Entropy Optimization of Degree Distributions against Security Threats in UASNs

Linfeng Liu [†,‡], Zhipeng Zhang [†,‡], Jiagao Wu [†,‡], and Jia Xu [†,‡]
[†] School of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
[‡] Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

***Abstract*—Underwater Acoustic Sensor Networks (UASNs) are deployed for various underwater applications, such as underwater creature tracking and tactical surveillance. Particularly, an UASN deployed in military applications could be invaded by some underwater spy-robots which act as eavesdroppers or hackers. An UASN is confronted with two typical security threats: the eavesdroppers move around the anchored nodes and eavesdrop on the communication channels silently, and the data messages disseminated from the anchored nodes are probably stolen by these eavesdroppers; the hackers disguise themselves and propagate the viruses to infect the anchored nodes in a cascading manner. To reduce the theft probability of data messages and the number of cascading failures (the number of infected nodes) while maintaining the required topology connectivity, an analysis framework is first formulated to investigate the relations between the entropies of degree distributions and the resistances of security threats, and then the entropies of degree distributions are optimized to resist the security threats through appropriately coordinating the communication ranges of anchored nodes. We propose a Topology Control Strategy based on Entropy Optimization (TCSEO). In TCSEO, each anchored node independently sets the initial communication range according to a binomial distribution, and then the communications ranges of anchored nodes are checked and adjusted to maintain the required topology connectivity. Simulation results demonstrate the preferable performance of TCSEO, i.e. TCSEO can reduce the theft probability of data messages and the number of infected nodes effectively, while the required topology connectivity is maintained as much as possible.**

***Index Terms*—underwater acoustic sensor network; topology control; entropy optimization; data theft; cascading failure.**

I. Introduction

Currently, Underwater Acoustic Sensor Networks (UASNs) [1] have been applied into various underwater applications, such as underwater creature tracking, tactical surveillance, deep sea surveillance [2], and mineral reconnaissance [3]. The sensor nodes are equipped with floating buoys and are anchored to the underwater bottom by ropes, as illustrated in Fig. 1. The measurements of environmental events are monitored by the anchored nodes and encapsulated into some data messages which will be transferred to one of the surface sinks through several transmission hops.

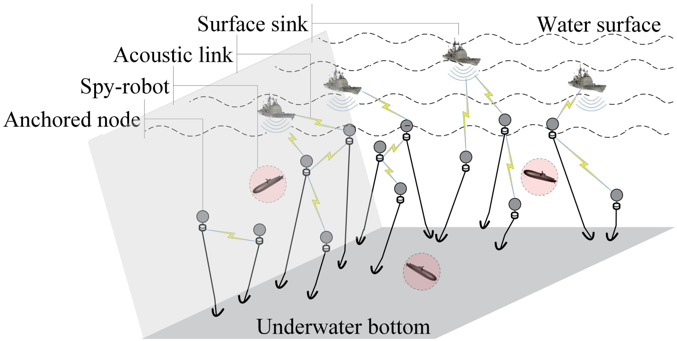
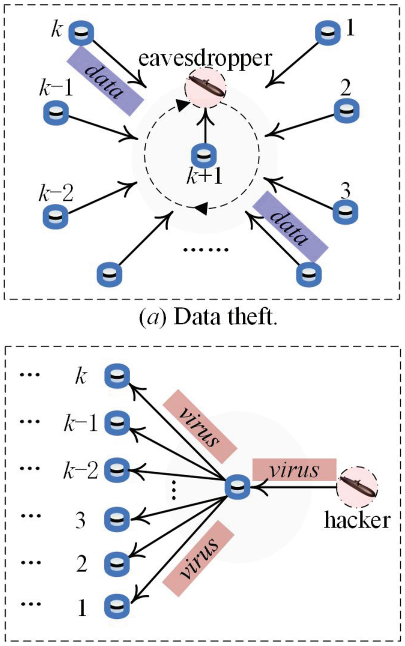


Fig. 1: Architecture of an UASN.

In underwater military applications, the data messages could contain some confidential information, such as the information regarding underwater tactical environments. Some underwater spy-robots [4] could be dispatched by enemy to invade an UASN, so as to steal the data messages or infect the nodes with some viruses (programs). There are two typical security threats in an UASN:

- Data thefts. Some underwater spy-robots termed $eavesdroppers$ move around anchored nodes and eavesdrop on their communication channels. The eavesdroppers are hardly perceived by anchored nodes because the eavesdroppers do not actively communicate with the anchored nodes and other eavesdroppers. As illustrated in Fig. $2(a)$, an eavesdropper is adjacent to an anchored node with in-degree k, and thus the eavesdropper falls into the communication ranges of $(k+1)$ different anchored nodes, i.e. the eavesdropper could steal the data messages disseminated from $(k+1)$ anchored nodes.
- Cascading failures. Some underwater spy-robots termed $hackers$ disguise themselves as ordinary nodes, and propagate some viruses to infect the anchored nodes. The viruses can be propagated by the infected nodes as well, which gives rise to a cascading failure phenomenon, as illustrated in Fig. $2(b)$.

To this end, the data messages should be prevented from being stolen by the eavesdroppers, and the anchored nodes should be protected from being infected by the hackers.



(a) Data theft.

(b) Cascading failure.

Fig. 2: Data theft and cascading failure.

Some existing methodologies can help to resist these security threats, e.g. the data messages could be encrypted before the dissemination. Besides, some identity authentication mechanisms (such as [5], [6]) could be introduced for anchored nodes to identify the underwater spy-robots and avoid the virus infections. However, a sophisticated encryption/decryption method or identity authentication mechanism will give rise to a large computational cost, which is typically intolerable due to the limited computational power of anchored nodes.

Different from the above methodologies, this paper investigates a topology control approach to resist the security threats including data thefts and cascading failures. The topology control technique of UASNs is defined as the art of coordinating the communication ranges of anchored nodes to generate a preferable network topology. **The topology control approach is taken as an alternative against the security threats, and it can be technically combined with some lightweight encryption methods and/or identity authentication mechanisms to resist the security threats more effectively**.

In this paper, we first analyze the probability distribution of communication ranges of anchored nodes. The relations between the entropies of degree distributions and the resistances of security threats are specially exploited, and then the entropies of degree distributions are optimized to resist the data thefts and cascading failures through appropriately coordinating the communication ranges of anchored nodes. Thus, the theft probability of data messages and the number of infected nodes can be reduced, while the required topology connectivity is maintained as much as possible.

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 gives an analysis framework for exploiting the relations between the entropies of degree distributions and the resistances of security threats. Section V presents a Topology Control Strategy based on Entropy Optimization (TCSEO). Simulation results for performance evaluation of TCSEO are reported in Section VI. Finally, Section VII concludes this paper.

II. Related Work

A. Topology Control Methods in WSNs and UASNs

The topology control technique for Wireless Sensor Networks (WSNs) has been extensively studied. For example, Tan *et al.* present a topology control approach for the energy-harvesting WSNs, and this approach allows each node to adaptively adjust its transmission power and effectively utilize the harvested energy [7]. Reference [8] investigates a Topology Control algorithm with Lifetime Extension (TCLE) which can construct dynamic network topologies. The topologies obtained by TCLE can improve the network lifetime significantly. In [9], a lightweight algorithm for Adaptive Transmission Power Control (ATPC) in WSNs is presented. In ATPC, the nodes build a model for each neighbour to measure the correlation between transmission power and link quality. Besides, a lifetime optimization problem is formulated to find the suitable topologies for network-coding-based WSNs [10].

Several topology control methods for UASNs that combine the characteristics of acoustic communications and underwater environments have been proposed, such as our early work [11], where a Topology Control Strategy based on Complex Network (TCSCN) is put forward to construct a double clustering topology, and there are two kinds of cluster-heads to ensure the connectivity and coverage, respectively. Reference [12] formulates a game-theoretic model for the topology control of nodes which are deployed in a sparse underwater sensor network, and it proves that the players can choose proper strategies to achieve a socially optimal Stackelberg-Nash-Cournot equilibrium. In [13], a single-leader and multi-followers Stackelberg game is utilized to formulate the topology control problem by exploiting the available communication opportunities. Specifically, the mobile nodes and anchored nodes act as leaders and followers, respectively. Zhang *et al.* propose a vulnerability repair algorithm [14], where the coverage matrix and the vulnerability edge nodes are applied to determine whether the overlay vulnerability needs to be repaired.

B. Topology Control Approaches against Security Threats

Some secure issues have been considered in the existing topology control methods, such as [15], a distributed fault-tolerant topology control algorithm is designed for heterogeneous WSNs. The ordinary nodes are connected with the resource-rich nodes, and hence a k-vertex supernode-connected network topology is generated to resist the malicious attacks. In [16], an efficient Topology Control algorithm for node Mobility (TCM) is proposed, and a digital signature authentication based on error correction code is adopted in TCM. In [17], an authenticated broadcasting mechanism is introduced into the topology formation stage, and two symmetric keys (a cluster key and a gateway key) are locally distributed in the cluster constructions.

The cascading failure phenomenon of a scale-free topology is investigated in [18], where the influence of nodal loads on the cascading failures is analyzed, and the critical load-triggering cascading failure can be obtained. Furthermore, a model based on the node degree and betweenness centrality is proposed in [19] to fortify the robustness of a scale-free network against the cascading failures. Lal *et al.* [20] provide a hybrid architecture including the physical layer security, software defined networking, node cooperation, cross-layering, context-awareness, and cognition. The network topology can be adjusted to counteract any on-going attacks. Reference [21] investigates the effects of topology parameters and loads on the cascading failures. It derives the structural features of scale-free topologies and the capacity limits, through which the cascading fault tolerance can be effectively enhanced. In [22], a robust optimization formulation is presented to guarantee the robustness of the network topologies against the uncertainty distributions. Reference [23] develops a construction algorithm to generate a k-connected communication topology. Furthermore, a distributed event-triggered controller is designed to guarantee the consensus under Mode-Switching DoS (MSDoS) attacks.

C. Motivation of Our Work

In an UASN invaded by some underwater spy-robots which act as eavesdroppers or hackers, the spy-robots can steal the data messages or infect the anchored nodes. As aforementioned above, the mechanisms of encryption and identity authentication can help to resist the security threats. However, the security threats cannot be completely avoided by these mechanisms, due to the following facts: (i) The anchored nodes are unconscious of the adjacent eavesdroppers, and thus the theft probability of data messages cannot be reduced through intentionally avoiding the message dissemination happening in the insecure areas around eavesdroppers. (ii) The hackers are difficult to be identified by the anchored nodes, because the hackers disguise themselves as ordinary nodes.

