Adaptive and Efficient Qubit Allocation Using Reinforcement Learning in Quantum Networks

Yanan Gao, Song Yang, Fan Li, and Xiaoming Fu

Abstract

Quantum entanglement brings high-speed and inherently privacy-preserving transmission for information communication in quantum networks. The qubit scarcity is an important issue that cannot be ignored in quantum networks due to the limited storage capacity of quantum devices, the short lifespan of qubits, and so on. In this article, we first formulate the qubit competition problem as the Cooperative-Qubit-Allocation-Problem (CQAP) by taking into account both the waiting time and the fidelity of end-to-end entanglement with the given transmission link set. We then model the CQAP as a Markov Decision Process (MDP) and adopt a Reinforcement Learning (RL) algorithm to self-adaptively and cooperatively allocate qubits among quantum repeaters. Further, we introduce an Active Learning (AL) algorithm to improve the efficiency of the RL algorithm by reducing its trial-error times. Simulation results demonstrate that our proposed algorithm outperforms the benchmark algorithms, with 23.5 ms reduction on the average waiting time and 19.2 improvement on the average path maturity degree, respectively.

Introduction

Quantum entanglement is a phenomenon occurring in the quantum system where several particles' properties synthesize into a whole and that can only be described through comprehensive system characteristics [1]. It is a crucial resource, costly to produce and that can implement valuable qubit transformation, for quantum communication and quantum computing. Assume "Alice" entangles with "Bob" and "Carol" entangles with "Dave." Bell-state measurement [2] can determine an entanglement state between "Alice" and "Carol," which will change the initial entanglement structure and produce new entanglement between "Bob" and "Dave" no matter how far away they are, called "Entanglement swapping" [3]. Then, an end-to-end entanglement can be implemented between source node and destination node ("Bob" and "Dave") through quantum repeaters ("Alice" and "Carol"). Multiple pairs of long-distance end-to-end entanglements sharing lots of repeaters compose a quantum network. With the faster-than-light transformation speed of entanglement, a quantum network is more efficient than a fiber network to complete the tasks that require more coordination, such as clock synchronization, telemedicine. The clock synchronization and telemedicine are representative quantum applications, which make use of the uncertain property of quantum state. Before the Bell measurement, all states are uncertainty. Once the state of one side is determined, the states of other sides are simultaneously determined. This property is the basis of clock synchronization, telemedicine, and anti-eavesdropping applications.

The non-repeatability of the qubit is the key challenge of end-to-end entanglement transmission (through repeaters) in a quantum network because it can cause permanent data loss. Quantum entanglement is one of the most effective properties of the quantum communications compared to the classical fiber network due to its high-speed and privacy-preserving transmission. The transmission in quantum networks depends on lots of repeaters and transmission path generation. Quantum entanglements can be established among quantum nodes via the forwarding in repeaters by using entanglement swapping operations. Then, the transmission paths can be generated by resorting to quantum entanglements. We therefore must ensure the availability and stability of the generation of transmission path (end-to-end entanglement) on the given paths in link layer [4].

Based on above analysis, some factors should be considered:

- Except source node and destination node, the end-to-end entanglement contains several quantum repeaters where some qubits are stored and entanglement swappings are successfully implemented. Only when each repeater successfully executes its entanglement swapping, the whole end-to-end entanglement can be established completely. The whole end-to-end entanglement can only wait until its repeaters complete entanglement swapping successfully at the same time. Hence, the end-to-end entanglement has different available times (the waiting time) for generation. Correspondingly, qubits stored in repeaters are occupied in different time duration, which makes the state information of quantum networks change dynamically as different entanglement swapping operations are executed in each time slot.
- Fidelity is a measure of two qubits' intimacy, which can evaluate the quality of their entan-

Digital Object Identifier:
10.1109/MNET.001.2200157

Yanan Gao, Song Yang (corresponding author) and Fan Li are with Beijing Institute of Technology, China; Xiaoming Fu is with the University of Göttingen, Germany.

0890-8044/22/$25.00 © 2022 IEEE

glement state. The stability of end-to-end entanglement depends on the summation of all fidelities produced by each entanglement link between two qubits on the whole end-to-end entanglement path. It has an important influence on end-to-end transmission latency (the time of responding requests) and throughput (the number of successful end-to-end entanglements) of a quantum network.

- Each entanglement transmission can consume a Bell-state measurement process with cost of a certain number of qubits due to the generation of particle decay [5]. It is more beneficial to timely generate a stable entanglement state rather than repeated attempts.

Usually, quantum networks need to serve multiple requests (multiple pairs of source and destination). Each request establishes more than one end-to-end entanglement to increase the network throughput. Based on the above considerations, there exists a severe resource scarcity and high consumption of repeaters' qubits in the whole quantum network. Here, qubit can be understood as quantum memory in the sense that qubit allocation can be accordingly regarded as quantum memory allocation. The requested end-to-end entanglement will enter into the waiting sequence if some qubits of repeaters that are on the given paths are occupied by other end-to-end entanglements. Hence, network information can continuously change with the dynamics of qubit allocation decisions, such as the number of qubits in each repeater, the fidelity of each two-qubit entanglement, and the waiting time of an end-to-end entanglement. Most of existing work about quantum networks focuses on throughput optimization, designing routing algorithm, or entanglement improvement. However, qubit scarcity (quantum memory scarcity) is also an important issue to be solved in quantum networks. There are many factors causing qubit scarcity: the limited storage capacity of quantum hardware device, the short lifespan of qubits due to the quantum decoherent, the limited transmission load of entanglement link, and the requirement of multiple requests' concurrent responses. The above reasoning leads to fierce competition for qubit allocation when quantum networks need to accommodate multiple requests concurrently. Some works present some straightforward algorithms to solve the qubit allocation problem, but they cannot work well with large network scale and multiple requests. Also, they cannot adapt to the dynamic changes of quantum networks.

