From Real-World Data to Synthetic Models: a Time-Varying Approach for Underwater Networks

Dimitri Sotnik
Fraunhofer FKIE
Wachtberg, Germany
dimitri.sotnik@fkie.fraunhofer.de

Roberto Francescon
DEI, University of Padova
Padova, Italy
frances1@dei.unipd.it

Filippo Campagnaro
DEI, University of Padova
Padova, Italy
campagn1@dei.unipd.it

Michael Goetz
Fraunhofer FKIE
Wachtberg, Germany

Ivor Nissen
Bundeswehr Technical Center for Ships and Naval Weapons, Maritime Technology and Research
Eckernförde, Germany
ivornissen@bundeswehr.org

Michele Zorzi
DEI, University of Padova
Padova, Italy
zorzi@dei.unipd.it

Abstract

The selection of the most appropriate method for modeling the underwater acoustic channel for underwater network simulation has always been a trade-off between computational cost, accuracy, and statistical significance. While some models, such as ray tracers [2], offer high physical accuracy, they are also highly demanding in terms of computational cost. Other models, such as Urick Thorp [17], are very lightweight but lack physical accuracy in some aspects. Recently, the research community has more often resorted to models based on data collected during sea-trials and experiments, to retain accuracy but using them to generate different statistical realizations to gain statistical variability. In this work, we describe and analyze an approach based on the Hidden Markov model and explore how to extract its parameters, based on real-world experimental data, for a better representation of the underwater acoustic channel to be used in underwater network simulations, and use it in an efficient and time-effective way. We then show the results of a simulation session, compared with the results retrieved from the experiment data.

Keywords

Underwater acoustic networks; underwater networks simulation; channel modeling; acoustic modeling; DESERT Underwater; experiments.

ACM Reference Format:
Dimitri Sotnik, Roberto Francescon, Filippo Campagnaro, Michael Goetz, Ivor Nissen, and Michele Zorzi. 2024. From Real-World Data to Synthetic Models: a Time-Varying Approach for Underwater Networks. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (WUWNet '24).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

WUWNet '24, October 28–31, 2024, Sibenik, HR

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction and related work

Underwater acoustic (UWA) wireless communications have become mature enough to allow for a wide variety of new applications [8]. In turn, new proposed scenarios pose new challenges to underwater communication networks. Once a new communications protocol or scheme has been devised, it is first evaluated from a theoretical point of view, by trying to highlight all the possible drawbacks and downsides and making the correct trade-offs. After the design is established, and once it is implemented and debugged, the new technology needs to be evaluated in a real-world deployment. Underwater acoustic communications suffer heavily from the dependence on the particular deployment setting [9, 14]: waves, rain, snapping shrimps, temperature, and thus the period of the year and the time of the day, are all factors that impact heavily on the UWA communication. For this reason, real-world sea-trials are still considered by the community the only way to meaningfully and completely assess the performance of a new UWA communication protocol. Nonetheless, deploying underwater acoustic networks for wireless communications has always been a demanding, expensive, and time- and resource-consuming task, even in the not-so-common case where all the facilities and equipment are available. The community has then relied on simulators to, at least, limit the number of field trials needed to evaluate and assess an underwater acoustic communication protocol stack or architecture: the simulations can highlight issues or possible inefficiencies earlier than the actual experiments, thus making the sea-trial more cost- and time-effective. The simulators employed in such evaluations must include a model for the underwater acoustic physical layer: these models have some desirable features, among which there is a light computational cost, and thus a short execution time, that would allow running many simulations with different network schemes, topologies, technologies, transmission patterns or, simply, to get more statistically significant results by multiple runs. It is understood, then, that the underwater acoustic physical model employed in such simulations does not need to be highly physically accurate like the ones needed by the sonar community but, instead, statistically accurate: for this reason the underwater networking community has relied on lightweight models such as the Urick-Thorp model [14, 17]. We point out that physically accurate models such as, e.g., those based on

ray tracing like Bellhop [2] do exist, but their computational cost limits their usage.