Note that the theft probability of data messages and the number of cascading failures (the number of infected nodes) are affected seriously by the network topology, e.g. the data messages are easier to be stolen when the communication ranges of anchored nodes cover more eavesdroppers, and the anchored nodes are easier to be infected when the communication links are with less heterogeneity. In this paper, we will prove that the network topology with homogeneous in-degrees of anchored nodes is beneficial to reduce the theft probability of data messages, and the network topology with heterogeneous out-degrees of anchored nodes helps to reduce the number of infected nodes.

Specifically, the entropies of degree distributions [24] (including in-degree distribution and out-degree distribution) are introduced to measure the heterogeneity of the network topology, and then the relations between the entropies of degree distributions and the resistances of security threats are exploited to coordinate the communication ranges of anchored nodes and generate a preferable network topology.

III. System Model and Problem Formulation

TABLE I shows the list of notations used in the formulation of topology control problem.

TABLE I: Main Notations

Parameter	Description
$r(i)$	Communication range of anchored node v_i
$d(i,j)$	Distance between two anchored nodes v_i and v_j
$P_c(r(i), d(i,j))$	Existence probability of the potential communication link (i,j)
$P_s(G(\boldsymbol{V},\boldsymbol{E}))$	Theft probability of data messages
$Conn(G(\boldsymbol{V},\boldsymbol{E}))$	Topology connectivity
$g(x \cdot r_0)$	Probability of the distance between two neighboring nodes being $x \cdot r_0$
$R(\cdot)$	Probability density function of communication ranges of anchored nodes
$I(k)$	Probability distribution function of in-degrees of anchored nodes
E_i	Entropy of in-degree distribution
$O(k)$	Probability distribution function of out-degrees of anchored nodes
E_o	Entropy of out-degree distribution
$N_f(t)$	Number of infected nodes after t virus propagations
$n_f(t,\hbar)$	Number of the $\hbar$-hop infected nodes after t virus propagations
λ_i	Setting of λ for reducing the theft probability
λ_o	Setting of λ for reducing the number of cascading failures
λ_c	Setting of λ for maintaining the required topology connectivity
$\widetilde{\lambda}$	Optimal setting of λ

A. Anchored Nodes and Surface Sinks

There are N anchored nodes and M surface sinks. The set of anchored nodes and surface sinks is denoted by $\boldsymbol{V} = \{v_1, \cdots, v_N, s_1, \cdots, s_M\}$, where the anchored nodes $v_1, \cdots, v_N$ are uniformly deployed in underwater space $\boldsymbol{D}$ ($\boldsymbol{D} \in \mathbb{R}^{+3}$), and the surface sinks $s_1, \cdots, s_M$ are uniformly berthed at the water surface. The topology of an UASN is represented by a graph $\boldsymbol{G}(\boldsymbol{V}, \boldsymbol{E})$, where the set of links $\boldsymbol{E} \subseteq \boldsymbol{V} \times \boldsymbol{V}$.

Suppose there are χ underwater spy-robots which can navigate freely in the underwater space $\boldsymbol{D}$. Each spy-robot could steal the data messages disseminated from the anchored nodes or propagate the viruses to infect the anchored nodes.

B. Potential Communication Links

The communication range of an anchored node v_i is denoted by $r(i)$, and $r(i)$ can be set to one of the communication range levels: $r_0, 2r_0, 3r_0, \cdots, r_{max}$, where r_0

and r_{max} denote the minimum communication range and the maximum communication range, respectively.

The distance between two anchored nodes v_i and v_j is denoted by $d(i,j)$. If $r_0 \leq d(i,j) \leq r(i)$, and then the communication link (i,j) is considered to be a potential communication link. A potential communication link is defined as a communication link which is possible to exist in the UASN, and the existence of the potential communication link is determined by the underwater acoustic channels, as introduced in Section III.C.

C. Existences of Potential Communication Links

With regard to an anchored node v_i, the set of neighboring nodes is expressed as $\{v_j | v_j \in \boldsymbol{V}\ and\ d(i,j) \leq r_{max}\}$, and the number of neighboring nodes is $K = \frac{4\pi \cdot {r_{max}}^3 \cdot N}{3|\boldsymbol{D}|} - 1$ when the anchored nodes are uniformly deployed.

The probabilistic communication is a typical phenomenon in underwater environments, and it is caused by various factors, such as antenna directions/gains, transmitting power, battery status, signal-to-noise ratio threshold, and underwater obstacles, which make the acoustic waves reflected, diffracted, or scattered underwater. According to [25], [26], [27], Rayleigh fading is appropriate for describing the underwater acoustic channels, and the existence of a potential communication link (i,j) is determined by the following probability $P_c(r(i), d(i,j))$:

$$P_c(r(i), d(i,j)) = \begin{cases} 0, & if \quad d(i,j) > r(i), \\ e^{-\left\{\frac{d(i,j)}{r(i)}\right\}^K \cdot \sigma^{-2}}, & otherwise, \end{cases} \qquad (1)$$

where σ denotes a Rayleigh fading parameter.

Fig. 3 illustrates the probabilistic communication phenomenon in underwater environments, which indicates that the existence probability of the potential communication link (i,j) is increased with the increase of $\frac{d(i,j)}{r(i)}$.

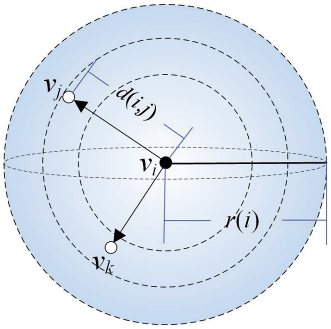


Fig. 3: Probabilistic communication phenomenon.

Equation (1) implies that a potential communication link could be unidirectional. Due to the unidirectional communication links between anchored nodes, we introduce the concepts of in-degree and out-degree. An in-link of an anchored node denotes a communication link from another anchored node to itself, and the in-degree denotes the number of its in-links. Likewise, an out-link of an anchored node denotes a communication link from itself to another anchored node, and the out-degree denotes the number of its out-links.

For example, in Fig. 4, with regard to an anchored node v_i, there are three in-links (v_1, v_i), (v_3, v_i) and (v_4, v_i), and four out-links (v_i, v_1), (v_i, v_2), (v_i, v_4) and (v_i, v_5). Thus, the in-degree and out-degree of v_i are equal to 3 and 4, respectively.

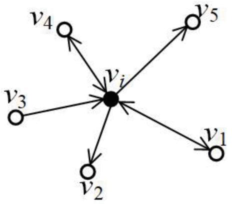


Fig. 4: In-degree and out-degree.

D. Objective Function

To reduce the theft probability of data messages and the number of infected nodes while maintaining the required topology connectivity, the objective function is formulated as follows:

$$\begin{cases} \min\ P_s(G(\boldsymbol{V}, \boldsymbol{E})), \\ \min\ N_f(t), \end{cases} \quad s.t.\ Conn(G(\boldsymbol{V}, \boldsymbol{E})) \geq \varphi, \qquad (2)$$

where $P_s(G(\boldsymbol{V}, \boldsymbol{E}))$ denotes the theft probability of data messages, and the expression of $P_s(G(\boldsymbol{V}, \boldsymbol{E}))$ will be given in the next section. $N_f(t)$ denotes the number of infected nodes after t virus propagations. $Conn(G(\boldsymbol{V}, \boldsymbol{E}))$ denotes the topology connectivity and is expressed as:

$$Conn(G(\boldsymbol{V}, \boldsymbol{E})) = \min_{v_i \in \boldsymbol{V}} Conn(i), \qquad (3)$$

where $Conn(i)$ denotes the maximum path connectivity from v_i to one of the surface sinks.

In (2), the constraint condition $Conn(G(\boldsymbol{V}, \boldsymbol{E})) \geq \varphi$ indicates that there is at least one available communication path (with the maximum path connectivity larger than φ) from each anchored node to one of the surface sinks, i.e. $\forall v_i \in \boldsymbol{V}$, there is $Conn(i) \geq \varphi$.

IV. Analysis Framework

To simplify the further analysis, the maximum communication range of each anchored node is divided into several layers with the radius r_0, as depicted in Fig. 5. In the communication range of r_{max}, the probability of the distance between two neighboring nodes being $x \cdot r_0$ is denoted by $g(x \cdot r_0)$:

$$g(x \cdot r_0) = \frac{(x \cdot r_0)^3 - [(x-1) \cdot r_0]^3}{{r_{max}}^3}, \qquad (4)$$

where $x = 1, 2, \cdots, \frac{r_{max}}{r_0} - 1$.

To guarantee the communication range of each anchored node fall into the range interval $[r_0, r_{max}]$, we enable the

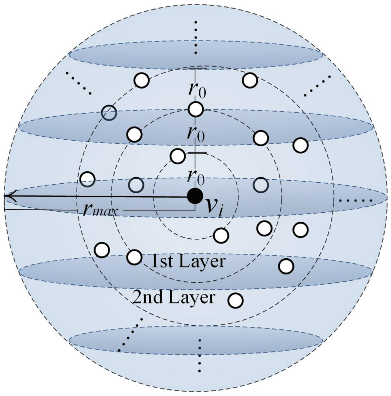


Fig. 5: Layers of maximum communication range of an anchored node.

communication ranges of anchored nodes to obey a binomial distribution $B\left(\frac{r_{max}}{r_0}, \lambda\right)$, and the probability density function $R(x \cdot r_0)$ $(1 \leq x \leq \frac{r_{max}}{r_0})$ is defined as:

$$R(x \cdot r_0) = \binom{\frac{r_{max}}{r_0} - 1}{x - 1} \cdot \lambda^{x-1} \cdot (1-\lambda)^{\frac{r_{max}}{r_0} - x}, \tag{5}$$

where $0 \leq \lambda \leq 1$.