In order to capture these dynamic changes occurring in quantum networks, we regard a two-qubit entanglement generation of an end-to-end entanglement as a unit time slot and serialize the request process as a Markov Decision Process (MDP). We introduce Reinforcement Learning (RL) algorithm to self-adaptively learn the qubit allocation decision in quantum networks. However, the qubit's scarcity does not allow a large number of the trial-and-error process for RL. We, therefore, resort to Active Learning (AL) algorithm to reduce the trial-and-error times during RL's training process. Considering the real-world implementation of quantum networks, we assume the lifespan of qubits in a repeater is long enough to support multiple entanglement generations. Also, we assume the end-to-end entanglement paths for multiple source-destination pairs to accommodate a given set of requests are given in advance.

The non-repeatability of the qubit is the key challenge of end-to-end entanglement transmission (through repeaters) in a quantum network because it can cause permanent data loss.

Based on the above assumption, we try to solve the qubit allocation problem for all quantum repeaters in quantum networks to satisfy the concurrent schedule of multiple end-to-end entanglements. As shown in Fig. 1, the central cloud platform of quantum computers possesses a global view of quantum networks through the information transmission via optical fiber link, including the generated end-to-end entanglements, the residual qubit resource of quantum repeaters, the fidelity of each entanglement state between two qubits, and so on. Each trusted repeater can also obtain network information from the central cloud and execute entanglement swapping to generate end-to-end entanglement. In all, our contributions are as the following:

- We define the qubit allocation problem in quantum networks to minimize the average waiting time.
- We devise an adaptive and efficient qubit allocation algorithm by combining the Reinforcement Learning (RL) algorithm and the Active Learning (AL) algorithm, that can simultaneously capture the dynamics of the network environment and speeds up the learning process.
- We conduct the simulations to verify the effectiveness of the proposed algorithm.

In the following content, we firstly define the qubit allocation problem in quantum networks based on the given end-to-end entanglement paths. Then, we introduce our designed framework of the whole algorithm model. After that, we detail the design of the proposed algorithm including RL and AL algorithms. Next, Performance evaluations on an opening simulation tool are described to demonstrate the superiority of our algorithm. Lastly, we conclude our work.

Related Work

Lots of existing works have explored the capacity optimization of quantum network respectively or synchronously from the two aspects: global view (throughput [6, 7], the path generation of end-to-end entanglement [6, 8]) and entanglement view (fidelity or stability for two-party or multi-party entangled state [6, 9], entanglement purification and error correction [6, 10]). Shi *et al.* [8] initially propose a novel routing algorithm to increase the number of end-to-end entanglement for one request. Then, Zhao *et al.* [7] explore the same problem for multiple end-to-end entanglements and improve the routing algorithm by filtering the failure entanglement link. Next, Zhao et al., propose an algorithm related to the routing and the purification by considering the fidelity of end-to-end entanglement link to maximize network throughput for multiple requests [6]. They simultaneously consider the stability, throughput, and routing factors. Lee *et al.* [9] also consider the entanglement distribution rate and photonic technology to maintain the fidelity over long-dis-