Another fruitful approach is to make use of data collected during sea experiments: sea-trials usually generate large amounts of data that, if time- and node-referenced, could be used in emulation or network-replay [12]. One approach to make use of experimental data is to use Lookup Tables (LUTs) to calculate the performance of a device and map it to the simulated one [11]: the LUTs can map performance such as Packet Delivery Rario (PDR) versus Signal-to-Noise Ratio (SNR) or distance and are dependent on the type of experiment and on the application the user is applying them to [15]. This approach, while being accurate in modeling the average characteristics, is limited in that it does not introduce any time variability. A better approach is to replay a real-field scenario into a simulated one by means of time reference. In the ASUNA emulation framework [7], the authors have collected a large amount of experimental data, made this data publicly available, and proposed a way to use it in a scenario-replay fashion. The idea is to allow trace-based simulation, or experiment replay, through a growing number of experimental traces, added by the users of the tool. A similar approach has also been used in the DESERT Underwater-based network simulator [4] employed in the EDA SALSA project [5].

Other models for simulating the underwater acoustic channel are presented in [10], highlighting the advantages and disadvantages of each of them and indicating statistical models as a promising solution for underwater network simulations.

To try to retain both accuracy and statistical variation, based on measured data, the authors in [16] have implemented a series of Markov chain-based models whose parameters were inferred from the *SubNet09* experimental data, in which JANUS packets were transmitted and recorded. The 2-state trained hidden Markov model (HMM) proved to be the most accurate. Recently, based on the work in [16], the authors in [6] have presented a new model based on the HMM statistical model, with a third additional state. This model captures the acoustic channel behavior better than the 2-state one. In this work, the three states represent values of Packet Error Rate (PER): the first issue the model presents is the choice of the states or, more precisely, the choice of the thresholds that define the states. Although the authors proved that a visual inspection can be sufficient to perform this task, an automatic and more robust method would be preferred. In this work, we are going to deploy a three-state HMM to model the fading of underwater acoustic links and illustrate an automatic way to infer the parameters for the selected HMM and evaluate the selection thus obtained. The second issue of the model in [6] is the fact that only the time-varying PER is mapped to the simulator, without information on received signal or noise, hence limiting the information to the receiver and its higher layers for performing channel adaptation. In this work, instead, we map a time-varying received power, hence allowing the system to change modulation and coding schemes (MCSs) according to the channel conditions.

2 Channel model

In this section we present the channel model that has been included in the DESERT Underwater Framework [4], and the methodology used to infer the channel parameters from the sea trial data.

2.1 Description

We selected the 3-state HMM described in [6] and [13] to model the underwater acoustic channel. The model features an observable random process that statistically describes the performance of an underwater acoustic link, on top of a hidden Markov Chain (MC) that statistically describes the intrinsic state of the same link: this state can be represented by measurements such as SNR, Received Signal Strength Indicator (RSSI) or other physical characteristics. The observable process is usually associated with network performance metrics such as PDR or Bit Error Rate (BER). We chose the output SNR (i.e., the SNR obtained after equalization) as a measure modeled by the HMM, with a three-state Markov chain. The three states are defined as GOOD, MEDIUM, and BAD.

As a visible process we chose to model the large-scale fading, or shadowing, for each link: each one of the three states corresponds to a value of loss (averaged) that is to be added to the total channel loss, accounting for the shadowing component.

The first issue encountered in the definition of the model is the definition of the thresholds, separating the three states, i.e., the SNR values that define the BAD state from the MEDIUM state and the one that defines the MEDIUM state from the GOOD state: this will be investigated in Section 2.3.

A second issue to tackle involves the choice of the shadowing value to assign to each of the three states. We decided to choose the median of all the values present in a state: because the distribution clearly showed that the average was moved towards the extreme by numerous outliers, especially in the BAD state, hence median seems to better represent the quantity we are looking for.