A. Data Theft vs. Entropy of In-degree Distribution

The probability distribution function of the in-degrees of anchored nodes is denoted by $I(k)$ $(0 \leq k \leq K)$. The expression of $I(k)$ is given by:

$$I(k) = \left\{ \sum_{x=1}^{\frac{r_{max}}{r_0}} \left[g(x \cdot r_0) \cdot \sum_{\kappa=x}^{\frac{r_{max}}{r_0}} P_c(\kappa \cdot r_0, x \cdot r_0) \cdot R(\kappa \cdot r_0) \right] \right\}^{k} \cdot$$
$$\left\{ \sum_{x=1}^{\frac{r_{max}}{r_0}} \left[g(x \cdot r_0) \cdot \left(1 - \sum_{\kappa=x}^{\frac{r_{max}}{r_0}} P_c(\kappa \cdot r_0, x \cdot r_0) \cdot R(\kappa \cdot r_0) \right) \right] \right\}^{K-k} \tag{6}$$

where $\sum_{\kappa=x}^{\frac{r_{max}}{r_0}} P_c(\kappa \cdot r_0, x \cdot r_0) \cdot R(\kappa \cdot r_0)$ denotes the probability of a unidirectional communication link (with the length equal to $x \cdot r_0$) existing in the UASN. The form of (6) motivates us to construct $I(k)$ as a geometric distribution:

$$I(k) = p_i \cdot {q_i}^k, \quad p_i, q_i \in (0,1), \tag{7}$$

and there is $\sum_{k=0}^{K} I(k) = 1$.

The probability that an eavesdropper around an anchored node (with in-degree k) cannot break any data messages disseminated from the $(k+1)$ neighboring nodes (as shown in Fig. 2(a)) is written as $(1-\alpha)^{k+1}$, where α denotes the probability of an eavesdropper successfully breaking (deciphering) a data message.

Besides, the eavesdroppers navigate freely in the underwater space, and thus each eavesdropper is assumed to appear at any position with an equivalent probability. Therefore, the theft probability (the probability of at least a data message being stolen by eavesdroppers) is written as:

$$P_s(G(\boldsymbol{V}, \boldsymbol{E})) = \frac{\chi \cdot K}{N} \cdot \sum_{k=0}^{K} \left\{ I(k) \cdot \left[1 - (1-\alpha)^{k+1} \right] \right\}. \tag{8}$$

Lemma 1 proves that both E_i and $P_s(G(\boldsymbol{V}, \boldsymbol{E}))$ are decreased with the reduction of q_i.

Lemma 1: The theft probability of data messages is increased with the increase of the entropy of in-degree distribution.

Proof: The entropy of in-degree distribution is first given by:

$$E_i = -\sum_{k=0}^{K} I(k) \cdot \ln I(k) = -p_i \cdot \sum_{k=0}^{K} {q_i}^k \cdot (\ln p_i + k \cdot \ln q_i)$$
$$= -p_i \cdot \left\{ \begin{array}{l} \ln p_i \cdot \frac{1-q_i^{K+1}}{1-q_i} + \\ \ln q_i \cdot \left[\frac{q_i \cdot \left(1-q_i^K\right)}{(1-q_i)^2} - \frac{K \cdot q_i^{K+1}}{1-q_i} \right] \end{array} \right\}, \tag{9}$$

and then the first-order partial derivative of E_i with respect to q_i is expressed as:

$$\frac{\partial E_i}{\partial q_i} =$$
$$-p_i \cdot \left\{ \begin{array}{l} \ln p_i \cdot \frac{-(K+1) \cdot q_i^K \cdot (1-q_i) + 1 - q_i^{K+1}}{(1-q_i)^2} + \\ \frac{1}{q_i} \cdot \left[\frac{q_i \cdot (1-q_i^K)}{(1-q_i)^2} - \frac{K \cdot q_i^{K+1}}{1-q_i} \right] + \ln q_i \cdot \\ \left[\begin{array}{l} \frac{(1-q_i)^2 \cdot (1-q_i^K - K \cdot q_i^K) + 2(1-q_i) \cdot q_i \cdot (1-q_i^K)}{(1-q_i)^4} \\ - \frac{K \cdot (K+1) \cdot q_i^K \cdot (1-q_i) + K \cdot q_i^{K+1}}{(1-q_i)^2} \end{array} \right] \end{array} \right\}. \tag{10}$$

When K is large enough (i.e. the anchored nodes are densely deployed), the equation $\sum_{k=0}^{K} I(k) = \sum_{k=0}^{K} p_i \cdot {q_i}^k = 1$ yields that: $p_i \approx 1 - q_i$. Besides, the value of ${q_i}^K$ can be approximated to 0. Then, $\frac{\partial E_i}{\partial q_i}$ is approximatively rewritten as:

$$\frac{\partial E_i}{\partial q_i} \approx -p_i \cdot \left[\frac{\ln p_i + 1}{(1-q_i)^2} + \ln q_i \cdot \frac{(1-q_i)^2 + 2(1-q_i) \cdot q_i}{(1-q_i)^4} \right]$$
$$= -\frac{1}{1-q_i} \cdot \left[\ln(1-q_i) + 1 + \ln q_i \cdot \frac{1+q_i}{1-q_i} \right]$$
$$> -\frac{1}{1-q_i} \cdot \left[\ln(1-q_i) + 1 + \ln q_i \right] > 0. \tag{11}$$

Note that the value of q_i reflects the possibility of generating communication links among neighboring nodes, which is related to the communication ranges of anchored nodes and the distances between anchored nodes, as depicted in (1). Thereby, (11) indicates that the entropy of in-degree distribution is increased with the increase of possibility of generating communication links.

Furthermore, the first-order partial derivative of

$P_s(G(\boldsymbol{V},\boldsymbol{E}))$ with respect to q_i is written as:

$$\frac{\partial P_s(G(\boldsymbol{V},\boldsymbol{E}))}{\partial q_i} = \frac{\chi \cdot K}{N} \cdot \sum_{k=0}^{K} \left\{ \begin{array}{l} \left[k \cdot q_i^{k-1} - (k+1) \cdot q_i^{k}\right] \\ \cdot \left[1 - (1-\alpha)^{k+1}\right] \end{array} \right\}$$

$$= \frac{\chi \cdot K}{N} \cdot \left\{ \begin{array}{l} \frac{1-q_i}{q_i} \cdot \left[\frac{q_i \cdot (1-q_i^K)}{(1-q_i)^2} - \frac{K \cdot q_i^{K+1}}{1-q_i} \right] - \frac{1-q_i}{q_i} \cdot (1-\alpha) \\ \cdot \left\{ \frac{q_i \cdot (1-\alpha) \cdot \left\{1-[q_i \cdot (1-\alpha)]^K\right\}}{[1-q_i \cdot (1-\alpha)]^2} - \frac{K \cdot [q_i \cdot (1-\alpha)]^{K+1}}{1-q_i \cdot (1-\alpha)} \right\} \\ +(1-\alpha) \cdot \frac{1-[q_i \cdot (1-\alpha)]^{K+1}}{1-q_i \cdot (1-\alpha)} - \frac{1-q_i^{K+1}}{1-q_i} \end{array} \right\} \tag{12}$$

By (12), we obtain that $\frac{\partial P_s(G(\boldsymbol{V},\boldsymbol{E}))}{\partial q_i} \approx \frac{(1-\alpha)-(1-\alpha)^2}{1-[q_i \cdot (1-\alpha)]^2} > 0$. Equations (11) and (12) imply that fewer data messages are stolen along with a smaller entropy of in-degree distribution. □

Lemma 1 also indicates that the network topology with the homogeneous in-degrees of anchored nodes is beneficial to reduce the theft probability of data messages. Thus, the optimal value of q_i for reducing the theft probability of data messages can be calculated as the minimum solution of the following equation set:

$$\left\{ \begin{array}{l} p_i = \frac{1-q_i}{1-q_i^{K+1}}, \\ q_i = \frac{1-\sqrt[K]{p_i}}{\sqrt[K]{p_i}}. \end{array} \right. \tag{13}$$

B. Cascading Failure vs. Entropy of Out-degree Distribution

The probability distribution function of the out-degrees of anchored nodes is denoted by $O(k)$:

$$O(k) = \hat{Q}^k \cdot \left(1-\hat{Q}\right)^{K-k}$$
$$= \left\{ \sum_{\kappa=\sqrt[3]{\frac{k}{K}} \cdot \frac{r_{max}}{r_0}}^{\frac{r_{max}}{r_0}} \left[\begin{array}{l} R(\kappa \cdot r_0) \cdot \\ \sum_{x=1}^{\kappa} P_c(\kappa \cdot r_0, x \cdot r_0) \cdot g(x \cdot r_0) \end{array} \right] \right\}^k$$
$$\cdot \left\{ 1 - \sum_{\kappa=\sqrt[3]{\frac{k}{K}} \cdot \frac{r_{max}}{r_0}}^{\frac{r_{max}}{r_0}} \left[\begin{array}{l} R(\kappa \cdot r_0) \cdot \\ \sum_{x=1}^{\kappa} P_c(\kappa \cdot r_0, x \cdot r_0) \cdot g(x \cdot r_0) \end{array} \right] \right\}^{K-k} \tag{14}$$

Note that the out-degree of each anchored node must be greater than or equal to 1, because there must be at least one available communication path to a surface sink, and there is $\sum_{k=1}^{K} O(k) = 1$. The entropy of out-degree distribution is defined by:

$$E_o = -\sum_{k=1}^{K} O(k) \cdot \ln O(k), \tag{15}$$

and there is

$$\frac{\partial E_o}{\partial \hat{Q}} = -\sum_{k=1}^{K} \left\{ \begin{array}{l} \hat{Q}^k \cdot \left(1-\hat{Q}\right)^{K-k} \cdot \\ \left[k \cdot \ln \hat{Q} + (K-k) \cdot \ln \left(1-\hat{Q}\right)\right] \end{array} \right\} > 0. \tag{16}$$

To measure the number of cascading failures, the number of infected nodes after t virus propagations is given by:

$$N_f(t) = \sum_{\hbar=1}^{t} n_f(t,\hbar). \tag{17}$$

With the virus propagations in the UASN, more anchored nodes are probably infected by the hackers or other infected ones. When $t > 1$, the expected number of the $\hbar$-hop infected nodes is expressed as:

$$\cdot n_f(t,\hbar) = \left\{ \begin{array}{ll} \gamma \cdot \chi \cdot K \cdot \left[1 - \frac{N_f(t-1)}{N}\right], & if\ \hbar = 1, \\ \gamma \cdot n_f(t-1,\hbar-1) \cdot \left[1 - \frac{N_f(t-1)}{N}\right] & \\ \cdot \sum_{k=1}^{K} k \cdot O(k) + n_f(t-1,\hbar), & if\ 1 < \hbar \leq t, \\ 0, & if\ \hbar > t, \end{array} \right. \tag{18}$$

where γ denotes the probability of an anchored node being infected by neighboring hackers, and $\left[1 - \frac{N_f(t-1)}{N}\right]$ denotes the proportion of uninfected nodes after $(t-1)$ virus propagations.