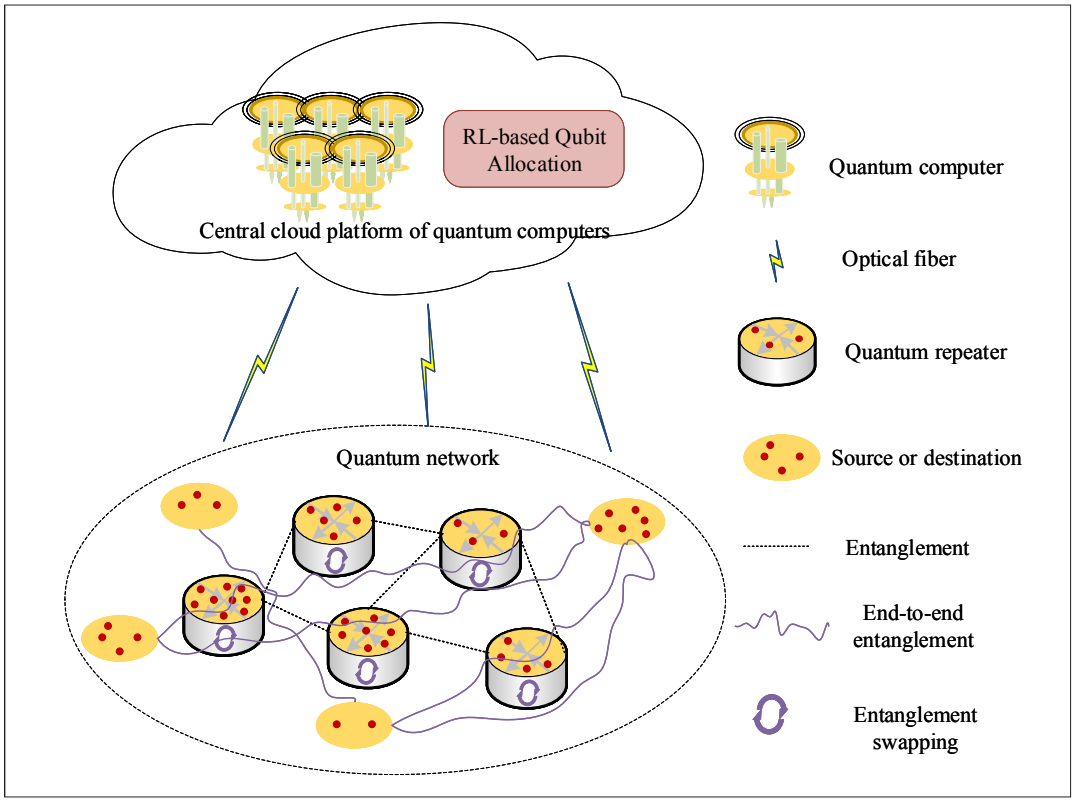

FIGURE 1. Overall structure of quantum networks.

tance end-to-end entanglement. Krastanov *et al.* [10] depart from the original purification for the entanglement state and consider the quantum hardware limitations to increase the stability for long-distance end-to-end entanglement.

For the qubit management in repeaters, the existing works all assume a fixed quantity or adequate storage. Moreover, they do not consider the cost of entanglement generation, which causes a large waste of photon resources in quantum nodes. In this article, we not only consider the efficiency and stability of long-distance end-to-end entanglement generation (fidelity, waiting time) but solve a qubit allocation problem among heterogeneous repeaters from a global view of the quantum network, based on the known throughput (end-to-end entanglement paths).

Preliminaries and Problem Definition

Preliminaries

Quantum Network: We consider a quantum network as a quantum communication system where a quantum processor (quantum computer) transfers information through qubits. It aims to satisfy multiple transmission requests each of which begins from the source quantum processor to the destination quantum processor via multiple media quantum processors (repeaters). Because of the inherent property of quantum entanglement (one-qubit transmission limitation), a request usually requires multiple end-to-end entanglement paths to achieve its task as fast as possible. The quantum network is a multi-photon and trusted-repeater. Each node stores a certain number of qubits to support entanglement generation. Repeaters have the privilege to obtain global entanglement information.

Entanglement Swapping: The entanglement swapping means that, a repeater with a certain number of qubits can turn two external entanglement links into an internal entanglement link between its belonged two qubits. Here, we only consider the bipartite entanglement swapping. The failures of entanglement generation/swapping are possibly caused by unsatisfied qubit state that is too weak to support the entanglement link, or when there is no idle qubit resource because of the competition of multiple entanglement chains at the same time. As shown below, we will use two indicators (the given constant t_{max} and the self-defined variable M) to account for the unsuccessful entanglement generation/swapping.

Entanglement Chain Protocol: We refer to an entanglement state between two quantum nodes without swapping operations as an "elementary link," and we refer to an end-to-end entanglement chain via multiple media repeaters as a "transmission link," which contains different number of segments of elementary links. To enhance the stability of one transmission link and save the resource cost, we do not consider the d-DIST-SWAP protocol, where transmission links with the same path are repeatedly produced and then turned them into a single high-stability transmission link by entanglement distillation operation [11]. Here, we only consider the SWAP-ONLY entanglement chain protocol that transmission links can only be produced by 2^d-segments elementary links with $2^d - 1$ entanglement swapping.

The Waiting Time and Fidelity: The waiting time of a transmission link with 2^d segments depends on the production of two 2^{d-1}-segments transmission link. Fidelity F, belonging to [0,1],

quantifies how different the output state is from the input state via the noisy quantum channel in entanglement transmission. It can decrease with the decay of quantum states because of the qubit's long-time storage in the memories [12].

Repeater's Qubit Competition: Costly limited by photon decay, each repeater is equipped with a certain number of qubits to support entanglement swapping. When multiple transmission links simultaneously occupy the same repeaters, a scenario of qubit competition occurs. Because a transmission link is an entanglement chain, it must be selected by all repeaters on the path.

Problem Definition

Given:

- An undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ regarded as the quantum network, where $\mathcal{V} = \{v_1, v_2, ..., v_n\}$ is the quantum node set with size of n and $\mathcal{E}$ is the edge set. Node j has its own memory capacity c_j.
- A transmission "link" set $\mathcal{L} = \{l_1, l_2, ..., l_m\}$ represents there are totally m transmission paths for multiple source-destination pairs to accommodate a given set of requests. $\mathcal{L}$ can be generated by using existing routing algorithms.[1] Because a transmission link can only transmit several qubits, a certain request requires different number of transmission links between a source-destination pair, due to its various requirement of information transmission. That is to say, m depends on the given set of requests and network conditions. Each transmission link can be denoted by its path information. We use a binary constant b_{ij}, $1 \le i \le m$, $1 \le j \le n$, which denotes whether the transmission link l_i passes through ($b_{ij} = 1$) the quantum repeater v_j or not ($b_{ij} = 0$). Then, the transmission link l_i can be described by $\{s_i, b_{i1}v_1, b_{i2}v_2, ..., b_{in}v_n, d_i\}$, where s_i and d_i are respectively the source and destination of its request. The response priority of a transmission link is also known to generate the elementary link.