2.2 Implementation

The time evolution of the state s of each network link is modeled as a three state MC, with $s \in S = \{G, M, B\}$, with G the GOOD, M the MEDIUM and B the BAD states. Following the procedure explained in [6], we considered a slotted time and, once the SNR thresholds are set to define whether, at a certain time, channel was G, M or B state, we can compute the transition probabilities p_{ij} from state i to state j in one time slot and arrange them in the transition matrix $\mathbf{P}$ defined as:

$$\mathbf{P} = \begin{pmatrix} p_{gg} & p_{gm} & p_{gb} \\ p_{mg} & p_{mm} & p_{mb} \\ p_{bg} & p_{bm} & p_{bb} \end{pmatrix}. \tag{1}$$

Finally, the transition probability from state i to state k in N time slots can simply be computed with the matrix exponentiation $\mathbf{P^N}$.

Once the transition probabilities and the SNR thresholds are obtained, we need to include them in the physical layer of the DESERT Underwater Framework. This operation was performed by adding to the SNR, computed according to the model in [14], a fading factor $e(s)$ that depends on the link state s. The resulting SNR is therefore computed as

$$SNR = \frac{P_{tx} \cdot a(f,d) \cdot e(s)}{N(f,w)} = SNR' \cdot e(s), \tag{2}$$

where P_{tx} is the transmitted acoustic source level, $a(f,d)$ the signal attenuation that depends on the center frequency f and the distance d, and $N(w)$ the acoustic noise in the center frequency f and the bandwidth w. $a(f,d)$ and $N(f,w)$ are computed as in [14], while the fading value $e(s) > 0$ is a zero-truncated Gaussian-distributed

random value whose mean and variance depend on the state s. SNR' is the SNR computed according to the original formula in [14]. We remark that, even if considering the HMM, we still compute the signal attenuation and noise from [14] to maintain a certain relation between SNR, distance and frequency band. Being $\overline{SNR[s]}$ the mean of the SNR observed in state s, we define the mean $\overline{e(s)}$ as follows:

$$\overline{e(s)} = \frac{\overline{SNR[s]}}{SNR'}. \tag{3}$$

Optionally, the standard deviation of $e(s)$, used to simulate channel variability in the same state, is obtained in the same way starting from the standard deviation of the SNR in the three states and dividing it by SNR'. For simplicity of analysis, the standard deviation can be set to 0: in this case the random variable $e(s)$ degenerates to a constant value equal to $\overline{e(s)}$.

Once computed the SNR, the BER is computed assuming the use of the Frequency-Repetition Spread Spectrum (FRSS) modulation schemes, obtained from the formula in [3]. FRSS provides four different profiles, with profile 4 the most robust, and profile 1 the one providing the highest bitrate.

The simulation scenario involves 9 nodes, of which some are underwater stationary nodes and others are moving Autonomous Underwater Vehicles (AUVs). The modulation employed is FRSS with a carrier frequency of 8 kHz. Each node has a transmission session every 180 s, where it transmits 5 times to 5 different nodes (randomly chosen) every 10 s. The transmitted packets are either received or not and are not forwarded by any other node.

2.3 Parameters inference

The underlying data were collected from extensive sea trials and harbor tests, where SNR measurements between different underwater nodes were recorded at a specified interval, thus giving a slotted measurement. The dataset extracted contains links between nodes characterized by SNR values and timestamps. To ensure that only valid and significant data were included in the analysis, invalid or interpolated values were filtered out.

To classify the SNR values, the KMeans algorithm [1] was utilized, dividing the SNR values into three clusters: BAD, MEDIUM, and GOOD. The choice of three clusters allows for the definition of states for the HMM based on the frequency and distribution of the SNR values. Additionally, the PDR was analyzed to ensure that the clusters correlated meaningfully with channel performance. The KMeans algorithm iteratively minimizes the squared distances between data points and their respective cluster centers, facilitating the grouping of SNR values by similarity and contributing to the identification of typical channel states.

