Efficient Remote Entanglement Distribution in Quantum Networks: A Segment-Based Method

Zhonghui Li, *Member, IEEE*, Jian Li, *Member, IEEE*, Kaiping Xue, *Senior Member, IEEE*, David S. L. Wei, *Life Senior Member, IEEE*, Nenghai Yu, Qibin Sun, *Fellow, IEEE*, and Jun Lu

***Abstract*—Entanglement distribution between distant quantum nodes plays an essential role in realizing quantum networks' capabilities. In addition to path selection, remote entanglement distribution involves two pivotal quantum operations, i.e., entanglement generation and entanglement swapping. The existing studies mainly adopt two methods, i.e., Tell-and-Generation (TAG) and Tell-and-Swapping (TAS), to manage these two quantum operations on a selected path. However, both methods fatally introduce redundant stop-and-wait processes, which are detrimental to the performance of remote entanglement distribution in terms of latency and fidelity. To achieve low-latency and high-fidelity entanglement distribution between far-off quantum nodes, we propose a segment-based method consisting of an entanglement generation algorithm and a segment design to diminish the unnecessary stop-and-wait processes. The entanglement generation algorithm adopts a concurrent design to establish entanglement links using the one-demand generation model, thus effectively reducing waiting time compared to hop-by-hop and parallel designs. The segment design is proposed to split a long-distance path into multiple short-haul segments with the similar ability to swap entanglement, and these segments build multi-hop entanglement connections in parallel. Extensive simulations show that the segment-based method significantly outperforms the existing methods, including TAG and TAS, in entanglement distribution latency and effectively mitigates fidelity attenuation.**

***Index Terms*—Quantum networks, entanglement distribution, entanglement swapping, segment-based method.**

I. Introduction

CURRENTLY, quantum information technology has progressively become a hot research topic because of quantum superiority [1], [2], [3]. As the heart of the future success of quantum information technology, quantum networks [4], [5], [6] provide a promising platform for ground-breaking quantum applications, such as secure communication known for quantum key distribution [7], distributed quantum computing [8], and high-precision quantum sensing [9]. Notably, most of these quantum applications require distant quantum end nodes to maintain a peculiar physical phenomenon called entanglement [10]. Thus, *remote entanglement distribution*, a quantum technology for distributing entangled pairs between distant quantum end nodes to establish end-to-end entanglement connections, becomes one of the building blocks of quantum networks.

Although distributing entangled pairs between adjacent quantum nodes has been experimentally demonstrated [11], [12], [13], it is still challenging to achieve remote entanglement distribution due to the inherent photon loss in quantum channels. Fortunately, the distance limitation can be effectively overcome with the assistance of quantum repeaters. A quantum repeater is a physical device that can swap entanglement, facilitating the extension of entanglement distribution distance [14]. Different from the signal regeneration and amplification techniques adopted in classical communications, the swap operation, known as *entanglement swapping* [15], is essentially a local joint measurement (or swap operation) that can "teleport" an entangled particle separated from an entangled pair from a node to another one without suffering from channel noise, thus creating entanglement connections between two past un-entangled quantum nodes. Hence, instead of directly distributing entangled pairs over a long-length quantum channel, an efficient solution to achieve remote entanglement distribution between a *Source-Destination* (SD) node pair is to perform swap operations along a swapping path [16] consisting of multiple quantum repeaters.

Until now, some valuable studies have been done for achieving remote entanglement distribution in repeater-assisted quantum networks. These studies mainly focus on solving the problem of selecting swapping paths and entanglement resource allocation. Although these valuable studies pave the road for establishing end-to-end entanglement connections in quantum networks, they ignore the effect of the management of swap operations on remote entanglement distribution. In order to achieve efficient remote entanglement distribution in quantum networks, the swap operation management design deserves to be studied due to the inherent features of entanglement swapping.

Manuscript received 3 February 2023; revised 9 May 2023; accepted 3 July 2023. Date of publication 18 July 2023; date of current version 7 February 2024. This work is supported in part by Innovation Program for Quantum Science and Technology under grant No. 2021ZD0301301, Anhui Initiative in Quantum Information Technologies under grant No. AHY150300, Youth Innovation Promotion Association Chinese Academy of Sciences (CAS) under grant No. Y202093, and National Natural Science Foundation of China (NSFC) under Grants No. 61972371 and No. U19B2023. The associate editor coordinating the review of this article and approving it for publication was M. J. Khabbaz. *(Corresponding author: Kaiping Xue.)*

Zhonghui Li, Jian Li, and Qibin Sun are with the School of Cyber Science and Technology and the School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China (e-mail: kpxue@ustc.edu.cn).

Kaiping Xue and Nenghai Yu are with the School of Cyber Science and Technology and the School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China, and also with Hefei National Laboratory, Hefei 230088, China (e-mail: kpxue@ustc.edu.cn).

David S. L. Wei is with the Department of Computer and Information Science, Fordham University, Bronx, NY 10458 USA.

Jun Lu is with the School of Cyber Science and Technology, Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China.

Digital Object Identifier 10.1109/TNSM.2023.3296672

Two important problems caused by the inherent features of entanglement swapping must be noted for remote entanglement distribution. First, entanglement swapping is an imperfect operation because of the inherent limitations of quantum hardware [17]. If a quantum repeater fails to swap entanglement and the others on the swapping path still apply joint measurement to its local entangled particles, the failed swap operation introduces errors to the final entangled systems, i.e., end-to-end entanglement connections are established with low quality. Notably, improving the entangled system's quality will consume precious entanglement resources, thus resulting in low network throughput during a time slot. Hence, remote entanglement distribution needs to deal effectively with the imperfection of swap operations. Second, the asynchronous entanglement swapping design, an effective method to track swap operations, will result in competition for entanglement resources between quantum repeaters. Although quantum repeaters in a swapping path can perform swap operations using a parallel method to establish end-to-end entanglement connections, it is hard to track entanglement relationships, thus wasting entanglement resources due to failed swap operations. In contrast to parallel entanglement swapping, two quantum repeaters will compete to use the shared entangled pair, and one swap operation will inevitably be blocked. Overall, the two problems mentioned above significantly affect the implementation of remote entanglement distribution. Thus, to efficiently establish end-to-end entanglement connections, the swap operation management design that can timely detect the failure of entanglement swapping and avoid potential competition for entanglement resources between quantum repeaters is required to control swap operations [18].

There are two methods adopted to manage asynchronous swap operations. We call the first method *Tell-and-Generation* (TAG), i.e., each quantum repeater attempts to establish link-level entanglement with its successor only after its predecessor successfully swaps entanglement. In a nutshell, both entanglement generation and entanglement swapping are performed hop by hop in the TAG method. The other method is called *Tell-and-Swapping* (TAS), i.e., entanglement generation is performed in advance, and swap operations are performed hop by hop along the swapping path. These two methods inevitably introduce stop-and-wait processes, negatively affecting remote entanglement distribution. For the TAG method, each swap operation needs to wait for the successful entanglement generation before being performed. However, it is hard to establish link-level entanglement connections [19]. Hence, hop-by-hop entanglement generation significantly results in high remote entanglement distribution latency. For the TAS method, entangled pairs are stored in quantum memory and wait to be measured. Notably, the quality of an entangled system decays during the interaction between entangled particles and noisy quantum memory [20], [21]. Consequently, the TAS method contributes to end-to-end entanglement connections with low fidelity, thus impairing the performance of quantum applications [22]. Summarily, the stop-and-wait processes in TAG and TAS methods introduce redundant waiting time, thus doing harm to remote entanglement distribution. Hence, it is necessary to eliminate the unnecessary stop-and-wait processes.

This paper aims to realize low-latency and high-fidelity remote entanglement distribution in quantum networks. To achieve this goal, we propose a segment-based method to manage entanglement generation and entanglement swapping on the selected swapping path. The segment-based method consists of two essential parts: an entanglement generation algorithm and a segment design. The entanglement generation algorithm can realize the concurrent establishment of entanglement links on a swapping path. The segment design aims at dividing a long-distance swapping path into short-haul segments with a similar ability to establish multi-hop entanglement connections. In this way, swap operations can be performed in parallel at different segments to reduce the decoherence time of entangled pairs. Extensive simulation results reveal that the segment-based method significantly reduces the remote entanglement distribution latency compared with the existing methods and contributes to establishing high-fidelity end-to-end entanglement connections. We summarize the main contributions of this paper as follows:

- We discuss the performance of TAG and TAS methods in latency and fidelity. Then a segment-based method is proposed to achieve low-latency and high-fidelity remote entanglement distribution in quantum networks.
- The segment-based method first adopts a concurrent entanglement generations design and then attempts to perform swap operations in parallel by dividing a long-distance swapping path into multiple short-haul segments, thus reducing the unnecessary waiting time.
- Extensive simulations are performed to demonstrate the segment-based method's advantages in terms of latency and fidelity. Besides, we investigate the effect of segments' length on the segment-based method to improve this work in our future studies.

The remainder of this paper is organized as follows. Section II presents the background of our work. After that, two methods (TAG and TAS) and the performance analysis are elaborated in Section III. Section IV presents a segment-based method for remote entanglement distribution. Finally, extensive performance evaluations are presented in Section V, and the conclusions of this paper are drawn in Section VI.

II. Background

This section first reviews the fundamental knowledge of remote entanglement distribution. Furthermore, we introduce the network model and some assumptions considered in this paper. Finally, some related works are introduced.

A. Entanglement Generation

Entanglement, the most remarkable difference between classical and quantized mechanics, is a striking correlation between a pair of quantum particles, i.e., entangled pairs. The process of distributing entangled pairs between adjacent quantum nodes is known as entanglement generation. Some entanglement generation schemes have been demonstrated [23], [24], [25]. Here, we introduce a common scheme, i.e., single-atom excitation by the laser beam [26], to realize entanglement generation and entanglement distribution

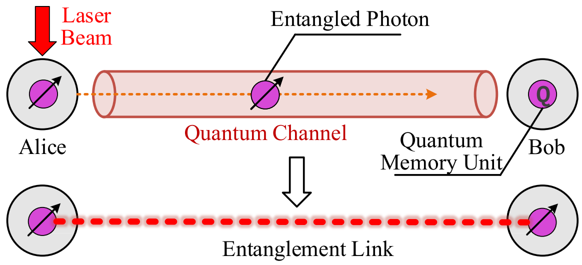


Fig. 1. The implementation of entanglement generation between adjacent quantum nodes based on single-atom excitation scheme.

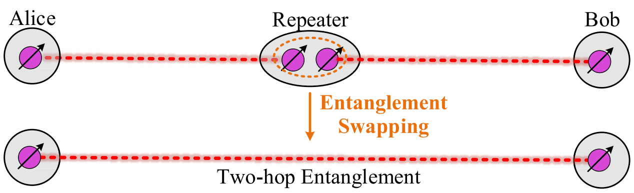


Fig. 2. Performing a swap operation to establish a two-hop entanglement connection between two distant quantum end nodes.