Definition 1: Define the response process of $\mathcal{L}$ as a time slot interval $\mathcal{T}$ serialized by a period of an elementary link generation, denoted as $\{t_1, t_2, ...\}$.

On a time slot t, there exist three types of operations: elementary link generation, entanglement swapping with $2^{d_i^t}$ segments of transmission link l_i, the qubit consumption. Elementary link generation is prepared for the entanglement swapping on time slot $t + 1$. The entanglement swapping with $2^{d_i^t}$ segments is prepare for the entanglement swapping with $2^{d_i^{t+1}}$ segments. Each entanglement swapping consumes two qubits of the passing repeater.

Definition 2: Define the waiting time of the transmission link l_i as the time slot t_{com} when $2^{d_i^0} - 1$ swapping operations are all completed. After that, the occupied qubits are released instantaneously.

According to the state decay model, $\mathcal{F}(\rho(\omega), |\Phi^+\rangle\langle\Phi^+|) = (1 + 3\omega)/4$ and $\omega_{decay} = \omega \cdot e^{-\Delta t/Tcoh}$, proposed in [12] to describe the fidelity between the maximally-mixed state and Bell states. $\rho(\omega)$ denotes the single-hop (elementary link) entangled state (the mixed state during entanglement generation), Φ^+ and $\langle\Phi^+|$ are respectively the Bell state of two qubits, ω is a single parameter belonging to [0, 1], and T_{coh} is the joint coherence time of the two quantum repeaters holding the qubits. We can see that, the fidelity decreases as the waiting times increase. The total waiting time of $\mathcal{L}$ should be as short as possible to verify the fidelity effectiveness of each transmission link.

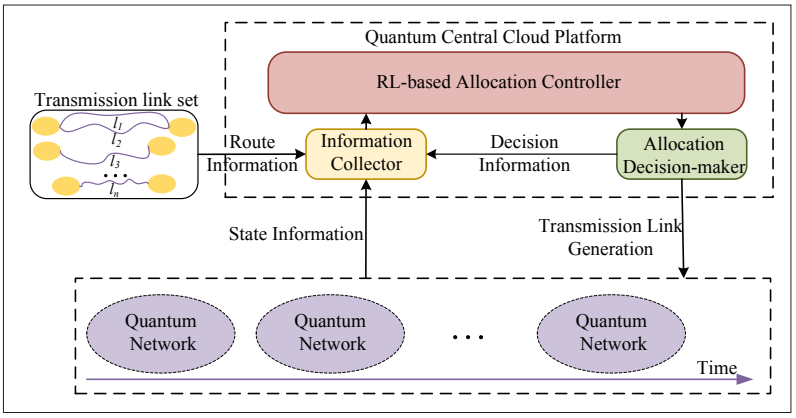

FIGURE 2. Overall structure of proposed qubit allocation model.

Based on the above definition, we define a problem to describe the qubit allocation in the quantum network as follows.

Problem 1: Based on the known transmission link set $\mathcal{L}$, all repeaters should determine how to cooperatively allocate qubits to deal with transmission links in $\mathcal{L}$ with the shortest average waiting time T. We call this problem as Cooperative-Qubit-Allocation-Problem (CQAP). There are two constraints in CQAP: satisfying the qubit capacity constraint for each quantum node, and determining the value of t_{com}^i, where if the transmission link l_i is not established successfully ($M < 1$), $t_{com}^i = t_{max}$. Otherwise ($M = 1$), t_{com}^i is set to be the current time t_{cur}. t_{max} is a given time threshold to terminate the solution process, and M is an intermediate variable.

Framework Design

Based on the known transmission link set $\mathcal{L}$, to capture the dynamics of the quantum network as the transmission links in $\mathcal{L}$ are responded, we serialize the whole response process into a series of time slots by taking an elementary link generation as a unit. Then, we model the CQAP as a Markov Decision Process and use RL algorithm to train a self-adaptive qubit allocation model in the central cloud which possesses a global view of each quantum node.

Figure 2 shows the overall structure of our proposed qubit allocation model. An RL-based Allocation Controller supported by Information Collector and Allocation Decision-maker runs in the quantum central cloud platform. In each time slot, state information combined with the given transmission link information and last-time-slot decision information is inputted into the information collector to integrate a data format for RL-based Allocation Controller. Then, Allocation Decision-maker transforms the output of RL-based Allocation Controller to the transmission link generation at the current moment, which changes the current state information into the next state information. Repeat the above process, until all the transmission links in $\mathcal{L}$ are responded with their $Ms = 1$ or the RL's updating process is compulsorily terminated.

[1] The adopted routing algorithms such as RL-based algorithms need to consider the properties of quantum networks, including the probability of successful path generation, the fidelity of elementary, and the transmission delay.

