Statistical ON-OFF Link Modeling Based on Sea Trial Data

Nils Morozs
School of Physics, Engineering & Technology
University of York
York, YO10 5DD, UK
nils.morozs@york.ac.uk

Filippo Campagnaro
Department of Information Engineering
University of Padova
Padova, PD, Italy
filippo.campagnaro@unipd.it

Lu Shen
School of Physics, Engineering & Technology
University of York
York, YO10 5DD, UK

Benjamin T. Henson
School of Physics, Engineering & Technology
University of York
York, YO10 5DD, UK

Florian Mahieu
School of Physics, Engineering & Technology
University of York
York, YO10 5DD, UK

Yuriy Zakharov
School of Physics, Engineering & Technology
University of York
York, YO10 5DD, UK

Paul D. Mitchell
School of Physics, Engineering & Technology
University of York
York, YO10 5DD, UK

ABSTRACT

This paper presents a statistical ON-OFF link modeling approach for underwater acoustic networks (UANs) based on sea trial data. It aims to enable computationally efficient UAN simulation models that capture the complex temporal characteristics of underwater acoustic links. The main idea of the proposed method is to synthesize realistic link availability and outage patterns from empirical cumulative distribution functions (CDFs) derived from sea trial data. The Werbellin lake experiment dataset from ASUNA is used as a case study, representative of short range shallow water environments. In this dataset, weak correlation between link quality and distance and weak cross-correlation between different links allows us to model each link as an independent random process. However, we also propose a way of extending this method to generate multiple CDFs representing different link types, distances, node depths etc., all exhibiting different link statistics. The proposed statistical approach provides UAN researchers with a valuable tool for more realistic and efficient network simulation, supporting the development and evaluation of UAN protocols and systems. Additionally, it offers the potential to generate reproducible benchmark test environments for standardized protocol design evaluation.

KEYWORDS

channel modeling, simulation, underwater acoustic network

WUWNET '24, October 28–31, 2024, Sibenik, Croatia

ACM ISBN 979-8-4007-1160-2/24/10.
https://doi.org/10.1145/3699432.3699468

ACM Reference Format:
Nils Morozs, Filippo Campagnaro, Lu Shen, Benjamin T. Henson, Florian Mahieu, Yuriy Zakharov, and Paul D. Mitchell. 2024. Statistical ON-OFF Link Modeling Based on Sea Trial Data. In *The 18th ACM International Conference on Underwater Networks & Systems (WUWNET '24), October 28–31, 2024, Sibenik, Croatia.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3699432.3699468

1 INTRODUCTION

Underwater acoustic networks (UANs) have a wide range of applications: water quality monitoring (e.g. in fish farms) [8], seismic monitoring [13], marine animal tracking [4], off-shore asset monitoring [1], coastal defense [17], etc. UAN protocol design is a challenging task due to the characteristics of the underwater acoustic (UWA) communication medium, such as slow propagation, low bandwidth and highly variable bit error and frame error rates often encountered at sea.

The development, testing and validation of UAN protocols involve two principal steps: simulations and sea experiments. In addition to circumventing the high cost and logistical challenges involved in performing sea experiments, the major advantage of simulation-based studies is that they enable researchers to test their network protocols under controlled, reproducible conditions, and obtain more comprehensive, statistically valid results, e.g. via parameter sweeps, Monte Carlo simulations etc. In contrast, implementing and testing the network protocols at sea is more suitable as a validation step to demonstrate that they work in a real deployment. It is usually not logistically feasible at sea to perform parameter sweeps, benchmark comparisons, and obtain large statistical samples of the network protocol performance. Instead, a UAN sea experiment is usually a demonstration of the network operating in a specific environment. Therefore, simulation is of particular importance in performing a thorough empirical evaluation.

One of the key challenges in developing a credible network simulation model is a realistic representation of the UWA channel. Therefore, significant efforts have been made to replicate in simulation the channels observed at sea. For example, detailed measurement campaigns have taken place to record the data on bathymetry, seabed composition and sound speed profile (SSP) [5, 20] and subsequently use BELLHOP beam tracing [19] to replicate the channels from such sea experiments. Alternatively, channel impulse responses (CIRs) recorded in sea trials, e.g. such as those included in the Watermark benchmark [22], can be directly used in UWA signal processing simulation experiments. However, this approach limits the simulation studies to only using the exact transmitter/receiver locations from the sea trials. In either case, fully modelling the interaction between the CIR and the transmit/receive signal processing chains is prohibitively computationally expensive in network simulation research, where potentially hundreds of time-varying UWA links need to be modelled.

An alternative, much more computationally efficient approach for UAN modelling is to directly import the data on packet success rates (PSRs) from UAN sea deployments, such as the MISSION 2012 [11] and MISSION 2013 [10] experiments carried out by Chitre et al. More recently, Casari et al. [9] published the ASUNA database of UAN sea experiments that includes several metrics, including the link connectivity matrices among the network nodes, based on the transmitted and received packet logs at every node. The key advantage of this approach is that the complex interaction between the UWA channel and the receiver signal processing chain is implicitly included in this data, i.e. the data provides the output of this process. In both of these examples [9, 10], the authors propose a method of using this sea trial data to derive a time-varying PSR for every link, which can then be used to randomly generate packet errors in network simulations. The same procedure has been applied for underwater network simulations in the EDA SALSA project, as explained in [6].

The approach proposed in this paper focuses on the temporal characteristics of random packet loss observed in real sea environments. For example, if the measured PSR on a communication link within a given time window is 70%, do 30% of packets get lost as independent random events, or is there significant time correlation, e.g. for a period of time (e,g. 1 minute) most/all packets get through, and then for a another period of time no packets get through, and so on? The typical behaviour observed in our previous lake and sea trials [16] and the data recorded in ASUNA [9] strongly suggest the latter. For example, this effect can be modelled analytically using a two-state Markov model proposed by Zorzi et al. [23], where a UWA link can be in two states: 1) "ON" – packets get received successfully, and 2) "OFF" – packets fail to be detected/decoded successfully. This model results in a link behaviour, where the duration for which the link stays in an ON or OFF state is random and exponentially distributed. Campagnaro et al. [7] extend this approach to a three-state Markov model, where a link can be in a "good", "medium" and "bad" state, each characterised by a PSR or bit error rate (BER) range. The transition probabilities are computed to fit the real sea experiment data from the ASUNA dataset, and the paper shows a close match between the PSR in simulation and the PSR observed at sea.