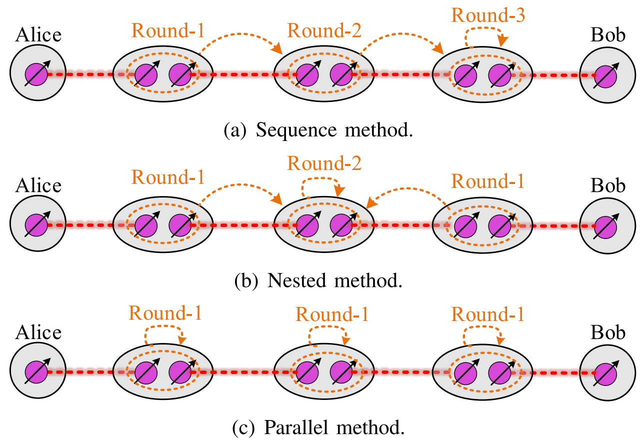


Fig. 3. Three methods for managing swap operations.

between directly linked quantum nodes using two cavities [27]. As shown in Fig. 1, atoms are excited by a laser beam, and then an atom-entangled photon departs from the cavity in Alice and is transmitted through a quantum channel, reaching another cavity placed in Bob. The photons are absorbed coherently in Bob, thus mapping the photon's polarization onto the state of remote atoms. As a result, a link-level entanglement connection (hereafter referred to as entanglement links) is created between Alice and Bob. Note that the success probability of link-level entanglement generation decays exponentially with the physical distance of a quantum channel [28]. Consequently, the process that a quantum node makes a series of attempts to establish an entanglement link introduces a non-negligible delay time.

B. Entanglement Swapping

Entanglement swapping is essential for building long-distance entanglement connections. Fig. 2 pictures the implementation of remote entanglement distribution between Alice and Bob with the aid of entanglement swapping. The intermediate quantum repeater is first entangled with Alice and Bob by sharing entangled photon pairs, respectively. Then the quantum repeater performs a local joint measurement on two un-entangled photons and sends two classical bits resulting from the joint measurement to Bob. According to the measurement outcome, Bob applies the corresponding quantum gate to manipulate the local entangled photon. In this way, the information of the entangled photon located at Alice is completely projected onto Bob's local entangled photon, i.e., Alice and Bob share an entangled pair. Consequently, Alice and Bob establish a two-hop entanglement connection. Hence, multiple entangled pairs on a swapping path can be "integrated" into long-distance entanglement connections by repeatedly performing swap operations [29]. Notably, entanglement swapping present the probability feature due to the imperfection of quantum devices and the failed swap operations introduce errors in entangled systems. Hence, a mechanism that can discover failed swap operations is required during remote entanglement distribution.

As shown in Fig. 3, there are three methods to manage swap operations along a swapping path consisting of multiple quantum repeaters. The first method is the sequence method, that is, swap operations are performed hop-by-hop along the swapping path. Hence, the entanglement swapping rounds equal the number of quantum repeaters on the swapping path. As shown in Fig. 3(a), establishing an end-to-end entanglement connection between Alice and Bob involves three rounds of entanglement swapping. The sequence method effectively avoids resource competition and facilitates the tracking of entanglement correlations, but it results in complex interactions during remote entanglement distribution. The second method, i.e., the nested method, is shown in Fig. 3(b). For this method, the predecessor and successor of a quantum repeater perform swap operations simultaneously to create two-hop entanglement in each swapping round. An end-to-end entanglement connection can be established after multiple rounds of nested entanglement swapping. This method can improve the performance of remote entanglement distribution while avoiding resource competition. The last method adopts a parallel manner in which all quantum repeaters perform local measurements to swap entanglement simultaneously, and then Bob establishes entanglement connections with Alice based on the convergent measurement results. Hence, only one round of entanglement swapping is required, as shown in Fig. 3(c). The parallel method significantly simplifies the management of swap operations but contributes to inefficient remote entanglement distribution due to the probability nature of entanglement swapping.

C. Quantum Decoherence and Fidelity

As one of the key obstacles in the future application of quantum information technology, *quantum decoherence* is a process in which quantum systems interact with noisy environments, causing the irreversible loss of quantum properties [30]. Generally, quantum decoherence appears whenever quantum systems are coupled with noisy environments such as quantum channels and quantum memory. In other words, quantum decoherence exists in the generation, transmission, storage, and measurement of quantum systems [31]. The most direct impact of quantum decoherence is introducing information errors in quantum systems, thus hindering quantum information technology from moving toward practical application.

A fundamental but meritorious figure, i.e., fidelity [32], can indirectly reflect the severity of errors in a quantum system. Fidelity is used to measure the information coincidence degree of two quantum systems, taking values between 0 and 1.0. Note that the fidelity will decay with the time the quantum system waits to be measured. The lower the fidelity, the more quantum properties are lost. When the fidelity of a quantum system equals zero, a quantum-to-classical transition will happen [33], i.e., the information of the quantum system becomes deterministic in the same way as a classical bit. In this paper, we use fidelity to measure the quality of an entangled system. The fidelity of the entangled systems established between two quantum end nodes affects the performance of quantum applications, e.g., low-fidelity entanglement results in inefficient quantum teleportation. Consequently, establishing high-fidelity end-to-end entanglement connections is an essential task in quantum networks, and it is also one of the goals of our work.

D. Network Model

To facilitate the statement of our work, notations and explanations are first shown in Table I. In this paper, we focus on the management of quantum operations (including entanglement generation and entanglement swapping) during the establishment of end-to-end entanglement connections in quantum networks. Most notably, remote entanglement distribution requires a large number of classical information interactions between quantum nodes. In order to simplify the complexity of remote entanglement distribution designs in quantum networks, we adopt a classical centralized controller to select the swapping path and manage quantum operations. The interaction between the classical controller and quantum nodes is realized through the classical Internet.

For a given request, a swapping path $p = \{v_0, v_1, \ldots, v_{s-1}, v_s\}$ is selected by the centralized controller, where v_0 and v_s is an SD pair while s denotes the length of the swapping path, i.e., $|p|$. Besides, entanglement generation is implemented by the single-atom excitation scheme, and uses a deterministic method [34]. Each entanglement generation endeavor independently succeeds with a constant probability determined by the physical medium. We denote the success probability of an entanglement generation endeavor between adjacent nodes v_u and v_v as $r_{(u,v)}$, and the total time spent for each endeavor is denoted as $t_{(u,v)}$. $t_{(u,v)}$ is a distance-dependent constant and is an expected value due to the imperfection of quantum hardware. Moreover, a quantum node can establish multiple entanglement links with its neighbors, and the number of entanglement links varies between different adjacent quantum node pairs.

We model an entangled pair shared by any two adjacent quantum nodes v_u and v_v as a Werner state $\rho(w_{(u,v)})$ with an initial Werner parameter $0 \leq w_{(u,v)} \leq 1$ [35] after the successful entanglement generation:

$$\rho\Big(w_{(u,v)}\Big) = w_{(u,v)}\big|\Phi^+\big\rangle\big\langle\Phi^+\big| + \Big(1 - w_{(u,v)}\Big)\frac{\mathbb{1}_4}{4},$$

where $\left|\Phi^+\right\rangle$ is a Bell state and $\mathbb{1}_4/4$ is the maximally-mixed state on two photons. In this paper, we use the Werner parameter to indicate the state of an entangled system since a Werner state is completely determined by its Werner parameter. The swapping operation successfully performed on two Werner states $\rho(w_{(u,v)})$ and $\rho(w_{(v,k)})$ will generate a new Werner state that can be written as

$$\rho\Big(w_{(u,k)}\Big) = \rho\Big(w_{(u,v)} \cdot w_{(v,k)}\Big). \quad (1)$$

The degree of information concordance between a Werner state $\rho(w_{u,v})$ and Bell state $\left|\Phi^+\right\rangle$ is defined as the fidelity and is equal to $\langle\Phi^+|\rho(w_{(u,v)})|\Phi^+\rangle$, i.e., $(1 + 3w_{(u,v)})/4$.

Moreover, we assume that the entangled pairs shared by the same pair of adjacent quantum nodes show the same fidelity after entanglement generation, but the entangled pairs shared by different pairs of adjacent quantum nodes are not identical different in entangled systems' fidelities due to differences in physical device configurations. Notably, the fidelity of the

TABLE I
NOTATIONS AND EXPLANATIONS

Notation	Explanation
p	The swapping path selected for an SD pair.
$\|p\|$	The length of the swapping path p.
v_i	The i-th quantum node in the selected swapping path.
$r_{(i,i+1)}$	The success probability of establishing an entanglement link between v_i and v_{i+1}.
$t_{(i,i+1)}$	The expected time spent for establishing an entanglement link between v_i and v_{i+1}.
$w_{(i,j)}$	The Werner parameter of the original Werner state shared by quantum nodes i and j.
Δt	The storage time of an entangled pair.
T_{coh}	The coherence time of two quantum memory units.
$w'_{(i,j)}$	The Werner parameter of the decoherent Werner state shared by quantum nodes i and j.
$T_{(i,i+1)}$	The total expected time spent for establishing multiple entanglement links between v_i and v_{i+1}.
T_{max}	The maximum value of the duration set $\{T_{(0,1)}, \cdots, T_{(s-1,s)}\}$.
t_{ret}	The time spent for retrieving an entangled photon from a quantum memory.
t_{mea}	The expected time spent for jointly measuring the local entangled photons to swap entanglement.
t_{max}	The maximum value of the expected time spent for an successful entanglement generation.
L_{TAG}^{single}	The time spent for extending entanglement distribution distance in each hop using the TAG method.
L_{TAG}	The latency of remote entanglement distribution between an SD pair using the TAG method.
L_{TAS}	The latency of remote entanglement distribution between an SD pair using the TAS method.
$w_{(0,s)}^{TAG}$	The Werner parameter of the final entangled system established with the TAG method.
$w_{(0,s)}^{TAS}$	The Werner parameter of the final entangled system established with the TAS method.
L_{SEG}	The latency of remote entanglement distribution between an SD pair using the segment-based.
$w_{(0,s)}^{SEG}$	The Werner parameter of the final entangled system established with the segment-based method.
N_p	The number of entanglement swapping rounds.
T_k	The total decoherence time of all entangled pairs in a swapping round with k segments.

Werner state will decay during interacting with a noisy environment. After a storage time Δt in an imperfect quantum memory, the Werner parameter of the Werner state can be expressed as [36]:

$$w'_{(u,v)} = w_{(u,v)} \cdot e^{-\Delta t / T_{coh}}, \tag{2}$$

where T_{coh} is the coherence time of two quantum memory units holding a pair of entangled photons and is a constant determined by the physical medium.