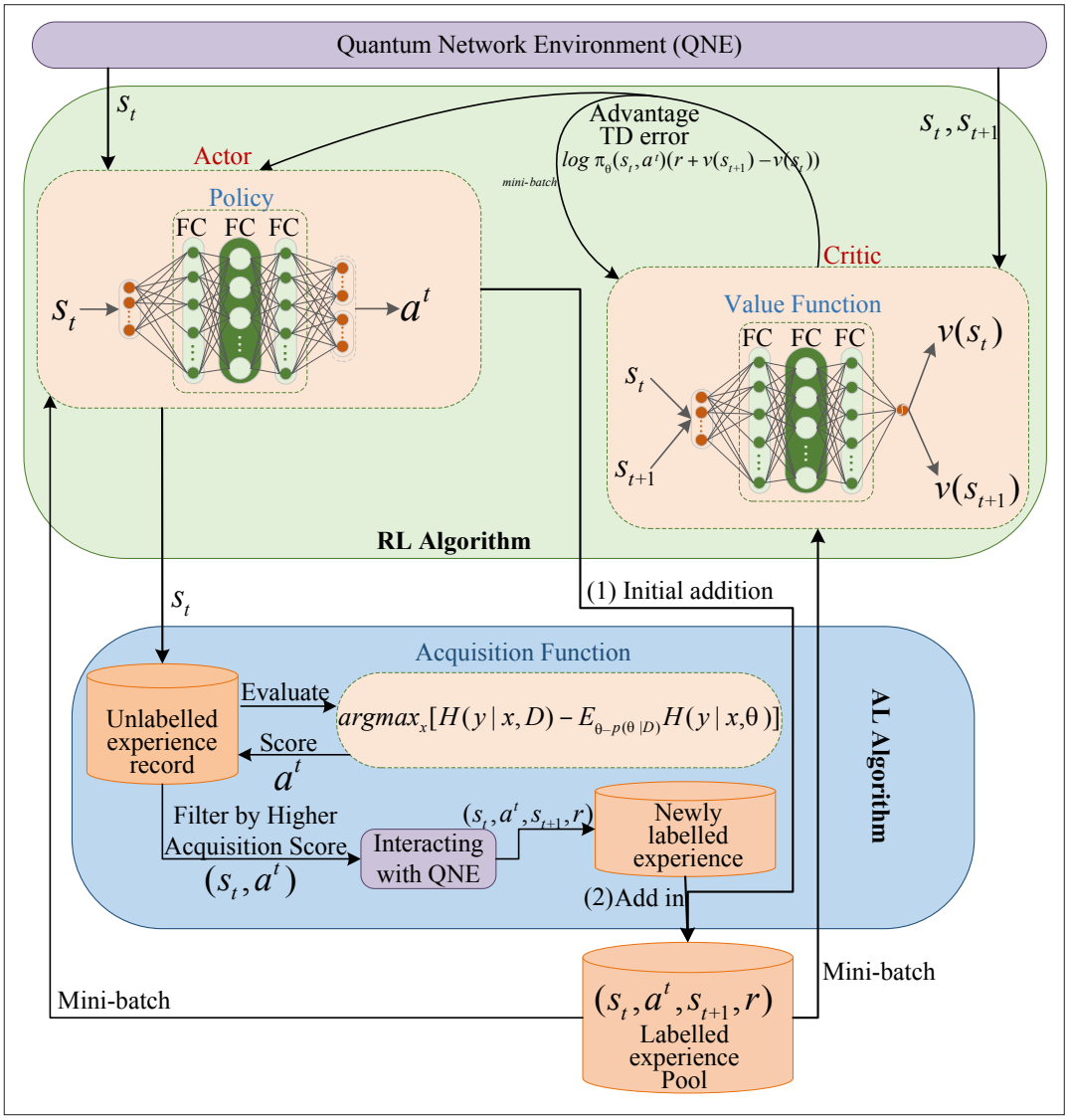

FIGURE 3. Structure of ALRL algorithm.

RL Details

RL algorithm accumulates the experience (state, action, next-time-slot state, and reward) through the interaction with the environment in each time slot. RL agent optimizes to maximize the accumulated reward to select the optimal action in an episode (lifespan). Taking the Advantage Actor-Critic (A2C) RL algorithm as an example, Fig. 3 shows the learning mechanism of RL. Actor network inputs s_t to obtain the action a^t, and Critic network inputs s_t and s'_t to obtain the value function $v(s_t)$ and $v(s'_t)$, which are used to calculate the advantage TD error for their gradient updating. Critic network can evaluate the performance of actor network through Advantage TD error loss. Their batch gradient descent operation depends on the mini-batch experience selected from the experience pool. Until an episode is completed, these two network models are concurrently updated. We provide an RL details for the defined CQAP as follows.

Action: We use an integer list $[a_{ij}^t]$, $1 \le j \le n$, $1 \le i \le m$ to describe the action space of RL agent, where a_{ij}^t is a binary variable which denotes whether repeater v_j supports transmission link $l_i (a_{ij}^t = 1)$ or not $(a_{ij}^t = 0)$ on time slot t. The size of action space is $n \cdot m$.