Figure 1 displays the histogram of the SNR values, showing a typical distribution of our measurement data. The cluster boundaries are marked and labeled as BAD, MEDIUM, and GOOD. These boundaries define the states used in the HMM and illustrate the distribution of SNR values across the three states.

After the clusters have been calculated, and accordingly the cluster thresholds, the transition probabilities P_{ij} for moving from state i to state j were calculated using the formula

$$P_{ij} = \frac{N_{ij}}{\sum_k N_{ik}}, \tag{4}$$

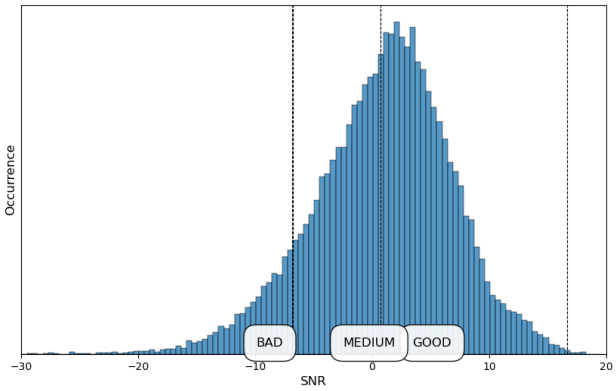

Figure 1: Histogram of SNR values with cluster boundaries (BAD, MEDIUM, GOOD).

where N_{ij} represents the number of observed transitions from state i to state j and $\sum_k N_{ik}$ is the total number of transitions from state i. This method allows for detailed modeling of channel transitions and forms the basis for predicting future channel states.

Additionally, the median SNR for each cluster was calculated to characterize the SNR quality in each state; the median SNR values provide a robust estimate of the central tendency and are less susceptible to outliers than mean values, allowing for a more accurate characterization of SNR quality. The median SNR values for the BAD, MEDIUM, and GOOD states were determined according to their distributions in the clusters.

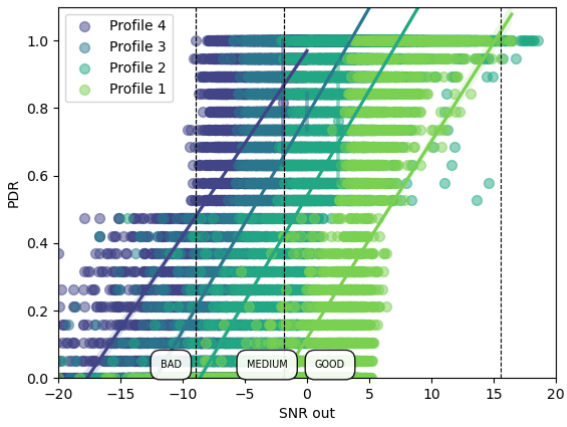

Figure 2: Relationship between SNR and PDR with cluster boundaries and trend lines.

Figure 2 visualizes the relationship between SNR and PDR. The four lines represent the regression lines for each profile of the FRSS modulation. This observation confirms the expected positive correlation between SNR and PDR, supporting the validity of our cluster classification. The cluster boundaries, as shown in Figure 1, are also

included to highlight the PDR variability within each cluster. Low PDR values predominantly characterize the BAD cluster, while the GOOD cluster typically has PDR values close to 1.0. The MEDIUM cluster encompasses a wide range of PDR values, indicating mixed channel quality. We highlight that the values were shifted to hide to correct values of the data while retaining the relationships, for non-disclosure reasons.

Furthermore, penalty factors were calculated to quantify the signal degradation in each state: we chose, for simplicity of analysis, to set the variance to 0 thus reducing the random variables $e(s)$ to constants, for each state. These factors are based on the differences in median SNR values, and represent the fading in the GOOD channel. The penalty factors quantify the additional signal degradation and were calculated for transitions between MEDIUM and GOOD, as well as BAD and GOOD.