E. Related Work

In recent years, some prominent works have been done on remote entanglement distribution in quantum networks. These studies mainly consider two entanglement generation models, i.e., on-demand generation and advance generation [37], [38]. For the on-demand generation model, the swapping path is first selected, and then entanglement generation is performed on demand along the selected path. For the advance generation model, adjacent quantum nodes establish entanglement links before path selection. Based on two models, [39] first introduces connection-oriented and connectionless remote entanglement distribution strategies inspired by classical communications. Besides, other studies have also explored the implementation of remote entanglement distribution based on these two entanglement generation models.

Based on the on-demand generation model, [40] attempts to apply the Dijkstra algorithm to select paths for remote entanglement distribution; [41] introduces a multi-path routing algorithm in a diamond topology; [42] adopts a greedy solution for path selection in grid topology; [43] introduces an opportunistic method to select quantum nodes, providing an easily-applicable, moderate-complexity solution for remote entanglement distribution. Reference [44] presents a design to select the swapping path and assign the established entanglement links to concurrent SD pairs' requests; Reference [45] adopts the idea of network flow for path selection and resource allocation to maximize the throughput for multiple SD pairs. Reference [46] introduces a fidelity-aware scheme for path selection and entanglement link allocation to optimize network throughput.

There are some highlighting works in remote entanglement distribution using the advance generation model. Reference [47] studies the request scheduling design to decide which SD pairs with different resource demands to be served in a resource-limited quantum network. Reference [48] extends the entanglement distribution design presented in [42] by allowing quantum memories to store onto entangled pairs for multiple time slots. Reference [49] proposes a fidelity-guaranteed design, Q-PATH, to select the swapping path and decide the number of entanglement links allocated to each SD pair by each hop based on fidelity requirements. Reference [50] discusses three resource allocation strategies for multi-path, multi-request, and multi-channel scenarios in a resource-limited quantum network and evaluates the performances of these strategies in terms of fairness, resource utilization, delay, and throughput. Reference [51] presents a swapping-based entanglement routing design that aims to select swapping paths and allocate entanglement link resources for multiple SD pairs to mitigate congestion.

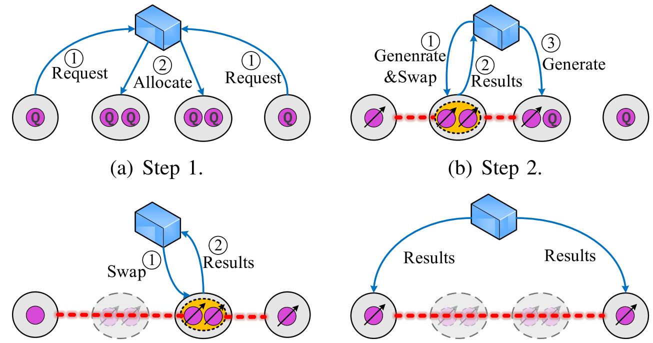


Fig. 4. The process of the TAG method for remote entanglement distribution.

Although these valuable efforts paved the road for remote entanglement distribution in quantum networks, they paid little attention to managing swap operations. The existing studies usually adopt the parallel method to realize entanglement swapping on the selected path. However, it is challenging to track entanglement connections and find failed swap operations timely, thus resulting in the waste of entanglement resources and high errors in the final entangled pairs shared by an SD pair. In order to effectively track entanglement connections, an asynchronous entanglement swapping design is first considered, and two famous methods are the TAG method using the on-demand generation model and the TAS method using the advance generation model. However, these methods inevitably introduce stop-and-wait processes, damaging remote entanglement distribution. In this paper, we introduce a segment-based method to provide an effective solution for low-latency and high-fidelity entanglement distribution in quantum networks.

III. TAG and TAS Remote Entanglement Distribution Methods

This section starts with an introduction to the TAG and TAS methods adopted for achieving remote entanglement distribution in quantum networks. Furthermore, a comprehensive performance analysis of the latency and fidelity of these two methods is presented using a mathematical method.

A. TAG Method

The TAG method adopts the on-demand generation model and the sequence entanglement swapping method to realize remote entanglement distribution. Entanglement generation and swap operations are performed hop by hop along the swapping path. For the TAG method, the establishment of k-hop entanglement connections between the source node v_0 and k-th quantum repeater v_k needs to wait for the successful entanglement generation between v_k and v_{k-1}. Consequently, the entanglement generation operation inevitably introduces a waiting time between two adjacent swap operations.

We elaborate on the process of the TAG remote entanglement distribution method as follows (shown in Fig. 4):

1) An SD pair sends a request to the centralized controller (the blue cuboid in Fig. 4(a)) for remote entanglement distribution. Then, a swapping path is selected by a routing algorithm running on the controller, and each quantum repeater allocates the dedicated quantum memory units for the SD pair to store entangled photons.
2) The centralized controller informs the first quantum repeater attempts to establish entanglement links with its two neighbors. After entanglement generation on both sides are successful, a swapping operation (yellow ellipse in Fig. 4(b)) is performed, and the results of the swap operation are fed back to the centralized controller. Then, the centralized controller tells the next quantum repeater to attempt to establish entanglement links with the destination node.
3) The second quantum repeater waits to measure its local particles after the following entanglement link is built (Fig. 4(c)) and sends measurement results to the centralized controller. If the number of quantum repeaters exceeds two in the swapping path, the remaining quantum repeaters perform the above operations repeatedly.
4) After receiving the feedback of the swap operation performed by the last quantum repeater, the centralized controller sends the outcomes to the SD pair. Finally, the source node can establish end-to-end entanglement connections with the destination node (Fig. 4(d)).

The TAG method facilitates the detection of failed swap operations and can effectively avoid the competition of entanglement resources between quantum repeaters. However, the hop-by-hop operation method contributes to high latency for remote entanglement distribution. Besides, to simply the description of the process of the TAG method, we do not take the fact that entanglement swapping is an imperfect operation into account. However, the failed swap operations significantly harm the performance of remote entanglement distribution. Suppose a situation in the TAG method where the k-th $(k \geq 1)$ quantum repeater in the swapping path fails to swap entanglement. In this case, entanglement generation and entanglement swapping would be re-executed along the selected swapping path. As a result, each SD pair takes more time to establish an end-to-end entanglement connection and consumes more entanglement links.

B. TAS Method

The TAS method adopts the advance generation model and the sequence entanglement swapping method to realize remote entanglement distribution. Unlike that entanglement generation and swap operations are performed hop-by-hop in the TAG method, swap operations and entanglement resource allocation are alternately performed hop-by-hop in the TAS method. After entanglement swapping, each quantum repeater in the swapping path will tell its successor through the centralized controller to allocate entanglement links resources for an SD pair. In a nutshell, entanglement swapping will wait to be executed until the entangled pairs on the following quantum link have been successfully assigned to an SD pair.

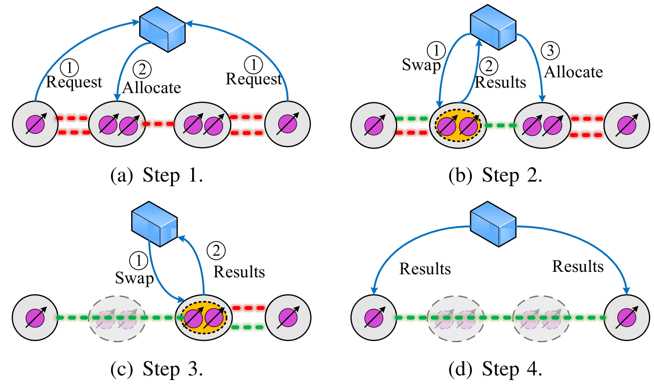


Fig. 5. The process of the TAS method for remote entanglement distribution.

Explicitly, as shown in Fig. 5, the process of the TAS remote entanglement distribution method is concluded as follows:
1) Initially, any pair of adjacent quantum nodes spontaneously attempts to build entanglement links. When the centralized controller receives a request from an SD pair, a swapping path connecting the source and destination nodes is determined. Then, the centralized controller notifies the first quantum repeater of the swapping path to allocate entanglement resources to the SD pair on both the left and right quantum links (Fig. 5(a)).
2) After the first repeater allocates entanglement resources (green dashed line in Fig. 5(b)) to the SD pair on both the left and right quantum links, the controller informs the first quantum repeater to swap entanglement. Then, a pair of un-entangled photons is retrieved from quantum memory units and measured jointly. The measurement results are sent to the centralized controller.
3) After successful entanglement swapping, the centralized controller tells the next quantum repeater to designate entanglement links to the SD pair. Then, the next quantum repeater performs entanglement swapping (Fig. 5(c)) and sends the results of swap operations to the centralized controller. The remaining quantum repeaters in the swapping path iteratively perform the above operations to swap entanglement.
4) After receiving the results of the swap operation performed by the last quantum repeater, the centralized controller sends the results of remote entanglement distribution to the SD pair. Fig. 5(d) finally, end-to-end entanglement connections can be established.

Compared to the TAG method, the TAS method can reduce the waiting time introduced by entanglement generation between adjacent quantum nodes, but at the cost of degraded entanglement fidelity. Besides, the imperfect swap operations also affect the implementation of the TAS method. Suppose the scenario in the process of TAS where the k-th $(k \geq 1)$ quantum repeater in the swapping path fails to swap entanglement. In this case, the first $k+1$ quantum repeaters need to reallocate link-level entanglement resources and re-perform swap operations. Besides, if there are no shared entangled pairs between adjacent quantum repeaters that can be reassigned to the SD pair, entanglement generation will be triggered to prepare entanglement links. As a result, the entangled pairs used for building end-to-end entanglement connections will experience long decoherence times, i.e., end-to-end entanglement connections are built with low fidelity.

C. Performance Analysis

We focus on the performances of the TAG and TAS remote entanglement distribution methods in terms of the latency and the Werner parameter of the final Werner state. For simplicity, we ignore the delay time of classical communication between quantum nodes in our work and mainly consider the delay time caused by quantum operations. Besides, the latency of remote entanglement distribution is calculated, assuming each swap operation is successfully performed.