The purpose of this paper is to propose a method of deriving a statistical model for the link ON/OFF state switching directly from sea trial data. As such, it follows the same philosophy as the Markov Model model approach [7] described above. The key difference of the methodology proposed in this paper is that it does not attempt to fit any distribution to the data, e.g. exponential as required in Markov models, but instead uses the empirical cumulative distribution function (CDF) of the sea trial data to generate UWA link behaviour. We use the data from the Werbellin lake trials in Northern Germany in June 2016, publicly available in the ASUNA database [9], as a case study in this paper.

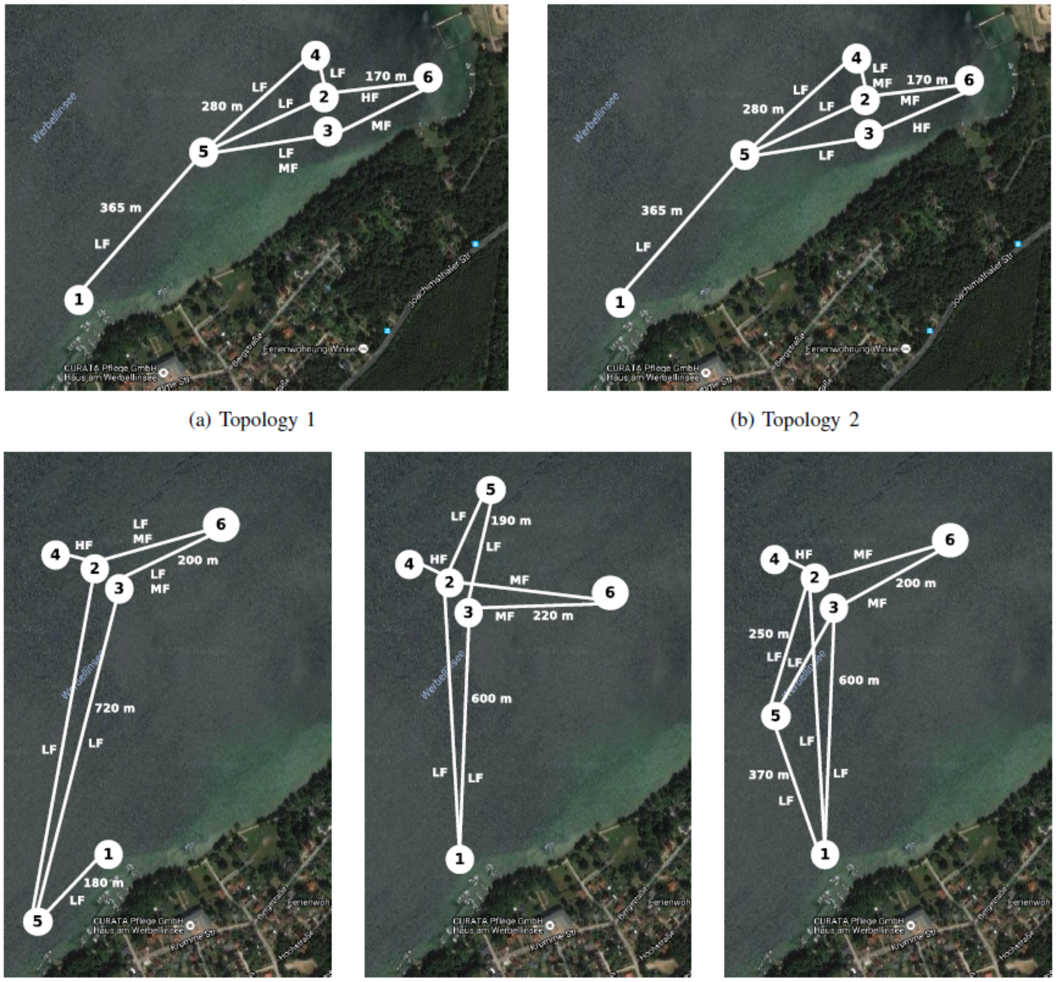
The structure of the paper is as follows: Section 2 describes the performance of UWA links typically observed in practice, using the data from Werbellin lake trials as an example; Section 3 shows how we use this data to derive a statistical channel model for network simulations; in Section 4 we validate our approach and show example link realisations obtained from it; finally, Section 5 concludes the paper.

2 UWA LINK PERFORMANCE AT SEA

UWA communication channels are typically highly time-varying due to several factors: significant changes in the channel multipath structure and Doppler distortion caused by transmitter/receiver movement and random sea surface motion, spatial and temporal changes in acoustic noise, sound speed profile, changes in the orientation of acoustic modems with a non-uniform beam pattern, etc. [21]. As a result, the typical performance of a UWA communication link can be characterized by sporadic and often very high packet loss. Detailed modeling of the physical properties of a UWA channel described above is not feasible for large-scale network simulation studies, which are crucial for the design and verification of network protocols. Furthermore, we argue that modeling the channel alone is insufficient to obtain accurate and realistic estimates of link-level performance; the receiver signal processing chain also needs to be modeled, as different modulation, coding, channel estimation, equalization, etc., algorithms will perform very differently in a given UWA channel.

To circumvent the need to model all the physical (PHY) layer properties discussed above, this paper investigates using link-level sea trial data to derive statistical models of link performance that provide realistic temporal packet loss patterns, in line with those often observed at sea. In this way, practical PHY layer performance is implicitly modelled within these link statistics, albeit specific to the modems used in these sea trials.

