
Seyed Mohammad Ghoreyshi, Alireza Shahrabi, and Tuleen Boutaleb

Abstract—From the view of routing protocols in underwater sensor networks (UWSNs), the presence of communication void, where the packet cannot be forwarded further using the greedy mode, is perhaps the most challenging issue. In this paper, we review the state of the art of void-handling techniques proposed by underwater geographic greedy routing protocols. To this, we first review the void problem and its negative impact on the category of the geographic greedy routing protocols, which does not entail any void recovery technique. It is followed by a discussion about the constraints, challenges, and features associated with the design of void-handling techniques in UWSNs. Afterward, currently available void-handling techniques in UWSNs are classified and investigated. They can be classified into two main categories: 1) location-based and 2) depth-based techniques. The advantages and disadvantages of each technique along with the recent advances are then presented. Finally, we present a qualitative comparison of these techniques and also discuss some possible future directions.

Index Terms—Underwater sensor networks, void-handling techniques, routing protocols.

I. INTRODUCTION

UNDERWATER acoustic sensor networks have obtained a considerable attention to support aquatic applications such as exploration of ocean resource, disaster prevention, intrusion detection, military applications, and pollution monitoring [1]–[4]. The sensors are distributed in different depths to collect information and forward them to a destination, which may be a sink, a group of sinks or an Autonomous Underwater Vehicle (AUV) [5]–[8].

Different routing protocols are proposed to improve the packet delivery with minimum energy and delay cost in UWSNs, in which greedy routing protocols are the most prominent approaches due to the simplicity of use in UWSNs [9]–[11]. With the aid of localization mechanisms, geographic greedy routing has become a promising scheme in the sensor and ad hoc networks. Geographic greedy routing (also called position-based, or location-based) is a routing principle, which relies on the geographic position information to forward the data packets closer to the destination in each hop [12]–[14].

In contrast to a table-driven (proactive) routing, which requires large communication overhead to establish end-to-end routes [15]–[17], geographic routing does not need to discover and maintain the full path from the source to the destination. In most cases, the geographic information of one or two hops has been held to route the packets. Thus, there is no need for the routing tables and routing messages to hold and update the route path. This unique feature makes them scalable to be used in the large networks with many nodes [9], [18]. Geocasting is another service which can be supported by the geographic routing to deliver a packet to an intended geographic region [19].

Geographic routing follows a greedy-forwarding strategy in which every node looks for the closest neighbouring node to the destination. However, greedy forwarding may fail because of communication voids or local maximum phenomenon [20]–[22]. In this case, the forwarding node cannot find any qualified node with a positive progress towards the destination, so the packet may be dropped even though there is a valid path from the sender to the destination.

As depicted in Fig. 1, node $e$ is a void node since it has no neighbouring node closer to the sinks $S_1$ and $S_2$ than itself. Thus, in a greedy-forwarding strategy, the packet is dropped if node $e$ is selected as the next forwarding node instead of node $d$, which has a valid path to a sink. Without resolving this issue, data packets may drop in the network, wasting the network resources such as energy and bandwidth. Moreover, the void problem is more challenging as it is unpredictable as to when and where a void may occur due to dynamic nature of operating environment.

A number of factors individually or a combination of them, cause the void phenomenon, such as the sparse topology, temporary obstacles, and unreliable nodes or links [20], [23]. Increasing the density of the network is a simple solution; however, it is not possible all the time and, even so, it cannot entirely eliminate the void problem [24]. Therefore, in order to improve the routing efficiency, many different techniques and recovery methods are proposed to handle the void problem in the wireless and ad hoc networks [21], [25]–[28].

However, due to the different characteristics of UWSNs, such as the differences in topological and environmental features, void-handling techniques proposed for terrestrial networks are impractical in the underwater environment.

Manuscript received May 5, 2016; revised October 27, 2016 and December 24, 2016; accepted January 17, 2017. Date of publication January 25, 2017; date of current version May 31, 2017.
The authors are with the School of Engineering and Built Environment, Glasgow Caledonian University, Glasgow G4 0BA, U.K. (e-mail: seyed.mohammadghoreyshi@gcu.ac.uk; a.shahrabi@gcu.ac.uk; t.boutaleb@gcu.ac.uk).
Digital Object Identifier 10.1109/COMST.2017.2657881

This is attributed to the fact that, first, all communications voids in UWSNs are three-dimensional, which requires different treatments than two-dimensional holes in the terrestrial networks [29]. Second, the mobility of most underwater nodes makes the void mobile. A mobile void can also result from the surrounding environment [30]. This can be the case of a ship that navigates over an underwater network, it blocks communications in the nearby area and thus generates a void that moves along with the ship.

Taking into account the different requirements of the three-dimensional and mobile voids, designing some efficient void-handling techniques for the routing protocols specific to UWSNs is necessary. The performance of these void-handling techniques depends on many factors, such as the number of void nodes, network dynamics, and number of destinations.

Generally, the routing strategies and void-handling techniques can be categorised into two main groups: location-based and depth-based. In the location-based category, the void node is determined based on the geographical advancement of the neighbouring nodes. A node is called a void node, if it cannot find any other node within its neighbourhood with shorter euclidean distance toward the destination. In the depth-based category, the void node is determined based on the depth advancement of the neighbouring nodes towards the water surface. Depth information indicates the vertical distance from each node to the water surface. A node is a void node if it cannot find any neighbouring node with the lower depth than itself. Because of different features in these categories, different void-handling techniques are required.

In this paper, our goal is to investigate the void-handling techniques reported in the literature. To achieve this, we first mention the characteristics of UWSNs in Section II. Then, in Section III, we investigate the void problem in 3D UWSNs and its ignorance impact on the protocols, which do not support any void-handling technique. Then, we indicate the challenges and required features to design and evaluate a void-handling technique. In Section IV, we propose a classification for all void-handling techniques in UWSNs. In Section V, almost all currently reported void-handling techniques in the literature are discussed in detail. The advantages and disadvantages of each void-handling technique are then shown using different examples and analysis. These void-handling techniques qualitatively are compared in terms of their performance regarding the void problem in Section VI. In Section VII, we identify some directions and guidelines for the future research on the void-handling techniques in UWSNs. Finally, in Section VIII, we conclude the paper.

II. CHARACTERISTICS OF UNDERWATER SENSOR NETWORKS

In this section, we present the characteristics of UWSNs. The underwater sensor networks pose more severe situation to cope with void regions. Due to the characteristics mentioned below, the terrestrial void-handling techniques are quite impractical and can not be employed directly in the underwater environment. Thus, it is required to develop the void-handling techniques suitable for underwater acoustic communications taking all the characteristics into account.

A. Three-Dimensionality and Node Movement

In contrast to the terrestrial networks, UWSNs are three-dimensional, and sensors move with the water current. The three-dimensional holes in the routing path can lead to more packet failures. The effect of hidden terminal problem [31] in 3D UWSNs is more intense due to the existence of more neighbouring nodes in different directions. Moreover, the topology continuously changes with the nodes movement. The speed of node depends on the water velocity, which varies at different times [9], [10].

B. High and Variable Propagation Delay

Underwater sensors use the acoustic waves which are preferred for the long distance communications. The speed of sound in underwater is about 1500 m/s [32]. Thus, it causes a large propagation delay, which is about to five orders of magnitude higher than that of radio frequency (RF). The sound velocity varies based on the different parameters such as temperature, salinity, and depth of water [33]. It is critical to taken into account the propagation delay in designing the void-handling techniques in UWSNs.

C. Limited Bandwidth

Due to features of acoustic waves and environmental noise, the acoustic bandwidth is severely limited in UWSNs. The available acoustic bandwidth depends on the communication range and acoustic frequency. As a result of the limited bandwidth, the data rate for underwater sensors can rarely exceed 100 kbps [34]. Therefore, the limited bandwidth of acoustic channels should be taken into consideration when designing a void-handling technique.
D. Path Loss

The underwater environment has higher path loss in comparison to terrestrial physical layer. The path loss results from the attenuation and geometric spreading [35]. The attenuation also results from the absorption of acoustic waves in water [36]. The path loss effect can be reduced by increasing the transmission power or decreasing the traversed distance. Thus, packet forwarding is more likely to be successful if packets are relayed over multiple short distances instead of traversing over long distances [31], [35].

E. Noises

There are two kinds of noise which affect the acoustic communications: man-made noise and ambient noise. Man-made noise is mainly generated by human activities like using pumps and shipping. Ambient noise refers to natural events such as seismic and tides [3], [36]. The main sources of the noise include turbulence, shipping, waves and thermal noise [35]. These noises lead to a lossy and noisy underwater environment, which should be considered carefully in the void-handling techniques.

F. Energy Consumption

Energy consumption is another primary concern in UWSNs since it is hard to replace or recharge the sensor batteries. In UWSNs, the energy consumed by the sensors is much more than what is consumed by the terrestrial sensors [37], [38]. Therefore, energy efficiency is an essential requirement of void-handling techniques in UWSNs.

III. VOID PROBLEM AND CHALLENGES

In this section, we introduce the void problem, the used terminology, and its challenges.

A. Definitions and Characteristics

An underwater sensor network has a 3D network topology in which one or more sinks are located on the water surface equipped with an acoustic modem for underwater communication and with a radio modem for out of water communication [5]. Anchored nodes are located at the bottom of the ocean in the predetermined locations to collect the information and deliver them to a sink by using the relay nodes which are located at different levels in between.

In a greedy-forwarding strategy, each forwarding node transfers packets to a node closer than itself to the destination [9]. Given a sender node \( m_i \) and a destination node \( S \), the advancement of a neighbouring node \( m_j \) is defined as

\[
ADV(m_j) = D(m_j, S) - D(m_i, S),
\]  

where \( D(m, S) \) denotes the Euclidean distance from node \( m \) to destination \( S \). In the location-based routing, destination \( S \) is considered as the closest sink to the node \( m \) [8], [39]. Whereas in pressure-based routing, \( D(m, S) \) is simplified to the Euclidean distance between node \( m \) and the water surface, i.e., the depth of node \( m \). Depending on the class of routing protocol, one or a few candidate nodes can participate in the packet forwarding, which are selected from the following set

\[
C_{m_i} = \{m_j \in N(m_i) : ADV(m_j) > 0\},
\]  

where \( N(m_i) \) includes all the neighbouring nodes within the transmission range of \( m_i \). An empty \( C_{m_i} \) means that node \( m_i \) has no qualified next-hop node to forward the packet in greedy mode. Such a node is called a void node. In the literature, a void node is also referred to as local maxima or stuck node, which are also used interchangeably throughout this paper.

During the greedy mode of packet forwarding in geographic routing, if a relay node cannot find any neighbouring node with positive advancement toward the sink(s), it should switch to the recovery mode to bypass the void area; otherwise, the packet will be dropped [11], [26], [40], [41]. The void nodes are generally located on the boundary of a void communication area. In UWSNs, a void communication area is a three-dimensional region between underwater nodes which is empty of any nodes inside (like the void areas shown in Figs. 2 and 3). A void area prevents communication between some of the nodes in the network. The path between the local maxima node and a non-local maxima node, where greedy routing can be resumed, is called the recovery path.

The forwarding direction specifies whether a hole is a communication void or not. In UWSNs, void areas are usually considered as the holes between the relay nodes and water surface where the sink(s) is located. For further clarification, Fig. 2 shows a case in which there is a void area between node \( e \) and sink \( a \) on the surface. If a greedy routing protocol does not include any void-handling technique, the packet...
Void areas in the terrestrial sensor network are usually fixed because they consist of a set of static sensor nodes [20]. However, in UWSNs, by the movement of floating nodes with the water current, void areas can gradually move to other regions. Hence, mobility of the void communication area is another feature which should be taken into consideration in this environment. Displacement speed of void area is dependent on the velocity of the underwater nodes which is not always so high. Nonetheless, high dynamics is not always a negative factor for the routing protocols, because sometimes the temporary voids can be vanished with the aid of the newly arrived nodes [30].

