Adaptive Algorithm for Resource Generation in a Quantum Network

Using a Markov Decision Process Approach to Optimize the Quantum Resource Generation Algorithm

Author:
Ioana-Lisandra Draganescu
I.L.Draganescu@student.tudelft.nl

Responsible Professor:
Gayane Vardoyan

Supervisor:
Bethany Davies

1 Research Question

"How does the optimal resource generation protocol vary with the threshold fidelity of the target protocol, when using different resource generation protocols is allowed?"

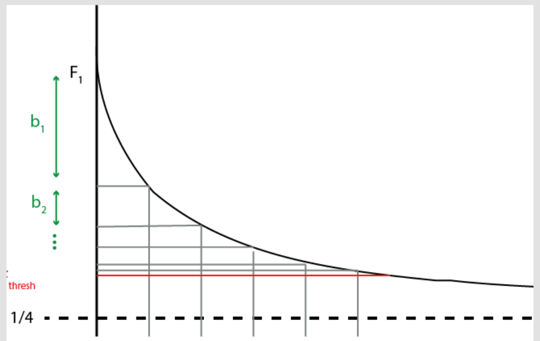


Figure 1: Fidelity bins. Adapted from B. Davies with permission

2 Problem Formulation

1. **Fidelity bins** → a way to discretize the time-to-live of links

Number of bins:

$$m = max_{link}(ttl(link))$$

Bin 0 → highest fidelity & longest time-to-live

2. Model the problem as a **Markov decision process**:

L = the goal number of simultaneously active links

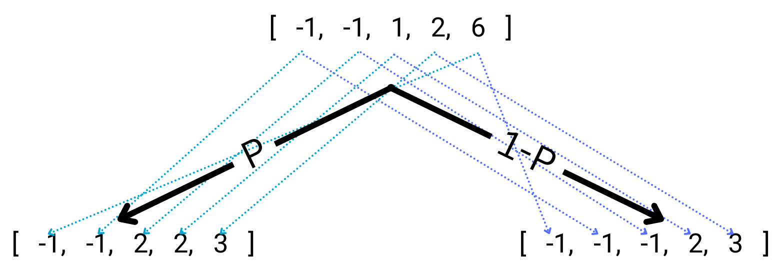
State space:

$$S = \{\{l_0, l_1, ..., l_{L-1}\} \,|\, l_0, l_1, ..., l_{L-1} \in \{-1, 0, ..., m-1\}\}$$

Action space:

$$A = \{a_0, a_1, ..., a_{n-1}\}$$

Example of a state transition for L = 5, m = 7, when using a protocol with probability P that generates a link in bin 2:



Bellman equation to find the optimal expected time T, where p is the probability to go from state s to state s' given an action:

$$T_\pi(s) = 1 + \sum_{s' \in S} p(s'|s, \pi(s)) T_\pi(s'), \forall s \in S \text{ [1]}$$

We use a value iteration algorithm to calculate T:

Algorithm 1 Value Iteration [2]

Initialize array T arbitrarily
repeat
  $\Delta \leftarrow 0$
  for each $s \in S$ **do**
    $v \leftarrow T(s)$
    $T(s) \leftarrow min_a(1 + \sum_{s' \in S} p(s'|s, a) T(s'))$
    $\Delta \leftarrow max(\Delta, |v - T(s)|)$
  end for
until $\Delta < \theta$ (a small positive number)
Output a deterministic optimal policy π such that $\pi(s) = argmin_a(1 + \sum_{s' \in S} p(s'|s, a) T(s'))$

The policy π represents the optimal policy.

References:
[1] Á. G. Iñesta, G. Vardoyan, L. Scavuzzo, and S. Wehner, "Optimal entanglement distribution policies in homogeneous repeater chains with cutoffs," npj Quantum Information, vol. 9, no. 1, p. 46, 2023.
[2] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

3 Results

In both cases, we use 7 protocols with probabilities and fidelities:

P = [0.025, 0.05, 0.075, 0.1, 0.2, 0.3, 0.4]
F = [0.975, 0.95, 0.925, 0.9, 0.8, 0.7, 0.6]

We introduce a **heuristic**:

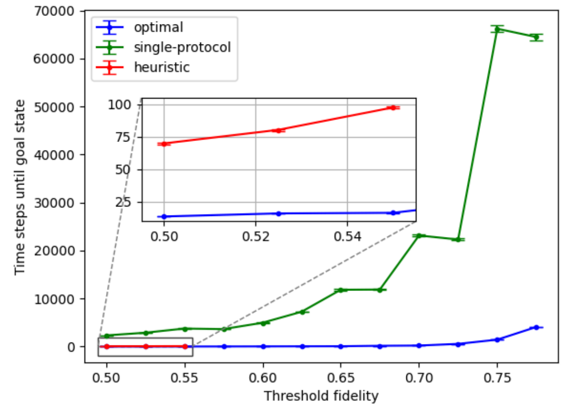
- Prefer protocols with higher fidelity for the first links
- Prefer protocols with higher probability for the last links

- **Case 1: Three links, few bins**

We want 3 simultaneously active links.
We use 0.1 as the rate of decoherence in memory → relatively few fidelity bins.

The optimal policy performs several orders of magnitude better than the single protocol policy.

The heuristic policy performs well up until a point.

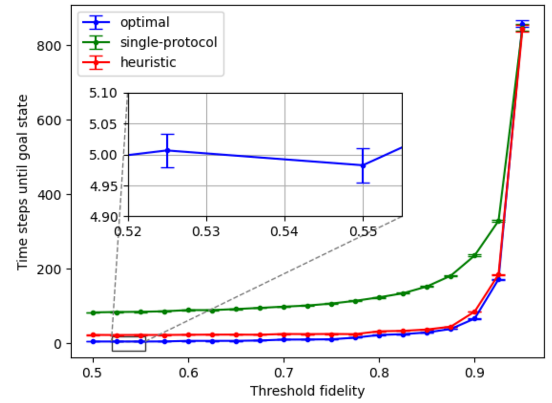


- **Case 2: Two links, many bins**

We want 2 simultaneously active links.
We use 0.01 as the rate of decoherence in memory → relatively many fidelity bins.

The optimal and heuristic policies perform comparably well.

The standard error is relatively small.



4 Conclusions and Future Work

- We modelled the problem as a Markov decision process.
- We computed the optimal policy and introduced a heuristic for the case when using different protocols is allowed.
- The optimal policy performs significantly better than the single-protocol policy for the analyzed cases.

Future work recommendations:

Test the findings in practical applications.