Latency: For the TAG method, all quantum operations are executed hop-by-hop along the swapping path. Hence, the total latency, L_{TAG}, of the TAG remote entanglement distribution method in the swapping path *p*, is a linear accumulation of link-level delay times. As mentioned above, the link-level latency consists of the time spent for entanglement generation, the time spent for local measurement, and the latency of entangled photons retrieval. We formulate the duration of the *k*-th TAG process, i.e., *k*-th quantum repeater performs entanglement swapping, and the next quantum repeater successfully builds an entanglement link, as

$$L_{TAG}^{single} = t_{(i,i+1)} + t_{mea} + t_{ret}. \tag{3}$$

Thus, the time spent for building an end-to-end entanglement connection via the swapping path *p* can be written as

$$L_{TAG} = \sum_{i=0}^{|p|-1} t_{(i,i+1)} + (|p|-1)\cdot(t_{mea} + t_{ret}). \tag{4}$$

In the TAS method, entanglement generation between any pair of adjacent quantum nodes is performed before selecting a swapping path, and entanglement swapping is executed hop-by-hop. In this way, the time spent for entanglement generation between adjacent quantum nodes can be ignored for remote entanglement distribution between each SD pair. Hence, the latency of remote entanglement distribution mainly consists of the time spent for entangled photon retrieval and joint measurement. Most notably, swap operations and entangled photons retrieval in the TAS method takes the same duration as the TAG method. As a result, we can write the latency of distributing an entangled pair between an SD pair via the swapping path *p* as

$$L_{TAS} = (|p|-1)\cdot(t_{mea} + t_{ret}). \tag{5}$$

Werner parameter: For the TAG method, the entanglement link established after entanglement generation will be measured immediately to swap entanglement. In this, the Werner parameters of entanglement links can be considered hypothetically unchanged. However, the multi-hop entanglement connection established after performing multiple swap operations needs to wait for the successful generation of an entanglement link. Hence, the attenuation of entanglement fidelity mainly comes from the quantum decoherence of the extended multi-hop entanglement connections in the TAG method. Before each swap operation, the decoherence time is the sum of the time spent for next-hop entanglement generation and entangled photons retrieval. According to Eq. (2), for the *k*-th swap operation, the Werner parameter of the entangled system shared by nodes v_0 and v_k is

$$w'_{(0,k)} = w_{(0,k)} \cdot e^{-\left(t_{(k,k+1)}+t_{ret}\right)/T_{coh}}.$$

After entanglement swapping, the new Werner state is $\rho(w'_{(0,k)} \cdot w_{(k,k+1)})$ according to Eq. (1). As a result, the Werner parameter of an end-to-end entanglement connection established by the TAG method via the swapping path *p* can be written as

$$w_{(0,s)}^{TAG} = \prod_{i=0}^{|p|-1} w_{(i,i+1)} \cdot e^{-\left[\sum_{j=1}^{|p|-1} t_{(j,j+1)}+(|p|-1)t_{ret}\right]/T_{coh}}. \tag{6}$$

For the TAS method, entanglement swapping is performed hop-by-hop along a swapping path after entanglement generation. Quantum decoherence mainly exists in two parts, i.e., before path selection and after path selection. Before path selection, the entangled pair distributed first will wait for all entangled pairs on the swapping path to be successfully distributed. Note that each entanglement generation endeavor is independent. Consequently, the quantum decoherence times of different entangled pairs shared by v_i and v_{i+1} are different. The interval time between two successful entanglement generation endeavors is $t_{(i,i+1)}$. Assume that the number of entangled pairs needs to be shared between adjacent nodes v_i and v_{i+1} is $n_{(i,i+1)}$, so the time spent for entanglement generation, $T_{(i,i+1)}$, is equal to $t_{(i,i+1)} \cdot n_{(i,i+1)}$, and T_{max} equals the maximum value of the set $\{T_{(0,1)}, \ldots, T_{(s-1,s)}\}$. For simplicity, we assume that the last created entanglement link is allocated to the SD pair for entanglement swapping in each hop. Therefore, the storage time Δt of the assigned entanglement link established between v_i and v_{i+1} before path selection is

$$\Delta t = T_{max} - T_{(i,i+1)}. \tag{7}$$

According to Eq. (2), the Werner parameter of the entanglement link after duration Δt equals

$$w'_{(i,i+1)} = w_{(i,i+1)} \cdot e^{-\left[\left(T_{max}-T_{(i,i+1)}\right)\right]/T_{coh}}. \tag{8}$$

After path selection, each entangled pair stored in quantum memory units will wait to be measured. According to the above analysis of the delay time in the TAS method, for *k*-th swap operation, the Werner parameter of the last created entanglement link between v_i and v_{i+1} is $w'_{(k,k+1)} \cdot e^{-[(k-1)\cdot t_{mea}+k\cdot t_{ret}]/T_{coh}}$.

Hence, integrating the quantum decoherence time of two parts, the Werner parameter of the end-to-end entanglement connection established by the TAS method is

$$w_{(0,s)}^{TAS} = \prod_{i=0}^{|p|-1} w'_{(i,i+1)} \cdot e^{-\left[(i-1)\cdot t_{mea}+i\cdot t_{ret}\right]/T_{coh}}. \tag{9}$$

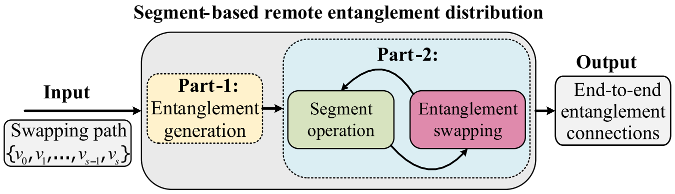


Fig. 6. The implementation of the segment-based method for establishing end-to-end entanglement connections.

D. Discussion

Comparing Eq. (4) and Eq. (5), we can find that the TAS method outperforms the TAG method in remote entanglement distribution latency since it is challenging to establish an entanglement link between adjacent quantum nodes. Nevertheless, it is hard to conclude which of the two methods performs better in the Werner parameter because the superiority of these two methods is associated with the number of entangled pairs that need to be distributed per hop in the TAS method. Notably, the TAG and TAS methods adopt the sequence entanglement swapping method, which will lead to unnecessary delay times caused by stop-and-wait processes. Consequently, although the two methods facilitate the management of quantum operations, they inevitably impair the performance of remote entanglement distribution in terms of latency and fidelity. The nested entanglement swapping method can effectively track entanglement connections while reducing unnecessary stop-and-wait processes. Hence, we propose a segment-based remote entanglement distribution method using the nested-based method to manage swap operations on a swapping path. Furthermore, we introduce an algorithm to realize concurrent entanglement generation based on the on-demand generation model, thus reducing latency and quantum decoherence compared to hop-by-hop and parallel entanglement generation methods. We elaborate on the segment-based method in the following section.

IV. Segment-Based Remote Entanglement Distribution Method

This section describes the segment-based remote entanglement distribution method, including an entanglement generation algorithm and a segment design. The entanglement generation algorithm can diminish unnecessary waiting time by creating entanglement links along a swapping path in parallel. The segment design is applied to split a long-distance swapping path into multiple short-haul segments, thus performing swap operations in parallel at different segments to reduce unnecessary stop-and-wait processes.

A. Overview

The segment-based method can reduce redundant stop-and-wait processes during remote entanglement distribution by adopting the concurrent entanglement generation design and the nested entanglement swapping method. After path selection, this method involves two essential parts, as shown in Fig. 6. The first part introduces the entanglement generation algorithm, which adopts the concurrent entanglement generation design to create entanglement links between adjacent quantum nodes along a swapping path. The second part consists of multiple swapping rounds. In each swapping round, a long-distance swapping path is split into multiple short-haul segments by performing a segment operation that takes the minimum component unit of entanglement swapping as a segment. Then, all segments perform swap operations in parallel to establish multi-hop entanglement connections, and quantum repeaters that perform swap operations in segments are removed from the swapping path to form the swapping path of the successor swapping round. Entanglement swapping and segment operations are performed alternately to establish end-to-end entanglement connections. If a segment fails to swap entanglement in a swapping round, the multi-hop entanglement connection within this segment can be timely re-established with the aid of a concurrent entanglement generation design and the nested entanglement swapping method. Hence, the segment-based method can effectively reduce unnecessary waiting times and fidelity attenuation in remote entanglement distribution between SD pairs.

B. Entanglement Generation Algorithm

Since the success probability of entanglement generation decays exponentially with the physical length of a quantum channel, it is challenging to distribute an entangled pair between adjacent quantum nodes. Hence, the hop-by-hop entanglement generation method contributes to high remote entanglement distribution latency. Besides, different pairs of adjacent quantum nodes vary in the expected time spent for successfully establishing an entanglement link. Hence, parallel entanglement generation will result in waiting events, i.e., the entanglement link established first needs to wait for its neighbors to be established, thus introducing unnecessary decoherence time. To efficiently establish end-to-end entanglement connections, it is required to minimize the decoherence time of entangled pairs before being measured as much as possible. Here, we adopt a concurrent entanglement generation design to diminish the unnecessary waiting time. The core idea of the entanglement generation algorithm is to enable the shared entangled pair between different pairs of adjacent quantum nodes to be successfully distributed simultaneously. In a nutshell, instead of using the store-and-wait manner in which the entanglement link created first is stored in quantum memory units and waits for the other entangled pairs to be successfully distributed, two un-entangled photons dispatched from two different entangled pairs can simultaneously reach the intermediate quantum node where local measurement is performed. In this way, link-level entanglement resources can be used directly for entanglement swapping after successful entanglement generation.

Here, we present the entanglement generation algorithm, as shown in Algorithm 1. The entanglement generation algorithm adopts the idea that the entanglement generation between adjacent quantum nodes with the longest expected duration is performed first to distribute entangled pairs. For a selected swapping path, the entanglement link with the highest

Algorithm 1: Entanglement Generation Algorithm

Input: $p = \{v_0, v_1, \ldots, v_s\}$, $t_{(i,i+1)}$ $(i = 0, 1, \ldots, s-1)$;
Output: The time when each entangled pair starts to try to be distributed;
1 $t_{max} \leftarrow$ calculate the maximum value of $t_{(i,i+1)}$;
2 $t_{int} \leftarrow$ set the initialization time to 0;
3 **if** $t_{(i,i+1)}$ *is equal to* t_{max} **then**
4   Attempt to create link-level entanglement between adajcent quantum nodes v_i and v_{i+1} first;
5 **end**
6 **while** $t_{int} \neq t_{max}$ **do**
7   **if** $t_{(i,i+1)}$ *is equal to* $t_{max} - t_{int}$ **then**
8   Attempt to create link-level entanglement between adajcent quantum nodes v_i and v_{i+1};
9   **end**
10     $t_{int} \leftarrow t_{int} + 1$;
11 **end**

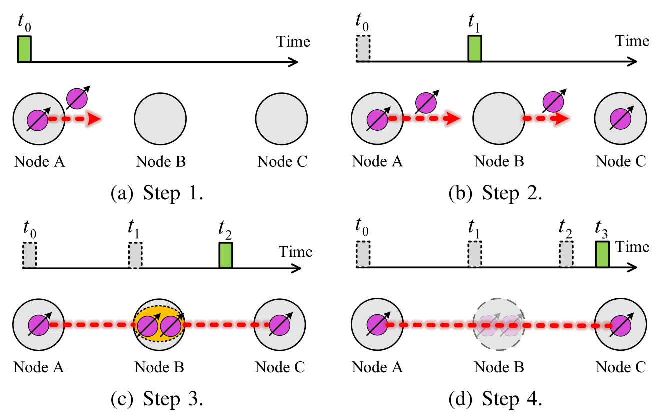


Fig. 7. Parallel entanglement generation.

establishment duration is first attempted to be created (Lines 3-5). Before the entanglement link is successfully built, other quantum nodes attempt to perform entanglement generation to create entanglement links in turn (Lines 6-10). The time at which each quantum node attempts to perform entanglement generation is decided based on the difference in the expected time spent for establishing an entanglement link between it and the node with the highest expected time. In this way, although entanglement generation between some pairs of adjacent quantum nodes is performed later, it takes less time to successfully establish an entanglement link. Consequently, entanglement links can be built between all adjacent quantum node pairs simultaneously in the selected swapping path.

