Quantum Virtual Link Generation via Reinforcement Learning

Ramon Aparicio-Pardo, Antoine Cousson, Redha A. Alliche

▶ To cite this version:

Ramon Aparicio-Pardo, Antoine Cousson, Redha A. Alliche. Quantum Virtual Link Generation via Reinforcement Learning. 23rd International Conference on Transparent Optical Networks (ICTON 2023), Jul 2023, Bucharest, Romania. hal-04136014

HAL Id: hal-04136014

https://hal.science/hal-04136014v1

Submitted on 21 Jun 2023

Quantum Virtual Link Generation via Reinforcement Learning

Ramon Aparicio-Pardo, Antoine Cousson, Redha A. Alliche
Université Côte d'Azur, CNRS, I3S,
2000 Rte des Lucioles, 06900 Sophia Antipolis, France
E-mail: raparicio@i3s.unice.fr

ABSTRACT Quantum networks make use of the quantum entanglement as building block. When two qubits are entangled, their state changes exhibit non-classical correlations used to design new applications not possible with classical communication, such us quantum key distribution or distributed quantum computation. Unfortunately, quantum entanglement is a probabilistic process strongly dependent on the features of involved devices (optical fibers, lasers, quantum memories, ...). The management decisions (i.e., the control policy) to set up and keep the entanglement as long as possible with the highest quality constitutes a stochastic control problem. This process can be modelled as Markov Decision Process (MDP) and solved via the Reinforcement Learning (RL) framework (a form of Machine Learning). In this work, we apply this RL framework to learn an entanglement management policy outperforming the State-of-the-Art policy when models characterising precisely the involved quantum devices are not known.

Keywords: Quantum communications, quantum swapping, deep reinforcement learning.

1. INTRODUCTION

In the last years, the application of quantum physics principles to computer networks is gaining momentum among the research and industry communities, as shown by the first attempts of standardisation of a so-called "Quantum Internet" [1], [2] by the Internet Engineering Task Force (IETF). Amongst these principle the *quantum entanglement* has been identified as fundamental resource for Quantum Communication [1], since it enables the Quantum Internet applications, as secure cryptographic key distribution and, distributed quantum computing [2]. But, *quantum entanglement* is a probabilistic process strongly dependent on the features of the involved communication devices. Consequently, the entanglement management constitutes a stochastic control problem that can be formulated as a Markov Decision Process (MDP) [3]. In this preliminary work, we investigate the capacity of Deep Reinforcement Learning (DRL) to solve these problems, in particular, when a quantum entanglement is set up between two remote communication nodes not directly connected by a link. In the paragraphs below, we will introduce the required background.

Qubit and entanglement. In quantum communication and quantum computing, the counterpart of a classical bit is the *quantum bit (or qubit)*. But, whereas the classic bit can take either the "0" state or the "1" state, the *qubit* can be in a *superposition* of the two, with a certain probability to be at one of the states. The *qubit* exists in this superposition until its eventual measurement. Afterwards, it will take the "0" value or "1" value according to the corresponding probability. When two qubits are *entangled*, their individual states cannot be described in a separated way: a state change, i.e., a qubit reading measurement, in one of them implicitly comes with a change in the other one, regardless of the physical distance between them. Thus, the measurements at the two entangled qubits exhibit non-classical correlations used to design new applications not possible with classical communication, such us quantum key distribution or distributed quantum computation.

Quantum network. A set of nodes able to exchange qubits and distribute entangled states amongst themselves is defined as a *quantum network* in the RFC [1]. These *quantum nodes* are connected each other by *optical fiber* or *satellite laser* links. In this paper, we assume fiber links. When, an entanglement is set up between two qubits located at two adjacent quantum nodes connected by a direct link (e.g., between nodes A and B in Fig. 1), the entanglement constitues an *elementary quantum link* [1]. Its success probability P_e exponentially decreases with distance, which means that short-distance entanglements (like A-B, in Fig. 1) are more likely to succeed than long-distance entanglements (like A-C, in Fig. 1). To overcome this issue, we can create a *virtual link* [1] over two elementary links via the so-called *entanglement swapping* [1], [4]. This process allows creating long-distance entangled pairs by consuming the previously generated elementary links on the path between two further end-points. In Fig. 1, the elementary links A-B and B-C are consumed to create a longer virtual link A-C. Quantum nodes (as B in Fig. 1) that create long-distance entangled pairs via entanglement swapping are called quantum repeaters [1] and they must store intermediate elementary links on the so-called *quantum memories* [1] to be consumed later.