We use an ASUNA [9] dataset from the Werbellin lake trials, which took place in June 2016 in Northern Germany, as an example case study in this paper. It involved five Evologics S2C 18-34 ("LF" – low frequency), three S2C 48-78 (("MF" – medium frequency)), and two S2CM HS ("HF" – high frequency) acoustic modems deployed in topologies shown in Figure 1. There was a total of six nodes, some of which were equipped with more than one modem (operating in different bands). The majority of the modems were deployed at 10 m depth, except Node 5 in Topologies 1 and 3 (5 m depth); and Node 3 (3 m depth in all topologies). Each topology was deployed for 10 minutes, with every modem logging received packets at 1-second resolution.



(a) Topology 1
(b) Topology 2
(c) Topology 3
(d) Topology 4
(e) Topology 5

Figure 1: Network topologies deployed in the Werbellin lake trials, Northern Germany, Jun 2016. (Figure reproduced from [9]).

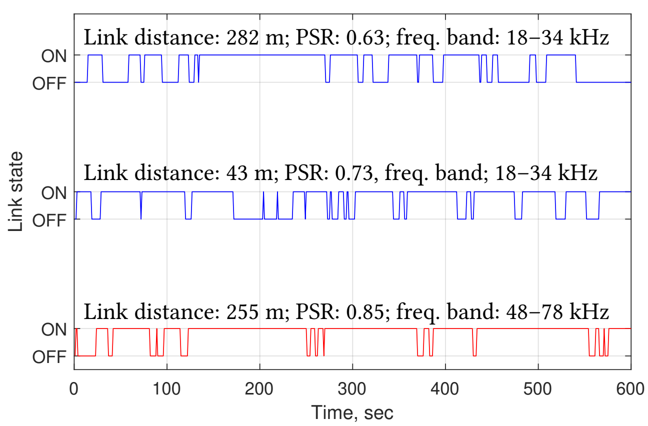


Figure 2: Examples of link ON-OFF state data recorded in the Werbellin lake trials. PSR – packet success rate.

The resulting data from the modem logs can be visualised as shown in Figure 2. For every pair of transmitter/receiver locations, a timeline of the link being in the "ON" and "OFF" state – i.e., when a packet gets through and when a packet gets lost, respectively – can be plotted for the duration that this topology is deployed. This provides a realistic picture of UWA link-level performance that *implicitly* includes the impact of all environmental factors present in that environment and their effect on signal processing performance.

For example, such data enables UAN protocol developers to exercise their protocols in a direct “replay" of the link-level performance observed in a real environment. However, a significant limitation of the replay approach is that such tests are limited to the exact number of nodes, topologies, and duration that were used in the trials. For example, in the Werbellin lake dataset, the topologies under test would be limited to five nodes, as this was the number of S2C 18-34 modems deployed in this trial. A much more powerful approach is to generalize this data to a statistical model that can then be used to simulate networks of arbitrary size, topology, and duration of deployment. In the next section, we propose a method of doing this.

3 STATISTICAL LINK MODELING

Figure 2 shows three examples of link ON/OFF state timelines from the the ASUNA Werbellin lake dataset. In the rest of this section, we firstly examine whether there are any significant correlations that need to be taken into account, and then propose a methodology for processing such ON/OFF link state data to derive statistical link models for the purpose of network simulation.

3.1 Correlation between link quality and distance

The *packet success rate (PSR)* can be estimated from the ON/OFF state data for link i as follows:

$$PSR_{\mathrm{i}} = \frac{N^{\mathrm{i}}_{\text{samples: on}}}{N^{\mathrm{i}}_{\text{samples}}} \tag{1}$$

where $N^{\mathrm{i}}_{\text{samples: on}}$ is the number of state samples where link i was ON, and $N^{\mathrm{i}}_{\text{samples}}$ is the total number of link state samples (e.g. in our case: 600) for this link. PSR provides an estimate of the link quality, i.e. the proportion of time this link can support a successful transmission.

Figure 3 plots the PSR for every link in the 18–34 kHz band from the ASUNA Werbellin lake trial dataset against the transmitter-receiver distance. Interestingly, it shows that there is very little correlation between the link distance and quality. This is in line with our observations from past trials [16], especially in shallow, relatively short-range environments, where a key factor for the physical layer performance is strong multipath self-interference, rather than attenuation of the main path with distance. This is an especially important consideration for routing protocol design; the data clearly shows that any routing strategies based on node locations and link distances are not preferable in such environments.

The key conclusion from the point of view of modelling this communication environment is that the link distance does not need to be considered as a key parameter to synthesise link ON/OFF state patterns in a simulation model, e.g. the same underlying statistical distribution can be used to simulate both shorter and longer range links. However, it may not be the case in other environments. For example, if a similar experiment was repeated in open deep sea, there would likely be some correlation between the link distance and PSR. In this case, multiple different statistical distributions can be derived for different communication ranges.

3.2 Spatial and temporal correlation

Next, we examine if it is a reasonable approximation to model every UWA link as an independent random process. Figure 4 plots the explained variance – the square of the correlation coefficient – between every pair of links in every deployed topology. Each point is plotted against the distance between the receivers for the given pair of links, and grouped by: a) whether the two links have the same receiver, b) the same transmitter, c) they are reverse links (i.e., the links between the same two nodes), d) these links involve four different nodes. The data shows generally very low correlation between most pairs of links. There are a few data points with stronger correlation between some links with the same transmitter, but this is not sufficiently consistent to draw significant conclusions from them. This data suggests that the main source of the link ON/OFF state variability is likely the time varying channel structure, which is unique for every link. Therefore, it is reasonable to model this environment using an independent random process for every link.

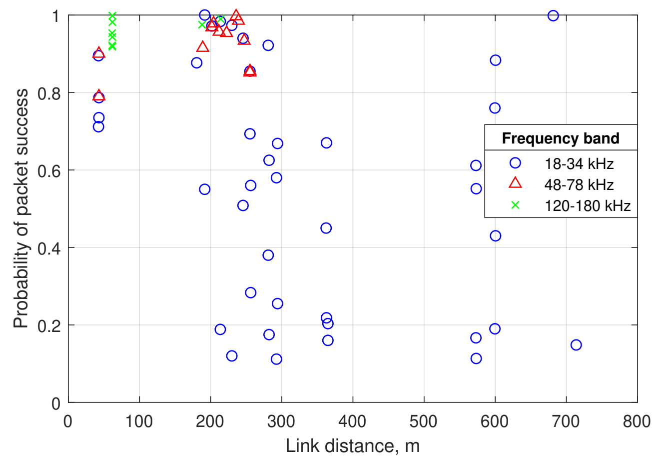


Figure 3: Packet success rates against link distance observed in the Werbellin lake trials, using three types of Evologics modem operating in the: 18–34 kHz band, 48–78 kHz band, and 120–180 kHz band.