B. Void-Ignorance Routing Protocols

In this section, we briefly discuss the negative impact of void on some well-known routing protocols which intentionally (for the sake of simplicity) or ignorantly do not consider it. Some routing protocols such as Vector-Based Forwarding (VBF), and Hop-by-Hop Vector-Based Forwarding (HH-VBF) [46], [47] are location-based greedy routing in which forwarding nodes are selected within a virtual pipeline faced toward the destination. However, no solution when facing with a void in the pipeline is provided.

In VBF, packets are forwarded within a fixed virtual pipeline between every pair of source (e.g., an anchored node at the bottom of ocean) and destination. The performance of VBF is dropped in the sparse networks, where candidate nodes inside the pipeline can barely be found. In order to increase the chance of finding nodes in the pipeline, HH-VBF requires a different pipeline at each hop originated from every intermediate (relay) node. However, the void occurrence in the pipeline toward the destination still remained as a problematic issue. This problem is shown in Fig. 4. As can be seen, the packets generated in source A can successfully be delivered to the sink; however, packets generated in source B is stuck in node v and dropped, though there exists a valid path like (v-w-x-y-s) to the sink.
As another protocol, we can consider RDBF [48], which similarly relies on the use of location-based coordinates but with no void-handling technique. In this protocol, packets are relayed through the nodes with the nearest geographical distance to the sink node. RDBF does not limit the forwarding nodes in a pipe, or other geometric shape; however, in facing a void area, RDBF does not present a recovery mode to deal with the packets which are stuck in a local maxima node.

Some routing protocols can also be found in the pressure-based category, with no particular solution for void problem. For instance, DBR [49] is the first pressure-based routing protocol which takes the advantage of the depth of each node to forward packets towards the surface. However, forwarding node selection is not performed in a way that packets do not become stuck in a void node, and neither proposes any recovery method after getting stuck in a local maxima node. This problem is shown in Fig. 5. When node \( k \) receives a packet, since it does not have any neighbour with lower depth or pressure than itself, the packet will be dropped, though there are some valid paths to one of the sinks (e.g., \( k-l-m-n-S1 \)). DBMR and EEDBR [50], [51] are also proposed in this category which consider the nodes residual energy in their forwarding set selection; however they still have no solution for the local maxima node problem.

All in all, the presence of void area in the routing path can dramatically decrease the performance of the network. High packet loss and wasting resources are the immediate consequence of not including an appropriate void-handling technique in the routing protocol. Specially, in delay sensitive applications, dropping a packet can lead to missing a critical event and failure of the network duty. Thus, an efficient geographic routing protocol should include a void-handling mode in addition to the greedy-forwarding mode.

The following section, we discuss the existing challenges to design an efficient void-handling technique in UWSNs.

### C. Void-Handling Techniques: Constraints and Challenges

For 3D sensor networks, it is proved that there is no local and memoryless routing algorithm which can guarantee the packet delivery [52]. Hence, packet delivery just can be guaranteed by utilising the updated reachability information, or using an expensive exhaustive search like flooding. In the remaining of this section, we investigate some techniques that might be used to overcome the void issue followed by a discussion about the constraints and challenges associated with each technique. The features of the void-handling techniques are also summarised in Table I.

1) *Increasing the Nodes Density:* Although the density of the network has, to some extent, an impact on the occurrence of the void area [24], increasing the network density cannot entirely eradicate this issue as it cannot be predicted when and where it can happen. In fact, many other factors, such as deployment pattern of the nodes, node movement, and unreliability of some links, are involved in the creation of the void areas. Hence, the unpredictable nature of void occurrence makes it a complicated task to be located and avoided. Moreover, underwater nodes are much more expensive than the almost costless RF sensor nodes. For these reasons, considering a dense topology to limit the number of void areas is an unrealistic solution.

2) *Flooding Techniques:* In the terrestrial networks, flooding is one of the simplest ways to deal with the void area. However, full flooding technique or original flooding is not a cost efficient method in the underwater environment. The basic idea behind the full flooding is to give a copy of the packet to all the nodes in the network to increase the packet delivery probability [53]. However, when a stuck node floods the packets to all its neighbours, it causes the reception of many duplicated packets by the destination. Furthermore, packets are not relayed in an optimal path with minimum distance and energy cost between the stuck node and destination. Three-dimensionality of the UWSNs also serves to exacerbate the situation by increasing the number of duplicated packets. Duplicated packets waste the network resources and deplete the energy of the nodes (especially nodes close to the sink).

In total, the full flooding wastes the network resources. Nevertheless, restricted flooding or partial flooding can be cost efficient to deal with the void areas. In this way, flooding range and rate are limited to prevent the packets from being distributed into the whole network. Thus, the void problem can be resolved with the minimum cost which is the key aspect of a workable flooding-based void-handling technique [54]–[56].

3) *Heuristic Techniques:* A heuristic technique uses the experience to find a satisfactory solution by employing a practical method. In order to handle the voids, heuristics techniques also can be used to find the recovery paths [54], [57]–[59]. Heuristic techniques have the ability to be customised for different network topologies and void areas. This group of techniques has their own advantages and disadvantages. Although heuristic techniques cannot guarantee the packet delivery, they can significantly decrease the complexity of a solution. These techniques do not follow strict rules to achieve more simplicity and easier applicability in UWSNs. In order to understand where heuristic techniques are applicable, some theoretical analyses usually are required. The derived results from the theoretical analyses can determine the effectiveness and efficiency of each heuristic technique and its various possible configuration.

4) *Transmission Power Adjustment:* Due to special features of the acoustic signal, the transmission power of each node...
TABLE I
SUMMARY OF THE VOID-HANDLING TECHNIQUES IN UWSNs

<table>
<thead>
<tr>
<th>Void-handling Technique</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increasing the nodes density</td>
<td>Reducing the void occurrence to some extent</td>
<td>A dense topology imposes high cost and it cannot entirely eradicate the void problem</td>
</tr>
<tr>
<td>Flooding techniques</td>
<td>Involving more nodes to increase the packet delivery probability</td>
<td>Increasing the number of duplicated packets and wasting the network resources</td>
</tr>
<tr>
<td>Heuristic techniques</td>
<td>Ability to be customised for different network topologies and decreasing the complexity of a solution</td>
<td>They cannot guarantee the packet delivery and not applicable in some configurations</td>
</tr>
<tr>
<td>Transmission power adjustment</td>
<td>Finding non-void nodes in the farther distance by increasing the transmission power</td>
<td>Energy dissipation, and interference in the MAC layer</td>
</tr>
<tr>
<td>Backward forwarding</td>
<td>Initiating an on-demand packet recovery which is reasonable when the number of void areas is limited</td>
<td>It may create a loop between the stuck node and some other nodes which waste the energy</td>
</tr>
<tr>
<td>Passive participation</td>
<td>Contributing to the self-healing property of network topology by the passive participation of the void nodes</td>
<td>An inappropriate approach to handle voids in a sparse underwater network</td>
</tr>
<tr>
<td>Void Avoidance</td>
<td>Minimising the possibility of encountering the void area during the packet forwarding</td>
<td>Exchanging beacons between nodes imposes communication overhead in the highly dynamic networks</td>
</tr>
<tr>
<td>Learning techniques</td>
<td>The packet delivery can be enhanced by learning from the past results</td>
<td>It is not applicable to the networks with a high dynamic</td>
</tr>
<tr>
<td>Network topology control</td>
<td>Reducing the number of void areas by using the depth adjustment mechanism</td>
<td>It consumes high energy for topology adjustment</td>
</tr>
<tr>
<td>Backup network</td>
<td>Obtaining more information about the void areas or collecting data directly from sensor nodes</td>
<td>High energy consumption and complexity</td>
</tr>
<tr>
<td>Hybrid technique</td>
<td>Improving the efficiency by using a combination of the void-handling techniques</td>
<td>More complexity</td>
</tr>
</tbody>
</table>

can be adjusted in a cross-layer fashion [60], [61]. Thus, with the hope of finding non-void nodes (with positive progress) in the farther distance, a local maxima node can increase its forwarding range by increasing the transmission power. It is even possible to use a very high power to connect the local maxima node to the destination directly. However, increasing the transmission power can lead to energy dissipation, and interference in the MAC layer. Thus, in order to exploit this technique, increasing the maximum power and transmission range should be limited at each node. After finding a node with positive progress toward the sink, relaying the packet with a predefined transmission power can be resumed.

5) Backward Forwarding: Some void-handling techniques allow a packet to get stuck and then initiate a recovery method to guide the packet to a non-void node. As a recovery technique suggested in [62], if a relaying node with positive advancement cannot be found, the packet can be forwarded back to a node with the least negative advancement to deliver the packet. However, this approach may create a loop between the stuck node and some other nodes. Void-handling technique should obviously be loop-free, otherwise it only wastes the resources.

6) Passive Participation: As another solution, on a voluntary basis, void nodes can take themselves out of the packet forwarding to provide the opportunity for other available nodes [63]. In this way, each node should be able to recognise whether it is a void node or not. Afterwards, void nodes simply drop the packets as soon as they could not find any neighbouring node with positive progress toward the sink. In this way, when a forwarding node does not receive any acknowledgement from the void node, it selects another node to relay the packet, as if the void node does not exist in the network. This technique contributes to the self-healing property of network topology. However, in some cases, removing a void node may lead to the disconnection with the sink node if the only valid path to the destination is via the void node. Thus, this technique is an inappropriate approach to handle voids in the sparse underwater networks.
7) Void Avoidance: Not sending packets to the stuck nodes is another strategy to cope with the void problem, while other techniques let it happen and then is handled by the local maxima node. This strategy can minimise the possibility of encountering the void area during the packet forwarding. There are different approaches to achieve this objective like VAPR [43], OVAR [37], and LLSR [64]. For instance, a local maxima node can inform its neighbouring nodes about its current state by sending beacons. Thus, neighbouring nodes can set the cost of void nodes to an infinite value to prevent sending packets to them. The difference with the passive participation is that void nodes actively inform other nodes about their current status. Thus, these techniques can reduce the long delay in the routing decision making. However, exchanging beacons between nodes still imposes communication overhead in the high dynamic networks.

8) Learning Techniques: Void-handling techniques can also be augmented using some learning methods [65]. As nodes are mobile, very old information may not be very useful. Furthermore, it should be noted that nodes usually move together with the same pattern. In this way, underwater nodes can learn from the past using the results that have been achieved so far. For instance, if a node has transmitted the packets to a specific route, after a while, it can observe that what percentage of packets have been successfully delivered. The node can then decide whether to send the packets as before or to select a new route. The stability of the decision is based on the dynamics of the underwater environment.

9) Network Topology Control: Underwater sensor nodes can be equipped with the depth adjustment mechanism which enable them to deal with the communication void problem [8], [66], [67]. All void nodes can then move vertically to establish a connection with at least one non-void node. If all topological void can be removed, then there is no need to any further technique to bypass the void area. However, this technique consumes high energy for topology adjustment which must be justified for long-term and non-time-critical applications.

10) Backup Network: Void-handling techniques can use some backup facilities like AUV to obtain extra information about the void boundary and local maxima nodes. Nodes can forward a data packet on the boundary of a void area by using information provided by a backup network. The in-depth knowledge about the characteristics of local maxima nodes and void areas can assist to design more efficient void-handling techniques than current approaches.

Nevertheless, high energy consumption and complexity should be considered as the expenditure side of these techniques. As another solution, AUVs also can identify the networks holes and act as a relay to cover holes between the nodes. Moreover, AUVs can be used as an alternate network to directly collect information from all nodes and deliver them to the sink on the surface. In this way, an AUV travels in a predefined path to collect information from sensor nodes and deliver them to the sink after each complete rotation [68], [69].