Quantum memory lifetimes. The probability that a qubit stored in a quantum memory is still, after a certain time, in its original state (e.g., an entangled state) decreases with time [5]. This probability is referred as to memory efficiency η_m [5], and its decay is known as decoherence. This process is the consequence of the progressive interactions of the quantum memory with the environment, since a memory cannot be perfectly

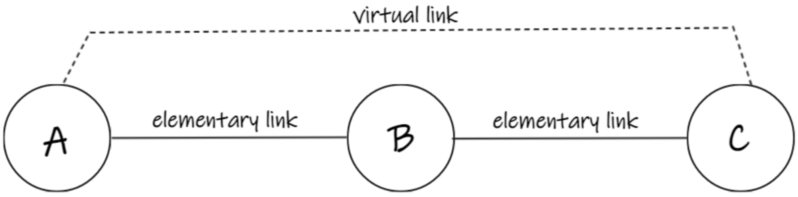


Figure 1: *Elementary link vs virtual link.*

isolated from it. The entanglement swapping success probability P_s depends on the memory efficiency η_m of the oldest loaded quantum memory involved in the swapping [6].

This paper, as far as we know, is one of the first works modelling a quantum virtual link generation process as a classical MDP and using a DRL algorithm to find an optimal generation policy tracking the age of the elementary links. *This supposes an innovative contribution with respect to the related works, where this age of the elementary are not used in the generation procedures.*

Related works are presented in Section 2. The MDP modelling the virtual link generation along with the DRL approach used to solve it are described in Section 3. Numerical results and experiment settings are shown in Section 4. Section 5 concludes the article.

2. RELATED WORKS

The idea that the management problem of *quantum elementary and virtual links* can be mathematically formalised as a quantum generalisation of a MDP was developed in the Khatri's PhD dissertation [3]. This dissertation assumes the Quantum Decision Process (QDP) framework, [7], the quantum analog of a MDP, where states are quantum and actions are quantum operators, which implies the usage of a quantum computer. In contrast to this work, we model the decision problem as a classical MDP with states described by measured physical properties and actions taken on a macroscopic level. We think that many of Khatri's ideas can be adapted to the current state-of-the-art without waiting for the development of a quantum computer. Then, in the this paper, we model as a classical MDP the *QDP with entanglement swapping* presented in the Appendix D in [3]. This process aims to generate a *virtual link* from two *elementary links* via *entanglement swapping* as explained in Section 1. The virtual link generation has been recently studied in two contexts: (i) quantum repeaters chains [8] and, (ii) quantum entanglement routing [9]. In both of them, we target to set up long-distance entanglements between non-adjacent communication nodes when topologies are Daisy chains and mesh networks, respectively. In these works, the "history" of the links is ignored, i.e., the timestamps at which the elementary (and virtual) link generation processes succeed are not used. Besides, the virtual link generation always follows a *infinity memory cutoff-time policy* when, once the early elementary link is successful set up, we keep it till the late also succeeds, regardless the impact of larger decoherence of the oldest one onto the swapping probability P_s.

3. REINFORCEMENT LEARNING FOR VIRTUAL LINK GENERATION

3.1 Problem Statement

As aforementioned, the management of a virtual link generation over two elementary links via entanglement swapping can be formulated as a MDP [3]. In this decision process, we aim to maximize the number of successful entanglement swaps per time unit, i.e., the virtual link generation rate. To generate the virtual link, two elementary links must have been successfully created before attempting the entanglement swapping with probability P_s. The older the first generated elementary link is, the more likely the swapping will fail. Then, after a cutoff time t_c, discarding the oldest elementary link and trying to generate it again (i.e., resetting it) becomes beneficial, since links freshly resets have always a higher swapping probability. Unfortunately, this reset (a new elementary link generation) comes with a cost, as it is also a stochastic process with success probability P_e, which must be repeated till success, delaying the eventual swapping attempt. Thus, the cutoff time t_c of the elementary link has to be carefully selected to reduce the time between two successful swaps and, hence, maximize the virtual link generation rate.

Now, we describe the process as a MDP. At each time step t, a control agent applies a certain action a_t after observing the current state s_t. The execution of this action will trigger a transition into a new state s_{t+1} with a certain probability $p(s_t, a_t, s_{t+1})$. The agent receives a reward r_{t+1} based on the "quality" of the pair (s_t, a_t) to maximize a long-term objective. Then, the *agent* observes the new state s_{t+1} and repeats these steps. Assuming an initial state s_0, the MDP gives rise to a *trajectory*: $s_0, a_0, r_1, s_1, a_1, r_2, s_2, a_2, r_3, , \ldots$. This *trajectory* is generated based on an agent *policy* $\pi(s, a)$ denoting the probability that action $a_t = a$ is taken at state $s_t = s$. In our case, the system state (action) consists of the concatenation of the states (actions) of the two elementary links and the virtual link. The state of each link is a vector $s = [x, m]$, where x is *1* if entanglement is active

(*0*, otherwise); and, m is the entanglement age (*-1*, if entanglement is inactive). Two actions can be taken over each link: either the link is *re(set)*, i.e., a link generation is (re)tried; or the link *waits*, i.e., no link generation is (re)tried. The former provokes a stochastic transition with probability P_e (P_s) to a state $s = [1, 0]$ for an elementary (virtual) link generation. The later leaves the current link state unchanged. Here, we assume a unique virtual link to be set up by swapping two elementary links, i.e., a space action of size $2^3 = 8$. The reward is simply *1* if the entanglement swapping succeeds; and *0*, otherwise.