Lemma 2 proves that $N_f(t)$ is decreased with the increase of $\hat{Q}$.

Lemma 2: The number of infected nodes is decreased with the increase of the entropy of out-degree distribution. **Proof**: When $\gamma << 1$, there is $\frac{N_f(t-1)}{N} << 1$. Thus, $\frac{\partial n_f(t,\hbar)}{\partial \hat{Q}}$ can be approximatively written as:

$$\frac{\partial n_f(t,\hbar)}{\partial \hat{Q}} \approx$$
$$\left\{ \begin{array}{ll} 0, & if\ \hbar = 1\ or\ \hbar > t, \\ \frac{n_f(t-1,\hbar)}{\partial \hat{Q}} + \gamma \cdot n_f(t-1,\hbar-1) & \\ \cdot \sum_{k=1}^{K} \left[k \cdot \hat{Q}^{k-1} \cdot (1-\hat{Q})^{K-k-1} \cdot \left(k - K \cdot \hat{Q}\right)\right] & \\ +\gamma \cdot \frac{\partial n_f(t-1,\hbar-1)}{\partial \hat{Q}} \cdot \sum_{k=1}^{K} k \cdot O(k), & if\ 1 < \hbar \leq t, \end{array} \right. \tag{19}$$

where $\sum_{k=1}^{K} \left[k \cdot \hat{Q}^{k-1} \cdot (1-\hat{Q})^{K-k-1} \cdot \left(k - K \cdot \hat{Q}\right)\right]$ is smaller than 0, and hence we have that $\frac{\partial n_f(t,\hbar)}{\partial \hat{Q}} \leq 0$. Likewise, when N is large enough, there is $\frac{\partial N_f(t)}{\partial \hat{Q}} < 0$, which implies that fewer anchored nodes are infected along with a larger entropy of out-degree distribution. □

.

Lemma 2 indicates that the network topology with the heterogeneous out-degrees of anchored nodes can help to reduce the number of infected nodes.

C. Optimal Setting of λ

According to the values of p_i and q_i obtained from (13), the optimal value of λ for reducing the theft probability of data messages is denoted by λ_i. λ_i can be obtained from the following equation set:

$$\left\{ \begin{array}{l} p_i = \left\{ 1 - \sum_{x=1}^{\frac{r_{max}}{r_0}} \left[\begin{array}{l} g(x \cdot r_0) \cdot \\ \sum_{\kappa=x}^{\frac{r_{max}}{r_0}} \left(\begin{array}{l} P_c(\kappa \cdot r_0, x \cdot r_0) \\ \cdot R(\kappa \cdot r_0) \end{array} \right) \end{array} \right] \right\}^K, \\ q_i = \frac{\sum_{x=1}^{\frac{r_{max}}{r_0}} \left[g(x \cdot r_0) \cdot \sum_{\kappa=x}^{\frac{r_{max}}{r_0}} P_c(\kappa \cdot r_0, x \cdot r_0) \cdot R(\kappa \cdot r_0) \right]}{\sum_{x=1}^{\frac{r_{max}}{r_0}} \left\{ g(x \cdot r_0) \cdot \left[1 - \sum_{\kappa=x}^{\frac{r_{max}}{r_0}} P_c(\kappa \cdot r_0, x \cdot r_0) \cdot R(\kappa \cdot r_0) \right] \right\}}, \\ \sum_{k=0}^{K} p_i \cdot {q_i}^k = 1. \end{array} \right. \tag{20}$$

Because $\hat{Q}$ is increased with the increase of λ, and hence the optimal value of λ for decreasing the number of cascading failures is denoted by λ_o, and λ_o is calculated as the maximum solution of the following equation set:

$$\begin{cases} \sum_{k=1}^{K} O(k) = 1, \\ \sum_{k=1}^{K} k \cdot O(k) = \sum_{k=0}^{K} k \cdot I(k). \end{cases} \tag{21}$$

Furthermore, λ_c denotes the minimum value of λ to maintain the required topology connectivity. As illustrated in Fig. 6, the lower anchored nodes typically have smaller path connectivity than those of upper anchored nodes. Therefore, the maximum path connectivity of a bottom node should be larger than the required topology connectivity φ:

$$\max_{x=1,\cdots,\frac{r_{max}}{r_0}} \left\{ \sum_{\kappa=x}^{\frac{r_{max}}{r_0}} R(\kappa \cdot r_0) \cdot P_c(\kappa \cdot r_0, x \cdot r_0) \right\}^{\frac{H}{x \cdot r_0}} \geq \varphi, \tag{22}$$

where H denotes the water depth.

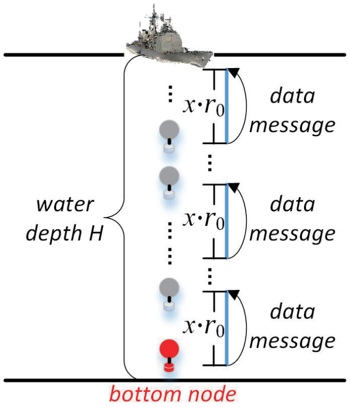


Fig. 6: Path connectivity of a bottom node.

Finally, the optimal setting of λ is denoted by $\widetilde{\lambda}$ and obtained by:

$$\widetilde{\lambda} = \max \left\{ \beta \cdot \lambda_i + (1-\beta) \cdot \lambda_o, \lambda_c \right\}, 0 \leq \beta \leq 1, \tag{23}$$

where β is a weight to make a tradeoff between the reduction of data thefts and the reduction of cascading failures. Some numerical values of $\widetilde{\lambda}$, λ_i, λ_o, and λ_c are provided in Fig. 7, where the parameters are set according to TABLE IV given in Section VI.

V. Topology Control Strategy based on Entropy Optimization

Topology Control Strategy based on Entropy Optimization (TCSEO) is specially designed to resist the security threats of data thefts and cascading failures, while the required topology connectivity is maintained as much as possible. Note that TCSEO is a completely distributed strategy, i.e. the global computations and message exchanges are not required, and thereby the execution cost of TCSEO is very low.

In TCSEO, each anchored node sets the initial communication range according to a binomial distribution. Then, the communications ranges of anchored nodes are checked and adjusted to maintain the required topology connectivity, so that there is at least an available communication path from each anchored node to a surface sink. TABLE II provides the symbols used in the description of TCSEO.

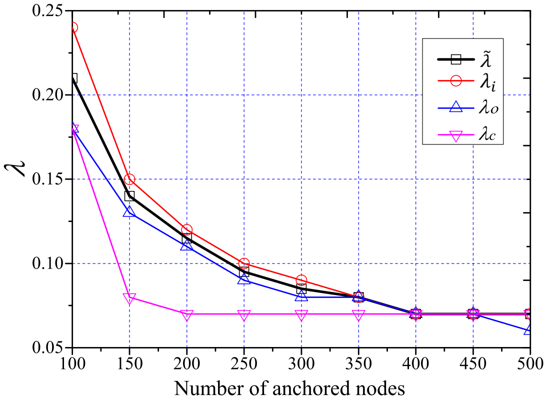


Fig. 7: Numerical values of $\widetilde{\lambda}$, λ_i, λ_o, and λ_c (when β=0.5).

TABLE II: Symbols in TCSEO Description

Symbol	Definition
$\widehat{r}(i)$	Initial communication range of anchored node v_i
$path\_list(i)$	Communication path list of anchored node v_i
$path(i,k)$	Communication path from anchored node v_i to surface sink s_k
$start\_msg$	Start message from a surface sink
$path\_list\_msg$	Path list message of an anchored node
$adjust\_msg$	Adjust message from a lower anchored node to an upper anchored node
$random(0,t_b)$	Random backoff time
t_w	Waiting time
t_r	Interval of topology reconstruction

A. Stage 1: Initial Communication Range Settings of Anchored Nodes

Before setting the initial communication ranges of anchored nodes, a time synchronization process is first accomplished among anchored nodes and surface sinks.

Then, $\widetilde{\lambda}$ is calculated by (23), and each anchored node v_i independently sets the initial communication range according to the binomial distribution $B\left(\frac{r_{max}}{r_0}, \widetilde{\lambda}\right)$, i.e. $\widehat{r}(i) \sim B\left(\frac{r_{max}}{r_0}, \widetilde{\lambda}\right)$, where $\widehat{r}(i)$ denotes the initial communication range of v_i. Besides, the communication range of each surface sink is set to r_{max}. An example of the binomial distribution is given in Fig. 8.

Although the initial communication ranges of anchored nodes are set to reduce the theft probability and the number of infected nodes while maintaining the required topology connectivity. However, due to the fact that each anchored node sets the initial communication range independently, the required topology connectivity of network topology could not be guaranteed. Therefore, the communication

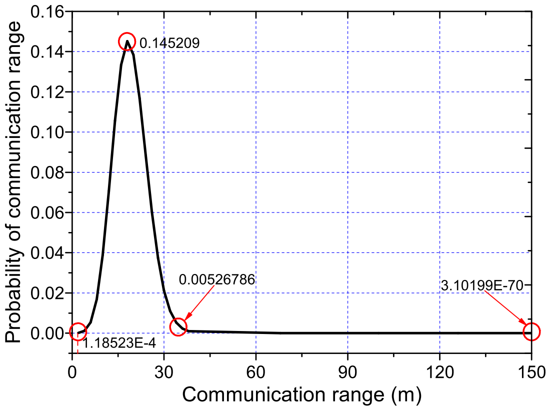


Fig. 8: An example of the binomial distribution of communication ranges (when $N = 200$, $\varphi = 0.8$).

ranges of anchored nodes should be further checked and adjusted, as introduced in Stage 3.