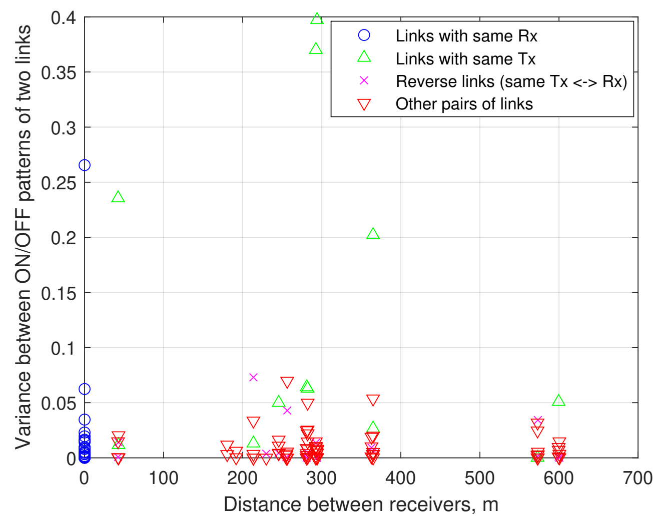


Figure 4: Weak correlation was found among different link ON/OFF state patterns over time, suggesting that each link can be modelled using an independent random process.

3.3 Statistical Link Model Derivation

The main objective of the model proposed in this paper is to simulate link ON/OFF state behaviour with the same statistics as that observed in sea trials, e.g. depicted in Figure 2. For that purpose, we first derive empirical cumulative distribution functions (CDFs) of the ON state and OFF state duration as follows.

Let $\boldsymbol{s}_\mathbf{i}$ be the vector of link state observations for a given link i in the sea trial data, where $s_\mathrm{i}[n] = 1$ if the n^th link state sample is ON, and $s_\mathrm{i}[n] = 0$ if it is OFF.

First, a vector of changes in the link state can be computed as:

$$c_\mathrm{i}[n] = s_\mathrm{i}[n+1] - s_\mathrm{i}[n], \text{ for } n = 1, 2, .., N^\mathrm{i}_\mathrm{samples} - 1, \tag{2}$$

where $c_\mathrm{i}[n] = 0$ if there is no change of link state, $c_\mathrm{i}[n] = 1$ if the state changed from OFF to ON, and $c_\mathrm{i}[n] = -1$ if the state changed from ON to OFF. Note that all the non-zero elements in $\boldsymbol{c}_\mathrm{i}$ are strictly alternating between -1 and 1, i.e. if the previous transition between states was ON→OFF, then the following one can only be OFF→ON.

Next, the times at which those link state changes occur can be computed as:

$$\boldsymbol{\tau}_\mathrm{i} = \left\{ nT_\mathrm{s} \,\middle|\, c_\mathrm{i}[n] \neq 0, \text{ for } n = 1, 2, .., N^\mathrm{i}_\mathrm{samples} - 1 \right\}, \tag{3}$$

where $\boldsymbol{\tau}_\mathrm{i}$ is a vector containing the times at which the state of link i changed (in seconds), and T_s is the link state sample period.

Now, a vector $\boldsymbol{t}^\mathrm{i}_\mathrm{on}$ that contains the duration of every ON state for link i can be computed as:

$$\boldsymbol{t}^\mathrm{i}_\mathrm{on} = \begin{cases} \left\{ \tau_\mathrm{i}[2k] - \tau_\mathrm{i}[2k-1] \,\middle|\, k = 1, 2, .., \left\lfloor \frac{K}{2} \right\rfloor \right\}, & s_\mathrm{i}[1] = 0 \\ \left\{ \tau_\mathrm{i}[2k+1] - \tau_\mathrm{i}[2k] \,\middle|\, k = 1, 2, .., \left\lfloor \frac{K-1}{2} \right\rfloor \right\}, & s_\mathrm{i}[1] = 1 \end{cases} \tag{4}$$

where K is the number of elements in vector $\boldsymbol{\tau}_\mathrm{i}$. Similarly, a vector $\boldsymbol{t}^\mathrm{i}_\mathrm{off}$ of OFF state durations for link i can be computed as:

$$\boldsymbol{t}^\mathrm{i}_\mathrm{off} = \begin{cases} \left\{ \tau_\mathrm{i}[2k+1] - \tau_\mathrm{i}[2k] \,\middle|\, k = 1, 2, .., \left\lfloor \frac{K-1}{2} \right\rfloor \right\}, & s_\mathrm{i}[1] = 0 \\ \left\{ \tau_\mathrm{i}[2k] - \tau_\mathrm{i}[2k-1] \,\middle|\, k = 1, 2, .., \left\lfloor \frac{K}{2} \right\rfloor \right\}, & s_\mathrm{i}[1] = 1 \end{cases} \tag{5}$$

Having calculated the vectors of ON and OFF state durations for every link in the dataset as described above, we can now derive an empirical CDF for the ON state duration as follows. First, we combine and sort the ON state duration data for all links in the data set:

$$\boldsymbol{t}^\mathrm{all}_\mathrm{on} = \mathrm{sort}\left(\left\{ t^\mathrm{i}_\mathrm{on}[m] \,\middle|\, \forall i, m \right\} \right) \tag{6}$$

where $\boldsymbol{t}^\mathrm{all}_\mathrm{on}$ is a sorted vector of ON state durations for all links. We also generate a vector of percentiles $\boldsymbol{p}_\mathrm{on}$:

$$\boldsymbol{p}_\mathrm{on} = \left\{ \frac{j}{N_\mathrm{t:on}} \,\middle|\, \text{for } j = 1, 2, .., N_\mathrm{t:on} \right\}, \tag{7}$$

such that $t^\mathrm{all}_\mathrm{on}[j]$ is the $p_\mathrm{on}[j]$'th percentile of the CDF; $N_\mathrm{t:on}$ is the number of elements in $\boldsymbol{t}^\mathrm{all}_\mathrm{on}$.