Using the example of entanglement generation in a segment, we analyze the parallel entanglement generation design. As shown in Fig. 7, this design consists of four steps. We assume that the expected duration $t_{(a,b)}$ of a successful entanglement generation endeavor between nodes A and B is larger than the expected duration $t_{(b,c)}$. In the initial step $(t = t_0)$, node A attempts to create an entanglement link with node B. Second, when t_1 equals the difference between $t_{(a,b)}$ and $t_{(b,c)}$, node C tries to create an entanglement link with node B. Thirdly, when t_2 is equal to $t_{(a,b)}$, two adjacent entangled pairs are successfully distributed simultaneously, which triggers node B to perform entanglement swapping. Finally, nodes A and C establish a two-hop entanglement connection. The duration of remote entanglement distribution is the sum of $t_{(a,b)}$ and t_m, and the decoherence time in this process can be ignored. Most notably, entanglement swapping is successfully performed with probability. If the failed swap operations cannot be detected in time in the swapping path, some entangled pairs will be wasted during remote entanglement distribution. Hence, in our design, entanglement swapping is not performed spontaneously after successful entanglement generation between each quantum node and its predecessor and successor. Entanglement swapping is controlled by the centralized controller, and only the intermediate quantum node of each segment swaps entanglement after each segment round. The details of entanglement swapping management, i.e., the segment scheme, are elaborated in the next section.

C. Segment Design

The segment scheme aims to reduce unnecessary stop-and-wait processes during remote entanglement distribution by performing swap operations in parallel. More concretely, this scheme consists of two essential operations, i.e., segment operations and swap operations. The segment operations aim to split a long-distance swapping path into multiple short-haul segments. Swap operations are performed to "glue" together entangled pairs within each segment into multi-hop entanglement connections. These two operations are repeated alternatively in the selected swapping path until an end-to-end entanglement connection is established.

The core problem of the segment scheme in this paper is determining each segment's components, i.e., which quantum repeaters each segment contains. Here, the principle of the segment scheme is to consider a minimum component unit of entanglement swapping consisting of three quantum nodes as a segment as far as possible. If there are not three sequential quantum nodes, i.e., these quantum nodes cannot form a basic unit of entanglement swapping, they are treated as a segment. Assume that a swapping path $p = \{v_0, v_1, \ldots, v_{s-1}, v_s\}$ is selected to establish end-to-end entanglement connections. In each swapping round, the segment scheme is responsible for selecting quantum repeaters from the swapping path p to form different segments. However, there are many strategies to realize the segmentation function for the segment scheme, and these strategies vary in the performance of remote entanglement distribution. It is worth noting that entanglement swapping is an imperfect operation. The failed swap operation negatively affects the performance of remote entanglement distribution. Hence, the segment scheme needs to fully consider the imperfection of entanglement swapping.

In order to reduce the negative influence of swap operations' imperfection on remote entanglement distribution, we introduce a strategy that can make the success probability of establishing multi-hop entanglement connections in all segments containing three quantum nodes approximately equal.

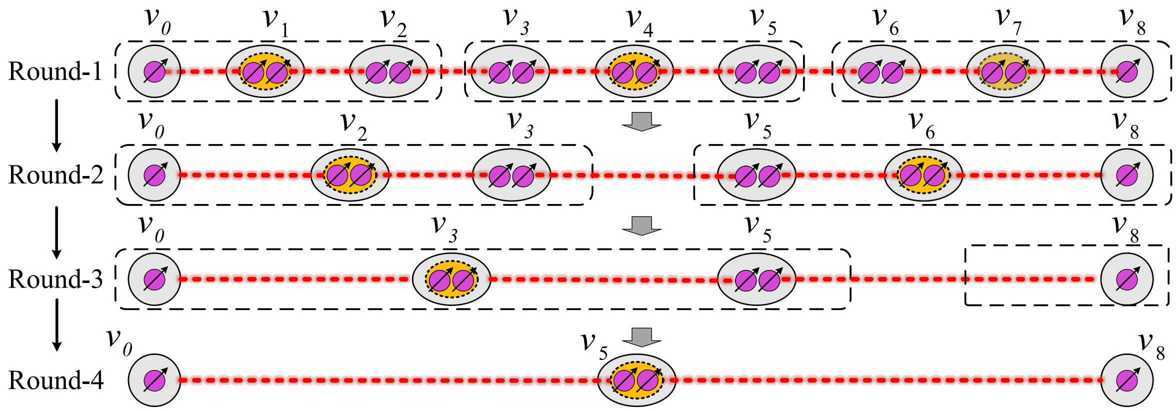


Fig. 8. Remote entanglement distribution achieved by segment-based method.

More concretely, for all the segmentation results, the segment scheme selects the result that all segments containing three nodes show the smallest variance in the success probability of establishing a multi-hop entanglement connection as the final set of segments in each entanglement swapping round. In this way, all segments that can perform swap operations show approximately consistent performance in the ability to extend entanglement distribution distance. Consequently, parallel swap operations in the segment scheme can be approximately considered to be performed successfully at the same time, thus effectively reducing the stop-and-wait delay during remote entanglement distribution. Besides, in order to reduce the number of segment rounds, the segment scheme requires building as many segments containing three quantum nodes as possible in each round.

We conclude the segment scheme as Algorithm 2. This algorithm consists of two essential steps. The first step determines the set of quantum repeaters that perform swap operations in each round. In order to realize this goal, we first calculate all the sets of segments in the swapping path and then calculate each set's variance in the success probability of entanglement swapping (Lines 5-8). Based on the calculation results, we select the set with the largest number of segments and a close-to-zero variance as the set of swapping repeaters in this round and then cut these quantum repeaters from the swapping path in preparation for determining the set of quantum repeaters that perform swap operations in the next round (Lines 9-12). The second step is responsible for performing swap operations in parallel in each round to establish multi-hop entanglement connections and output end-to-end entanglement connections. We retrieve a set of quantum repeaters from the queue obtained in step one and perform swap operations in parallel until the queue is empty (Lines 15-19). Consequently, end-to-end entanglement connections can be established after multiple segmentation rounds in the swapping path. Here, we take a complexity analysis of the segmentation algorithm and assume that the selected swapping path contains *n* quantum nodes. In the first step, the time complexity of calculating all the sets of segments in the selected path is $O(n^2)$. In the second step, the time complexity of performing swap operations is $O(\lceil |n|/3 \rceil)$. As a result, the time complexity of the segmentation algorithm is $O(n^2)$.

Algorithm 2: Segmentation Algorithm

Input: $p = \{v_0, v_1, \ldots, v_s\}$;
Output: end-to-end entanglement connection;
1 Initialize the set of quantum repeaters that performing swap operations in each round, S_r;
2 Initialize the queue of S_r in the swapping path *p*, Q_p;
3 **Step 1: Determining the set of swapping repeaters in each segmentation round.**
4 **for** $|p| - 2 \neq 0$ **do**
5 Calculate all the sets of segments in *p*;
6 **for** *each set of segments* **do**
7 Calculate the variance in the success probability of swap operations;
8 **end**
9 $S_r \leftarrow$ the set with the largest number of segments and a close-to-zero variance;
10 Add S_r into the queue Q_p;
11 $S_r \leftarrow \varnothing$;
12 $p \leftarrow$ cut the swapping repeaters from the path *p*;
13 **end**
14 **Step 2: Performing swap operations to establish multi-hop entanglement connections.**
15 **for** Q_p *is not null* **do**
16 $S_r \leftarrow$ retrieve the header element of Q_p;
17 Performing swap operations in paralle;
18 Reduce the header index of Q_p by one;
19 **end**

Fig. 8 shows an example of the segment scheme, and there are nine quantum nodes in the swapping path $\{v_0, v_1, \ldots, v_7, v_8\}$. In the first round, three segments are determined, and three swap operations can be performed at quantum repeaters v_1, v_4, and v_7 in parallel to establish two-hop entanglement connections. In this round, entanglement swapping is triggered by the event that two adjacent entangled pairs are successfully distributed simultaneously. After the first round of entanglement swapping, we delete each segment's intermediate quantum node from the swapping path to reconstruct a new swapping path $\{v_0, v_2, v_3, v_5, v_6, v_8\}$. In the second round, swap operations are performed simultaneously at quantum repeaters v_2 and v_5 to establish three-hop entanglement connections. After entanglement swapping, quantum nodes v_0, v_3, and v_5 form a segment to establish five-hop entanglement connections, and the remaining quantum node v_8 forms an independent segment. In the last round, v_0, v_5, and v_8 form a segment, and v_5 performs entanglement swapping to establish an end-to-end entanglement connection. In summary, segmentation operations and swap operations are repeatedly executed in the segment scheme until the reconstructed path only consists of three quantum nodes, i.e., the source node, a quantum repeater, and the destination node.

D. Performance Analysis

Latency: For the segment-based method, the total latency is the sum of the delay times introduced by entanglement generation and entanglement swapping. According to Algorithm 1, the duration of entanglement generation is the maximum delay time spent for establishing an entanglement link on the initial swapping path. Notably, in each segmentation round, swap operations are performed in parallel in all segments. Hence, the total measurement time spent for one round is t_{mea}. As a result, the duration of the iterative entanglement swapping during remote entanglement distribution is $N_p \cdot t_{mea}$, where N_p is the number of rounds required to perform parallel swap operations. Note that each entangled pair used for entanglement swapping is measured directly after entanglement generation in the first round. Hence, the total delay times caused by the retrieval of entangled photons is $(N_p - 1) \cdot t_{ret}$. Consequently, we can get the total latency, L_{SEG}, of the segment-based remote entanglement distribution method as

$$L_{SEG} = t_{max} + N_p \cdot t_{mea} + \left(N_p - 1\right) \cdot t_{ret}, \quad (10)$$

where t_{max} is the maximum value of the set $\{t_{(0,1)}, t_{(1,2)}, \ldots, t_{(s-1,s)}\}$. Compare to Eq. (4) and Eq. (5) under the same assumption that each swap operation is successfully performed (i.e., $N_p = \lceil |p|/3 \rceil + 1$), the segment-based method outperforms both TAG and TAS methods in remote entanglement distribution latency.

