17–24, Dalian, China, 2014.

[19] D. Zhao, H. Ma, S. Tang, *et al*, "COUPON: A Cooperative Framework for Building Sensing Maps in Mobile Opportunistic Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 2, pp. 392–402, 2015.

[20] J. Herrera-Tapia, E. Hernández-Orallo, A. Tomás, *et al*, "Evaluating the Use of Sub-gigahertz Wireless Technologies to Improve Message Delivery in Opportunistic Networks," *IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*, Calabria, Italy, 2017.

[21] L. Chancay-García, E. Hernández-Orallo, P. Manzoni, *et al*, "Evaluating and Enhancing Information Dissemination in Urban Areas of Interest Using Opportunistic Networks," *IEEE Access*, vol. 6, pp. 32514–32531, 2018.

[22] R. W. L. Coutinho, A. Boukerche, L. F. M. Vieira, *et al*, "Performance Modeling and Analysis of Void-handling Methodologies in Underwater Wireless Sensor Networks," *Computer Networks (Elsevier)*, vol. 126, pp. 1–14, 2017.

[23] G. Han, L. Liu, N. Bao, *et al*, "AREP: An Asymmetric Link-based Reverse Routing Protocol for Underwater Acoustic Sensor Networks," *Journal of Network and Computer Applications (Elsevier)*, vol. 92, pp. 51–58, 2017.

[24] F. Ahmed, Z. Wadud, N. Javaid, *et al*, "Mobile Sinks Assisted Geographic and Opportunistic Routing Based Interference Avoidance for Underwater Wireless Sensor Network," *Sensors*, vol. 18, no. 4, 2018.

[25] Q. Kang, X. Liu, Y. Yao, *et al*, "Efficient Authentication and Access Control of Message Dissemination over Vehicular Ad Hoc Network," *Neurocomputing*, vol. 181, pp. 132–138, 2016.

[26] X. Liu, Y. Xia, W. Chen, *et al*, "SEMD: Secure and Efficient Message Dissemination with Policy Enforcement in VANET," *Journal of Computer and System Sciences*, vol. 82, no. 8, pp. 1316–1328, 2016.

[27] D. Wu, F. Zhang, H. Wang, *et al*, "Security-oriented Opportunistic Data Forwarding in Mobile Social Networks," *Future Generation Computer Systems (Elsevier)*, vol. 87, pp. 803–815, 2018.

[28] T. Osuki, K. Sakai, and S. Fukumoto, "Contact Avoidance Routing in Delay Tolerant Networks," *IEEE Conference on Computer Communications (INFOCOM)*, Atlanta, USA, 2017.

[29] L. Xiao, G. Sheng, S. Liu, *et al*, "Deep Reinforcement Learning-Enabled Secure Visible Light Communication Against Eavesdropping," *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 6994–7005, 2019.

[30] Y. Qin, Y. Cao, W. Zhang, *et al*, "Research of Wireless Sensor Network Data Fusion Technology Based on RMSMTV," *4th International Conference on Information Science and Control Engineering (ICISCE)*, pp. 1622–1626, Changsha, China, 2017.

[31] R. Sarathy, and K. Muralidhar, "Evaluating Laplace Noise Addition to Satisfy Differential Privacy for Numeric Data," *ACM Transactions on Data Privacy*, vol. 4, no. 1, pp. 1–17, 2011.

[32] M Rodriguez-Garcia, M. Batet, and D. Sánchez, "A Semantic Framework for Noise Addition with Nominal Data," *Knowledge-Based Systems*, vol. 122, pp. 103–118, 2017.

[33] Z. Li, H. Lv, and Z. Liu, "Noise-added Selection Method for Location-based Service using Differential Privacy in Internet of Things," *Advances in Mechanical Engineering*, vol. 11, no. 1, pp. 1–13, 2019.