B. Stage 2: Information Exchanges

In Stage 2, the information regarding the initial communication ranges of anchored nodes is exchanged, and this stage is launched by surface sinks. Each surface sink broadcasts a $start\_msg$ in the maximum communication range. Then, two cases are discussed as follows:

Case *i*: Suppose an anchored node v_i receives the first $start\_msg$ from a surface sink s_k, and then the communication path list of v_i is initialized as:

$$path\_list(i) \leftarrow \{(s_k, v_i), r_{max}, \widehat{r}(i)\}. \tag{24}$$

After a waiting period of t_w, $path\_list(i)$ should be updated when the $start_msg$s from other surface sinks or the $path_list_msg$s from other anchored nodes are received by v_i.

Note that each $path\_list\_msg$ includes the available communication paths from the sender to surface sinks. Then, after a random backoff time $random(0, t_b)$ to avoid the communication collisions (t_b denotes the maximum backoff time), v_i broadcasts a $path\_list\_msg$ to the neighboring nodes in the maximum communication range.

Case *ii*: If an anchored node does not receive any $start_msg$s, and when it receives the first $path\_list\_msg$ from a neighboring node (suppose v_j receives the first $path\_list\_msg$ from v_i), the communication path list of v_j is initialized as:

$$path\_list(j) \leftarrow path\_list(i) \bigcup \{(v_i, v_j), \widehat{r}(i), \widehat{r}(j)\}. \tag{25}$$

Likewise, $path\_list(j)$ will be updated after a waiting period of t_w.

The above process will be continued until the $path\_list\_msg$ of each anchored node has be initialized and updated. The process of information exchanges in TCSEO is shown in Fig. 9.

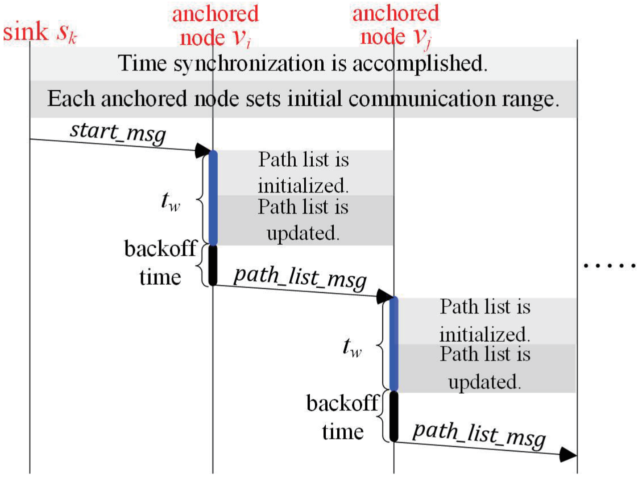


Fig. 9: Information exchanges.

C. Stage 3: Communication Range Adjustments

According the received $path_list_msg$s, each anchored node selects the communication path with the maximum path connectivity from itself to a surface sink. The process of communication range adjustments is launched by $leaf\ node$s which do not receive any $adjust_msg$s during a waiting period of t_w. An $adjust\_msg$ is taken as a request to upper anchored nodes for the adjustments of their communication ranges.

Each leaf node calculates the maximum path connectivity. For example, with regard to a leaf node v_l, suppose a communication path $v_l, v_j, v_i, \cdots, s_k$ is found in $path\_list(l)$ and is denoted by $path(l, k)$. The maximum path connectivity of v_l is calculated by:

$$Conn(l) \leftarrow \max_{path(l,k) \in path\_list(l)} \prod_{(j,i) \in path(l,k)} P_c(\widehat{r}(j), d(j,i)). \tag{26}$$

Then, three cases are discussed according to the value of $Conn(l)$:

Case *i*: If $Conn(l) \geq \varphi$. We set $r(l) \leftarrow \widehat{r}(l)$, where $r(l)$ denotes the adjusted communication range of v_l.

Case *ii*: If $Conn(l) < \varphi$ and $Conn(l) \cdot \frac{P_c(r_{max}, d(l,j))}{P_c(\widehat{r}(l), d(l,j))} \geq \varphi$. Then, $r(l)$ is updated by:

$$r(l) \leftarrow \min \left\{ \kappa \cdot r_0 \middle| \; Conn(l) \cdot \frac{P_c(\kappa \cdot r_0, d(l,j))}{P_c(\widehat{r}(l), d(l,j))} \geq \varphi \right\}. \tag{27}$$

Case *iii*: If $\frac{Conn(l) \cdot P_c(r_{max}, d(l,j))}{P_c(\widehat{r}(l), d(l,j))} < \varphi$. We set $r(l) \leftarrow r_{max}$, and $Conn(l)$ is updated by:

$$Conn(l) \leftarrow Conn(l) \cdot \frac{P_c(r_{max}, d(l,j))}{P_c(\widehat{r}(l), d(l,j))}. \tag{28}$$

After that, an $adjust\_msg$ is sent from v_l to v_j. Once receiving the $adjust\_msg$ from v_l, if $Conn(l) \cdot \frac{P_c(r_{max},d(j,i))}{P_c(\widehat{r}(j),d(j,i))} \geq \varphi$, and then $r(j)$ is updated by:

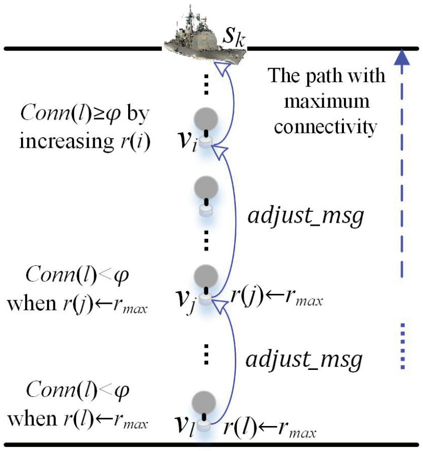
$$r(j) \leftarrow \min\left\{\kappa \cdot r_0 \middle| Conn(l) \cdot \frac{P_c(\kappa \cdot r_0, d(j,i))}{P_c(\widehat{r}(j), d(j,i))} \geq \varphi\right\}. \tag{29}$$

If $\frac{Conn(l) \cdot P_c(r_{max},d(j,i))}{P_c(\widehat{r}(j),d(j,i))} < \varphi$, we set $r(j) \leftarrow r_{max}$, and then an $adjust\_msg$ is sent from v_j to v_i for the further adjustments of communication ranges of upper anchored nodes.

The adjustment process will be repeated, until the maximum path connectivity of each leaf node is not smaller than φ, and thus the required topology connectivity is maintained, as proven in Lemma 3.

Lemma 3: If the maximum path connectivity of each leaf node is not smaller than φ, and then the required topology connectivity is maintained.
Proof: With regard to each anchored node v_o which is not a leaf node, there is at least a leaf node (denoted by v_l) whose communication path passes through v_o according to the definition of leaf nodes, which indicates that there is $Conn(o) > Conn(l) \geq \varphi$. Hence, the maximum path connectivity of each anchored node is not smaller than φ. □

An illustration of the communication range adjustments is provided in Fig. 10, the communication path with the maximum path connectivity is selected by a leaf node v_l, and the communication ranges of the anchored nodes along this path are adjusted from v_l to the surface sink s_k sequentially, until the inequality $Conn(l) \geq \varphi$ has been satisfied.



s_k
The path with maximum connectivity
Conn(l)≥φ by increasing r(i)
v_i
adjust_msg
Conn(l)<φ when r(j)←r_{max}
v_j
r(j)←r_{max}
adjust_msg
Conn(l)<φ when r(l)←r_{max}
v_l
r(l)←r_{max}

Fig. 10: An illustration of communication range adjustments.

Moreover, note that TCSEO will be re-executed at a constant interval of t_r, and thus the communication ranges of anchored nodes are periodically reset and re-adjusted. Such mechanism can balance the energy consumption of anchored nodes and avoid the intentional threats of eavesdroppers or hacks towards the anchored nodes with large communication ranges. The pseudo-code of TCSEO is depicted in Algorithm 1.

Algorithm 1 Pseudo-code of TCSEO

```
Input: : $\boldsymbol{D}, \boldsymbol{V}, r_{max}$.
Output: : $G(\boldsymbol{V}, \boldsymbol{E})$.
  while at a constant interval of $t_r$ do
    Time synchronization is accomplished.
    for each surface sink $s_k \in \{s_1, \cdots, s_M\}$ do
      $r(k) \leftarrow r_{max}$.
    end for
    for each anchored node $v_i \in \{v_1, \cdots, v_N\}$ do
      $\widehat{r}(i) \sim B\left(\frac{r_{max}}{r_0}, \widetilde{\lambda}\right)$.
    end for
    for each surface sink $s_k \in \{s_1, \cdots, s_M\}$ do
      A $start\_msg$ is broadcasted.
    end for
    for each anchored node $v_i \in \{v_1, \cdots, v_N\}$ do
      $path\_list(i)$ is initialized and updated.
      A $path\_list\_msg$ is sent to neighboring nodes.
    end for
    for each leaf node $v_l$ do
      The communication path with the maximum path connectivity is found.
      if $Conn(l) \geq \varphi$ then
        $r(l) \leftarrow \widehat{r}(l)$.
      else
        if $Conn(l) \cdot \frac{P_c(r_{max},d(l,j))}{P_c(\widehat{r}(l),d(l,j))} \geq \varphi$ then
          $r(l)$ is adjusted to the minimum communication range for $Conn(l) \geq \varphi$.
        else
          $r(l) \leftarrow r_{max}$.
          An $adjust\_msg$ is sent from $v_l$ to upper anchored nodes along this communication path.
          while $v_j$ receives the $adjust\_msg$ do
            if $Conn(l) \cdot \frac{P_c(r_{max},d(j,i))}{P_c(\widehat{r}(j),d(j,i))} \geq \varphi$ then
              $r(j)$ is adjusted to the minimum communication range for $Conn(l) \geq \varphi$.
            else
              $r(j) \leftarrow r_{max}$.
              An $adjust\_msg$ is sent from $v_j$ to upper nodes along this communication path.
            end if
            $G(\boldsymbol{V}, \boldsymbol{E})$ is constructed.
          end while
        end if
      end if
    end for
  end while
```

D. Complexity of TCSEO

TABLE III shows the communication complexity and computational complexity of the proposed TCSEO.

In Stage 1, the time synchronization such as TPSN (Timing-sync Protocol for Sensor Networks) [28] is executed among anchored nodes and surface sinks, and then each anchored node independently sets the initial communication range. There are $\mathrm{O}(N+M)^2$ messages for the time synchronization, and $\mathrm{O}(N+M)$ computations for the communication range settings of anchored nodes.