Similarly, we derive the sorted vector of all OFF state durations and their corresponding percentile values:

$$\boldsymbol{t}^\mathrm{all}_\mathrm{off} = \mathrm{sort}\left(\left\{ t^\mathrm{i}_\mathrm{off}[m] \,\middle|\, \forall i, m \right\} \right) \tag{8}$$

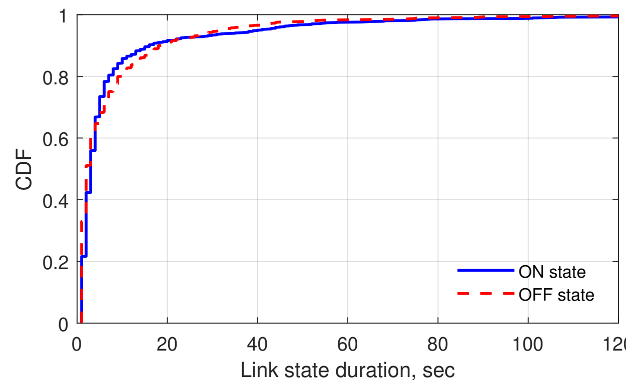


Figure 5: Cumulative distribution functions (CDFs) of the link ON/OFF state duration.

$$\boldsymbol{p}_\mathrm{off} = \left\{ \frac{j}{N_\mathrm{t:off}} \,\middle|\, \text{for } j = 1, 2, .., N_\mathrm{t:off} \right\}, \tag{9}$$

where $N_\mathrm{t:off}$ is the number of elements in $\boldsymbol{t}^\mathrm{all}_\mathrm{off}$.

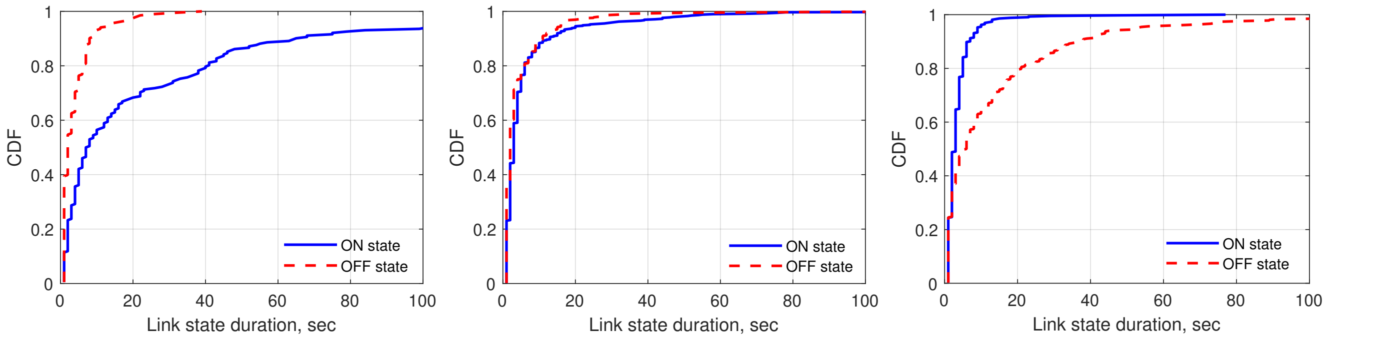
Figure 5 plots the resulting empirical CDFs, computed using the method described above using the ASUNA Werbellin lake dataset. As expected, both of these CDFs follow a long-tailed distribution. We found that these CDFs fit a Pareto distribution significantly better than the negative exponential distribution (e.g. as assumed in theoretical Markov models [23]), but the general shape is similar. Furthermore, one of the key features of the approach proposed here is that it is not necessary to fit any particular distribution, instead the empirical CDFs computed from sea trial data are used directly to synthesize link ON/OFF state behaviour.

3.4 Modeling multiple types of link

In Subsections 3.1 and 3.2, we established that there was little correlation between the link quality and distance, and little spatial or temporal correlation between different links. Therefore, in the context of the Werbellin lake environment (i.e. relatively short range and shallow water), it is a reasonable approximation to model the ON/OFF state switching on every link as an independent random process, regardless of the distance between transmitter and receiver. However, for other environments, e.g. open deep sea, there is likely to be some correlation between the link distance and quality. In those cases, the approach described above can be extended to derive empirical CDF models for multiple distance ranges, thus reflecting different link statistics observed at different communication distances. Similarly, this approach can be extended to derive multiple CDF models classified by different parameters, e.g. node depth, packet duration etc.

In this section, we provide a different example; we filter the links from the Werbellin lake dataset into three categories: "poor", "average" and "good" – with PSR below 0.4, between 0.4 and 0.7, and above 0.7, respectively. Each of these link types follows their own statistical distribution derived from sea trial data.