[34] J. Du, G. Han, C. Lin, *et al*, "ITrust: An Anomaly-resilient Trust Model Based on Isolation Forest for Underwater Acoustic Sensor Networks," *IEEE Transactions on Mobile Computing*, DOI: 10.1109/TMC.2020.3028369, 2020.

[35] Y. He, G. Han, J. Jiang, *et al*, "A Trust Update Mechanism Based on Reinforcement Learning in Underwater Acoustic Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 21, no. 3, pp. 811–821, 2022.

[36] G. Zhou, T. He, S. Krishnamurthy, *et al*, "Models and Solutions for Radio Irregularity in Wireless Senosr Networks," *ACM Transactions on Sensor Networks*, vol. 2, no. 2, pp. 221–262, 2006.

[37] M. Hussain, and N. Trigoni, "Distributed Localization in Cluttered Underwater Environments," *the Fifth ACM International Workshop on UnderWater Networks (WUWNet'10)*, Article no. 8, Massachusetts, USA, 2010.

[38] L. Liu, P. Wang, and R. Wang, "Propagation Control of Data Forwarding in Opportunistic Underwater Sensor Networks," *Computer Networks (Elsevier)*, vol. 114, pp. 80–94, 2017.

[39] J. Daemen, and V. Rijmen, "The Design of Rijndael: AES–The Advanced Encryption Standard," *Information Security and Cryptography*, Springer, vol. 26, no. 3, 2001.

[40] D. Santos-Domínguez, S. Torres-Guijarro, A. Cardenal-López, *et al*, "ShipsEar: An Underwater Vessel Noise Database," *Applied Acoustics (Elsevier)*, vol. 113, pp. 64–69, 2016.

[41] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa, "Stochastic Properties of the Random Waypoint Mobility Model," *ACM/Kluwer Wireless Networks: Special Issue on Modeling and Analysis of MobileNetworks*, vol. 10, no. 5, pp. 555–567, 2004.

[42] M. Quinzan, J. Castro, M. Marin, *et al*, "Unveiling the Influence of the Environment on the Migration Pattern of the Atlantic Pomfret (Brama brama) in North-eastern Atlantic Waters," *Fisheries Oceanography*, vol. 25, no. 6, pp. 610–623, 2016.

AUTHOR BIOGRAPHY

Linfeng Liu received the B. S. and Ph. D. degrees in computer science from the Southeast University, Nanjing, China, in 2003 and 2008, respectively. At present, he is a professor in the School of Computer Science and Technology, Nanjing University of Posts and Telecommunications, China. His main research interests include the areas of vehicular ad hoc networks, wireless sensor networks and multi-hop mobile wireless networks. He has published more than 80 peer-reviewed papers in some technical journals or conference proceedings, such as IEEE TMC, IEEE TPDS, IEEE TSC, IEEE TITS, IEEE TVT, ACM TAAS, ACM TOIT.

Zhiyuan Xi received the B. S. degree in communication engineering from the Nanjing University of Posts and Telecommunications in 2018. At present, he is a master student of Nanjing University of Posts and Telecommunications. His current research interest includes the areas of mobile opportunistic networks and vehicular ad-hoc networks.

Jiagao Wu received the Ph. D. degree in computer science from the Southeast University, Nanjing, China, in 2006. At present, he is an associate professor of the School of Computer Science and Technology, Nanjing University of Posts and Telecommunications. His current research interest includes the areas of mobile social networks and opportunistic networks.

Jia Xu received the Ph. D. Degree in School of Computer Science and Engineering from Nanjing University of Science and Technology, Jiangsu, China, in 2010. He is currently a professor in Jiangsu Key Laboratory of Big Data Security and Intelligent Processing at Nanjing University of Posts and Telecommunications. His main research interests include crowdsourcing, edge computing and wireless sensor networks. Prof. Xu has served as the PC Co-Chair of SciSec 2019, Organizing Chair of ISKE 2017, TPC member of Globecom, ICC, MASS, ICNC, EDGE. He currently serves as the Publicity Co-Chair of SciSec 2021.