Werner Parameter: For the segment-based entanglement distribution method, the entangled pairs shared by all adjacent quantum nodes are successfully distributed at the same time using the entanglement generation algorithm. Hence, the attenuation of fidelity in this part can be ignored, and entangled systems' fidelities mainly decay in the second part. For each entanglement swapping round, two entangled pairs within each segment are measured to build a multi-hop entanglement connection, and the entangled pairs shared by adjacent segments remain in the initial state. Hence, the decoherence time of entangled pairs within each segment only equals t_{ret}. However, the decoherence time of the entangled pairs shared by adjacent segments is the duration of a swap operation, including the retrieval and measurement of entangled pairs. Besides, swap operations are performed in parallel in each segmentation round, and all the entangled pairs in the same position (within a segment or between adjacent segments) experience the same decoherence time. Most notably, if there are only one or two nodes in a segment, the decoherence time of the entangled pair within this segment is the same as the entangled pairs shared by adjacent segments. Therefore, for a segmentation round with *m* segments containing three quantum nodes, *n* segments containing two quantum nodes, and *k* segments containing one quantum node, the total decoherence time of all entangled pairs is

$$\begin{aligned} T_k &= 2mt_{ret} + 2n(t_{ret} + t_{mea}) + k(t_{ret} + t_{mea}) \\ &= (2m + 2n + k)t_{ret} + (2n + k)t_{mea}. \end{aligned} \quad (11)$$

Combining Eq. (2) and Eq. (11), we can get that the Werner parameter of the end-to-end entanglement connection established by the segment-based method is determined by the features of the swapping path *p* (e.g., the path's length and the success probability of the swap operation performed by each quantum repeater) and the initial Werner parameters of all Werner-state entangled pairs.

E. Implementation and Discussion

The segment-based remote entanglement distribution method is implemented with the aid of a classical controller. First, SD pairs send requests to the controller to establish end-to-end entanglement connections, and swapping paths are selected for them based on the routing algorithm running on the controller. Then quantum nodes on each selected path attempt to establish entanglement links following the entanglement generation algorithm. Under the management of the controller, multiple segments attempt to establish multi-hop entanglement connections by performing swap operations in parallel in each swapping round and thus generating end-to-end entanglement. If some quantum nodes are conflicting when performing entanglement generation or swap operations for different SD pairs, the controller can adopt a suitable scheduling design to control generation and swap operations performed for different SD pairs' requests.

Compared to the hop-by-hop and parallel entanglement generation methods, the segment-based method can effectively reduce unnecessary waiting time during remote entanglement distribution. The reason is that the segment-based method allows all entanglement links on a swapping path to be established at the same time. Moreover, since the segment-based method enables swap operations to be performed in parallel in each swapping round, it simplifies the interaction between quantum nodes caused by the feedback of measurement results and quantum decoherence time compared to the sequence (or hop-by-hop) entanglement swapping method. Besides, our segment design outperforms the general nested entanglement swapping method in terms of dealing with the probability feature of swap operations since all segments present the approximate capabilities to swap entanglement in each swapping round. Moreover, the segment-based method outperforms the nested entanglement swapping method in terms of resource consumption because it can detect failed swap operations in time to avoid the re-establishment of entangled links. To sum up, the combination of the concurrent entanglement generation design and the nested entanglement swapping method enables the segment-based method to provide an efficient solution for remote entanglement distribution in quantum networks.

The fact that entanglement swapping is an imperfect operation must be considered in the segment-based method. Suppose a quantum repeater v_i fails to swap entanglement in an entanglement swapping round. In this case, the segment containing v_i re-executes concurrent entanglement generation and then re-attempts to establish multiple-hop entanglement connections. The additional delay time introduced by re-establishing a multiple-hop entanglement connection in this segment is the sum of the time spent for concurrent entanglement generation and the time spent for multiple swapping rounds. In our segment design, we limited the maximum length of the segment to three, i.e., the minimum unit of entanglement swapping, which is not conducive to reducing the latency caused by the failed swap operation. In order to reduce the additional delay time caused by the failed swap operations, the parallel entanglement swapping method can be adopted in the segment-based method. That is, we can extend the length of each segment to reduce the number of swapping rounds, thus reducing the latency of the re-establishment of a multiple entanglement connection in a segment. Moreover, the more quantum nodes a segment contains, the more entangled pairs are redistributed with initial fidelity if a quantum node fails to swap entanglement. As a result, the total decoherence time will decline with the increase in the segments' length.

However, the success probability of establishing multi-hop entanglement connections in each segment is negatively correlated with the length of the segments. Although the additional delay times introduced by the failed swap operations slightly affect remote entanglement distribution latency due to the concurrent entanglement generation method and parallel entanglement swapping method, a segment successfully establishing a multiple-hop entanglement connection with a low probability will lead to high entanglement distribution latency. Besides, the end-to-end entanglement connection will be built with low fidelity because the successfully established multiple-hop entanglement connection in other segments will undergo longer waiting times. Hence, the performance of the segment-based method is not proportional to the length of each segment. There is a trade-off between the length of each segment and the performance of the segment-based method. We perform simulations to investigate the impact of segments' length on the segment-based method in Section V.

V. Performance Evaluation

In this section, we evaluate the performance of the segment-based remote entanglement distribution method. Simulations are performed in a randomly generated network topology with a certain amount of quantum nodes and the success probability of entanglement generation.

A. Evaluation Methodology

In a quantum network, we randomly select multiple pairs of non-adjacent quantum nodes as SD pairs and use the shortest path routing algorithm to select swapping paths for these SD pairs. To demonstrate the superiority of the segment-based method, we compare the effects of different factors, e.g., the success probability of entanglement swapping, the coherence time of quantum memory, on the performance of different remote entanglement distribution methods. Furthermore, we investigated the performance of the segment-based method for different segments' lengths. For a given set of parameters used in each simulation scenario, we run the simulation 100 trials and show the averaged results.

Network Topology: To generate a network topology, we randomly place a given number of quantum nodes into a $10^4\mathrm{km}$ by $10^4\mathrm{km}$ square area. We adopt a Waxman-based improved Salama model [52], [53] to determine the probability that nodes *u* and *v* are directly linked, and the probability is

$$p_{(u,v)} = \frac{k\bar{e}}{|V|}\gamma e^{-l_{(u,v)}/\beta \cdot L_{max}}, \tag{12}$$

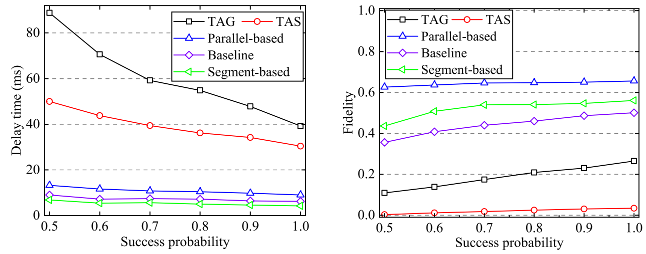
where $\bar{e}$ is a parameter used to control the average node degree of the randomly generated network, *k* is a constant, $l_{(u,v)}$ is Euclidean distance between node *u* and *v*, $|V|$ is the size of nodes, and L_{max} is the largest distance between any pair of nodes. An increase in either γ and β can improve network connectivity. Each entanglement link can be successfully built between two nodes v_u and v_u with probability $r_{(u,v)} \approx e^{-\alpha l_{(u,v)}}$, where α mainly determines the success probability of entanglement generation between adjacent quantum nodes. Each entanglement generation attempt takes duration $l_{(u,v)}/c$, where $l_{(u,v)}$ is the physical length of a quantum channel and *c* is the speed of light in the used transmission medium. We get the expected time spent for successfully establishing an entanglement link:

$$t_{(u,v)} = \frac{l_{(u,v)}}{c \cdot r_{(u,v)}}.$$

Besides, the initial fidelity of each entangled pair is uniformly picked from 0.8 to 1.0.

Default Settings: By default, we set $\gamma = 0.9$, $\beta = 0.001$, $k = 2$, and $\bar{e} = 5$ in Salama model to generate simulation network topology; Moreover, the duration of retrieving entangled photons from quantum memory units is about 1.04ms; Finally, the time spent for jointly measuring local entangled photons in each swap operation is 0.1ms [54].

Comparison Schemes: We compare the segment-based remote entanglement distribution method with four typical methods: TAG, TAS, Parallel-based, and Baseline methods. Remote entanglement distribution based on the TAG method is implemented by performing two quantum operations, entanglement generation and entanglement swapping, hop by hop along the swapping path. The TAS method is implemented based on the advance generation model and uses the parallel entanglement generation design and the sequence entanglement swapping method to establish end-to-end entanglement connections. The parallel-based method is generally adopted in existing studies. In the parallel-based method, entanglement generation and swap operations are performed in parallel to realize remote entanglement distribution. The baseline method is realized based on the on-demand generation model. This method uses parallel entanglement generation design and the general nested entanglement swapping method. However, in the baseline method, multiple segments are determined based on a simple idea, i.e., three consecutive quantum nodes



(a) Delay time vs the success probability of entanglement generation.

(b) Fidelity vs the success probability of entanglement generation.

Fig. 9. Impact of the success probability of entanglement generation.

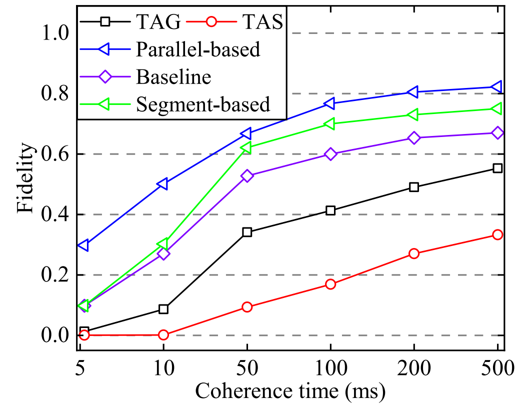


Fig. 10. The impact of the coherence time of quantum memory units on entanglement fidelity.

(starting from the source node) on a swapping path form a segment.