State: State information mainly consists of three parts:

- The information of transmission links $P_1^t \cup P_2^t \ldots \cup P_m^t$, where $P_i^t = \{b_{i1}v_1, b_{i2}v_2, \ldots, b_{in}v_n\}$, which is updated with time slots. When transmission link l_i has responded, $b_{ij} = 0$, $1 \le j \le n$.
- The residual qubits of each repeater $[\bar{c}_j^t]$, $1 \le j \le n$, where

$$\overline{c_j^t} = c_j - 2 \cdot \sum_{i=1}^{m} a_{ij}^t.$$

- The decision information of last time slot $[a_{ij}^{t-1}]$, which is used to record the agent information to supplement state feature. The size of state space is $2n \cdot (m + 1)$.

Reward: We define the reward as

$$\frac{1}{m}\sum_{i=1}^{m} M_i^t$$

to denote the average maturity of the transmission link set $\mathcal{L}$ in time slot t. Because the inter-

mediate variable M is related to the real-time variable a_{ij}, which is more elaborate than the waiting time to describe the real-time information for the elementary link generation.

AL-Based Improvement for RL

One-time transmission of an end-to-end entanglement consumes a certain number of qubit resources because the entanglement generation depends on the Bell-State Measurement of two decayed photons. More accurate generation of elementary links can bring considerable gains for network capacity. Hence, we try to decrease trial-error times and improve the utilization rate of history experience during the training process of RL algorithm through controlling its interaction with the quantum network environment.

Active Learning (AL) algorithm is commonly used to assist experts in labeling new images when the original images with their labels are insufficient for the training process of a deep learning algorithm in CV field [13], which can strongly reduce artificial cost. It guides the experts to label the images with more information, rather than the whole unlabelled data set. Here, we introduce a kind of AL algorithm based on uncertainty theory to assist the training process of RL, called Bayesian Active Learning by Disagreement (BALD) [14] AL algorithm. BALD focuses on minimizing the uncertainty of model parameter θ under the given labelled data. The uncertainty is measured by Bayesian conditional entropy (referred "Acquisition Function" in Fig. 3).

Taking A2C RL algorithm and BALD AL algorithm (we call it as "ALRL" algorithm) as an example, we demonstrate the AL-based improvement for RL in Fig. 3. Firstly, we regard the output (actions a_t) of actor network as the "label" and the input (state s_t) of actor network as the "image features" in AL. So the replay buffer pool of RL is transformed into Labelled experience pool with (s_t, a^t, s_{t+1}, r). Before t_{thd} time slot, the actor network is pre-trained with a mini-batch data selected from the initial-batch-size labelled experience pool. Then, s_t is taken as an input to the actor network with parameter θ_t to obtain its label a^t. Differently, we do not interact with the quantum network to implement the "state-action-state" transition and return reward r. We add a checkout before the interaction. We adopt Acquisition Function to calculate the uncertainty score of this unlabeled data s_t. Here, the "dropout" network structure is used to calculate the entropy by setting the dropout rate for each FC layer in actor network. The interaction can be implemented if the uncertainty score is bigger than the historical average score. Then, this record (s_t, a^t, s_{t+1}, r) is added in the "Newly Labelled Experience." Therefore, except for the initial addition, the "Labelled Experience Pool" can only be supplemented by uncertainty checkout.

Performance Evaluation

We execute the ALRL algorithm on the QUantum NETwork SIMulator (QuNetSim) platform, an opening quantum network model inspired by the OSI model [15], to solve the defined CQAP. The simulations are all executed with NVIDIA GeForce GTX 1660 SUPER in *Pytorch*.

One-time transmission of an end-to-end entanglement consumes a certain number of qubit resources because the entanglement generation depends on the Bell-State Measurement of two decayed photons.

Simulation Setup

For A2C, the learning rate is 0.001, the discounted factor is 0.99, the size of the minibatch is 128, and the capacity of labelled experience pool is 2000. For BALD, the initial batch size of labeled experience pool t_{thd} is 1000, the cell numbers of hidden FC layers are 128,256,128, and the dropout rates of there FC layers are 0.2, 0.4, 0.2.

In CQAP, we randomly select the priorities for all requests from [1, 9] using a uniform distribution. Multiple transmission links share the same priority of their common source and destination. The Poisson Distribution

$$P(X = k)\frac{\lambda^k}{k!}e^{-\lambda}$$

is adopted to generate the elementary link, where we regard the priority as λ and set k = 10. Then, we can obtain the generation probability of elementary link for each request with multiple transmission links. The memory capacities of repeaters are randomly selected from [10, 20] by a uniform distribution. We randomly generate 10-pair requests with 30 transmission links in a topology with 30 repeaters. We randomly generate the hop number of these links from $[2^2, 2^5]$. t_{max} is set to be 100, which also limits the iteration in an episode.

We evaluate the performance on the average maturity and the average waiting time for transmission link set $\mathcal{L}$ of ALRL, RL and Greedy-RL algorithms.

RL: We use the A2C algorithm to represent RL-class algorithm. Different from ALRL, it does not have the checkout process before the newly interacting record is added in experience pool.

Greedy-RL: In an iteration of policy network, repeaters compulsorily select the transmission link with the highest priority if the output of network does not conform the priority-greedy idea. Then, the data record generated by interacting with environment is stored in labelled experience pool.