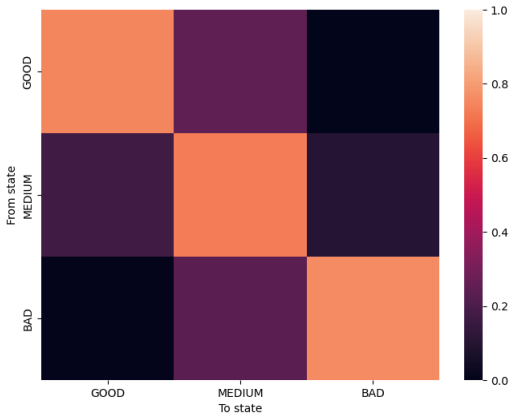

Figure 3: Heatmap of transition probabilities between states.

As a result of the evaluated data, Figure 3 presents a heatmap of the transition probabilities between the various states. This matrix shows the probabilities for transitions from one state to another, color-coded to highlight the transition probabilities visually. It is evident that the probability of remaining in the current state is approximately 70%, whereas the probability of directly skipping an intermediate state is nearly zero. This heatmap indicates that the channel tends to remain stable and that abrupt state changes are rare.

3 Results

In this section, we present the results of our simulations, which aim to replicate the conditions observed during the sea trials. We utilized our HMM-based implementation to generate results and compare them with the analysis of the sea trial data. These simulations were conducted over 24-hour periods to match the duration of the sea trials, and they were repeated multiple times with different random seed initializations to ensure robustness and reliability and ran the simulations over the four profiles of the modulations.

From our simulations, we expect to observe several key patterns that align with the empirical data from the sea trials. First, the transition probabilities from the simulations should converge to those observed in the sea trials, confirming the robustness of our HMM approach. Second, the SNR distributions across different states should closely match the sea trial data, indicating accurate modeling of underwater communication channel variability. Third, the relationship between SNR and PDR should show a similar trend to the sea trial data, validating the clustering approach and the impact of SNR on communication performance. Last, the penalties (fading) calculated from the simulation data should match those derived from the empirical data, demonstrating accurate quantification of environmental factors on signal quality.

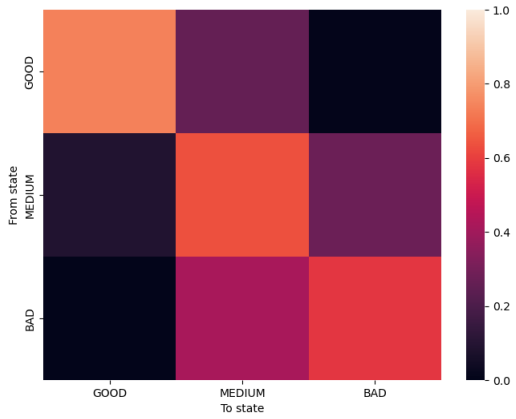

Figure 4: Heatmap of transition probabilities between states from simulations.

Figure 4 shows the heatmap of transition probabilities between the various states, derived from the simulation data. According to the law of large numbers, with sufficient simulation duration and repetitions, the heatmap should converge to the one observed from the sea trial data. This convergence is evident in our results, where both the transition probabilities and the SNR-PDR relationship (Fig. 5) closely mirror the empirical data. The probability of remaining in the same state p_{ii} remains approximately 60 - 70%, the transition probability between adjiacent states (i.e., $p_{g,m}$, $p_{m,g}$, $p_{m,b}$ and $p_{b,m}$) is about 20 - 30%, and the transition probability between GOOD and BAD states is nearly zero. This consistency between the simulation and the real-world data underscores the accuracy and reliability of our HMM approach.