Besides, $\widetilde{\lambda}$ should be calculated by (13), where λ_i, λ_o and λ_c are obtained from (20), (21) and (22), respectively. Newton-Raphson method can be applied to obtain λ_i, λ_o and λ_c. Thus, the computational complexity for obtaining λ_i, λ_o and λ_c is written as $\mathrm{O}\left(\log\left[\left(\frac{r_{max}}{r_0}-1\right)\cdot K\right]\right)$, $\mathrm{O}\left(\log\left[\left(\frac{r_{max}}{r_0}-1\right)\cdot K\right]\right)$ and $\mathrm{O}\left(\log\left[\left(\frac{r_{max}}{r_0}-1\right)\cdot \frac{H}{r_0}\right]\right)$, respectively. Typically, there is $\frac{H}{r_0}, K << M, N$, indicating that the computational complexity of calculating $\widetilde{\lambda}$ can be taken as a constant. Therefore, the computational complexity of Stage 1 is expressed as $\mathrm{O}(N+M)$.

In Stage 2, each surface sink broadcasts a $start\_msg$, and each anchored node broadcasts a $path\_list\_msg$, which indicates that a total of $\mathrm{O}(N+M)$ messages are broadcasted. Because each anchored node is assumed to have K neighboring nodes, in the worst case, an anchored node could update its communication path list once it receives a new $path\_list\_msg$ from a neighboring node. Due to the fact $K << N$, the computational complexity of Stage 2 is written as $\mathrm{O}(N)$.

In Stage 3, each anchored node receives at most K $adjust_msg$s, which gives rise to a total of $\mathrm{O}(N)$ message exchanges. With regard to each anchored node, the communication range is adjusted from r_0 to r_{max} in the worst case. There is $\frac{r_{max}}{r_0} < \frac{H}{r_0} << N$, and thus the computational complexity of Stage 3 is expressed as $\mathrm{O}(N)$.

Therefore, the communication complexity and computational complexity of TCSEO are written as $\mathrm{O}(N+M)^2$ and $\mathrm{O}(N+M)$, respectively.

TABLE III: Complexity of TCSEO

Stage	Communication Complexity	Computational Complexity
1	$\mathrm{O}(N+M)^2$	$\mathrm{O}(N+M)$
2	$\mathrm{O}(N+M)$	$\mathrm{O}(N)$
3	$\mathrm{O}(N)$	$\mathrm{O}(N)$
Total	$\mathrm{O}(N+M)^2$	$\mathrm{O}(N+M)$

VI. Simulations

In this section, TCSEO is evaluated by observing the performance variations with respect to different parameters and by comparing with other algorithms (TCLE, TCSCN, and TCM). We develop a simulator using Python language to assess the performance of TCSEO, and the main parameter settings are shown in TABLE IV.

A topology constructed by TCSEO is given in Fig. 11, where the surface sinks are marked in black, and each anchored node with the sum of in-degree and out-degree larger than 6 is marked in red. We observe that the proportion of the anchored nodes with large in-degrees and out-degrees is very small in Fig. 11.

TABLE IV: Simulation Parameters

Parameter	Description	Value
N	Number of anchored nodes	200
M	Number of surface sinks	15
χ	Number of underwater spy-robots	5
$\|\boldsymbol{D}\|$	Size of deployment space	750×200 ×25 m^3
r_{max}	Maximum communication range	150 m
r_0	Minimum communication range	2 m
σ	Rayleigh fading parameter	1.0
φ	Required topology connectivity	0.8
β	Tradeoff weight	0.5
t_b	Maximum backoff time	0.5 s
t_w	Waiting time	1.5 s
P_0	Minimum signal power	0.07 w
ε	Energy spreading factor	1.5
f	Acoustic frequency	10 kHz
S_{uw}	Propagation speed of acoustic sound	1,500 m/s
L_m	Size of each data message	500 B
B	Channel capacity	8 kbps
α	Probability of an eavesdropper breaking the captured data messages	0.45
γ	Probability of an anchored node being infected by viruses	0.1
T	Number of virus propagations	5
t_r	Interval of topology reconstruction	1,000 s

A. Execution Time and Topology Connectivity

For each anchored node, the neighboring nodes must be located in its maximum communication range, which implies that the number of neighboring nodes K is not large, and the number of anchored nodes on a communication path is not very large as well. Moreover, each anchored node independently coordinates (sets and adjusts) the communication range. To measure the execution time of TCSEO, we observe the results of the average time for an anchored node coordinating its communication range.

Fig. 12 illustrates that the average time for an anchored node coordinating its communication range is slowly increased with the increase of N, because more updates of communication path lists and more adjustments of communication ranges could be required when the anchored nodes are deployed more densely (i.e. there are more anchored nodes on communication paths). Besides, a larger M prolongs the average time, and the reason is that more surface sinks give rise to more available communication paths. The above phenomena tally with the analysis conclusion of Section V.D. Especially, note that the results in Fig. 12 fall into a small numerical interval [3.465,3.579], which indicates that TCSEO has a preferable practicability in terms of execution time.

Although TCSEO attempts to guarantee that each anchored node has at least one available communication path to a surface sink (i.e. the maximum path connectivity is larger than φ), the available communication paths may be invalid due to underwater probabilistic communications. To

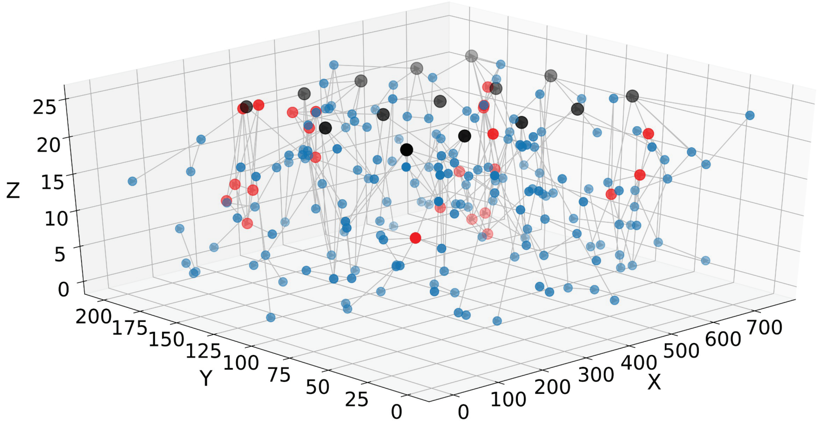


Fig. 11: A topology constructed by TCSEO (when $N = 200, M = 15$).

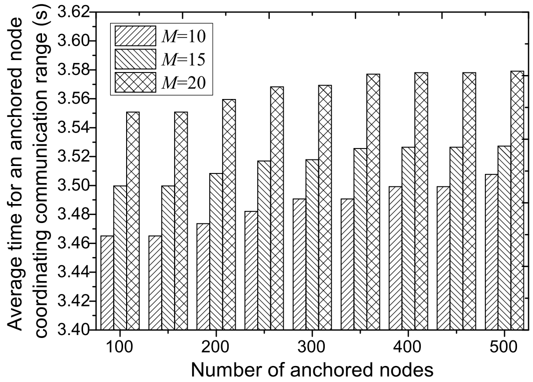
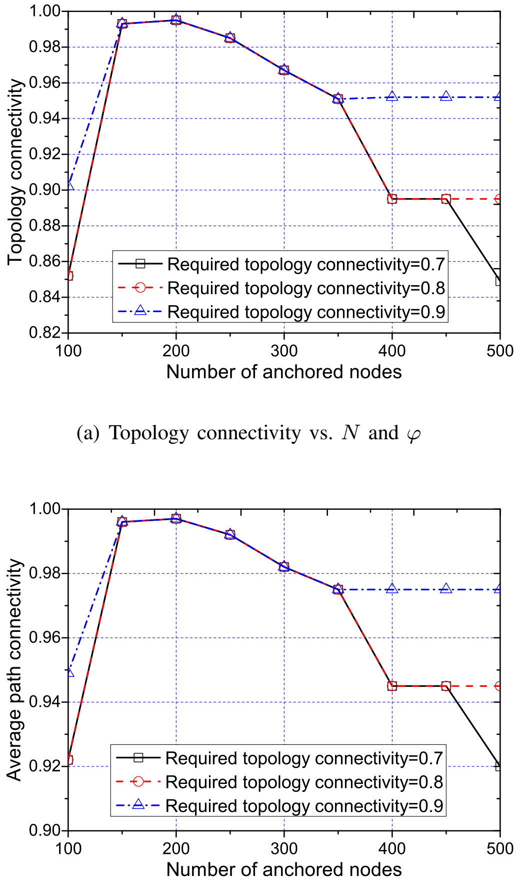


Fig. 12: Average time for an anchored node coordinating communication range.

measure the quality of topology connectivity, the indexes *topology connectivity* and *average path connectivity* are observed. Specifically, the average path connectivity is calculated as $\frac{\sum_{v_i \in V} Conn(i)}{N}$. The simulation results are given in Fig. 13.

Fig. 13 shows that the curve with a larger φ is generally higher than that with a smaller one, which is attributed to the fact that a larger φ gives rise to a larger $\widetilde{\lambda}$, indicating that the anchored nodes are endowed with larger communication ranges, and thus the path connectivity is improved.

Besides, the curves in Fig. 13(a) and Fig. 13(b) fluctuate with the increase of N, and this is because when the anchored nodes are deployed more densely (N becomes larger), more available communication paths from each anchored node to surface sinks can be found to improve the path connectivity. However, when the required topology connectivity has been maintained, TCSEO will shorten the communication ranges of anchored nodes to reduce the theft probability of data messages and the number of infected nodes.