11) Hybrid Technique: If a void-handling technique is not able to handle all kinds of voids itself, it can benefits from a hybrid technique by combining void-handling techniques together to improve the efficiency [54], [70]. Sometimes, just following a single void-handling technique can be very expensive in terms of the required resources. Thus, sometimes a combination of void-handling techniques can be used to reduce the excessive use of the resources.

D. Features of the Void-Handling Techniques for UWSNs

A wide variety of routing protocols have been proposed and developed under different assumptions over the past few years. Hence, it is critical to specify a set of criteria that enables us to properly evaluate them. Without such criteria, it is impossible to form an objective judgement, a qualitative comparison, and a comprehensible understanding of all different factors affecting the efficiency and effectiveness of current void-handling techniques.

1) Long or Short-Term Application: First of all, the type of network application has a significant impact on the model that is supposed to be selected [71]. It should therefore be into consideration whether the void-handling technique is designed for a short-term or long-term application. In short-term applications, there is usually no need to use some complicated and expensive void-handling methods. However, for long-term applications, detecting and handling the void areas with a long-term effective strategy is of high importance.

Using the backup networks or the network topology control approaches can only be justified in the long-term applications because of imposing high communication overhead. On the other hand, some void-handling techniques like heuristics, flooding, and passive participation are suitable for the short-term applications.

2) Guaranteed Delivery: One of the most important criteria in the void-handling techniques is whether it can guarantee the packet delivery in theory. Some void-handling techniques can guarantee the packet delivery [37], [43], [64], [72], as long as a topologically valid path exists between every local maxima node and the sink. Guaranteed delivery should be proved using the proposed void-handling mechanism and network topological properties assuming that other factors, such as physical links, MAC layer are in their ideal states.

The void-handling techniques using the reachability information from the destination, can guarantee the packet delivery by avoiding the void and trapped nodes. However, other techniques which are blind to the network topology may fail to find the valid path.

3) Involving Nodes: Involving as less as possible nodes to bypass a void area is another desirable factor. An efficient void-handling technique should find its path toward the destination with the least possible number of relay nodes (i.e., minimum number of transmissions). Perhaps the ideal condition is that packets are not transferred to the void nodes, which is not always possible, or to return to greedy mode as soon as possible using the minimum number of nodes.

The strategies like flooding, dense networks, and backwards forwarding increase the number of involving nodes. However, some strategies like passive participation, transmission power...
adjustment, and void avoidance techniques are able to reduce the number of involving nodes.

4) Communication Overhead: As another metric, void-handling technique should incur minimal overhead in terms of the number of control packets. Some techniques require to exchange a large amount of information between nodes which is not an appropriate manner in UWSNs with long propagation delay and limited energy. Furthermore, some techniques are not efficient in terms of the packet size. For instance, a void-handling technique may include all receiver’s IDs in the header of the packet which can increase the packet size.

The communication overhead is high for the void-handling techniques such as flooding, network topology control, backup network, and learning techniques. In contrast, the heuristic and passive participation techniques may reduce this overhead, if designed properly.

5) Computational Complexity: A void-handling technique should be simple enough to perform its tasks under the practical situations. High complexity can increase the deployment and computational cost of the routing protocols. Thus, an efficient void-handling technique should find an acceptable solution for the void problem within the minimal time and cost range.

The transmission power adjustment techniques need a cross-layer design to cope with the collisions in the MAC layer, which make them a complex approach to be used in UWSNs. The learning techniques and backup networks also increase the complexity. However, the heuristic, the passive participation, and the void avoidance techniques are simple enough to be used in UWSNs.

6) Locality and Scalability: Similar to the greedy routing strategy, a void-handling technique should be able to bypass a void area only by using the local information and in a distributed manner, so that the scalability of the routing protocol can always be preserved [14], [73]. For a void-handling technique, scalability means that how many local maxima nodes can be handled without any significant reduction in performance. It should be noted that centralised techniques are only suitable for small size stationary networks but impractical in the vast underwater environment.

7) States: A void-handling technique is able to obtain a better performance and scalability if fewer number of states is required to be held at each void node [20]. The less information to store, the longer they remain valid. Thus, the void-handling technique should rely on as less as possible number of states. According to this feature, void-handling techniques can be classified into four categories of stateful, stateless, partial stateful, and soft-state.
   a) Stateful model: Stateful routing protocols which should discover and hold a routing path from source to the destination, are suitable only for the static networks [74]–[76]. Underwater sensors continuously move with the water currents and consequently the discovered routing path becomes invalid over the time.
   b) Stateless model: In stateless techniques, nodes are almost blind to the network topology, but the overhead is reduced significantly.
   c) Partial stateful model: In the partial stateful model, nodes maintain a path not fully but limited to a few number of hops. In order to discover and hold the neighbour’s statuses in the partial stateful model, the impact of obstacles and noises on the connectivity should be evaluated periodically. It should be noted that there is always a trade-off between routing efficiency and route maintenance cost.
   d) Soft-state model: In soft-state techniques, no routing path is maintained in each node, but they use some reachability information (e.g., hop count distance, forwarding direction) which are useful, but not essential for efficiency as they can be regenerated or updated when needed [77]. Although this information gives a general view to each node, all routing decisions should be made locally to hold the scalability of the protocol. The void nodes then can be avoided in advance, using the reachability information from the destination. The majority of soft-state techniques are beacon-based to distribute the information between underwater nodes. Beaconing should be performed in regular time periods to update the information [37], [43].

8) Path Optimality: The path from a void node to a non-stuck node (a node which can resume greedy forwarding) should not be much worse than the optimal route [78]–[80]. In other words, the best node out of many available non-stuck nodes should be ideally selected to provide the closest path to the shortest path. In the geographic distance context, the shortest path is expressed as the length of straight line between two nodes.

The void-handling techniques using the reachability information from the destination may discover the near-optimal paths by avoiding the void and trapped nodes. However, other techniques, which are blind to the network topology may fail to find the optimal paths.

9) Path Reliability: An efficient void-handling technique can decrease the overall transmission cost per packet by avoiding lossy links and thereby attain higher throughput. To find a high-quality path, the void node should select the candidate nodes with the highest packet delivery probability. In order to calculate the packet delivery probability, some factors like the attenuation, the ambient noise, and the distance between nodes should be taken into account [81].

10) Opportunistic Forwarding: Any new proposed void-handling technique should take into account the high bit error rate issue in the underwater environment. Opportunistic routing is a promising solution to deal with lossy environments. In this way, packet forwarding is enhanced by taking advantage of simultaneous packet reception among one node’s neighbours and their collaboration to forward the packet [82]–[85]. Reliability and throughput can be increased by using the opportunistic forwarding in which packet is relayed by nodes collaboration in each hop.

11) Sender-Based or Receiver-Based: The void-handling technique should determine whether it is the void node which decides to whom the packet is transmitted to (sender-based), or it just broadcasts the packet and then each receiver decides whether to include itself in the forwarding process (receiver-based). In sender-based techniques, the forwarding node puts the ID numbers of all candidate nodes in the packet’s header.
The receiving node accepts the packet if its ID is included in the packet header.

In receiver-based techniques, when a neighbouring node receives a packet, it can accept or drop the packet according to its current status and the void-handling criteria (e.g., whether or not to be placed in the forwarding area).

12) Energy Efficiency: Moreover, a void-handling method should be energy efficient to prolong the life of the nodes and network [86]. Underwater sensors consume more energy than terrestrial sensors due to acoustic signals used as their communication medium [87]–[89]. The packet transmission by the void node wastes the energy [86]. As an efficient approach, a local maxima node can inform other nodes about its current status to prevent them from sending packets to it. This strategy can reserve energy in the local maxima node. The implicit ACK can also reduce the energy consumption of the void-handling technique since no extra packet are required to confirm the delivery. In this model, when a node overhears that one of its neighbouring node forwards a packet which is already in its buffer, it can consider it as an ACK [90].

The uniform energy consumption of the nodes is also of high significance [45], [91], [92]. Each local maxima node should therefore consider the residual energy of the neighbouring nodes in order to preserve the uniform energy consumption in the network. In the case of having many destinations, nodes with lower residual energy can be easily avoided during the packet forwarding.

Increasing the node density or using the full flooding techniques can exacerbate the energy consumption; however, the transmission power adjustment, the void avoidance, and passive participation approaches can reduce the energy consumption.

13) End-to-End Delay: It is defined as the average time taken from the moment of the creation of packets at the source node until successfully being delivered to the sink node. This parameter is critical for the delay-sensitive applications. Thus, an efficient void-handling technique should decrease the packet delivery time. The delay is dependent on various factors such as the packet holding time, the void-handling strategy, number of hops between the source and destination, network density, and communication overhead [30], [37].

The full flooding, backwards forwarding, and network topology control can increase the latency; however, increasing the node density, void avoidance, and passive participation techniques may reduce the end-to-end delay. The network coding can also be used by the void-handling techniques to increase the network throughput, and to reduce the delay [90].

14) Quality of Service: In some applications, packets are transmitted with different levels of priority. Some data can be more critical than other normal packets which only carry information about the ordinary events [93]–[95]. Thus, an efficient void-handling technique should deal with the critical packets differently. In order to maintain vital resources for the critical packets and deliver them on time, ordinary packets can be delivered using the best effort approach (e.g., using a longer path or with more delay).

15) Activeness: Depending on application requirements, a void-handling technique can be divided into four categories of reactive, proactive, preventative, and hybrid.

a) Reactive model: In a reactive model, the void-handling technique is triggered when a packet is stuck in a local maxima node. The path discovery for each stuck packet is performed on demand. After bypassing the void and finding a node with positive progress toward the destination, the greedy strategy is resumed.

b) Proactive model: In a proactive strategy, stuck nodes in the network are discovered in a preprocessing phase and the path to bypass the void is stored in each stuck node. When a packet is stuck at a void node, it follows a predefined path to bypass the void area.

c) Preventative model: In the preventative model, encountering a void area is prevented with the aid of some precautionary measures such as excluding the void and trapped nodes from the packet forwarding. These techniques try to send the packets to the non-void nodes before encountering a void area. Nevertheless, in some of these techniques, local maxima nodes cannot be avoided thoroughly. Moreover, these techniques need some extra information about the network topology which should be obtained and updated periodically.

d) Hybrid model: In the hybrid model, void-handling techniques consist of at least two void-handling techniques together to obtain more reliability. For instance, void-handling technique can try to avoid the holes with a preventative technique as far as it is possible, and also applies a reactive or proactive technique to deal with the packets may be stuck in a local maxima node.

IV. CLASSIFICATION OF VOID HANDLING TECHNIQUES

Several routing protocols have been proposed for UWSNs over the past few years. There are many ways to classify geographic routing strategies and void-handling techniques in UWSNs. Generally, they can be categorised into two groups: location-based and depth-based. The main difference between these protocols is related to the location service which is responsible for determining the position of the nodes [5], [9].

In the location-based category, all nodes are aware of their 3D location information by the aid of some localisation services [96]–[99]. However, it should be mentioned that Global Positioning System (GPS) cannot be used in underwater environment as a localisation system because of quick attenuation of its waves in water [5].

During the data forwarding of location-based routing protocols, each node can decide about relaying the packet based on its position, the position of the destination, and routing criteria. In some of the location-based protocols, each node should have a table to hold the positions of the neighbouring nodes [30], [100]. The main difference between these protocols is that each protocol tries to apply different fitness factor for selecting the next forwarding nodes.