Finally, we conclude this subsection by detailing the assumptions about the quantum environment model. In our work, entanglements are created in a "heralded" fashion, i.e., we know when the entanglement has been successfully established (therefore, cutoff times t_c can be measured). One created, entanglements must be stored in quantum memories to eventually be swapped to create virtual links. Amongst the different methods of entanglement generation, the DLCZ-based protocols [6] are an option satisfying these requirements. When DLCZ-based protocols are used the time becomes slotted with time slots L_0/v long, where L_0 is the length of the elementary link (fiber) and v is the light propagation speed at the fiber. This slot duration represents the time required to know if an elementary link generation has succeeded: the duration of the time step before observing a new state and taking a new action (retry or not a link generation). The elementary link generation success probabilities P_e exponentially decreases with fiber distance L_0 according to the work [6]. Whereas, the memory efficiency η_m, main factor defining the swapping success probability P_s, falls with the storage time following the Mims' model described in [5]. The exact values of these probabilities depend on the precise characterisation of the optical fiber and quantum memories involved. We assume that we do not know them.

3.2 Reinforcement Learning based approach

If we do not possess models characterizing precisely the elementary link generation probability P_e and the memory efficiency η_m (and, thus, the entanglement swapping success probability P_s), the state transition probabilities $p(s_t, a_t, s_{t+1})$ are unknown. In this case, we can apply Deep Reinforcement Learning (DRL) [10], [11] to find a policy maximizing the virtual link generation rate. In Reinforcement Learning (RL), we define a *Q-value function* $Q^\pi(s, a)$ as the *expected discounted return* from a given state-action pair (s, a) when following a policy π thereafter. The discounted return is the sum of the discounted rewards the agent receives over the future in a trajectory: $sum_{k=0}^{\infty}\gamma^k r_{t+k+1}$, with $\gamma \in [0, 1]$. Therefore, an agent solving a RL problem searches an optimal policy π^* maximizing this *Q-value function*. Here, this policy consists of finding the best cutoff time t_c for the oldest elementary link (see Section 3.1). When state transition probabilities are unknown (our case), this *Q-value function* can be estimated by using statistical learning with the Deep Q-Network (DQN) algorithm [11]. In the present work, the learning routine simply follows the classical DQN [11] but adapted to our virtual link generation problem.

4. SIMULATION EXPERIMENTS

4.1 Experiment Settings

We use OpenAI™`Gym` [12], one of the most popular toolkits to define RL environments in `Python`, to program a quantum environment modelling the MDP as described in Section 3.1. We consider the length of the fiber links L_0 and the light speed are 100 km and $200,000$ km/s, respectively, which yields to a time step duration of $0,5$ ms. We assume fiber losses of 0.2 dB/km achievable around $1,550$ nm. We use the Mims' equation corresponding to the DD sequence 'XX' in study [5], but considering a zero-time efficiency of 1.

The agent is implemented in `TensorFlow` [13] as a three-layer neural network with two dense layers of 32 neurones, each having a `tanh` activation function, and output layer without activation with as many neurones as actions (8, here since we consider two elementary links and one virtual link). The neural network is trained using the DRL algorithm called DQN [11] provided by the OpenAI™ `Baselines` library [14].

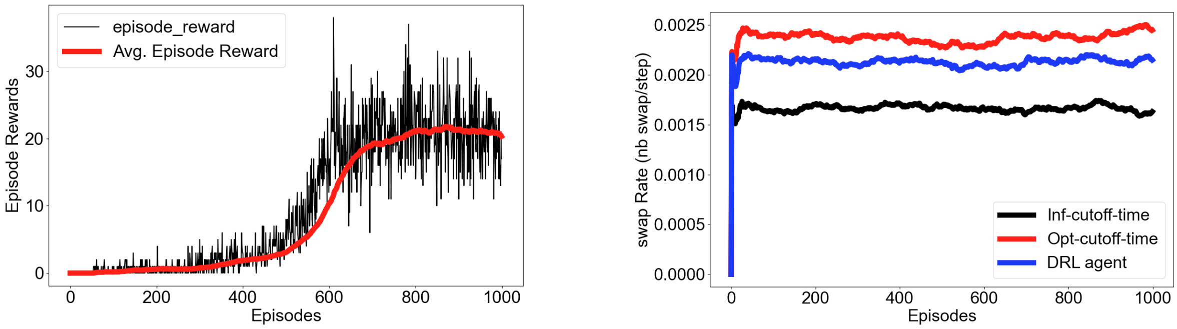
Fig. 2a depicts the episode reward evolution along the training time. An episode reward is the sum of the rewards r produced during all the steps in a training episode. We observe that the average episode reward increases with time, stabilizing after 600 episodes, where an episode is $10,000$ steps long.