(a) Topology connectivity vs. N and φ

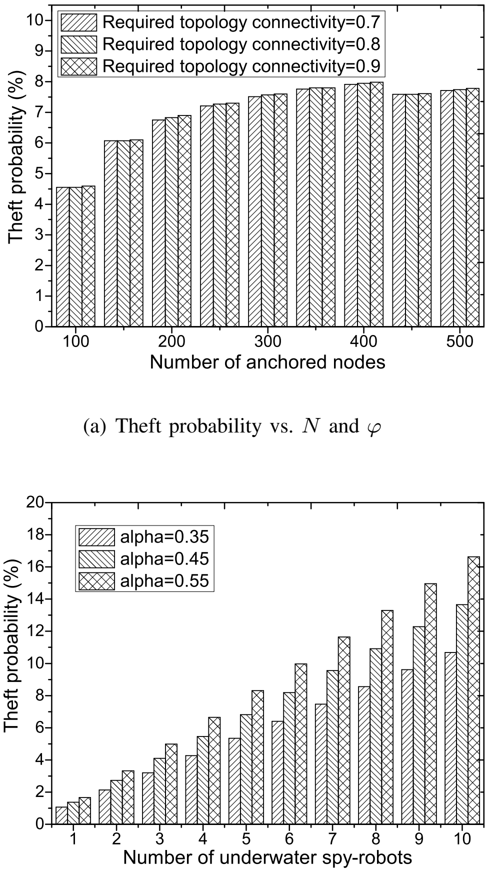
(b) Average path connectivity vs. N and φ

Fig. 13: Topology connectivity and average path connectivity.

B. *Theft Probability*

As depicted in Fig. 14(a), when N is fixed, the bar with a larger φ is slightly higher than that with a smaller φ. The reason is that a larger φ compels the anchored nodes to increase their communication ranges to maintain the larger required path connectivity, and the data messages disseminated from these anchored nodes are easier to be stolen by eavesdroppers. Moreover, the bars first ascend with the increase of N until $N = 400$, and then the bars descend with the further increase of N, which indicates that the communication ranges of anchored nodes will be shortened when the number of anchored nodes is large enough.

In Fig. 14(b), we observe that the theft probability of data messages increases with the increase of χ or α, due to the following reasons: (i) A larger χ implies that more underwater spy-robots could invade the UASN and navigate around the anchored nodes, and hence more data messages could be stolen. (ii) A larger α indicates that the underwater spy-robots are easier to break the captured data messages. For example, the underwater spy-robots have stronger computation power, or the data messages are encrypted by a simpler rule.



(a) Theft probability vs. N and φ

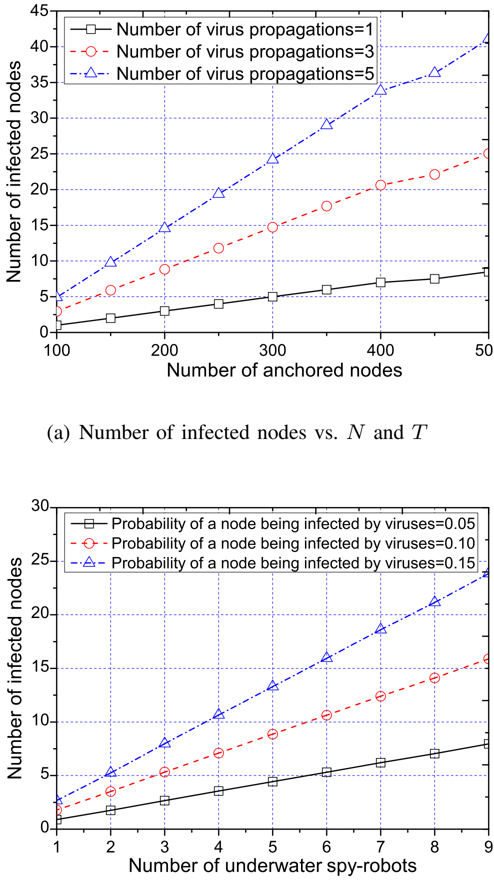
(b) Theft probability vs. χ and α

Fig. 14: Theft probability.

C. Number of Infected Nodes

In Fig. 15, two observations are obtained as follows: (i) The number of infected nodes is almost linearly increased with the increase of N or χ, which is attributed to the fact that the anchored nodes or the underwater spy-robots are assumed to be uniformly distributed in the UASN. (ii) When the viruses are allowed to propagated during a longer period (a larger T), or the anchored nodes are infected by the viruses more easily (a larger γ), more anchored nodes are infected.

Note that a tiny increase in γ leads to a remarkable increase in the number of infected nodes, which suggests that it is very vital for the anchored nodes to identify the hackers or the viruses, i.e. the number of infected nodes can be reduced significantly through combining TCSEO with an effective lightweight identity authentication mechanism.



(a) Number of infected nodes vs. N and T

(b) Number of infected nodes vs. χ and γ

Fig. 15: Number of infected nodes.

D. Algorithm Comparisons

To further evaluate the merits of TCSEO, we compare TCSEO with other algorithms (TCLE, TCSCN, and TCM). These algorithms are compared in terms of average path connectivity, theft probability, number of infected nodes, energy consumption (calculated as [29]), and propagation delay (calculated as [30]). The simulation results are presented in Fig. 16 and Fig. 17.

Fig. 16 indicates that TCSEO outperforms other algorithms in terms of theft probability (in Fig. 16(b)) and number of infected nodes (in Fig. 16(c)) by an obvious margin. The reason for these phenomena is that TCSEO attempts to shorten the communication ranges of anchored nodes, so as to reduce the theft probability of data messages and the number of infected nodes through optimizing the entropies of degree distributions, when the required topology connectivity has been maintained. This mechanism can prevent the data messages from being stolen by eavesdroppers and protect the anchored nodes from being infected by hackers as much as possible. Thus, the theft probability of data messages and the number of infected nodes can be reduced simultaneously, and the energy consumption is decreased due to the shortened communication ranges of anchored nodes, as shown in Fig. 17(a).

Fig. 16(a) and Fig. 17(b) illustrate the average path connectivity and propagation delay, respectively. In order to reduce the theft probability of data messages and the number of infected nodes, TCSEO does not achieve the best results in terms of average path connectivity and propagation delay. For example, in Fig. 16(a), the curve of TCSCN is higher than that of TCSEO when $N \leq 300$, since TCSCN allows some anchored nodes to adopt much larger communication ranges, especially when the anchored nodes are deployed sparsely. Hence, more available communication paths can be found in the topology constructed by TCSCN, and the average path connectivity is accordingly improved. Likewise, in Fig. 17(b), the propagation delay of TCSCN is shorter than that of other algorithms.

Besides, the theft probability curve of TCM is much lower than those of TCLE and TCSCN, and this is because a digital signature authentication is applied in TCM.

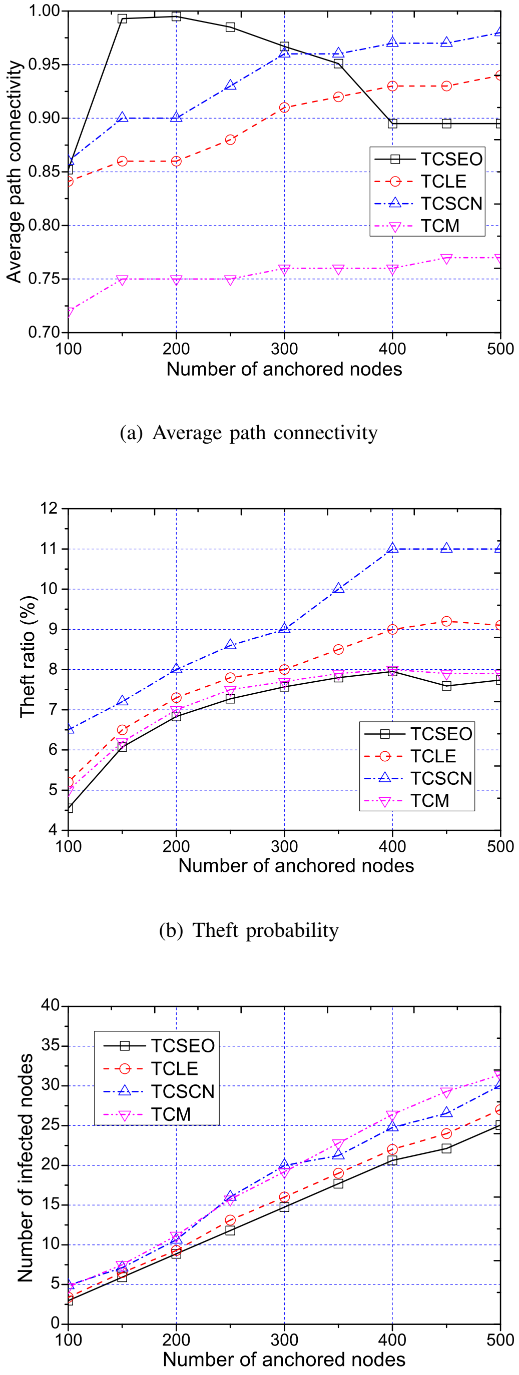
The above simulation results suggest that the topology constructed by TCSEO can reduce the theft ratio of data messages and the number of infected nodes effectively, while maintaining the required topology connectivity as much as possible. Especially, the proper tradeoffs between energy consumption and propagation delay can be made by TCSEO.

VII. Conclusions

This study explores the topology control problem for the UASNs invaded by some underwater spy-robots. To reduce the theft probability of data messages and the number of infected nodes while maintaining the required topology connectivity, the relations between the entropies of degree distributions and the resistances of security threats are carefully investigated. Then, the entropies of degree distributions are optimized to resist the security threats through appropriately coordinating the communication ranges of anchored nodes.

In the proposed Topology Control Strategy based on Entropy Optimization (TCSEO), each anchored node independently sets the initial communication range according to a binomial distribution. Then, the communications ranges of anchored nodes are checked and adjusted to maintain the required topology connectivity as much as possible.

However, the movements of underwater spy-robots could be more intelligent than the random movement assumption in this work, and the data messages will be stolen more easily. Moreover, as mentioned above, an effective authentication mechanism can help the anchored nodes to identify the underwater spy-robots and avoid the virus infections. Our future research will focus on investigating these issues.

