Joint optimization of sonar waveform selection and sonobuoy placement

- Dr. Christopher Taylor
- Prof. Simon Maskell
- Dr. Alexey Narykov
- Prof. Jason Ralph

UNIVERSITY OF LIVERPOOL

[dstl] The Science Inside

Agenda

- Overview
- Recap of the sonobuoy placement problem
- Sonar waveform selection
- Modelling the joint problem
- Creating benchmarks
- The custom environment simulation
- Experiments with deep reinforcement learning
- Discussion and conclusions
- Q&A

[dstl] The Science Inside

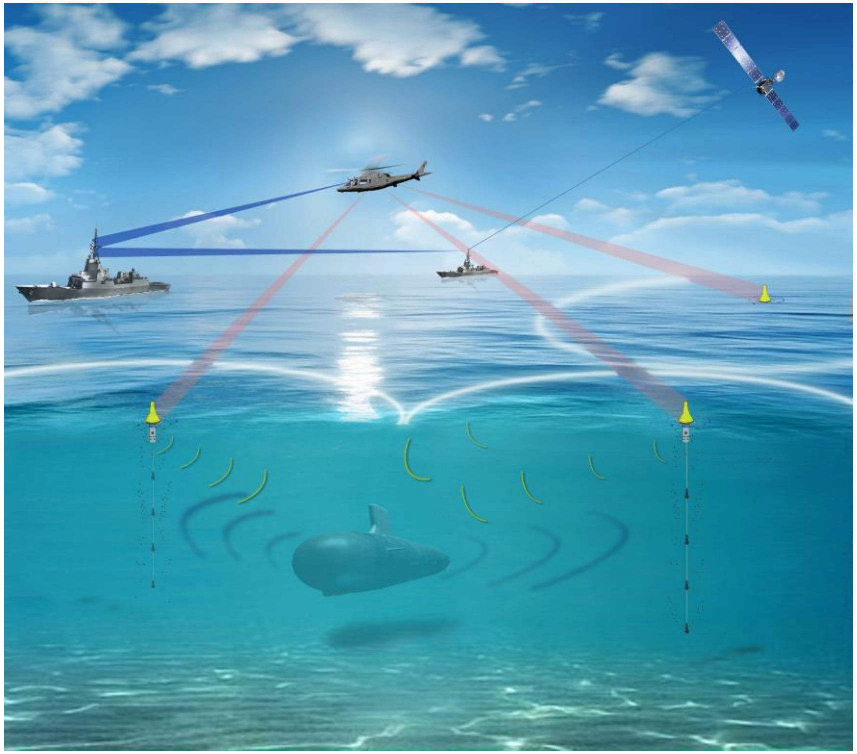
UNIVERSITY OF LIVERPOOL

Overview

- In previous work we looked at the problem of optimally placing a field of passive sonobuoys in a complex undersea environment to track an uncooperative and possibly stealthy target
- We now consider a joint problem for <u>active</u> sonar:
  - optimally place a field of sonobuoy receivers AND
  - optimally select waveform pulse train, in real time
- We wanted to establish benchmarks using fixed, predetermined placement patterns and pulse trains
- Then use deep reinforcement learning (DRL) to train a model that could respond to updated information as sonobuoys are placed and measurements are taken, and compare
- A particular algorithmic issue is that the sensor placement and waveform selection decisions have different timescales

Recap of the sonobuoy placement problem

- Sonobuoys are portable, expendable sonar systems
- Sonobuoys consist of a flotation device and an array of hydrophones that unfurls on deployment
- Sonobuoys are placed sequentially by an agent such as a helicopter, aircraft or UAV, typically directed by a surface vessel
- Sonobuoys may be active or passive, with monostatic, bistatic or multistatic operation



https://www.militaryaerospace.com/sensors/article/14198901/antisubmarine-warfare-asw-sonobuoys-multistatic

Waveform selection

- As with radar, active sonar systems can transmit different types of waveforms with various characteristics
- Much of the literature concentrates on a few simple waveforms, in particular continuous wave (CW) and linear frequency modulated (LFM)
- CW transmits with high power and has better range-rate (Doppler) performance
- LFM offers superior range performance compared to CW
- It is possible to combine waveforms sequentially in a pulse train
- For example, a pulse train [*LFM,LFM,CW,LFM,LFM,CW*] has length 6 and a proportion $\mu_{pt} = 0.66$ of LFM pulses