In this category, a node is called a local maxima node, if it cannot find any other node with positive progress toward the destination in terms of geographical distance (Euclidean distance). After encountering a void area, local maxima node may initiate a local search, within multi-hop neighbouring nodes vicinity, to find a node which is geographically closer to the destination than itself, or just drop the received packet.
Depth-based routing is another class of geographic routing protocols which is simplified to use only depth information to route the packets [49], [101], [102]. The depth of each node in water can be estimated through a pressure gauge which is embedded on it [43]. The final destination is located on the surface and, therefore, packets are forwarded to lower depth at each hop.

A node is a local maxima node if all of its neighbouring nodes make negative progress towards the surface. In this case, the void-handling problem is simplified to a searching process for a node with a lower depth value than that of forwarding node (but not necessarily a node with the closest geographical distance to the sink node) [42]. After finding such a node, depth-based greedy routing can be reactivated [42]. The features of location-based and depth-based techniques are also summarised in Table II.

Moreover, geographic routing protocols in UWSNs also can be classified based on the number and position of destinations into three groups: unicast, anycast, and geocast. In the unicast model, all forwarding nodes should deliver the packets to a specific sink on the surface (single sink architecture). According to the characteristics of unicast model, void occurs with respect to the position of the single sink on the surface.

On the other hand, in the anycast model (multi-sink architecture), there are a number of destinations (sinks or buoys) on the surface which can be utilised during the packet collecting phase [42]. In this way, all packets can be delivered via anycast routing to any sink or buoy on the surface. Due to the existence of different paths, a node can be considered as a local maxima node with respect to one of the sinks, but at the same time, it is a non-void node from another sink viewpoint.

Thus, a local maxima node simply can select another sink as its final destination to solve the void problem.

In the geocast model, a group of nodes in a particular geographical area and in a specific time interval will be selected as the destinations [19]. In this case, a packet should be received by all the nodes inside the target region to have a successful delivery. In surveillance applications, when an anonymous sensor or any underwater vehicle is sensed, it may be required that a data packet is generated and forwarded to a group of sensor nodes in a specific geographical region [7], [103]. Geocasting also can be used to initiate a query asking for needed information from the subset of underwater sensors [103]. In this group of protocols, the void within the geocast area should be addressed and resolved [7], [104]. The void is determined with respect to the packet entry area to the geocast region and often is resolved by considering a covering area around the geocast region [7], [103]. The features of unicast, anycast, and geocast models are also summarised in Table III.

Depending on the availability of location service and also the destination status, a classification of the state-of-the-art void-handling techniques in UWSNs is presented in Fig. 6. Under the class of location-based protocols, different routing protocol with void-handling techniques have been proposed in the literature for different architectures such as unicast (single sink), and geocast. For the pressure-based protocols, anycast (multi-sink) architecture also has been used. However, to the best of our knowledge, no geocast protocol has been proposed so far for pressure-based class of protocols, as it is really challenging to identify a group of nodes as the temporary destination without relying on a robust localisation system. The next section is presented to review the void-handling techniques in details.

### Table II

**Comparison Between the Location-Based and Pressure-Based Routing Protocol Classes**

<table>
<thead>
<tr>
<th>Location-based</th>
<th>Pressure-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>All nodes are aware of their 3D location information</td>
<td>All nodes are aware of their depth using a pressure gauge</td>
</tr>
<tr>
<td>Dependent on the localisation services</td>
<td>No need to any localisation services</td>
</tr>
<tr>
<td>The destination location should be known to all nodes</td>
<td>No need to the destination location knowledge</td>
</tr>
<tr>
<td>In some protocols under this class, each node holds the positions of the</td>
<td>In some protocols under this class, each node holds the depths of the</td>
</tr>
<tr>
<td>neighbouring nodes</td>
<td>neighbouring nodes</td>
</tr>
<tr>
<td>Every packet carries the position of the destination</td>
<td>No need to carry the destination location in each packet</td>
</tr>
<tr>
<td>Routing decision is based on the geographical locations</td>
<td>Routing protocols is simplified to use only the depth information</td>
</tr>
<tr>
<td>The final destination can be placed everywhere</td>
<td>The final destination is assumed to be on the surface</td>
</tr>
<tr>
<td>A void node is determined based on the geographical positions of the</td>
<td>A void node is determined based on the depth of the neighbouring nodes</td>
</tr>
<tr>
<td>neighbouring nodes</td>
<td></td>
</tr>
<tr>
<td>A recovery node is geographically closer to the destination than void node</td>
<td>A recovery node has a lower depth than the void node</td>
</tr>
</tbody>
</table>


TABLE III
COMPARISON BETWEEN THE UNICAST, ANYCAST, AND GEOCAST MODELS

<table>
<thead>
<tr>
<th>Unicast</th>
<th>Anycast</th>
<th>Geocast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single sink architecture</td>
<td>Multi-sink architecture</td>
<td>Nodes in a particular geographical area as destination</td>
</tr>
<tr>
<td>Suitable for small networks</td>
<td>Suitable for large networks</td>
<td>Suitable for large networks</td>
</tr>
<tr>
<td>Packet delivery is successful if packet is received by the single sink</td>
<td>Packet delivery is successful if packet is received by any sink</td>
<td>Packet delivery is successful if packet is received by all the nodes within the geocast region</td>
</tr>
<tr>
<td>Void determined with respect to the single sink</td>
<td>Void is determined with respect to all the sinks</td>
<td>Void is determined with respect to the packet entry area to the geocast region</td>
</tr>
<tr>
<td>Limited number of paths from a source to destination</td>
<td>More available paths from a source to sinks</td>
<td>Available paths depend on the covering area around the geocast region</td>
</tr>
</tbody>
</table>

V. VOID-HANDLING TECHNIQUES

In this section, we review the basic principles of UWSNs void-handling techniques reported in the literature. To be more specific, we focus mainly on the void-handling technique rather than the routing strategy. Thus, for each routing protocol, the routing strategy is briefly explained and then the void-handling technique is comprehensively analysed. In this analysis, our main focus is to study the void-handling techniques without considering other unrelated information such as network characteristics. According to our taxonomy presented in Fig. 6, existing void-handling techniques are classified into two main categories of location-based and depth-based. We present all void-handling techniques under each category along with a qualitative discussion. The features of all void-handling techniques presented in this paper are also summarised in Table IV. Our discussion relies on the features presented in Section III-D.

A. Location-Based Void-Handling Techniques

In this section, we classify existing location-based void-handling techniques into two subcategories of unicast, and geocast and individually discuss each of them. First, we discuss the void-handling techniques proposed under unicast category (single-sink architecture) like VBVA [29], AHH-VBF [30], DFR [54], and FBR [57]. Then, it is followed by discussing the void-handling techniques proposed under geocast category like RMTG [103], and Mobicast [7]).

1) Vector-Based Void Avoidance (VBVA): VBVA [29] is a reactive, stateless, and receiver-based technique which proposed to mitigate the negative impact of void communications on the vector-based routing protocols such as VBF [46], and HH-VBF [47]. In VBVA, each node knows the location of the sink, the source node (via the packet header), and itself. VBVA exploits two approaches, vector-shift and back-pressure for dealing with the convex voids and concave voids, respectively. In the packet forwarding section, VBVA exactly follows a vector-based approach (VBF) to forward the packets toward the sink.

When the routing procedure faces a convex void like Fig. 7, VBVA tries to route the packets along the boundary of the void with the aid of the vector-shift mechanism. To do this, local maxima node (node S) broadcasts a recovery packet called
vector-shift to all its neighbours (nodes $a$ and $d$). The vector-shift packet enables the nodes outside the current pipeline to participate in the packet forwarding by creating the new vectors emitted from them toward the sink. This procedure can be repeated by other nodes till packet is delivered to the destination.

When VBVA cannot find any neighbouring node by shifting method because of placing in a concave void, it initiates the back-pressure mechanism to route the packet backward to find some suitable nodes to do vector-shift (like the procedure shown in Fig. 8). To do this, local maxima node broadcasts a control packet, called back-pressure to let the other nodes with negative progress perform the vector-shift, in the hope that any available path toward the destination can be found. In the case of not finding any path, the back-pressure procedure is continued in the receiving nodes until the vector-shift mechanism can successfully be accomplished. For instance, in Fig. 8, the packet will be forwarded back from node $c$ to the node $S$ where the vector-shift mechanism can be successfully applied.

VBVA initiates the void-handling mechanism on demand while no extra information (e.g., neighbouring information, void characteristics) is required to be stored in each node. This feature increases the scalability and robustness of VBVA for highly dynamic and large network applications. However, the recovery procedure of VBVA is too complicated to be performed in the real underwater environment. VBVA lets the packets to be trapped in a concave hole and then tries to recover them using a time-consuming procedure which increases the end-to-end delay. As a further matter, it is obvious that vector-based protocols suffer from duplicated paths, and vector-shift mechanism can exacerbate the problem as can be seen in Fig. 8. By receiving the packets in two different sides of the void node, packets are subsequently delivered along both boundaries of the void area, resulting in more energy expenditure.

2) Adaptive Hop-by-Hop Vector-Based Forwarding (AHH-VBF): AHH-VBF [30] routing protocol applies a preventative technique to cope with the void problem. In AHH-VBF, every node knows the locations of the sink, the sender node (via the packet header), one-hop neighbours, and itself. This routing protocol is on the basis of HH-VBF, in which direction of the forwarding pipeline is changed hop by hop. In terms of dealing with the void, AHH-VBF is equipped with an adaptive approach which not only changes the direction of the pipeline but also the radius of the pipeline based on the neighbouring nodes distribution. For instance, when the region ahead is sparse, the radius of virtual pipeline (between the forwarding node and sink) is increased to cover a broader and larger area. In this way, more candidate nodes may be found to relay the packet and the transmission reliability is enhanced.

Unlike HH-VBF which uses a constant power level, AHH-VBF is able to adaptively adjust the power level according to the density of neighbouring nodes. Thus, the transmission power level can be increased to cover longer distance in sparse networks, or decreased to save more energy in dense networks. In AHH-VBF, each node has a local knowledge about its neighbours which is utilised to select the proper power level. In order to update information about neighbours, each node sends control packets in periodic times which are determined according to the speed of network topology changes. For instance, when the network topology changes very fast, neighbouring information should be exchanged within a shorter period of time.

AHH-VBF does not always guarantee the delivery of packets to the destination. Although adaptively changing of the forwarding area in a hop-by-hop manner and adjusting transmission power can handle small void areas within the pipeline, it is not flexible enough to change the forwarding direction when confronting with large holes. It is always possible that the pipeline at its maximum size excludes any node inside (like
the example in Fig. 4). Thus, an efficient void-handling technique should be able to shift the packet to another area except the predefined area by the protocol. AHH-VBF, however, lacks such an ability.

Furthermore, the decision on the size of the forwarding region only depends on the distribution of candidate nodes which is not an appropriate approach when nodes follow an irregular distribution. For instance, majority of neighbours may be gathered in a location close to the forwarding node, but because of sensing a neighbour in a far distance, power level should be set at its maximum level which is clearly a waste of energy. As an another problem, if the pipeline radius is set too large to bypass a void, forwarding nodes on the opposite corners of the pipeline will not be able to hear each other and probably forward the same packet concurrently. As a consequence, many duplicated paths are established between the sender and receiver which causes more energy consumption and collision.

3) Directional Flooding-Based Routing (DFR): DFR [54] is a location-based, stateless, and receiver-based routing protocol in which every node knows the locations of the sink, one-hop neighbours, and itself. DFR takes benefit of a controlled flooding approach to achieve more reliability in confronting with the various link qualities in UWSNs. For this purpose, each node adjusts the flooding zone based on the link quality of the area ahead. The intended progress area is considered toward the only existing sink on the surface. For instance, when facing a poor link toward the sink, flooding zone will be set in a way that more nodes can participate in the packet forwarding. On the other hand, if the network is strongly connected, packets can be relayed with the collaboration of few nodes.