Performances

Figure 4 shows the accumulated reward

$$\left(\sum_{t=0}^{t_{max}}\frac{1}{m}\sum_{i=1}^{m}M_i^t,\ \text{that's L} - \text{average } M,\right.$$
$$\left.\text{where } M \text{ denotes the maturity of a transmission link.}\right)$$

performance with the training episodes. We can see that, in Fig. 4a, ALRL has a similar reward performance with Greedy-RL on the initial episode state. Greedy-Rl performs a slight advantage over ALRL before 42 episodes. It is reasoned that Greedy action selection of request priority has some advantages when the actor-critic networks are not trained sufficiently, especially for RL. But Greedy-RL is quickly lost in a locally optimal solution even though critic network still assists to update actor (policy) network. Resorting to the experience check of Acquisition Function, ALRL performs better than RL before 200 episodes. But there has a similar convergence result at the end of the training process in Fig. 4a, which illustrates

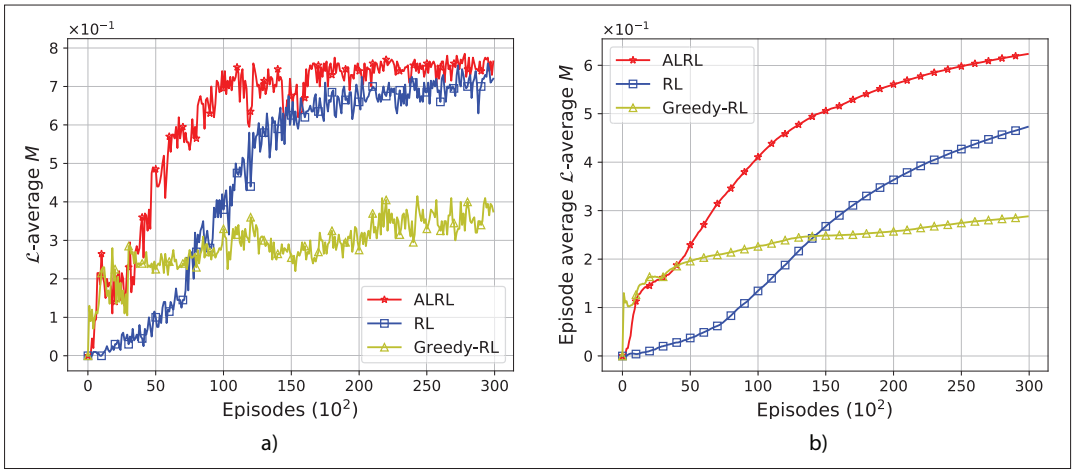

FIGURE 4. The performance of maturity: a) average maturity; b) episode average maturity.

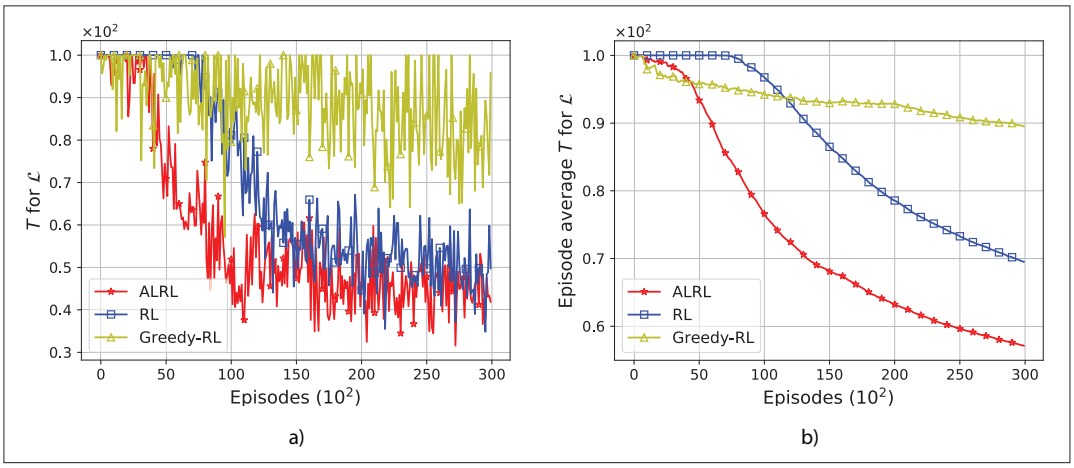

FIGURE 5. The performance of waiting time: a) average waiting time; b) episode average waiting time.

that ALRL can not enlarge the searching range of solution space. Overall, ALRL has a better average learning result as is shown in Fig. 4b.

We also evaluate the average waiting time

$$T = \frac{1}{m}\sum_{i=1}^{m} t_{com}^{i}$$

of the transmission link set L, shown in Fig. 5. We can see that, in Fig. 5a, T gradually decreases with the increase of training episode, which verifies the RL-based algorithms can capture the dynamic changes of the quantum network when the transmission link is responded in each time slot. Limited to the t_{max}, there exist some unideal episodes where all the transmission links are not responded with $T = t_{max}$ in a lifespan of the initial training stage. T fluctuates more violently than $\mathcal{L}$-average M, which has a negative impact on the algorithm convergence. This is the main reason why we define the intermediate index M. Figure 5b demonstrates a global view of the episode average T, from which we can see ALRL performs better than RL and Greedy-RL.