Modelling the joint optimization problem

- The mission goal is to maximize the amount of time during which position estimates have errors below a certain threshold value
- Errors are defined as the difference between mean position estimates for a single target of interest (TOI) and ground truth
- Constraints on the optimization problem:
  - Placing agent (e.g., a helicopter or UAV) has a limited payload of sonobuoys
  - Mission ends when the TOI leaves a predefined area of interest (AOI)
- It may not be possible to place all sonobuoys before the target leaves the AOI

Modelling the joint optimization problem

- The simulated environment has several stochastic, randomly generated elements that are not known by the algorithm:
  - The undersea environment has spatially varying noise/clutter
  - Initial position, course and speed of the target of interest (TOI)
- Simplifying assumptions:
  - After placement, sonobuoys do not drift, and positions are known with certainty
  - There is a single TOI which is detected by all sonobuoys
  - No false alarms or sensor failure
- We use an unscented Kalman filter (UKF)-based tracker for sensor fusion and localization
- Sensors that are a large distance from the TOI will yield little or no useful information

Creating benchmarks

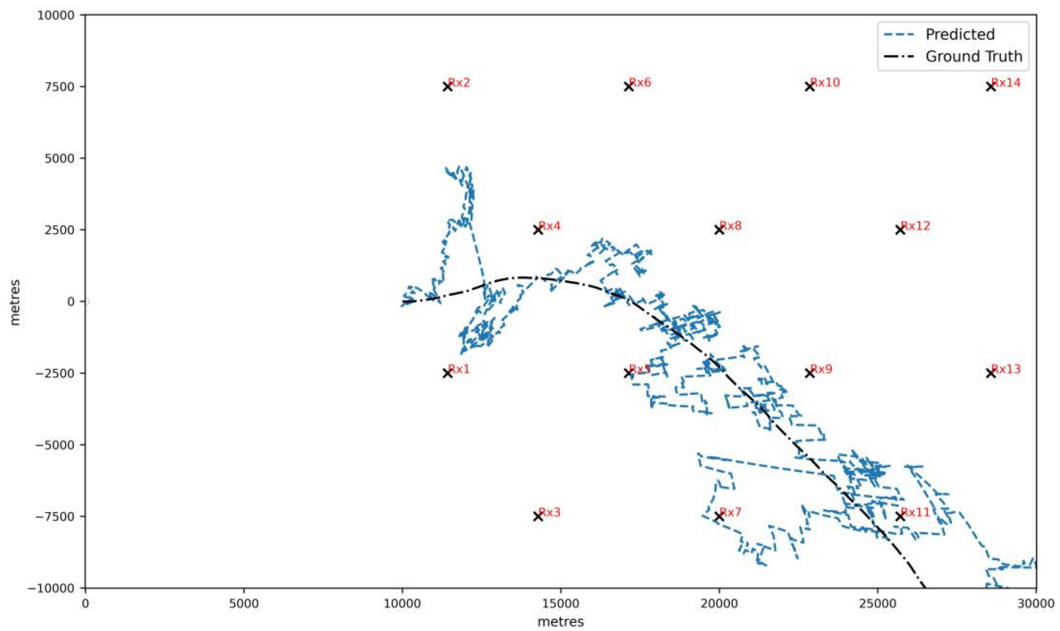
- We wanted to create benchmarks for comparison with our machine learning approach
- We investigated a scenario where:
  - a placing agent follows a fixed flight plan, dropping sonobuoys in a predetermined pattern
  - A fixed platform transmits a repeating, predetermined sonar pulse train consisting of a sequence of CW and LFM waveforms
- The difference between LFM and CW is modelled by appropriately parameterizing the respective UKF models
- We simulated all combinations of:
  - the 55 possible waveform combinations in a pulse train of length 6
  - 30 different randomly pre-generated noise/clutter maps
  - patterns of between 8 to 20 sonobuoys, arranged as staggered lattices