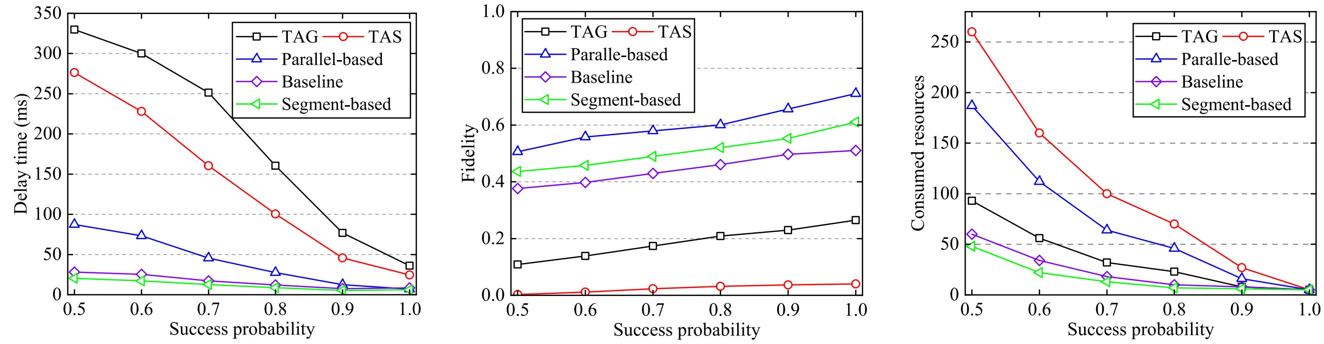
Performance Metrics: We compare the performance of different methods with respect to three metrics, i.e., delay time, fidelity, and the number of consumed entanglement links. The first metric presents the time spent for distributing an entangled pair between two distant quantum nodes along a selected swapping path. The shorter the delay time, the higher the remote entanglement distribution rate of the method. Fidelity reflects the quality of the established end-to-end entanglement connections, and an end-to-end entanglement connection with high fidelity facilitates the improvement of quantum applications' performance. The last performance metric is the number of consumed entanglement links, which reflects the influence of the probability characteristic of entanglement swapping on different remote entanglement distribution methods.

B. Evaluation Results

Main observations: Compared to the hop-by-hop entanglement generation design (i.e., TAG), the parallel entanglement generation design can effectively reduce remote entanglement distribution latency since it is challenging to establish an entanglement link between adjacent quantum nodes. However, parallel entanglement generation design contributes to a lower quality of end-to-end entanglement connections. Bedsides, although the parallel entanglement swapping method is conducive to establishing high-fidelity end-to-end entanglement connections, it results in high entanglement resource costs due to the failed swap operations. The segment-based method outperforms the other four methods in end-to-end entanglement distribution latency and can significantly mitigate the attenuation of entanglement fidelity. This is because we adopt the concurrent entanglement generation design to establish entanglement links and the nested entanglement swapping method to establish end-to-end entanglement connections. Moreover, appropriately increasing the length of each segment can effectively reduce fidelity attenuation at the cost of additional delay time.

Effect of the success probability of entanglement generation: In this set of simulations, we vary edge capacity from one to three and set the coherence time of quantum memory to 20ms to evaluate how different methods perform with the success probability of entanglement generation and the simulation results are shown in Fig. 9. As shown in Fig. 9(a), the delay time of remote entanglement distribution for five methods decreases as the success probability of entanglement generation increases. The reason is that the delay time introduced by entanglement generation can be effectively reduced with the increased success probability. Besides, the segment-based method performs better than the parallel entanglement generation design represented by TAS, Parallel-based, and Baseline methods in terms of delay time and significantly better than the TAG method using the hop-by-hop entanglement generation design. As shown in Fig. 9(b), in addition to the Parallel-based method, the fidelity of end-to-end entanglement for four methods is positively correlated with the success probability of entanglement generation since the increase in success probability of entanglement generation can reduce the total decoherence time of all entangled pairs, especially in the case where quantum nodes fail to swap entanglement. The segment-based method outperforms TAG, TAS, and Baseline methods in fidelity due to the concurrent entanglement generation design. Although the Parallel-based method adopts the parallel entanglement generation design to establish entanglement links similar to TAG and Baseline methods, it presents the superiority of end-to-end entanglement connections' fidelities. The reason is that a failed swap operation results in the re-establishment of all entanglement links on the swapping path, i.e., all entanglement links are measured with high quality.

Effect of the coherence time of quantum memory: When the joint coherence time of two quantum memories holding entangled pairs varies from 5 to 500ms, the performance of entanglement fidelity under different methods is shown in Fig. 10. The fidelity in all five methods will increase with the coherence time of quantum memory because the increase in coherence time can mitigate the attenuation of fidelity. Since swap operations are performed simultaneously in the Parallel-based method, i.e., only a small fidelity attenuation occurs for each entanglement link, the Parallel-based method performs best in this simulation. Besides, the increase of the fidelity in the segment-based method is greater than TAG, TAS, and Baseline methods since there are fewer stop-and-wait processes during remote entanglement distribution. Moreover, the negative effect of decoherence can be ignored in the



(a) Delay time vs the success probability of entanglement swapping.

(b) Fidelity vs the success probability of entanglement swapping.

(c) Consumed resources vs the success probability of entanglement swapping.

Fig. 11. The impact of the success probability of entanglement swapping on different methods.

segment-based method when the coherence time is greater than 500ms, i.e., the duration of entanglement distribution in the segment-based method is significantly less than the coherence time of quantum memory.

Effect of the success probability of entanglement swapping: We randomly select the SD pair in the generated network and set the success probability of the entanglement swapping of each node from 0.5 to 1.0. Under these settings, we repeat simulations to study the influence of the success probability of entanglement swapping on five methods in terms of delay time, fidelity, and the number of entanglement links consumed for remote entanglement distribution. The simulation results are shown in Fig. 11. As shown in Fig. 11(a), the segment-based method takes less time to implement remote entanglement distribution than the other four methods because the concurrent entanglement generation design and the improved nested entanglement swapping method can reduce redundant stop-and-wait processes. As the success probability increases, the time spent for establishing end-to-end entanglement connections in all five methods degrades because redundant and repeated quantum operations are not required when entanglement swapping is perfect. As shown in Fig. 11(b), the segment method performs better than TAG, TAS, and Baseline methods in terms of fidelity. Although the segment method is not as good as the Parallel-based method, the difference between them is very small since each segment can timely re-establish a multi-hop entanglement connection to reduce the decoherence time of other multi-hop entanglement connections. Fig. 11(c) presents the relationship between the number of entanglement links consumed by different methods to establish an entanglement connection and the success probability of entanglement swapping. As the success probability of entanglement swapping increases, the consumed entanglement resources decreases. When this probability equals 1.0, these five methods present the same performance in this simulation. Compared to the other four methods, the segment-based method can effectively deal with failed swap operations, so it consumes the fewest entanglement links to realize remote entanglement distribution.

Effect of the edge capacity: To study how the performance of the three methods will be impacted by the number of entanglement links each pair of adjacent quantum nodes holds

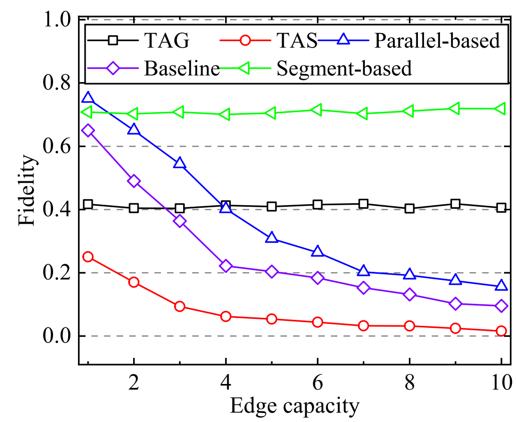


Fig. 12. The impact of the edge capacity on fidelity.

(i.e., edge capacity), we vary edge capacity from 1 to 10 and test the performances of different methods under the assumption that entanglement swapping is a perfect operation. Fig. 12 shows that the segment-based method outperforms the other four methods. The increase in edge capacity does not affect TAG and segment-based methods using the on-demand generation model in entanglement fidelity. This is because each entangled pair is distributed and measured independently of others in other methods. For these methods using the advance generation model, the larger the edge capacity, the longer the decoherence time suffered by entangled pairs before path selection. However, the TAS method performs worst in entanglement fidelity due to the hop-by-hop entanglement swapping method. As a result, for the TAS method, the fidelity of an end-to-end entanglement connection will decay to zero with the increase in edge capacity.

Effect of the length of each segment: Suppose a swapping path consisting of 15 quantum nodes, we vary the size of each segment from 3 to 7 to investigate the effect of segments' length on the performance of the segment-based method. As shown in Fig. 13, increasing each segment's length can reduce the delay time to some extent due to the number of entanglement swapping rounds is reduced. However, when the length of each segment exceeds 5, the delay time will increase with each segment's length. The reason is that the success probability of establishing multi-hop entanglement connections in a

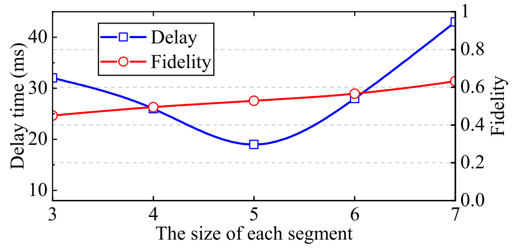


Fig. 13. The impact of the length of segments on the segment-based method.

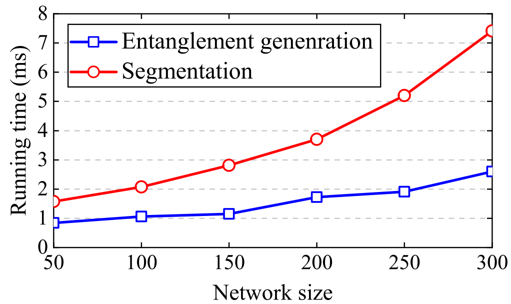


Fig. 14. Algorithm running time.

segment is inversely correlated with the length of the segment. The performance of the segment-based method in entanglement fidelity is positively correlated with the length of each segment. This is because that all entangled pairs will be redistributed if quantum nodes fail to perform swap operations, i.e., the decoherence time of all entangled pairs in a segment roughly equals zero. According to the results of these simulations, we can appropriately increase the length of segments to balance the performance metrics of delay time and entanglement fidelity.

Algorithm running time: Fig. 14 shows how much time the segment-based remote entanglement distribution method will spend to run the entanglement generation algorithm in phase-1 and the segment scheme in phase-2. The results are collected from a desktop carrying an AMD R5-5600 CPU with 16 GB memory. As shown in Fig. 14, we can observe that even when there are 300 nodes in a quantum network, only 2.60ms and 7.41ms are needed by each phase, respectively.