4.2 Numerical Results

In Fig. 2b, we compare the policy learned by the DRL agent with two benchmarks:

- ***Inf-cutoff-time*** **policy**: the cutoff time of the oldest link is set to infinity, then, when the first elementary link to succeed is set up, we keep it till the second one also succeeds regardless the decoherence level of the first one. This is the by-default policy considered by the State-of-the-Art [8], [9]. It represents a *lower bound* on the optimal policy.
- ***Opt-cutoff-time*** **policy**: the optimal cutoff time of the oldest link is found by *brute force*. This process cannot scale up with larger problem instances. It trivially represents an *upper bound* on the policy learned by the DRL agent.

We test the policies by simulating the MDP process 1000 times. An *episode* is a simulation instance. Each episode again consist of $10,000$ steps. We observe that the DRL agent clearly outperforms the *Inf-cutoff-time* policy and is close to the *opt-cutoff-time* policy.The found cutoff times are 146.0 steps and 108 steps for the DRL policy and *opt-cutoff-time* policy, respectively.



(a)

(b)

Figure 2: *(a) Episode reward evolution during training. (b) Swap rate of DRL agent vs benchmarks during test.*

5. CONCLUSIONS

In this study, we explore if DRL can be used to learn policies maximizing the virtual link generation rate via entanglement swapping in scenarios where a model of the entanglement success probabilities is not previously known. In such scenario, we have obtained some first results pointing out that we can discover a close-to-optimal policy outperforming the State-of-the-Art *Inf-cutoff-time policy*.

ACKNOWLEDGMENTS

This work has been supported by the French government, through the UCA DS4H Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-17-EURE-0004.

REFERENCES

[1] W. Kozlowski, S. Wehner *et al.*, "Architectural Principles for a Quantum Internet," RFC 9340, Mar. 2023. [Online]. Available: https://www.rfc-editor.org/info/rfc9340

[2] C. Wang, A. Rahman *et al.*, "Application Scenarios for the Quantum Internet," Internet Engineering Task Force, Internet-Draft draft-irtf-qirg-quantum-internet-use-cases-15, Mar. 2023, work in Progress. [Online]. Available: https://datatracker.ietf.org/doc/draft-irtf-qirg-quantum-internet-use-cases/15/

[3] S. Khatri, "Towards a general framework for practical quantum network protocols." [Online]. Available: https://digitalcommons.lsu.edu/gradschool_dissertations/5456

[4] L. Gyongyosi and S. Imre, "Advances in the quantum internet," *Communications of the ACM*, vol. 65, no. 8, pp. 52–63, 2022.

[5] A. Ortu, A. Holzpfel *et al.*, "Storage of photonic time-bin qubits for up to 20 ms in a rare-earth doped crystal," *npj Quantum Information*, vol. 8, no. 1, pp. 1–7, 2022.

[6] N. Sangouard, C. Simon *et al.*, "Quantum repeaters based on atomic ensembles and linear optics," *Reviews of Modern Physics*, vol. 83, no. 1, pp. 33–80.

[7] J. Barry, D. T. Barry *et al.*, "Quantum partially observable markov decision processes," *Physical Review A*, vol. 90, no. 3, p. 032311.

[8] B. Li, T. Coopmans *et al.*, "Efficient optimization of cutoffs in quantum repeater chains," vol. 2, pp. 1–15, conference Name: IEEE Transactions on Quantum Engineering.

[9] C. Li, T. Li *et al.*, "Effective routing design for remote entanglement generation on quantum networks," *npj Quantum Information*, vol. 7, no. 1, pp. 1–12, 2021.

[10] Y. Bengio, *Learning deep architectures for AI*. Now Publishers, 2009.

[11] V. Mnih, K. Kavukcuoglu *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540.

[12] "Gym: Gym toolkit for creating reinforcement learning environments," 19/04/23. [Online]. Available: https://www.gymlibrary.dev/

[13] "Tensorflow: An end-to-end open source machine learning platform," 19/04/23. [Online]. Available: https://www.tensorflow.org/

[14] "Openai baselines: high-quality implementations of reinforcement learning algorithms," 19/04/23. [Online]. Available: https://github.com/openai/baselines