Figure 5 depicts the relationship between SNR and PDR from the simulation data, including the cluster boundaries for BAD, MEDIUM, and GOOD. These clusters were chosen this time based on the simulation data also using the KMeans algorithm, resulting in clusters similarly distributed to those observed in the sea trial data. As in the empirical data, the PDR generally increases with rising SNR. The distribution of PDR values across the three clusters closely mirrors the distribution observed in the sea trial data. In the BAD cluster, low PDR values dominate, while the GOOD

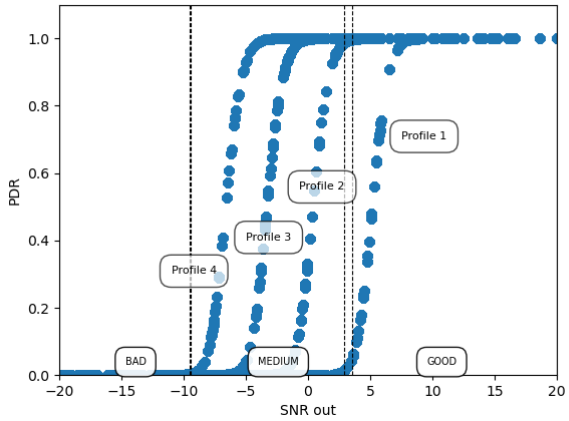

Figure 5: Relationship between SNR and PDR with cluster boundaries from simulations.

cluster is characterized by PDR values close to 1.0. The MEDIUM cluster again shows a mixed range of PDR values. In the figure, the same 4 curves associated with the 4 profiles of the chosen adaptive modulation, as in Figure 2, emerge more clearly. This is because the PDR was obtained using the logistic regression formulas presented in [3], directly from the obtained output SNR:

$$PDR = \frac{1}{[1 + e^{-k(SNR_{out} - x_0)}]}, \tag{5}$$

where x_0 is the value of a sigmoid midpoint, while k is the logistic growth.

4 Conclusions

These results demonstrate that our simulation model successfully replicates the dynamics observed in real-world underwater communication channels. The simulated channel includes the Urick-Thorp formula with additional fading, modeled with a Hidden Markov model. The transition probabilities and state distributions from the simulations reflect those seen in the empirical data, validating the effectiveness of our HMM approach. The consistency between the SNR-PDR relationship in the simulation and the sea trials further confirms the model's accuracy in capturing the behavior of underwater communication channels. This alignment of results not only validates our model but also enhances confidence in its predictive capabilities for future studies. The penalty values, which represent the signal attenuation due to fading, are particularly challenging to predict and must be calibrated against real-world data to ensure accuracy. Moving forward, future research will aim to automatically infer the penalties induced by fading and incorporate these into the propagation loss model.

Acknowledgments

This work was supported by the the Bundeswehr Technical Center for Ships and Naval Weapons, Maritime Technology and Research (WTD 71), Eckernforde, Germany, under the contract E/E71S/K1291/CF081, and the FSE REACT EU, PON Research and Innovation 2014-2020 (DM 1062/2021).