However, the void problem is still unresolved where no node can be found in the flooding zone. Accordingly, two types of void problems can appear during the packet forwarding. The first type of void is when a flooding zone without any node is established which causes the packet delivery failure. This phenomenon happens when the flooding zone is continuously decreased due to the good link quality among neighbours while no node can be located in the zone. Thus, DFR exploits a preventative void-handling technique to adjust the flooding zone to cover at least one node to relay the packet.

As shown in Fig. 9, in order to determine the flooding zone at each forwarding node, DFR considers two angles including a Reference-Angle value which is selected by the forwarding node (node $p$), and a Current-Angle which is determined by the geographic location of the receiving node respect to the source and destination (angle between $fs$ and $fd$). The nodes with the higher Current-Angle values than Reference-Angle value are placed in the forwarding zone and qualified to participate in the packet forwarding. Reference-Angle value at least should be smaller than one of the candidates’ Current-Angle values, to meet the void-handling requirements. Thus, it is only sufficient that Reference-Angle value is set smaller than the maximum angle value (amongst the candidates’ Current-Angle values) to cover at least one node in the flooding zone, which helps to prevent voids.

The second model of the void is that none of the forwarding node’s neighbours is closer to the destination than itself. So the flooding zone cannot be established in any way. However, it is possible that a topological detour path can be found via a neighbour with negative progress. To solve this problem, DFR ceases the greedy forwarding phase and tries to bypass the void by finding a detour path. Thus, DFR initiates a reactive void-handling technique which is presented as follows.

As shown in Fig. 10, if forwarding node $f$ cannot find any node with positive progress, it seeks to find a neighbour which has the smallest Current-Angle value among neighbours. After detecting the eligible node $j$, forwarding node unicasts a RTD (request to detour) packet including the original data for this intended node. Upon receiving the RTD packet, receiving node sends an acknowledgement in response to the forwarding node, while it updates the variables in the data packet and forwards it toward the sink. If forwarding node does not receive an acknowledgement from the target node, it continues unicasting to other candidates until finding a detour path.

DFR exploits a hybrid void-handling technique which is simple enough to be implemented and also scalable to be used for large networks. However, in the face of poor link quality and void communications, flooding zone becomes so big which makes it prone to have the hidden terminal problem [37], [42]. Thus, multiple nodes can forward the same packet due to the hidden terminal problem and waste the network resources. Furthermore, when a closer node to the sink cannot be found, the recovery technique is not reliable because of using a unicast approach. This approach does not comply with the opportunistic characteristics of the routing technique and also increases the end-to-end delay and communication overhead. The recovery model is not loop-free and
detour path is not optimal, because DFR has no knowledge about the void boundaries and shapes.

4) Focused Beam Routing (FBR): FBR [57] exploits a preventative void-handling technique to avoid void communications areas. In FBR, each node knows the location of the sink node, the sender node (via the packet header), and itself. The main objective of FBR, as a cross-layer protocol, is to minimise energy consumption by controlling the forwarding nodes transmission power. The main assumption is that each node can adjust its transmission power with a choice between a finite number of power levels as well as having the ability to perform beam forming.

In order to conserve energy, each forwarding node initiates an exploration for available candidates by multicasting a RTS (Request to Send) packet, at the lowest power level. If no neighbour responds to this request, forwarding node is able to increase the transmission power stepwise, until at last a suitable candidate can be located. Those nodes are eligible candidates which lie within a cone emanating from the forwarding node towards the final destination. All the receiving nodes are able to determine that they are within the transmitter’s cone or not, based on the provided location information in the packet (source location, destination location) and their own locations.

Considering the case in Fig. 11, node A intends to send a packet to node B. At the first packet transmission with lowest power level, node A cannot find any candidate node. However, by increasing the power level, it can detect two candidate nodes in its cone (node D and C). Eventually, node D is selected to relay the packet because it is closer to the destination among candidate nodes. If no relay node can be located toward the destination at the maximum transmission power, the main cone will be shifted to the left or right for bypassing the void ahead. By using this strategy, data packet observes a minimum deviation from the straight line between the source and destination (minimum amount of zigzagging), and it also gives another chance to the packet for being forwarded through other available paths.

The advantage of transmissions in the short geographic distances is twofold. First, it can increase the reliability due to access higher bandwidth. Second, it decreases the energy consumption because of using low power levels. If needed, transmission power can be increased to propagate the signal beyond of the current transmission range.

However, variable transmission ranges can interfere to the other nodes’ activities while requiring complicated MAC protocols to handle it. In FBR, sending and receiving the control packets in each hop to establish a connection is very time-consuming. The long-delay problem is exacerbated in the sparse networks, because nodes are usually far away from each other and only high power transmissions can connect them together, while sending the control packets already have to be done for all the lower power levels. In the case of cone shifting, although FBR can efficiently bypass the convex voids with slim shape, it has difficulties with other kinds of voids (e.g., concave, fat shape), due to the lack of a proper mechanism. For instance, when a packet is trapped in a concave void, cone should be rotated 180 degrees to forward the packet back, which seems impractical in the existing approach.

5) Routing and Multicast Tree Based Geocasting (RMTG): RMTG [103] is a 2-D geocast technique in UWSNs with the hole detection ability to distribute data in a specified geographical area covering a group of sensors (geocast region). In RMTG, each node knows the location of destination area (via the packet header), neighbouring nodes, and itself. Unlike many of the geocast techniques which use the flooding approach in the target region, RMTG utilises a covering area around the geocast region for data dissemination.

With the aid of greedy forwarding, the packet is relayed toward the target region. During the process of finding a route...
to reach the geocast region, if a packet is stuck in a local max-
ima node, the packet will be forwarded back to the previous
hop node. For instance, in Fig. 12, node b cannot find any
neighbouring node with positive progress toward the target
region, so the packet is forwarded back to node a. By receiv-
ing the error packet, the previous hop node divides the region
into four quadrants and selects the best next node from the
quadrant which is nearer to the geocast region (like select-
ing node c in Fig. 12). This procedure will continue until the
packet is delivered to the geocast region.

Upon receipt of the packet in the target region, the first
receiving node acts as a root node and creates a multicast
shortest path tree to disseminate the packet within this area.
Sometimes the constructed tree in the geocast region is not
able to cover all destination nodes as the leaf nodes due to
the presence of a void area (e.g., nodes e, f, g in Fig. 13).
Nevertheless, RMTG is able to detect hole inside the geocast
region where a packet cannot be forwarded any longer and
the boundary of the target region is not still reached. In this
situation, a virtual area of radius \( r \) is established around the
boundaries of the target region to involve more forwarding
nodes to handle the void problem.

In order to enter the geocast region from the other faces,
the root node initiates boundary traversing around the geo-
cast region, passing through the virtual area by selecting the
clockwise or counter-clockwise direction. During boundary
traversing, if the packet is received by a node within the tar-
get region for the first time, the packet will be delivered to
this node and subsequently its neighbouring nodes. Again,
boundary traversing is resumed until all the selected faces
are traversed and the packet is delivered to all remaining
nodes in the target area. A sample of this procedure is shown
in Fig. 13.

In RMTG, the assumptions about two-dimensionality of the
underwater environment and using GPS in UWSNs are clearly
improper. The route discovery and route maintenance are also
inappropriate with respect to the nodes movement and rapid
changes in the network topology. Moreover, a larger virtual
area around the target region may involve extra relaying nodes
which causes more energy consumption and also a smaller area
resulting in the lower chance of packet delivery.

6) Mobicast Routing Protocol: Mobicast [7] is a mobile
gecasting approach which aims to collect data from a 3D
underwater area in the presence of various water currents
and void areas. In Mobicast, each node knows its current speed, the
location of itself in different time stamps, and geacast region
(via the control packet by AUV). In this approach, there is
an AUV as a mobile sink which traverses a predetermined
route (usually a circle path) to collect data from sensor nodes
in different geographic regions called 3-D zone of references
(3-D ZOR). Nodes usually stay in sleep mode to save energy.
As shown in Fig. 14, when AUV collects data packets from
sensors within 3-D ZOR\(_t\), it should notify the sensors within
3-D ZOR\(_{t+1}\) to enter the active mode to be ready for the arrival
of AUV. However, message delivery to the next group of nodes
(within 3-D ZOR\(_{t+1}\)) that is supposed to be investigated at
the next time slot is a challenging task due to the presence
of topology holes. If a topological hole blocks the routing
path between AUV and next target region, sensor nodes cannot wake up on time to send their packets to AUV.

To overcome the hole problem, Mobicast routing protocol creates a covering area for surrounding the hole to find alternative paths to deliver the packet to all nodes within the 3-D \( ZOR_t \). As can be seen in Fig. 14, a 3-D zone of forwarding (3-D \( ZOF_{t+1} \)) around the region of interest is considered which is larger than or equal to the size of 3-D \( ZOR_t \). If a feasible routing path cannot be found in the region of interest (3-D \( ZOR_{t+1} \)) due to the void problem, it is possible that an alternative path can be discovered in 3-D \( ZOF_{t+1} \) (e.g., the discovered path via 3-D \( ZOF_{t+1} \) in Fig. 14).

The size of covering area (3-D \( ZOF_{t+1} \)) depends on the network density and the velocity of ocean current. For instance, if there is no topological hole and water current, the size of 3-D \( ZOF_{t+1} \) is exactly equal to the size of 3-D \( ZOR_t \). On the contrary, when there is a hole along with the ocean current and low network density, the covering area should be enlarged to cover more sensor nodes for route discovery. Nevertheless, estimating the accurate size of the covering area according to the available information of the current area (3-D \( ZOR_t \)) is unreliable. Hence, with the aid of the real-time information, Mobicast exploits an adjustment scheme to determine a proper size for the covering area.

Another important issue is that how many sensor nodes within the covering area should be used to deliver the packet. The AUV takes into account the impact of water current on the successful delivery rate within different parts of the covering area and only wakes up the sensor nodes with high successful delivery rate. Inspired by the apple slice concept, the AUV divides the covering area into several identical parts (segments) and only selects those parts which are able to deliver the packet successfully.

Some constraints can confine the performance of Mobicast. First of all, collecting information in this manner is suitable for short-term applications; however, in the long-term applications, AUVs and fixed underwater sensors should collaborate to monitor the underwater environment. Secondly, Mobicast cannot always find a topological valid path from AUV to the nodes within the 3-D \( ZOR_{t+1} \), if any. This is due to the fact that the alternative path may be passed from another area rather than “hold and forward zone” which is an overlapping area between \( ZOR_r \) and \( ZOR_{t+1} \). This problem is shown in Fig. 15, in which the topological valid paths of \( A-B-C-D \) and \( E-F-G-H \) cannot be discovered using the Mobicast protocol. As the last point, the performance of Mobicast highly depends on some parameters which can be set by the user such as user-defined path and response time, the radius of hold and forward zone, and expected range for data collection of AUV.

B. Depth-Based Void-Handling Techniques

In this category, greedy routing can be accomplished by utilising the depth information (or along with some extra information) without knowing the full geographical coordinates. Hence, there is no need to support the routing protocol with a costly distributed localisation mechanism in order to provide each node’s coordinates [42], [49], [105].

![Fig. 15. Problem in void-handling in Mobicast protocol.](image)

In depth-based protocols, a node is considered as a local maxima node, if it cannot find any node with lower depth in its neighbourhood. Hence, the void-handling problem can be simplified to a route discovery method to find a node whose depth is lower than that of the current node to resume the greedy approach. Some routing protocols (e.g., DCR [66], and GR+DTC [8]) are location-based in the routing phase; however, their void-handling techniques are mostly based on the depth values. Therefore, we classified them in the pressure-based category in our analysis.

Again, we first discuss the depth-based void-handling techniques with unicast objective (single-sink architecture) like LLSR [64], IVAR [72], and OVAR [37]. Afterwards, we will discuss the void-handling techniques with anycast objective like DCR [66], GR+DTC [8], HydroCast [42], VAPR [43], and WDFAD-DBR [63].