Instead of a single duration vector for all links in the dataset, the ON state duration data is placed into three vectors, $\boldsymbol{t}^\mathrm{poor}_\mathrm{on}$, $\boldsymbol{t}^\mathrm{avg}_\mathrm{on}$ and $\boldsymbol{t}^\mathrm{good}_\mathrm{on}$, containing filtered data for the "poor", "average" and "good"



(a) "Good" links: PSR $\geq$ 0.7

(b) "Average" links: PSR $\in$ [0.4, 0.7)

(c) "Poor" links: PSR < 0.4

Figure 6: CDFs of the link ON/OFF state duration, grouped by the link quality – "good", "average" and "poor" – based on the packet success rates (PSR).

links, respectively:

$$t_{\text{on}}^{\text{poor}} = \text{sort}\left(\left\{t_{\text{on}}^{\text{i}}[m] \,\middle|\, \forall i, m\,,\ PSR_{\text{i}} < 0.4\right\}\right) \tag{10}$$

$$t_{\text{on}}^{\text{avg}} = \text{sort}\left(\left\{t_{\text{on}}^{\text{i}}[m] \,\middle|\, \forall i, m\,,\ 0.4 \leq PSR_{\text{i}} < 0.7\right\}\right) \tag{11}$$

$$t_{\text{on}}^{\text{good}} = \text{sort}\left(\left\{t_{\text{on}}^{\text{i}}[m] \,\middle|\, \forall i, m\,,\ PSR_{\text{i}} \geq 0.7\right\}\right) \tag{12}$$

The boundaries between what constitutes a "poor", "average" or "good" link here were based in reasonable practical expectations, as an example. The data can also be split in the same way using different threshold values, into more categories, or using other criteria.

The OFF state duration data is split into three vectors in the same way as the ON data, and the percentile value vectors for all six sorted ON/OFF duration vectors are calculated using the method in (9). The resulting empirical CDFs for these three types of link are depicted in Figure 6.

3.5 Dequantization

The last step in our proposed method is to *dequantize* the empirical CDFs which, in their discrete form, have a step size of T_s. Figure 7 shows an example of this, where linear interpolation is applied between the centre points of the steps in the discrete CDFs from Figure 6a to yield continuous CDFs interpolated, in this case, at 0.1 percentile resolution. Note that the state duration data was not extrapolated below the link state sampling period of $T_s = 1$ second, as this would involve an assumption on the statistics that were not captured in the data (due to the sampling period limitation).

Finally, these continuous CDFs for every type of link q can be stored as a quadruple $(\boldsymbol{t}_{\text{on}}^{\text{q}}, \boldsymbol{p}_{\text{on}}^{\text{q}}, \boldsymbol{t}_{\text{off}}^{\text{q}}, \boldsymbol{p}_{\text{off}}^{\text{q}})$, e.g. containing the value of the CDF at every 0.1 percentile point. These can then be used to generate random link realisations in network simulations. An example of how it can be done is given in the next section.

4 EXAMPLE USE AND MODEL VALIDATION

The CDF models of link ON/OFF state duration derived in the previous section can be used to simulate UWA communication

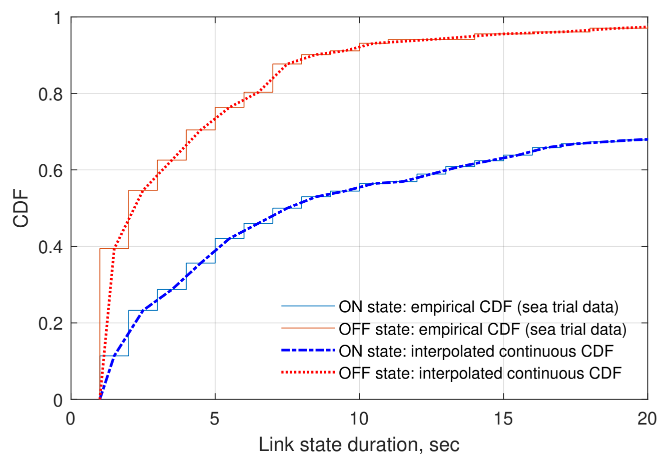


Figure 7: The discrete CDF of the sea trial data can be interpolated to provide a continuous CDF model of the link ON/OFF state duration.

links as shown in Algorithm 1. It iterates over every link in the network and synthesizes a list of ON/OFF state switch events for each of them according to the CDF derived from the sea trial data.

Figure 8 shows examples of synthesized link timelines using this approach. The random realizations of "good" links are characterized by long periods of ON state, i.e., when packets can be successfully transmitted across them, with link outages (OFF state) occurring at random times and for random durations. In comparison, the "average" link realizations are characterized by much more frequent switching between the ON and OFF states, which is consistent with the CDF model shown in Figure 6b, where the mean ON state duration is much shorter than for "good" links (Figure 6a). Lastly, the "poor" link realizations are characterized by longer periods of OFF time with even shorter bursts of ON time, again consistent with the CDF model in Figure 6c.

Finally, to verify that the link ON/OFF state patterns synthesized using this method are consistent with the empirical CDF model derived from sea trial data, Figure 9 shows a very close match

Algorithm 1 Generating random link realisations from the proposed statistical ON/OFF link model.

Set maximum duration of the simulation τ_{sim}
Initialise list of link state switching events L
for every node in the network i **do**
  for every node in the network j **do**
    if $i \neq j$ **then**
      Select statistical model to use for this link
      Set initial state s_{ij} of link $i \longrightarrow j$ at random
      Initialise previous event time $t_{\text{prev}} = 0$
      while $t_{\text{prev}} \leq \tau_{\text{sim}}$ **do**
        Draw r from uniform rand. distr. between 0 and 1
        if $s_{\text{ij}} = 0$ **then**
          Fetch CDF for the OFF state duration
        else
          Fetch CDF for the ON state duration
        end if
        Calculate the r^{th} percentile of the CDF: τ_{dur}
        Next state switch event time is $t = t_{\text{prev}} + \tau_{\text{dur}}$
        Add state switch event at t for link $i \longrightarrow j$ to list
        $t_{\text{prev}} = t$
      end while
    end if
  end for
end for

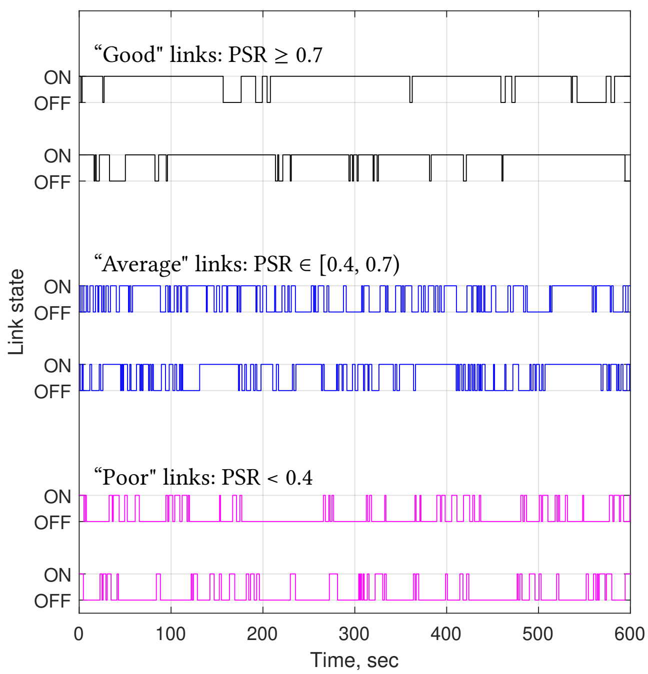


Figure 8: Examples of random link realisations of different qualities ("good", "average", "poor") that follow the ON/OFF state duration distributions derived from lake trial data.