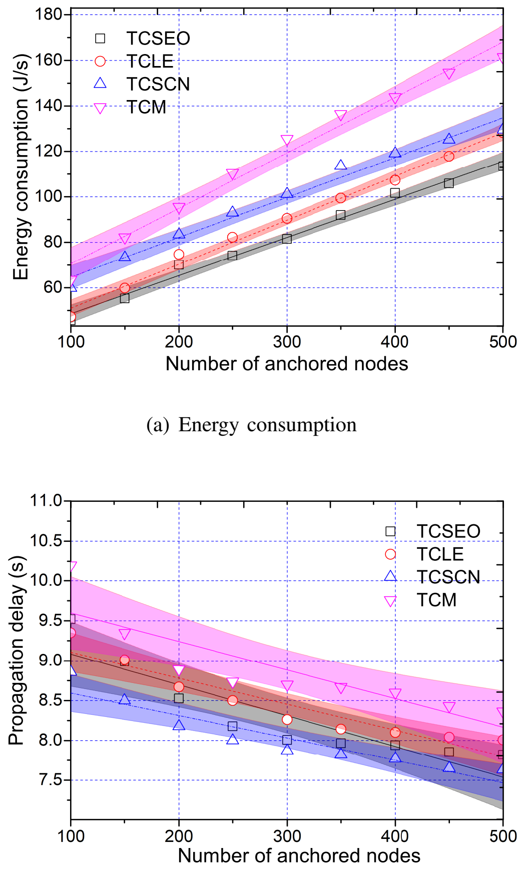


(a) Average path connectivity

(b) Theft probability

(c) Number of infected nodes

Fig. 16: Algorithm comparisons on average path connectivity, theft probability, and number of infected nodes.



(a) Energy consumption

(b) Propagation delay

Fig. 17: Algorithm comparisons on energy consumption, and propagation delay (fitted curves with 95% confidence intervals).

Appendix

A. Energy Consumption

The energy consumption is mainly produced by the sonars of anchored nodes due to the transmissions of acoustic waves, and the energy consumption of each message dissemination (in Joule per byte) is calculated by:

$$E(r(i)) = \frac{P_0 \cdot r(i)^{\varepsilon+1} \cdot 10^{\frac{r(i)\cdot\alpha(f)}{10}}}{S_{uw}}, \tag{30}$$

where P_0 denotes the minimum received power level to guarantee the quality of reception [29]. The energy spreading factor and absorption coefficient are denoted by $\varepsilon(\varepsilon \in [1,2])$ and $\alpha(f)$, respectively. S_{uw} denotes the propagation speed of acoustic sound.

The absorption coefficient for the frequency range of interest is calculated according to Thorp's expression [29]:

$$\alpha(f) = \frac{0.11f^2}{1+f^2} + \frac{44f^2}{4100+f^2} + 2.75\cdot 10^{-4}f^2 + 0.003, \tag{31}$$

where $\alpha(f)$ is in dB/km, and f is in kHz.

The total energy consumption is calculated as the sum of energy consumption of all anchored nodes:

$$E_{total} = \sum_{i=1}^{N} E(r(i)). \tag{32}$$

B. Propagation Delay

The propagation delay on a communication link (i,j) is expressed as [30]:

$$TD(i,j) = \frac{L_m}{B} + \frac{d(i,j)}{S_{uw}}, \tag{33}$$

where L_m denotes the size of a data message, and B denotes the channel capacity which is in bits per second. The propagation delay consists of channel preparation delay and transmission delay. $\frac{L_m}{B}$ denotes the channel preparation delay which is the period of a data message being prepared on channels, and $\frac{d(i,j)}{S_{uw}}$ denotes the transmission delay of acoustic wave propagation.

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] L. Liu, Z. Xi, and J. Wu, "Strengthening the Achilles' Heel: An AUV-aided Message Ferry Approach against Dissemination Vulnerability in UASNs," *IEEE Internet of Things Journal*, DOI: 10.1109/JIOT.2021.3072920, 2021.

[2] E. Felemban, F. K. Shaikh, U. M. Qureshi, A. A. Sheikh, and S. B. Qaisar, "Underwater Sensor Network Applications: A Comprehensive Survey," *International Journal of Distributed Sensor Networks*, vol. 2015, Article ID 896832, 2015.

[3] G. Tuna, and V. C. Gungor, "A Survey on Deployment Techniques, Localization Algorithms, and Research Challenges for Underwater Acoustic Sensor Networks," *International Journal of Communication Systems*, vol. 30, no. 3, 2017.

[4] Q. Wang, and H. N. Dai, "On Modeling of Eavesdropping Behavior in Underwater Acoustic Sensor Networks," *IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Macau, China, 2017.

[5] D. He, N. Kumar, and N. Chilamkurti, "A Secure Temporal-credential-based Mutual Authentication and Key Agreement Scheme with Pseudo Identity for Wireless Sensor Networks," *Information Sciences*, vol. 321, pp. 263–277, 2015.

[6] G. Fragkos, C. Minwalla, J. Plusquellic, and E. E. Tsiropoulou, "Artificially Intelligent Electronic Money," *IEEE Consumer Electronics Magazine*, vol. 10, no. 4, pp. 81–89, 2021.

[7] Q. Tan, W. An, Y. Han, Y. Liu, S. Ci, F. M. Shao, and H. Tang, "Energy Harvesting Aware Topology Control with Power Adaptation in Wireless Sensor Networks," *Ad Hoc Networks*, vol. 27, pp. 44–56, 2015.

[8] M. Xu, Q. Yang, and K. S. Kwak, "Distributed Topology Control With Lifetime Extension Based on Non-Cooperative Game for Wireless Sensor Networks," *IEEE Sensors Journal*, vol. 16, no. 9, pp. 3332–3342, 2016.

[9] S. Lin, F. Miao, J. Zhang, G. Zhou, L. Gu, T. He, J. A. Stankovic, S. Son, and G. J. Pappas, "ATPC: Adaptive Transmission Power Control for Wireless Sensor Networks," *ACM Transactions on Sensor Networks*, vol. 12, no. 1, Article 6, 2016.

[10] M. Khalily-Dermany, M. J. Nadjafi-Arani, and S. Doostali, "Combining Topology Control and Network Coding to Optimize Lifetime in Wireless Sensor Networks," *Computer Networks*, vol. 162, 2019.

[11] L. Liu, Y. Liu, and N. Zhang, "A Complex Network Approach to Topology Control Problem in Underwater Acoustic Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 12, pp. 3046–3055, 2014.

[12] S. Misra, T. Ojha, and A. Mondal, "Game-Theoretic Topology Control for Opportunistic Localization in Sparse Underwater Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 14, no. 5, pp. 990–1003, 2015.

[13] Y. Yuan, C. Liang, M. Kaneko, X. Chen, and D. Hogrefe, "Topology Control for Energy-Efficient Localization in Mobile Underwater Sensor Networks Using Stackelberg Game," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1487–1500, 2019.

[14] W. Zhang, G. Han, Y. Liu, and J. Wang "A Coverage Vulnerability Repair Algorithm Based on Clustering in Underwater Wireless Sensor Networks," *Mobile Networks and Applications*, DOI:10.1007/s11036-020-01621-4, 2020.

[15] F. Deniz, H. Bagci, I. Korpeoglu, and A. Yazici, "An Adaptive, Energy-aware and Distributed Fault-tolerant Topology-control Algorithm for Heterogeneous Wireless Sensor Networks," *Ad Hoc Networks*, vol. 44, pp. 104–117, 2016.

[16] Y. Zhang, W. Chen, J. Liang, B. Zheng, and S. Jiang, "A Network Topology Control and Identity Authentication Protocol with Support for Movable Sensor Nodes," *Sensors*, vol. 15, no. 12, pp. 29958–29969, 2015.

[17] C. T. Hsueh, C. Y. Wen, and Y. C. Ouyang, "A Secure Scheme Against Power Exhausting Attacks in Hierarchical Wireless Sensor Networks," *IEEE Sensors Journal*, vol. 15, no. 6, pp. 3590–3602, 2015.

[18] H. Liu, M. Dong, R. Yin, and L. Han, "Cascading Failure in the Wireless Sensor Scale-free Networks," *Chinese Physics B*, vol. 24, no. 5, pp. 050506: 1–6, 2015.

[19] H. Liu, Y. Hu, R. Yin, and Y. Deng, "Cascading Failure Model of Scale-free Topology for Avoiding Node Failure," *Neurocomputing*, vol. 260, pp. 443–448, 2017.

[20] C. Lal, R. Petroccia, K. Pelekanakis, M. Conti, and J. Alves, "Toward the Development of Secure Underwater Acoustic Networks," *IEEE Journal of Oceanic Engineering*, vol. 42, no. 4, pp. 1075–1087, 2017.

[21] Y. Xiao, "Dynamic Fault Tolerant Topology Control for Wireless Sensor Network Based on Node Cascading Failure," *International Journal of Online Engineering*, vol. 14, no. 5, pp. 118–128, 2018.

[22] M. Nazemi, M. Nazemi, and M. Lejeune, "A Mixed-Integer Distributionally Robust Chance-Constrained Model for Optimal Topology Control in Power Grids with Uncertain Renewables," *2019 IEEE Milan PowerTech*, Milan, Italy, 2019.

[23] T. Zhang, and D. Ye, "Distributed Secure Control Against Denial-of-Service Attacks in Cyber-Physical Systems Based on K-Connected Communication Topology," *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 3094–3103, 2020.

[24] B. Wang, H. Tang, C. Guo, and Z. Xiu, "Entropy Optimization of Scale-free Networks' Robustness to Random Failures," *Physica A: Statistical Mechanics and its Applications*, vol. 363, no. 2, pp. 591–596, 2006.

[25] G. Zhou, and T. Shim, "Simulation Analysis of High Speed Underwater Acoustic Communication Based on a Statistical Channel Model," *Proc. Congress on Image and Signal Processing 2008*, pp. 512–517, Sanya, China, 2008.

[26] M. Stojanovic and J. Preisig, "Underwater Acoustic Communication Channels: Propagation Models and Statistical Characterization," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 84–89, 2009.

[27] H. Nouri, M. Uysal and E. Panayirci, "Information Theoretical Performance Analysis and Optimisation of Cooperative Underwater Acoustic Communication Systems," *IET Communications*, vol. 10, no. 7, pp. 812–823, 2016.

[28] G. Seth, and A. Harisha, "Energy Efficient Timing-sync Protocol for Sensor Network," *International Conference on Computing and Network Communications (CoCoNet)*, Trivandrum, India, 2015.

[29] E. Sozer, M. Stojanovic, and J. Proakis, "Underwater Acoustic Networks," *IEEE Journal of Oceanic Engineering*, vol. 25, no. 1, pp. 72–83, 2000.

[30] L. Liu, R. Wang, G. Xiao, and D. Guo, "On the Throughput Optimization for Message Dissemination in Opportunistic Underwater Sensor Networks," *Computer Networks (Elsevier)*, vol. 169, pp. 1–16, 2020.