1) Location-Free Link State Routing (LLSR): LLSR [64] uses a greedy hop-by-hop routing by relying on the parameters such as hop count, path quality, and pressure. Hop count shows the proximity of node to the sink which enables LLSR to bypass void area in a preventative way. In this protocol, path quality indicates redundancy of routes which is measured by counting the number of neighbouring nodes with lower hop count values. This location-free approach is placed under the pressure-based and beacon-based categories.

In beacon dissemination phase, each node periodically broadcasts a beacon including hop count, path quality, and pressure. The beaconing starts from the sink node instead of a source node. The hop count value of sink is equal to zero and other nodes gradually obtain their hop distance and path quality toward the sink by receiving the beacons. In the routing phase, each node selects a one-hop neighbour with lowest hop count value as its next hop node. In the case of a tie, the node with greatest path quality is selected to relay the packet. If the tie persists after considering the path quality, the neighbouring node with lowest pressure is selected as the next hop. In LLSR, selecting a node with the lowest pressure contributes to
higher progress toward the surface where the sink is located, and it can decrease the routing distance between the source node and sink.

By moving the nodes over time, some established paths are not valid any more and should be updated. So nodes are able to update their tables and broadcast a beacon according to the network topology changes. When a node recognises that it is a void node, it should send a beacon with a hop count value equal to infinity for its neighbours. Upon receiving a beacon from a void node, receivers are aware of the lost connection and change their routing path, if necessary. When a void node finds a new path toward the sink, it changes its hop count value according to the newly discovered path.

The strength of this approach is that it is a loop-free strategy with a mechanism to address the network topology changes (e.g., broken links). By exploiting the reachability information from the sink, the void nodes can be detected and bypassed, and by using the pressure information, a positive progress toward the sink can be obtained. One major drawback of this approach is that LLSR does not take into account the opportunistic data forwarding in UWSNs. In this way, only one node is selected to relay the packet in each step which increases the chance of packet loss. Also when the path quality (redundancy of routes) is prioritised over the pressure metric (closeness to the surface) in the forwarding node selection, the packet advancement (toward the sink) may be sacrificed.

Furthermore, this method is not perfectly optimal in terms of energy consumption because beacons are always being sent even for the isolated nodes and the void nodes with no connections to the sink. Although LSSR can bypass all kinds of voids by using reachability information, the void-handling efficiency depends on different parameters such as becoming period, node movement speed, the network topology dimensions. If a node becomes a void node in the upper layers of water, it may affect many nodes status in deeper layers.

2) Void Avoidance Routing Protocols (IVAR, OVAR): Inherently Void Avoidance Routing (IVAR) [72] proposes a soft-state routing protocol which inherently excludes all the routes leading to a void area and therefore does not need to switch to recovery mode. IVAR can transfer a packet around the boundary of a hole and deliver it to the destination only by using depth and hop count information in each node.

This protocol initiates a beaconing process from the destination node instead of a source node. In this way, sensor nodes can obtain reachability information via periodic beaconing by the sink and relay nodes. Each beacon includes the hop count information, which shows the proximity of nodes to the sink.

As can be seen in Fig. 16, when a node broadcasts a data packet, all the receiving nodes with smaller hop count are potentially a candidate node to forward the packet. To avoid multiple forwarding of a packet by more than a node, each node considers its depth as the second metric to set a relaying timer. Each receiving node individually sets a relaying timer based on its depth advancement toward the water surface. The relaying timer of the node with the lowest depth is expired first allowing the node to forward the packet. Other candidate nodes in the vicinity of the forwarding node should discard the packet after hearing this packet transmission.

IVAR is a receiver-based forwarding model in which no neighbouring node information is required to be held by a forwarding node. Each receiver node is able to locally decide whether to participate in packet forwarding only by comparing its hop count value with the sender node. As can be concluded from Fig. 16, when the trapped and void nodes receive a data packet, they simply drop the packet having a hop count value equal or higher than that of the sender node. However, IVAR is unable to suppress all duplicated paths and transmissions resulting from broadcast nature of this method. The candidate nodes may be located in different direction of the forwarding node which leads to the hidden terminal problem and subsequently energy dissipation.

Opportunistic Void Avoidance Routing (OVAR) [37] is proposed to overcome the drawbacks of IVAR in dealing with the duplicated packets and hidden node problem in the forwarding set. OVAR takes the advantage of beaconing procedure, similar to IVAR, to handle the void communication issue. The only difference is that, in OVAR beaconing procedure, the one-hop neighbouring information is held to construct an adjacency graph at each forwarding node. OVAR is a sender-based approach, in which the forwarding node selects the candidate nodes and put their IDs in the packet header. In order to suppress the duplicated packets, the candidate nodes are selected in the vicinity of each other to exclude any hidden node in the forwarding set. With a view to managing the energy, the number of collaborative nodes can be adjusted according to the density of the network.

In IVAR and OVAR, the near-optimal path is selected in almost all cases and void areas are smoothly and efficiently bypassed while its path selection is almost insensitive to node mobility. In dense scenarios, these protocols can deliver packets using a shorter distance compared to other unicast routing protocols. In sparse networks with many possible void areas, these protocols are also able to find the best, or close to, path, if any, with minimum communication overhead [37].

However, there are some limitations which can confine their performance. Selecting an appropriate value for beaconing intervals also has great impact on the network performance in
terms of communication efficiency and information validity. Beaconing intervals with lower values impose high communication overhead to the network, but nodes have more accurate information about the network topology. Furthermore, higher values of beaconing intervals lead to unreliability and inaccuracy of information hold at each node. Despite the fact that the scalability of stateful routing protocols is much better than that of stateful protocols, it is still less scalable compared to the stateless protocols.

Although OVAR has eliminated drawbacks of IVAR in the packet forwarding, OVAR is slightly more complex than IVAR due to its mechanisms to remove the effects of hidden nodes and also to hold energy-reliability trade-off.

3) Depth-Controlled Routing (DCR): DCR [66] is the first geographic routing protocol which exploits a network topology control scheme (as described in Section III-C9), to deal with the void problem in UWSNs. In DCR, each node knows its pressure and the location of all sinks, the neighbouring nodes, and itself. Network topology control improves the network connectivity and diminishes the impact of the void problem by utilising the vertical movement capability of the nodes. Due to the fact that underwater nodes can move vertically, void nodes and disconnected nodes are able to change their depth to be connected to other nodes with an available path to a sonobuoy. In this protocol, the AUVs and on-board hydraulic pressure gauge are used to provide 3D location information for the underwater nodes.

In the routing stage, each node forwards the packet to the nearest sonobuoy via a greedy approach. However, some nodes fail to locate a next-hop node to reach any destination on the surface which makes them eligible for depth adjustment. First, all void nodes are identified by the DCR protocol using a centralised algorithm to determine the set of void nodes and calculate new depth values for them where greedy routing becomes possible. In depth-control stage, the Depth-First Search algorithm is initiated by all surface sonobuoys as root to identify all connected and disconnected nodes. Afterwards, all disconnected nodes are sorted from the shallowest to the deepest nodes and depth adjustment will be performed with this prioritisation.

Following this approach, each void node considers a set of candidate nodes with an available path to a sonobuoy and inside the cylinder shape, centred in the void node with a specified radius (e.g., Fig. 17). The new depth of the void node is examined with respect to all candidate neighbours and eventually void node moves to a new depth where it can be connected to a candidate node with minimum displacement. Void nodes are informed about their new depths via AUVs. When the depth adjustment phase begins, the network operation will be stopped until the topology is reformed. During the packet forwarding phase, if a node realises that it cannot locate a next-hop node, it broadcasts a message to all neighbours to exclude it from the routing path.

By using the topology-control approach, void nodes are reduced or even eliminated without relying on any recovery technique. Nevertheless, there are some limitations when using DCR: i) the high cost of localisation by AUVs, ii) ignoring the movement of nodes with the water current, iii) the impossibility of finding connected nodes in the deeper holes, and iv) the high delay caused by the topology control procedure.

4) Greedy Routing With Distributed Topology Control (GR+DTC): In order to improve the robustness of DCR in dealing with the void, GR+DTC [8] proposes a distributed algorithm for topology control which can react to any change in the network topology caused by nodes mobility. In GR+DTC, each node knows its pressure and the location of all sinks, the neighbouring nodes, and itself. Each node locally is able to determine if it is in a void area and accordingly selects a new depth value, if necessary. The network model (multi-sink architecture) and routing model (greedy forwarding strategy) is exactly the same as the one presented in DCR.

During the topology control, each void node initiates node adjustment on a priority basis on its distance to the nearest sink node. In this way, the nodes with the shorter distance to the surface have shorter waiting time to decide about depth adjustment. The key aspect of this prioritisation is that all nodes with smaller distance to the surface already performed their depth adjustment and other nodes can rely on them as next-hop nodes, if there exists any path between them and any sink. Whenever a void node adjusts its depth, it periodically broadcasts an adjusting information message to detect whether it is now connected to a non-void node. At the same time, non-void nodes can respond by a reply message in order that void node can stop its depth adjustment and set the responder as the next hop node. Upon finding a next hop node, the void node updates its status and broadcasts a beacon for all neighbours to inform its connectivity to a sink node.

An important aspect of GR+DTC is that void nodes do not have to wait for receiving the optimal location information from the monitoring centre and they can react immediately to any change in the topology. Furthermore, it is notable that beacon dissemination is only performed by the connected nodes (which have any available path to any sink) and not as a periodic mechanism for all nodes. Thus, unnecessary energy
consumption will be controlled by excluding the void nodes from the periodic beaconing.

A serious weakness with this routing protocol, however, is that it cannot deal with all kinds of void areas in UWSNs (similar to the drawback of DCR) and also the greedy routing is not equipped with an efficient void-handling technique to resolve this problem. Furthermore, node replacement is not performed in an optimal way.

5) HydroCast: HydroCast [42] protocol is consisted of two parts: greedy pressure-based routing algorithm and a local lower-depth-first recovery method. In HydroCast, each node knows the pressure of itself and neighbouring nodes, and two hop neighbouring distances. In the routing part, HydroCast tries to select a subset of neighbouring nodes with maximum greedy progress towards the destination taking also into account the hidden terminal problem.

In the void-handling mode, local-maxima nodes and recovery paths will be discovered in advance to bypass the void areas during the packet forwarding. The main idea in this technique is to identify stuck nodes by making use of the depth properties of deployed nodes. In this scenario, each node is able to determine if it is a void node or not, only by searching for a neighbouring node with lower depth than itself. If it cannot find any neighbouring node with lower depth, the node is counted as a void node.

In a proactive way, local maxima nodes try to discover a recovery path to a node with lower pressure by using a flooding approach. The discovered node (with lower depth) may itself be another void node which has a new recovery path to the sink. In Fig. 18, for instance, LM1 is a void node which has a recovery path to another void node LM2. As can be seen, node LM2 is also connected to the node S with another recovery path.

The discovered path from a stuck node to a non-void node with lower depth is stored in each local maxima node for the future use. After applying this method, each void node knows an alternative path to a non-void node or directly to a sonobuoy on the surface, if any. After reaching a packet in a void node, it exploits an opportunistic data forwarding over recovery path to deliver packets to a non-void node or a new void node which has connectivity to a destination.

In order to minimise the flooding energy cost, HydroCast uses 2D surface flooding instead of 3D flooding. This is due to the fact that 3D flooding can involve a large number of sensors in the network; however, 2D flooding is very manageable in terms of the number of involved nodes. In this way, only the nodes which are not dominated by the surface neighbouring nodes are able to participate in the route discovery.

In HydroCast, the packet forwarding on the recovery path is performed in an energy efficient manner by suppressing the duplicated transmissions. Void-handling technique used in HydroCast is a loop-free technique which also guarantees the packet delivery.