References

[1] Mohiuddin Ahmed, Raihan Seraj, and Syed Mohammed Shamsul Islam. 2020. The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics* 9, 8 (2020), 1295.
[2] Bellhop Ray Tracer. 2024. http://oalib.hlsresearch.com/Rays/ Last accessed: June 2024.
[3] Koen CH Blom, Henry S Dol, Frank Berning, and Paul A Van Walree. 2022. Development of a physical layer for adaptive underwater acoustic communications. In *2022 Sixth Underwater Communications and Networking Conference (UComms)*. IEEE, Lerici, Italy.
[4] Filippo Campagnaro, Roberto Francescon, Federico Guerra, Federico Favaro, Paolo Casari, Roee Diamant, and Michele Zorzi. 2016. The DESERT underwater framework v2: Improved capabilities and extension tools. In *2016 IEEE Third Underwater Communications and Networking Conference (UComms)*. IEEE.
[5] Filippo Campagnaro, Alberto Signori, Roald Otnes, Michael Goetz, Dimitri Sotnik, Arwid Komulainen, Ivor Nissen, Federico Favaro, Federico Guerra, and Michele Zorzi. 2021. A Simulation Framework for Smart Adaptive Long and Short-range Acoustic Networks. In *IEEE MTS OCEANS 2021: San Diego – Porto*.
[6] Filippo Campagnaro, Nicola Toffolo, and Michele Zorzi. 2022. Modeling Acoustic Channel Variability in Underwater Network Simulators from Real Field Experiment Data. *Electronics* 11, 14 (2022), 1–7. https://www.mdpi.com/2079-9292/11/14/2262
[7] Paolo Casari, Filippo Campagnaro, Elizaveta Dubrovinskaya, Roberto Francescon, Amir Dagan, Shlomo Dahan, Michele Zorzi, and Roee Diamant. 2021. ASUNA: A Topology Data Set for Underwater Network Emulation. *IEEE Journal of Oceanic Engineering* 46, 1 (2021), 307–318.
[8] Henry Dol. 2019. EDA-SALSA: Towards smart adaptive underwater acoustic networking. In *IEEE MTS OCEANS 2019 - Marseille*. Marseille, France.
[9] Ahmed Mahmood, Mandar Chitre, and Hari Vishnu. 2017. Locally Optimal Inspired Detection in Snapping Shrimp Noise. *IEEE Journal of Oceanic Engineering* 42, 4 (2017), 1049–1062.
[10] Nils Morozs, Wael Gorma, Benjamin T. Henson, Lu Shen, Paul D. Mitchell, and Yuriy V. Zakharov. 2020. Channel Modeling for Underwater Acoustic Network Simulation. *IEEE Access* 8 (2020), 136151–136175.
[11] Roald Otnes, Paul A van Walree, Helge Buen, and Heechun Song. 2015. Underwater acoustic network simulation with lookup tables from physical-layer replay. *IEEE Journal of Oceanic Engineering* 40, 4 (2015), 822–840.
[12] Roald Otnes, Paul A van Walree, and Trond Jenserud. 2013. Validation of replay-based underwater acoustic communication channel simulation. *IEEE Journal of Oceanic Engineering* 38, 4 (2013), 689–700.
[13] Alberto Signori, Filippo Campagnaro, Ivor Nissen, and Michele Zorzi. 2022. Channel-Based Trust Model for Security in Underwater Acoustic Networks. *IEEE Internet of Things Journal* 9, 20 (2022), 20479–20491.
[14] Milica Stojanovic. 2007. On the Relationship Between Capacity and Distance in an Underwater Acoustic Communication Channel. *ACM SIGMOBILE Mobile Computing and Communications Review (MC2R)* 11, 4 (October 2007), 34–43.
[15] Cristiano Tapparello, Paolo Casari, Giovanni Toso, Ivano Calabrese, Roald Otnes, Paul van Walree, Michael Goetz, Ivor Nissen, and Michele Zorzi. 2013. Performance evaluation of forwarding protocols for the RACUN network. In *Proceedings of the 8th International Conference on Underwater Networks & Systems* (Kaohsiung, Taiwan) *(WUWNet '13)*. Association for Computing Machinery, New York, NY, USA, Article 36, 8 pages.
[16] Beatrice Tomasi, Paolo Casari, Lorenzo Finesso, Giovanni Zappa, Kim McCoy, and Michele Zorzi. 2010. On modeling JANUS packet errors over a shallow water acoustic channel using Markov and hidden Markov models. In *2010-MILCOM 2010 Military Communications Conference*. IEEE, San Jose, CA, USA, 2406–2411.
[17] Robert J. Urick. 1983. *Principles of Underwater Sound* (3rd ed.). McGraw Hill, New York, US.

Received XX July 2024; revised XX July 2024; accepted XX July 2024