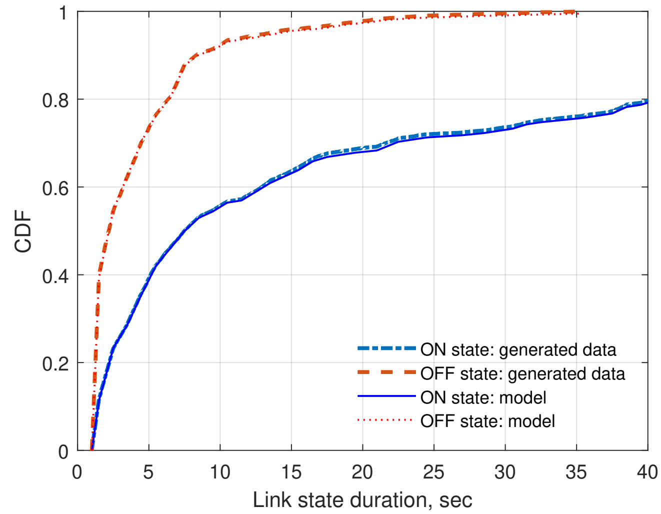


Figure 9: Comparison the synthetic link ON/OFF state duration data with the underlying statistical model shows a close match between the desired and the obtained CDF.

between the CDFs of 10,000 randomly generated ON/OFF state durations and their underlying CDF models. A similarly close match was observed when testing other types of links and CDF models.

4.1 Discussion

The key consideration for the statistical model proposed in this paper is to replicate the temporal link level behaviour observed in real UAN deployments. It can have a significant impact on protocol performance in simulation, and is crucial in identifying and fixing bugs and vulnerabilities in protocol logic. For example, for UANs that employ Automatic Repeat Request (ARQ) [2] at the Link Layer, i.e. acknowledgements (ACKs) and retransmissions, this model will provide significantly different (and more realistic) results than more simplistic models that treat packet loss as independent random events. In the latter case, allowing the source node to retransmit the packet multiple times will lead to successful packet delivery a vast majority of the time: with a probability $P(success) = 1 - (1 - PSR)^{N_{\text{tries}}}$, to be precise, where N_{tries} is the maximum number of ARQ tries. Whereas in a real environment random link outage occurs for a time period, that may potentially last longer than the time it takes to retransmit a packet multiple times. For example, if a link is in the OFF state for 20 seconds, retransmitting the same packet five times will not yield a successful packet delivery.

The effect described above can have a profound impact on higher layer protocols, e.g. routing [14] or network discovery [18], as learnt by us in our previous lake and sea trials [15, 16]. Temporary link outage can lead to failures in multi-hop routing or unsuccessful network setup procedures, e.g. if a node needs to be assigned an address, a time slot for transmission, or a master node for routing, etc. In those cases, it is crucial to design protocols that can adapt to such link outage, and the proposed statistical link model can facilitate this.

A key phenomenon that this model does not address is *inter-node interference* in networks, i.e. what happens if transmissions from two or more nodes overlap at the receiver. This will require the implementation of additional logic in the network simulation model. In the simplest (pessimistic) case, if any transmissions overlap in time at the receiver, all packets in question are dropped. Alternatively, if the receiver has successive interference cancellation capabilities [12] or spread spectrum processing gain [3], the network simulation model can pass through at least one of the packets in question, e.g. the first one arriving at the receiver or the one with the stronger received signal strength.

5 CONCLUSIONS AND FURTHER WORK

We presented a novel approach to statistical ON-OFF link modeling for UANs based on sea trial data, that enables computationally efficient UAN simulation modelling that captures the complex temporal characteristics of UWA links. The main idea of this method is to synthesize realistic link availability and outage patterns from empirical CDFs of ON and OFF state durations obtained directly from sea trial data, avoiding the need to fit analytical distributions to the data. We provided a detailed description of how sea trial data from acoustic modem logs can be converted to empirical CDFs of the ON and OFF state durations, and how those CDFs can be used to synthesize realistic link behaviour in network simulations.

The Werbellin lake experiment dataset available in ASUNA was used as a case study in this paper. In this dataset, we found a weak correlation between link quality and distance, and weak cross-correlation between different links. This is representative of shorter range, shallow water environments, and allowed us to model each link as an independent random process. However, in future work on other sea trial datasets, there may be stronger correlation between the link ON/OFF state statistics and environmental factors such as link distance, node depths, etc. In those cases, our proposed model can be extended to generate any required number of CDFs, each representing the statistics of different links. The example we presented in the paper is to filter the data for "good", "average", and "poor" links based on their PSR.

In conclusion, this statistical ON-OFF link modeling approach provides a valuable tool for UAN researchers, enabling more realistic and efficient simulations to support the development and evaluation of UAN protocols and systems. It also has the potential to generate easily reproducible benchmark test environments, thus enabling standardized test cases for new protocol design.

ACKNOWLEDGMENTS

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) through the COUSIN project grant (EP/V009591/1) and the FSE REACT EU, PON Research and Innovation 2014-2020 (DM 1062/2021).