HydroCast only can deal with the void areas near to the water surface. However, the void areas can appear in deeper regions of the water which are not addressed in this protocol. Furthermore, when there is a large bubble shape void area near to the water surface, more nodes may be involved in the path discovery, which wastes the energy. HydroCast also imposes high communication overhead to obtain two-hop neighbouring nodes information. Also, route discovery and maintenance in the void-handling mode incurs high overhead. Each node requires additional resources such as memory storage to record the discovered path, especially when the recovery path is very long. In terms of path optimality, 2D surface flooding cannot ensure the expected quality of discovered paths in all cases.

6) Void Aware Pressure Routing (VAPR): VAPR [43] is a preventive and soft-state technique which keeps packets away from the voids during the packet forwarding. In VAPR, each node knows the pressure of itself and neighbouring nodes, hop count information, and two hops neighbouring distances. VAPR benefits from enhanced beaconing and opportunistic directional data forwarding in order to handle the void problem.

In the beaconing phase, each beacon includes a sequence number, hop count and depth information which are used to determine the next hop direction (up or down) to reach the closest sink on the surface. The reachability information is propagated across the whole network by using the periodic beaconing initiated by the sink nodes. The sequence number is used to update nodes based on the most recent beacons. Upon receiving a beacon from a neighbour, the receiving node updates its data forwarding direction (DF_dir) and hop distance based to the closest available sink. If a beacon with lower hop count is received from a shallower depth, data forwarding direction should be set as up and otherwise down. The receiving node also extracts the sender’s data forwarding direction from the received beacon and sets it as the next-hop data forwarding direction (NDF_dir). Accordingly, each node knows the direction of packet forwarding for only two hops. Any direction change is a sign of the void area existence.

As can be seen in Fig. 19, when node b receives a packet from node c, its DF_dir and NDF_dir are equal to Up-Up. Thus, it only sends the packet to node a because it is in shallower depth than node b and also node a’s DF_dir (Up) matches with that of NDF_dir (Up). As can be seen, node x is a trapped node which is filtered out by node b because its DF_dir is set to down.

In this protocol, opportunistic data forwarding is only performed based on the directional trails and not on the basis of hop count values. VAPR can filter out the trapped nodes by...
checking the next-hop data forwarding direction. If void appears in the routing path (e.g., observing any direction change), the data forwarding direction of two hops is used to determine the correct routing path. In facing a void area, the forwarding node only considers the neighbouring nodes with a change in the routing direction (up-down or down-up) as candidate nodes. Generally, the forwarding direction is exactly equal to the opposite direction of the beacon reception direction. After selecting the candidate nodes, V APR uses HydroCast approach to select a forwarding set without hidden terminal problem which can maximise expected packet advancement in upward or downward direction (according to the selected direction).

V APR can provide nodes with a partial view of the network topology to diminish the impact of nodes blindness to the network topology. By propagating surface reachability information, there is no need to use any void recovery technique which imposes an extra cost of route discovery and maintenance. The strength of this technique is to guarantee the packet delivery by using an opportunistic directional forwarding. V APR also has loop-free property in a static and dynamic 3D environment.

However, V APR tries to bypass void areas by holding information of up to two-hop neighbouring nodes which impose high overhead to the system. Moreover, the beaconing procedure in V APR (for multi-sink architecture) is not properly utilised in a way that beacons carry some useful information in addition to the hop count. For this reason, each node in V APR is forced to periodically measure the distance to every neighbour and broadcast the measured information to all other one-hop neighbours.

As another problem, packet can only be forwarded up or down depending on the selected direction which cannot utilise subsets of nodes in the horizontal direction (including nodes with lower depth and higher depth together in the forwarding set). Because of this, in facing a convex void (similar to Fig. 20), packets will traverse longer distance because they cannot be forwarded in a horizontal direction to bypass the void and also the number of available candidates will be reduced which can increase the packet failure probability. Thus, to bypass a convex void region, data packets are routed along the longer route leading to an increase in the energy consumption, if this route is highly used.

7) Weighting Depth and Forwarding Area Division DBR (WDFAD-DBR): WDFAD-DBR [63] is a pressure-based routing protocol in which void nodes can take themselves out of the packet forwarding set to provide the opportunity for other available candidate nodes. In this protocol, each node knows its depth, the depth difference of two hops, and neighbouring distances. In WDFAD-DBR, the forwarding area is divided into a primary forwarding area (Reuleaux triangle) and two auxiliary forwarding areas. The primary forwarding area is constant all the time, but auxiliary forwarding areas may adaptively expand based on the node density and channel quality. This feature is useful to suppress the duplicate packets in a dense topology or increase the chance of packet delivery in a sparse network.

WDFAD-DBR also tries to estimate the relative position of neighbouring nodes with the aid of nodes movement pattern and speed, which then assists the underwater nodes to find out whether a neighbouring node is still located in the transmission range or not. The position prediction mechanism also aids to prolong the network lifetime by increasing the updating request interval. To address the void issue, WDFAD-DBR follows a preventive void-handling technique, which can avoid void nodes in advance by considering two hops information and also suppressing the packets in the void nodes. By considering the depth difference of two hops, the chance of encountering a void node is reduced.

As shown in Fig. 21, node $S$ is a sending node and node $A$ and node $B$ are two forwarding candidate nodes because they are located above the sending node. In DBR strategy, node $A$ will first forward the packet having the lowest depth. The packet transmitted by node $A$ will suppress the transmission of other neighbours like node $B$. However, after forwarding the packet by node $A$, there exist no nodes above node $A$ to continue the packet forwarding leading to
void communication issue. This case indicates the weakness of greedy strategies to fall in a local optimal solution. In WDFAD-DBR, however, node $B$ is selected to forward the packet rather than node $A$. This is due to the fact that the depth difference between node $E$ and $B$ is also considered in addition to the depth difference between node $S$ and $B$. Moreover, node $A$ drops the receiving packet after realising that there is no node above itself. This strategy efficiently decreases the probability of packet loss due to the void problem.

However, the issue with WDFAD-DBR is that it cannot identify the trapped nodes in advance. Trapped nodes may lead a forwarding packet to a void node. Thus a forwarding packet does not have any chance to bypass a convex void area. Moreover, in WDFAD-DBR, the packet is dropped if a void or trapped node senses an event and intends to send a packet toward the sink node. This is due to the fact that the forwarding direction in WDFAD-DBR is considered only upward which makes it impossible to consider a void or trapped node as a source node. As another problem, it needs to periodically update the neighbouring information resulting in generating more control packets and corresponding Acknowledgement Packets (ACKs).

VI. QUALITATIVE COMPARISONS

In this section, we compare all existing void-handling techniques in terms of quality to evaluate their effectiveness to deal with the void problem. Our comparison is founded on some criteria presented in Section III-D including activeness, opportunistic forwarding, states, guaranteed delivery, path optimality, end-to-end delay, communication overhead, scalability, and energy efficiency.

Table IV shows how different void-handling techniques are located under different categories while also highlighting their significant features. The majority of these techniques have already been evaluated with the aid of network simulators or using a real testbed. Obviously, our analysis is also consistent with those reported in the literature. To the best of our knowledge, this is the first research study which compares all void-handling techniques from different categories altogether.

Before comparison, it should be mentioned that all void-handling techniques have their own advantages and disadvantages. Since the metrics are not fully independent from each other, sometimes improving one metric (e.g., guaranteed delivery) may adversely affect another metric (e.g., lower complexity and cost). Thus, in order to obtain the maximum efficiency, some key points such as the environmental characteristics, intended application, and unique characteristics of the routing protocol should be considered when designing a new void-handling technique or selecting an existing one [20], [106].

Furthermore, note that we only evaluate the void-handling technique proposed for each protocol independently regardless of their routing algorithms. As another important point, some void-handling techniques are very similar when qualitatively compared; however, they may perform differently when all details are applied in quantitative studies.

A. Activeness

This feature indicates whether a void-handling technique is able to handle void communications reactively (on demand) or proactively (with a previous plan). Also, some approaches have no recovery method which requires them to use a preventative technique.

As can be seen in Table IV, most of the underwater void-handling techniques use a preventative approach. This is due to the fact that in UWSNs, reactive and proactive techniques often impose high communication overhead to the network [5], [107]. For instance, VBVA and RMTG are reactive techniques which are only activated when a packet is stuck at a void node. However, they should tolerate high cost to be able to recover packets from the void area. Nonetheless, some of the preventative techniques still suffer from the packet loss because they cannot efficiently bypass a void area. In order to increase robustness, DFR exploits a hybrid technique including a preventative (flooding zone adjustment) and reactive (finding detour path) approach. HydroCast is the only technique which uses a proactive technique to hold the recovery path from the void nodes to the non-void nodes. However, permanently keeping a recovery path at each node is very costly [108].

In general, preventative techniques can be less costly in terms of resource consumption when the topology changes are slow and a large number of control packets is not required to find the detour paths.

B. Opportunistic Forwarding

This metric specifies whether a void-handling technique exploits a subset of neighbouring nodes to relay a packet in order to increase the transmission reliability [109], [110]. The opportunistic data forwarding is usually neglected in some void-handling techniques.

For instance, DFR uses opportunistic data forwarding in the routing algorithm with a non-opportunistic void-handling technique. DFR void-handling technique will be satisfied just
with placing one node in the flooding zone or only by finding one of the best neighbouring nodes in the search for the detour path. This feature is in contradiction with the opportunistic nature. DCR, GR+DTC, RMTG and LLSR are also not opportunistic neither in routing nor in the void-handling technique. For instance, DCR and GR+DTC will be satisfied if only one node covers the void area which can make it a bottleneck node in the routing path. Other techniques listed in the table are opportunistic; however, some of them suffer from the hidden terminal problem [37], [42]. Only OVAR,
HydroCast and VAPR remove the hidden nodes from their forwarding set and therefore have no duplicated packet transmissions. WDFAD-DBR also suppress all duplicated packets by using a Reuleaux triangle as the primary forwarding area, but it may intentionally let to have duplicated packets in the sparse density by including the auxiliary forwarding areas.

In some void-handling techniques like AHH-VBF, DFR, OVAR, and WDFAD-DBR, the number of candidate nodes is adjusted based on the density of area ahead which makes them more suitable to be used in UWSNs in terms of energy and reliability tradeoff.

C. States

It is desirable to hold fewer states at each node in order to increase the scalability and performance of the void-handling technique [20], [111]. The majority of void-handling techniques in UWSNs are stateless. This means that each node only holds the states of one hop or up to two hops [5].

RMTG is the only stateful approach, since a path is established between the source and geocast region and also root node holds the states of all nodes in the geocast region. HydroCast is considered as a partial-path state approach which needs to maintain states with a partial path from a void node to a non-void node or final destination. LLSR, IVAR, OVAR and VAPR are considered as soft-state models, because they rely on the beaconing information which is useful for the routing efficiency, but not essential, as it can be updated or replaced if needed. They do not maintain a path toward the sink, but they use reachability information to decide about the routing path [43], [72].

According to the speed of the nodes and topology changes, the appropriate approach should be selected. The stateful, soft-state, and stateless approaches are suitable for stationary, semi-dynamic, and high-dynamic topologies, respectively.

D. Guaranteed Delivery

We measure this feature when a void node is able to deliver a packet to the destination if there exists a valid topological path or simply valid path between them. It is proved that no local and memoryless technique exists to guarantee the packet delivery in 3D networks [52].