[dstl] The Science Inside

UNIVERSITY OF LIVERPOOL

Sonobuoy placement patterns

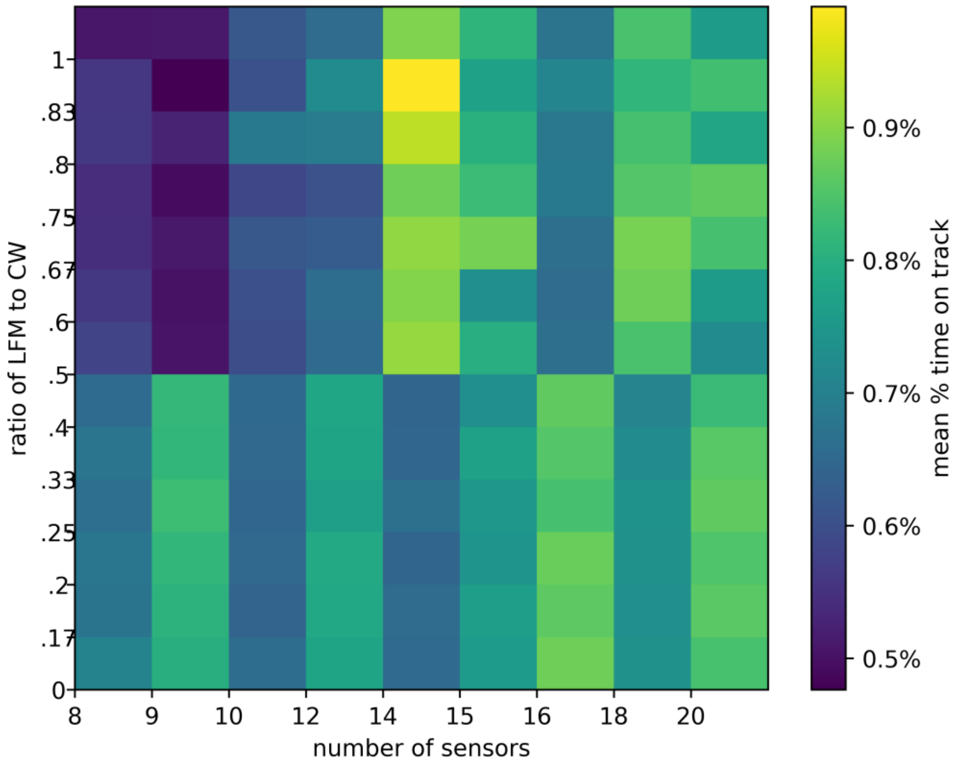
- Fidelity of the track varies considerably depending on the combination of pulse train and sonobuoy placement pattern, as well as the chosen noise/clutter map
- In some simulations, the tracking is even poorer than the one shown here!



- Each numbered **x** represents an ordered sonobuoy placement position
- Active sonar transmitter is located at the left-hand edge of the AOI

Benchmark results

- Pulse trains with higher proportions of LFM do better in combination with placement patterns with larger numbers of sonobuoys
- For smaller numbers, a lower proportion of LFM (and higher proportion of CW) performs better
- The "sweet spot" in these simulations is 14-15 sonobuoys with around 60-80% LFM
- Increased deployment time means more sensors is not always better



Mean % time-on-track is the average proportion of time localization error falls below a required threshold

The custom environment simulation

- Created a custom simulation environment using the OpenAI Gym standard – widely used by the RL community
- Effectively creates a customized “function” that:
  - takes an action (such as placing a sonobuoy or transmitting a sonar waveform) as its argument
  - returns observations (such as sonar measurements) and a reward for the action
- An RL algorithm can use observations and rewards in training to learn an approximately optimal policy
- The trained model will take the optimal action for any state of nature
- In DRL algorithms such as Proximal Policy Optimization (PPO):
  - inputs to the (policy) neural network are observations
  - outputs form a probability distribution over possible actions

Observation space

- With a maximum payload of *N* sonobuoys, the observation space consists of the following:
- x-position of each sensor (size:*N*)
- y-position of each sensor (size:*N*)
- bearing prediction from each sensor (size:*N*)
- range from each sensor (size:*N*)
- range-rate from each sensor (size:*N*)
- predicted state vector for the target from the tracker
- (size:4)
- position covariance from the tracker (size:4)
- prediction variance from the tracker (size:2)
- steps remaining to time limit (size:1)