Conclusion

In this article, we study the Cooperative Qubit Allocation Problem (CQAP) in a multiple-repeaters quantum network to deal with the multiple end-to-end entanglements for different requests by considering the average waiting time and entanglement fidelity. We first define this problem as an optimization problem to minimize the average waiting time. Then, focusing on the dynamic changes in quantum networks caused by the segments of end-to-end entanglement, we serialize the response process by a unit of elementary link generation and adopt RL algorithm to adaptively learn the qubit allocation regulation of each quantum repeaters. To decrease the insignificant entanglement generation, we next use AL algorithm to filter the training data before updating the RL networks to improve the training efficiency. Simulation results demonstrate the effectiveness of our proposed algorithm for qubit allocation in adaptively and efficiently achieving high-stability entanglement communication and low-cost resource consumption.

Acknowledgments

The work of Song Yang is partially supported by the National Natural Science Foundation of China (NSFC) under Grant No. 62172038, 61802018. The work of Fan Li is partially supported by the NSFC under Grant No. 62072040. The work of Xiaoming Fu is partially supported by the EU H2020 RISE COSAFE project (No. 824019). Song Yang is the corresponding author.

References

[1] R. Horodecki *et al.*, "Quantum Entanglement," *Reviews of Modern Physics*, vol. 81, no. 2, 2009, p. 865.
[2] N. Lütkenhaus, J. Calsamiglia, and K.-A. Suominen, "Bell

Measurements for Teleportation," *Physical Review A*, vol. 59, no. 5, 1999, p. 3295.
[3] M. Zukowski *et al.*, "'Event-Ready-Detectors' Bell Experiment via Entanglement Swapping," *Physical Review Letters*, vol. 71, no. 26, 1993.
[4] A. Dahlberg *et al.*, "A Link Layer Protocol for Quantum Networks," *Proc. ACM Special Interest Group on Data Commun.*, 2019, pp. 159–73.
[5] X. Wang and M. M. Wilde, "Cost of Quantum Entanglement Simplified," *Physical Review Letters*, vol. 125, no. 4, 2020, p. 040502.
[6] Y. Zhao, G. Zhao, and C. Qiao, "E2e Fidelity Aware Routing and Purification for Throughput Maximization in Quantum Networks," *Proc. IEEE INFOCOM*, 2022.
[7] Y. Zhao and C. Qiao, "Redundant Entanglement Provisioning and Selection for Throughput Maximization in Quantum Networks," *IEEE Infocom 2021*, IEEE, 2021, pp. 1–10.
[8] S. Shi and C. Qian, "Concurrent Entanglement Routing for Quantum Networks: Model and Designs," *Proc. Annual Conf. ACM Special Interest Group on Data Commun. on the Applications, Technologies, Architectures, and Protocols for Computer Commun.*, 2020, pp. 62–75.
[9] Y. Lee *et al.*, "A Quantum Router Architecture for High-Fidelity Entanglement Flows in Quantum Networks," arXiv preprint arXiv:2005.01852, 2020.
[10] S. Krastanov, A. S. de la Cerda, and P. Narang, "Heterogeneous Multipartite Entanglement Purification for Size-Constrained Quantum Devices," *Physical Review Research*, vol. 3, no. 3, 2021, p. 033164.
[11] H.-J. Briegel *et al.*, "Quantum Repeaters: The Role of Imperfect Local Operations in Quantum Communication," *Physical Review Letters*, vol. 81, no. 26, 1998, p. 5932.
[12] S. Brand, T. Coopmans, and D. Elkouss, "Efficient Computation of the Waiting Time and Fidelity in Quantum Repeater Chains," *IEEE JSAC*, vol. 38, no. 3, 2020, pp. 619–39.
[13] S. Budd, E. C. Robinson, and B. Kainz, "A Survey on Active Learning and Human-in-the-Loop Deep Learning for Medical Image Analysis," *Medical Image Analysis*, vol. 71, 2021, p. 102062.
[14] N. Houlsby *et al.*, "Bayesian Active Learning for Classification and Preference Learning," arXiv preprint arXiv:1112.5745, 2011.
[15] S. DiAdamo *et al.*, "Qunetsim: A Software Framework for Quantum Networks," *IEEE Trans. Quantum Engineering*, 2021.

Biographies

Yanan Gao (yanangao@bit.edu.cn) is currently a Ph.D. student at the School of Computer Science and Technology, Beijing Institute of Technology. Her research interests include deep reinforcement learning, and its applications to cloud/edge computing, and quantum networks.

Song Yang (S.Yang@bit.edu.cn) is currently an associate professor in the School of Computer Science at Beijing Institute of Technology. He received his Ph.D. degree from Delft University of Technology, The Netherlands, in 2015. His research interests focus data communication networks, cloud/edge computing, and network function virtualization.

Fan Li (fli@bit.edu.cn) received her Ph.D. degree in computer science from the University of North Carolina at Charlotte in 2008. She is currently a professor with the School of Computer Science, Beijing Institute of Technology. Her current research focuses on wireless networks, smart sensing, crowd sensing, and mobile computing.

Xiaoming Fu (fu@cs.uni-goettingen.de) received his Ph.D. in computer science from Tsinghua University, Beijing, China, in 2000. He has been a professor in computer science in University of Göttingen since 2007. His research interests include network architectures, protocols, and applications. He is a fellow of IEEE, IET and member of the Academia Europaea.