Nevertheless, LLSR, IVAR, OVAR and VAPR by using the reachability information, are able to bypass all kinds of voids. Although VBVA also has a solution for both kinds of voids including convex and concave, this technique is not loop-free. Back pressure in VBVA may send the packet to a non-void node and resume vector-based forwarding, but packet again can be stuck in another void node which has no valid path toward the sink except the previous traversed non-void node. On the other hand, the main goal of DCR and GR+DTC is to change the network topology in a way that all void nodes are moved to a non-void position. When there is no void node, in theory, the packet can be delivered by each node toward the destination [21], [112]. However, these techniques cannot eliminate all void nodes just by changing their depth, so they cannot deliver the packet in some cases. FBR and HydroCast only can guarantee the packet delivery if they are not faced with any concave void. Mobicast is successful in packet delivery if there is a valid path started from a node inside the “hold and forward zone”, otherwise it fails. DFR and RMTG are not loop-free techniques and AHH-VBF is not able to change its pipeline direction if it is required, so they may fail in the operation. WDFAD-DBR also finds no path toward the surface if the only valid topological path is via a void node.

It can be observed that only those techniques which have used the hop count distance information for the packet forwarding, are successful in bypassing all the void areas.

E. Path Optimality

Path optimality shows whether the traversed distance by a packet is close to the optimal path which is expressed as the length of straight line between the source node and destination or not [113]. In general, none of the existing void-handling techniques can ensure that an optimal path is always discovered.

Nevertheless, LLSR, IVAR, OVAR and VAPR are able to find a near-optimal path when the updated reachability information is properly supplied. Note that packets are not forwarded on an optimal path when VAPR confronts a wide convex void area, as shown in Fig. 20. Although other techniques are unable to find the optimal path, some of them may have other advantages such as simplicity of implementation, reliability of the proposed path, and also resource management efficiency.

As can be observed, in a 3D environment, finding a near-optimal path depends on the nodes viewpoint of the network topology [43], [114]. Thus, the soft-state techniques may route the packets in a shorter path.

F. End-to-End Delay

This metric shows the time required to deliver a packet from a source node to the sink [115]. The void-handling technique used by VBVA imposes high delay because it lets the packets be stuck in a void node and then tries to recover them using a time-consuming procedure. FBR also has a high end-to-end delay because it sends and receives the control packets in each hop to establish a connection. Although DFR tries to decrease the delay by using a preventative approach at the first step, the recovery technique may be required, which causes more delay. RMTG has a high delay because of using a reactive void-handling technique and also need to the route discovery between the source node and geocast region. Mobicast has less delay by using a preventative approach. Sof-state techniques such as LLSR, IVAR, OVAR, and VAPR have low latency by forwarding packets in a near-optimal path. DCR and GR-DTC are not suitable for the delay-sensitive applications due to the high delay caused by the topology control procedure. Although Hydrocast uses the predefined recovery paths to decrease the delay, it lets the packets be stuck in a void node which is still time-consuming. AHH-VBF and WDFAD-DBR have lower delay by using a preventative approach and without any need to perform a recovery phase.

It can be concluded that the reactive techniques impose a higher delay, by letting the packet being stuck in a void node at
the first stage [116]. Among the preventative techniques, those are performed with a lower delay, which collect the required routeing information before the forwarding phase.

G. Communication Overhead

This metric shows the amount of communication overhead is imposed when handling a void. Generally, the void-handling techniques in the underwater environment have higher overhead in comparison to the terrestrial void-handling techniques because of the dynamic nature of UWSNs [1], [117].

VBVA has high overhead due to a large number of control packets generated in a chaotic manner. DFR and AHH-VBF can adjust their forwarding zone hop-by-hop which enables them to control the protocol overhead to some extent. Thus, these approaches are considered as the low overhead techniques in UWSNs. FBR can be considered as a high overhead technique, due to sending and receiving Request To Send (RTS) and Clear to Send (CTS) in each forwarding node and also for each power level setting. DCR needs Depth-first Search to identify all void nodes, and also an AUV should gather information from relay nodes and then inform them about their new positions which make this technique a high overhead approach. However, GR+DTC decreases this overhead by using a distributed approach in which each node individually starts their depth adjustment initiated from the shallowest depth. In geocasting, RMTG uses a large number of control packets, while Mobicast relies on its estimation about the covering area around the geocast region and does not need to exchange control packets. LLSR, IVAR, OVAR and VAPR are not considered as high overhead techniques. Although beaconing technique inherently imposes communication overhead, it can be applied over long intervals due to the fact that nodes move slowly with water current [37], [43]. Moreover, the communication overhead in the beaconing can be justified against localisation overhead in the location-based techniques [96], [97]. WDFAD-DBR incurs low overhead during the routing phase but the updating phase for obtaining one-hop information still imposes overhead to the protocol. Finally, HydroCast is categorised as a high overhead technique because of its proactive approach and also recovery path discovery (2D flooding) and maintenance. Generally, the stateful, reactive, and heuristic techniques using multiple control packets, are prone to having the higher overhead [118].

H. Scalability

This feature shows that the performance of each void-handling technique is not affected by increasing the number of void nodes [20]. DCR is not scalable to the number of void nodes, since it follows a centralised approach; however, GR+DTC obtains more scalability by performing a localised approach for depth adjustment. RMTG is not scalable because the routing path between the source and geocast root should be held which is costly in a 3D dynamic environment. The opportunistic data forwarding in HydroCast is scalable, but its 2D surface flooding is not scalable because every void node should search and find a recovery path. Note that recovery paths are actually not used most of the time. Soft-state protocols such as LLSR, IVAR, OVAR, and VAPR are scalable, but not as well as the stateless routing protocols. Although VBVA uses a stateless approach, its scalability is not high because of using a reactive approach. Other void-handling techniques can be considered as high scalable because they use a stateless and preventative approach. In whole, the majority of stateless and distributed void-handling techniques are more scalable than stateful, soft-state, and centralised approaches [53], [106].

I. Energy Efficiency

This metric reflects how void-handling technique is energy-efficient by considering all influential factors such as number of transmissions, communication overhead, involving nodes, hidden terminal problem and so forth [87], [119], [120]. VBVA is not energy efficient due to a large number of generated control packets. DFR, FBR, and AHH-VBF exploit a forwarding zone which can prevent packets to be flooded in the unnecessary areas of the network. FBR and AHH-VBF also have the ability to adjust their transmission ranges to further control the energy consumption. Vertical movement of nodes is an energy-consuming task which only can be justified when used less frequently [8]. DCR is not considered as an energy efficient technique for exchanging a large number of control packets with monitoring centre; however, GR+DTC can diminish this energy dissipation by reducing the number of control packets. RMTG does not consider the energy issue in its geocasting, while Mobicast tries to precisely estimate the covering area to wake up only the required nodes at the right times.

Beaconing-based techniques are able to compensate beaconing energy consumption by traversing the optimal path in the routing (reducing the number of transmissions), and also this cost can be justified by considering the localisation energy cost in the location-based techniques. VAPR and OVAR are more energy efficient among them because of taking advantages of an opportunistic data forwarding which efficiently addresses the hidden terminal problem. WDFAD-DBR also resolves the duplicated packets in a dense network which turns it into an energy-efficient routing protocol. Flooding techniques usually consume high energy; nonetheless, HydroCast controls energy consumption by utilising a 2D flooding instead of 3D flooding and also uses opportunistic data forwarding on the recovery path to suppress duplicated packets. Overall, the void-handling techniques with lower communication overhead, power level adjustment, and resizeable forwarding area based on the network density, have achieved a higher energy efficiency.

VII. Open Research Challenges

From all discussions in the preceding sections, it seems that void-handling techniques have received great attention when designing efficient routing protocols for UWSNs. However, some technical, practical, and environmental issues are still remained for further investigation. Based on the literature surveyed above, we present the following potential directions that can be considered in the void-handling techniques.
First, hybrid void-handling techniques should be devised based on the new deployment environment. It would be interesting to know what kinds of characteristics can be inherited from the previous techniques and which of them should be devised based on the new deployment environment. Thus, with previous knowledge of void-handling techniques presented in the literature, it becomes increasingly clear why a hybrid approach in UWSNs is vital.

Second, the majority of current void-handling techniques are designed for the shallow waters with a limited number of void areas. However, it is interesting to study whether these techniques are still effective in a deep multi-holes environment under varying pressure, temperature, and salinity. In some cases, very specific solutions are required to deal with multi-holes in UWSNs [28].

Third, the trapped nodes issue is deserved to receive higher attention in void-handling techniques due to its direct impact on the performance of void-handling techniques. As an efficient solution, a void-handling technique should proactively discover the trapped nodes in a preprocessing phase and avoid them during the packet forwarding.

Fourth, to our best knowledge, the impact of nodes movement on the void area have not been investigated thoroughly in the literature. The void area is continuously reshaped or move with the water current [30]. The void-handling techniques also suffer from lack of a realistic model for node mobility. Most of the existing protocols assume that nodes are mobile at a low rate or they are stationary. Therefore, investigating the impact of node movement on the void-handling techniques seems to be a challenging issue.

Fifth, designing the void-handling techniques with a cross-layer view can enhance the performance of the routing protocols during packet delivery. The existing void-handling techniques have only focused on the network layer. However, with a cross-layer design, the number of collisions can be managed more efficiently over the MAC layer, while the results of some tasks, such as beaconing, can be shared between layers. [57], [121].

Sixth, dealing with a void area within a geocast region is still a challenging issue. The existing model involves many relay nodes to cover the geocast region with a larger area. However, it is necessary to design the new void-handling techniques to further decrease the number of involving nodes.

Finally, some existing void-handling techniques have been proposed under unrealistic assumptions about the underwater environment (e.g., availability of precise full coordinates information, noise-free environment, etc.). Thus, conducting a realistic study of the proposed void-handling techniques using a real testbed can easily enlighten their weakness as well as their strengths.

VIII. Conclusion

In this paper, we investigated the state of the art of void-handling techniques in UWSNs. First of all, we discussed the different features of void communications in the terrestrial and underwater environments and mentioned the unique challenges of designing void-handling techniques in UWSNs. Afterwards, the main features for designing efficient void-handling techniques have been introduced. To facilitate comparison of different techniques, we classified the current void-handling techniques into two main categories of location-based and depth-based techniques. For each category, all existing void-handling techniques have individually been explored in detail. Then, a comprehensive comparison of the currently available techniques has been proposed. It is shown that each void-handling technique is designed for a specific environment which has its own strengths as well as its constraints. Finally, some open research challenges are mentioned to deal with the void problem.

References


Seyed Mohammad Ghoreyshi received the B.E. degree from the Mazandaran University of Science and Technology in 2009 and the M.E. degree from the University of Tehran in 2013. He is currently pursuing the Ph.D. degree with Glasgow Caledonian University, U.K. His current research interests include underwater sensor networks, designing the routing protocols, and void-handling techniques.

Alireza Shahrabi received the B.Sc. and M.Sc. degrees in computer engineering from the Sharif University of Technology, Tehran, Iran, in 1991 and 1994, respectively, and the Ph.D. degree in computing science from the University of Glasgow, Glasgow, U.K., in 2003. Since 2001, he has been with the School of Engineering and Built Environment, Glasgow Caledonian University, where he is currently a Reader. His current research interests are network protocols, wireless, and mobile and sensor networks and also performance modeling and evaluation of parallel and distributed systems. He has published extensively in leading journals and well-established conferences. He has been on the editorial board of some international journals and also served as the organizing and program committees of many international conferences and workshops. He is a member of the IEEE Computer Society.

Tuleen Boutaleb received the B.Eng. degree in electronics engineering and the Ph.D. degree in cellular mobile networks for telemetry from Glasgow Caledonian University, U.K., in 1995 and 2006, respectively. From 1995 to 1998, she was a Researcher on a European project researching on the communications infrastructure. From 1998 to 2001, she was a Research Assistant researching on a European project. Since 2001, she has been a Telecommunications Engineering Lecturer within the School of Engineering and Built Environment, Glasgow Caledonian University. Her research interests include cellular communication networks performance enhancement, messaging in VANET, WSN, and satellite communications for remote monitoring.