Action space and action masking

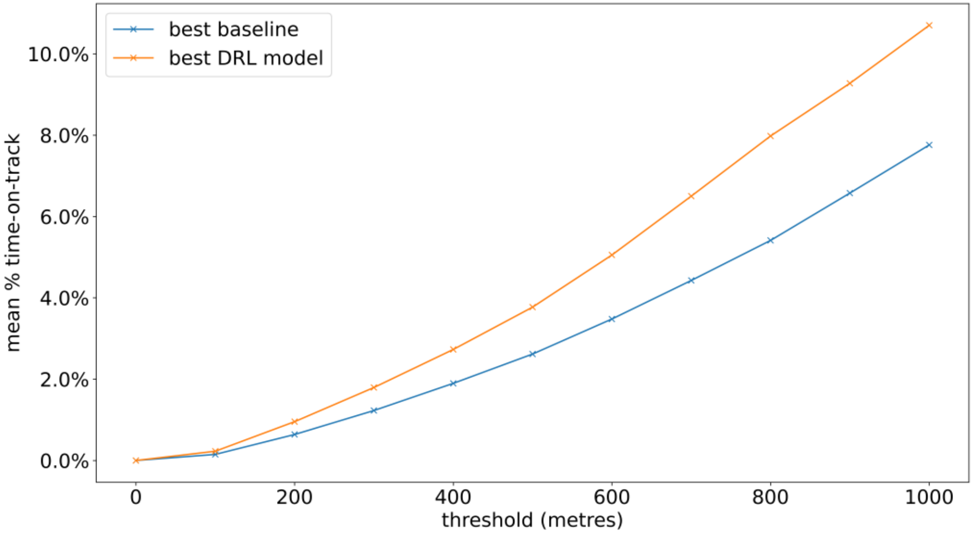
- Use a discretized grid with 5000 possible placement locations
- At each timestep once the first sonobuoy is placed and operational, can choose to transmit a CW or LFM waveform
- Only at timesteps where the placing agent (e.g. a UAV) has no course currently plotted and sonobuoys remain to be placed, a new sonobuoy placement location is chosen, and a course plotted
- Also take into consideration constraints on how close to each other sonobuoys can be placed
- Since not all actions are necessarily available at a given timestep, we use action masking to cope with this
- Unavailable actions are “masked” from the output of the neural network

Experiments with deep reinforcement learning

- We trained using the PPO algorithm, adapted for action masking
- Initial payload of 14 sonobuoys
- New random noise/clutter map generated each time the environment resets during training
- Since the algorithm has no access to ground truth, we cannot train directly using time-on-track
- Use a proxy reward function, where the reward is received at each timestep only if the localization uncertainty calculated by the tracker falls below a predetermined threshold
- Although we trained for as much as 20 million timesteps, we found the best results were from the model trained for 8 million timesteps
- Likely that the model overfits beyond a certain point in training

Results

- We compared our model to the benchmark using 100 randomly generated environments
- We looked at performance using different error thresholds up to 1km in 100m increments
- The best DRL model outperformed the best benchmark at all thresholds



The algorithm is considered “on track” if the predicted position is within a given distance threshold

Discussion & conclusions

- Investigations of fixed patterns reveal that:
  - Patterns with different numbers of sonobuoys benefit from being combined with pulse trains with different mixes of CW and LFM
  - More is not always better – increased deployment time for patterns with larger numbers of sonobuoys leads to diminishing returns last a certain number
- Using a placing agent with the ability to react to updated information and make both placement decisions and waveform choices accordingly can produce improved performance over fixed placement patterns and sonar pulse trains
- Further improvements might be made by changing the observation space or improving/changing the DRL algorithm

Q&A

- Any questions?

The authors would like to thank DSTL Grant no. 1000143726 for financial support.

References:

- Taylor CM, Maskell S, Narykov A.,Ralph JF. "Joint optimization of sonar waveform selection and sonobuoy placement". SSPD 2023 (forthcoming)
- Taylor CM, Maskell S, Ralph JF. "Using hybrid multiobjective machine learning to optimize sonobuoy placement patterns," IET Radar, Sonar and Navigation,vol. 17, no. 3, pp. 374–387, 2023
- Taylor CM, Maskell S, Ralph JF. "Optimizing sonobuoy placement using multiobjective machine Learning". In: SSPD 2022 pp. 86-90.

[dstl] The Science Inside

UNIVERSITY OF LIVERPOOL