REFERENCES

[1] S. Ali, A. Ashraf, S. B. Qaisar, M. Kamran Afridi, H. Saeed, S. Rashid, E. A. Felemban, and A. A. Sheikh. 2018. SimpliMote: A Wireless Sensor Network Monitoring Platform for Oil and Gas Pipelines. *IEEE Syst. J.* 12, 1 (2018), 778–789. https://doi.org/10.1109/JSYST.2016.2597171

[2] S. Azad, P. Casari, and M. Zorzi. 2013. The underwater selective repeat error control protocol for multiuser acoustic networks: Design and parameter optimization. *IEEE Trans. Wireless Commun.* 12, 10 (2013), 4866–4877.

[3] C. Bernard, P.-J. Bouvet, A. Pottier, and P. Forjonel. 2021. Multiple access acoustic communication in underwater mobile networks. In *Proceedings of UComms'21*. IEEE, 1–4.

[4] B.J. Boom, Jiyin He, S. Palazzo, P.X. Huang, C. Beyan, Hsiu-Mei Chou, Fang-Pang Lin, C. Spampinato, and R.B. Fisher. 2014. A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage. *Ecol. Inform.* 23, Supplement C (2014), 83–97.

[5] A. Caiti, K. Grythe, J. M. Hovem, S. M. Jesus, A. Lie, A. Munafò, T. A. Reinen, A. Silva, and F. Zabel. 2013. Linking Acoustic Communications and Network Performance: Integration and Experimentation of an Underwater Acoustic Network. *IEEE J. Ocean. Eng.* 38, 4 (2013), 758–771.

[6] 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*. https://doi.org/10.23919/OCEANS44145.2021.9706058

[7] 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). https://doi.org/10.3390/electronics11142262

[8] G. Cario, A. Casavola, P. Gjanci, M. Lupia, C. Petrioli, and D. Spaccini. 2017. Long lasting underwater wireless sensors network for water quality monitoring in fish farms. In *Proc. of IEEE OCEANS'17*. 1–6.

[9] P. Casari, F> Campagnaro, E. Dubrovinskaya, R. Francescon, A. Dagan, S. Dahan, M. Zorzi, and R. Diamant. 2021. ASUNA: A Topology Data Set for Underwater Network Emulation. *IEEE Journal of Oceanic Engineering* 46, 1 (2021), 307–318. https://doi.org/10.1109/JOE.2020.2968104

[10] M. Chitre and G. Chua. 2014. Modeling realistic underwater acoustic networks using experimental data. In *Proc. of the Asilomar Conference on Signals, Systems and Computers*. 39–43.

[11] M. Chitre, I. Topor, R. Bhatnagar, and V. Pallayil. 2013. Variability in link performance of an underwater acoustic network. In *Proc. of IEEE OCEANS'13*. 1–7.

[12] S. E> Cho, H. C. Song, and W. S. Hodgkiss. 2011. Successive interference cancellation for underwater acoustic communications. *IEEE Journal of Oceanic Engineering* 36, 4 (2011), 490–501.

[13] A. K. Mohapatra, N. Gautam, and R. L. Gibson. 2013. Combined Routing and Node Replacement in Energy-Efficient Underwater Sensor Networks for Seismic Monitoring. *IEEE J. Ocean. Eng.* 38, 1 (2013), 80–90.

[14] N. Morozs, P.D. Mitchell, and Y. Zakharov. 2019. Dual-Hop TDA-MAC and Routing for Underwater Acoustic Sensor Networks. *IEEE J. Oceanic. Eng.* 44, 4 (2019), 865–880.

[15] N. Morozs, P.D. Mitchell, Y. Zakharov, R. Mourya, Y.R. Petillot, T. Gibney, M. Dragone, B. Sherlock, J.A. Neasham, C.C. Tsimenidis, M.E. Sayed, A.C. McConnell, S. Aracri, and A.A. Stokes. 2018. Robust TDA-MAC for Practical Underwater Sensor Network Deployment: Lessons from USMART Sea Trials. In *Proc. of ACM WUWNet'18*. Article 11.

[16] N. Morozs, B. Sherlock, B. T. Henson, J. A. Neasham, P. D. Mitchell, and Y. Zakharov. 2022. Data Gathering in UWA Sensor Networks: Practical Considerations and Lessons from Sea Trials. *J. Mar. Sci. Eng.* 10, 9 (2022).

[17] S. Palisetti, B. R. Chandavarkar, and A. V. Gadagkar. 2021. Intrusion Detection of Sinkhole Attack in Underwater Acoustic Sensor Networks. In *Proc. of ICCCNT'21*. 1–7. https://doi.org/10.1109/ICCCNT51525.2021.9580148

[18] R. Petroccia. 2021. DIVE: A distributed ID assignment and topology discovery protocol for underwater acoustic networks. *Ad Hoc Networks* 122 (2021), 102610. https://doi.org/10.1016/j.adhoc.2021.102610

[19] M.B. Porter. 2019. Beam tracing for two- and three-dimensional problems in ocean acoustics. *J. Acoust. Soc. Am.* 146, 3 (2019), 2016–2029.

[20] B. Tomasi, G. Zappa, K. McCoy, P. Casari, and M. Zorzi. 2010. Experimental study of the space-time properties of acoustic channels for underwater communications. In *OCEANS'10 IEEE SYDNEY*. 1–9.

[21] P. A. van Walree. 2013. Propagation and Scattering Effects in Underwater Acoustic Communication Channels. *IEEE Journal of Oceanic Engineering* 38, 4 (2013), 614–631. https://doi.org/10.1109/JOE.2013.2278913

[22] P. A. van Walree, F. X. Socheleau, R. Otnes, and T. Jenserud. 2017. The Watermark Benchmark for Underwater Acoustic Modulation Schemes. *IEEE Journal of Oceanic Engineering* 42, 4 (2017), 1007–1018. https://doi.org/10.1109/JOE.2017.2699078

[23] M. Zorzi, R. R. Rao, and L. B. Milstein. 1995. On the accuracy of a first-order Markov model for data transmission on fading channels. In *Proc. of the IEEE International Conference on Universal Personal Communications (ICUPC'95)*. 211–215.