VI. Conclusion

In this paper, we proposed a segment-based method for establishing end-to-end entanglement connections along a swapping path. This method facilitates performing quantum operations in parallel, thus effectively reducing the redundant waiting times caused by stop-and-wait processes to achieve low-latency and high-fidelity remote entanglement distribution. The segment-based method includes two essential parts, i.e., the entanglement generation algorithm and the segment design. The entanglement generation algorithm is responsible for establishing entanglement links between all pairs of adjacent quantum nodes simultaneously. The segment design attempts to split a long-distance swapping path into multiple segments to perform swap operations using the nested entanglement swapping method. Extensive simulation results reveal that the segment-based remote entanglement distribution method outperforms both TAG, TAS, Parallel-based, and Baseline methods in terms of remote entanglement distribution latency and end-to-end entanglement fidelity.

In this work, we only consider entanglement generation and entanglement swapping in remote entanglement distribution. Entanglement purification is also an essential operation for high-fidelity entanglement distribution. In our future work, we plan to take entanglement purification into consideration when designing an improved entanglement generation algorithm. Besides, we intuitively set the maximum length of each segment to three, i.e., the minimum component unit of entanglement swapping. We plan to optimize the segment design to improve the segment-based remote entanglement distribution method in our future work.

References

[1] W. K. Wootters and W. H. Zurek, “A single quantum cannot be cloned,” *Nature*, vol. 299, no. 5886, pp. 802–803, 1982.

[2] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, “Quantum entanglement,” *Rev. Modern Phys.*, vol. 81, no. 2, p. 865, 2009.

[3] H. P. Robertson, “The uncertainty principle,” *Phys. Rev.*, vol. 34, pp. 163–164, Jul. 1929.

[4] S. Wehner, D. Elkouss, and R. Hanson, “Quantum Internet: A vision for the road ahead,” *Science*, vol. 362, no. 6412, 2018, Art. no. eaam9288.

[5] Z. Li et al., “Building a large-scale and wide-area quantum Internet based on an OSI-alike model,” *China Commun.*, vol. 18, no. 10, pp. 1–14, Oct. 2021.

[6] S.-H. Wei et al., “Towards real-world quantum networks: A review,” *Laser Photon. Rev.*, vol. 16, no. 3, 2022, Art. no. 2100219.

[7] H.-K. Lo and H. F. Chau, “Unconditional security of quantum key distribution over arbitrarily long distances,” *Science*, vol. 283, no. 5410, pp. 2050–2056, 1999.

[8] V. S. Denchev and G. Pandurangan, “Distributed quantum computing: A new frontier in distributed systems or science fiction?” *ACM SIGACT News*, vol. 39, no. 3, pp. 77–95, 2008.

[9] C. L. Degen, F. Reinhard, and P. Cappellaro, “Quantum sensing,” *Rev. Modern Phys.*, vol. 89, no. 3, 2017, Art. no. 35002.

[10] A. Einstein, B. Podolsky, and N. Rosen, “Can quantum-mechanical description of physical reality be considered complete?” *Phys. Rev.*, vol. 47, no. 10, p. 777, 1935.

[11] J. Yin et al., “Satellite-based entanglement distribution over 1200 kilometers,” *Science*, vol. 356, no. 6343, pp. 1140–1144, 2017.

[12] M. Lipka, M. Mazelanik, and M. Parniak, “Entanglement distribution with wavevector-multiplexed quantum memory,” *New J. Phys.*, vol. 23, no. 5, 2021, Art. no. 53012.

[13] Y. Wang et al. “Remote entanglement distribution in a quantum network via multinode indistinguishability of photons.” 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2107.03999

[14] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, “Quantum repeaters: The role of imperfect local operations in quantum communication,” *Phys. Rev. Lett.*, vol. 81, pp. 5932–5935, Dec. 1998.

[15] M. Zukowski, A. Zeilinger, M. Horne, and A. Ekert, ““Event-ready-detectors” bell experiment via entanglement swapping,” *Phys. Rev. Lett.*, vol. 71, no. 26, pp. 4287–4290, 1993.

[16] N. Gisin and R. Thew, “Quantum communication,” *Nature Photon.*, vol. 1, no. 3, pp. 165–171, 2007.

[17] L. Gyongyosi and S. Imre, “Advances in the quantum Internet,” *Commun. ACM*, vol. 65, no. 8, pp. 52–63, 2022.

[18] R. Van Meter, T. D. Ladd, W. J. Munro, and K. Nemoto, “System design for a long-line quantum repeater,” *IEEE/ACM Trans. Netw.*, vol. 17, no. 3, pp. 1002–1013, Jun. 2009.

[19] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, “Fundamental limits of repeaterless quantum communications,” *Nat. Commun.*, vol. 8, no. 1, pp. 1–15, 2017.

[20] M. Brune et al., “Observing the progressive decoherence of the “meter” in a quantum measurement,” *Phys. Rev. Lett.*, vol. 77, no. 24, p. 4887, 1996.

[21] E. Joos et al., *Decoherence and the Appearance of a Classical World in Quantum Theory*. Heidelberg, Germany: Springer, 2013.

[22] A. S. Cacciapuoti, M. Caleffi, F. Tafuri, F. S. Cataliotti, S. Gherardini, and G. Bianchi, "Quantum Internet: Networking challenges in distributed quantum computing," *IEEE Netw.*, vol. 34, no. 1, pp. 137–143, Jan./Feb. 2020.

[23] C. Cabrillo, J. I. Cirac, P. García-Fernández, and P. Zoller, "Creation of entangled states of distant atoms by interference," *Phys. Rev. A*, vol. 59, pp. 1025–1033, Oct. 1999.

[24] C. Ferrari and B. Braunecker, "Entanglement, which-way measurements, and a quantum erasure," *Amer. J. Phys.*, vol. 78, no. 8, pp. 792–795, 2010.

[25] S. Welte, B. Hacker, S. Daiss, S. Ritter, and G. Rempe, "Photon-mediated quantum gate between two neutral atoms in an optical cavity," *Phys. Rev. X*, vol. 8, no. 1, 2018, Art. no. 11018.

[26] S. Ritter et al., "An elementary quantum network of single atoms in optical cavities," *Nature*, vol. 484, no. 7393, pp. 195–200, 2012.

[27] A. Reiserer, "Colloquium: Cavity-enhanced quantum network nodes," *Rev. Modern Phys.*, vol. 94, no. 4, 2022, Art. no. 41003.

[28] M. Takeoka, S. Guha, and M. M. Wilde, "Fundamental rate-loss trade-off for optical quantum key distribution," *Nat. Commun.*, vol. 5, no. 1, pp. 1–7, 2014.

[29] K. Azuma and G. Kato, "Aggregating quantum repeaters for the quantum Internet," *Physical Rev. A*, vol. 96, no. 3, 2017, Art. no. 32332.

[30] H. E. Brandt, "Qubit devices and the issue of quantum decoherence," *Prog. Quant. Electron.*, vol. 22, nos. 5–6, pp. 257–370, 1999.

[31] M. Schlosshauer, "Quantum decoherence," *Phys. Rep.*, vol. 831, pp. 1–57, Oct. 2019.

[32] R. Jozsa, "Fidelity for mixed quantum states," *J. Modern Opt.*, vol. 41, no. 12, pp. 2315–2323, 1994.

[33] M. A. Schlosshauer, *Decoherence and the Quantum-to-Classical Transition*. Berlin, Germany: Springer, 2007.

[34] Q. Turchette et al., "Deterministic entanglement of two trapped ions," *Phys. Rev. Lett.*, vol. 81, no. 17, p. 3631, 1998.

[35] R. F. Werner, "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model," *Phys. Rev. A*, vol. 40, no. 8, p. 4277, 1989.

[36] S. Brand, T. Coopmans, and D. Elkouss, "Efficient computation of the waiting time and fidelity in quantum repeater chains," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 3, pp. 619–639, Mar. 2020.

[37] K. Chakraborty, F. Rozpedek, A. Dahlberg, and S. Wehner. "Distributed routing in a quantum Internet." 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1907.11630.

[38] F. Dupuy, C. Goursaud, and F. Guillemin, "A survey of quantum entanglement routing protocols—Challenges for wide-area networks," *Adv. Quant. Technol.*, vol. 6, Mar. 2023, Art. no. 2200180.

[39] Z. Li, K. Xue, J. Li, N. Yu, D. S. Wei, and R. Li, "Connection-oriented and connectionless remote entanglement distribution strategies in quantum networks," *IEEE Netw.*, vol. 36, no. 6, pp. 150–156, Nov./Dec. 2022.

[40] R. Van Meter, T. Satoh, T. D. Ladd, W. J. Munro, and K. Nemoto, "Path selection for quantum repeater networks," *Netw. Sci.*, vol. 3, no. 1, pp. 82–95, 2013.

[41] S. Pirandola, "End-to-end capacities of a quantum communication network," *Commun. Phys.*, vol. 2, no. 1, pp. 1–10, 2019.

[42] M. Pant et al., "Routing entanglement in the quantum Internet," *npj Quant. Inf.*, vol. 5, pp. 25–34, Mar. 2019.

[43] L. Gyongyosi and S. Imre, "Opportunistic entanglement distribution for the quantum Internet," *Sci. Rep.*, vol. 9, no. 1, p. 2219, 2019.

[44] S. Shi and C. Qian, "Concurrent entanglement routing for quantum networks: Model and designs," in *Proc. Conf. ACM Spec. Interest Group Data Commun. (SIGCOMM)*, 2020, pp. 62–75.

[45] Y. Zhao and C. Qiao, "Redundant entanglement provisioning and selection for throughput maximization in quantum networks," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, 2021, pp. 1–10.

[46] Y. Zhao, G. Zhao, and C. Qiao, "E2E fidelity aware routing and purification for throughput maximization in quantum networks," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, 2022, pp. 480–489.

[47] C. Cicconetti, M. Conti, and A. Passarella, "Request scheduling in quantum networks," *IEEE Trans. Quant. Eng.*, vol. 2, pp. 2–17, 2021.

[48] A. Patil, J. I. Jacobson, E. Van Milligen, D. Towsley, and S. Guha, "Distance-independent entanglement generation in a quantum network using space-time multiplexed Greenberger–Horne–Zeilinger (GHZ) measurements," in *Proc. IEEE Int. Conf. Quant. Comput. Eng. (QCE)*, 2021, pp. 334–345.

[49] J. Li et al., "Fidelity-guaranteed entanglement routing in quantum networks," *IEEE Trans. Commun.*, vol. 70, no. 10, pp. 6748–6763, Oct. 2022.

[50] C. Li, T. Li, Y.-X. Liu, and P. Cappellaro, "Effective routing design for remote entanglement generation on quantum networks," *npj Quant. Inf.*, vol. 7, no. 1, pp. 1–12, 2021.

[51] Z. Li et al., "Swapping-based entanglement routing design for congestion mitigation in quantum networks," *IEEE Trans. Netw. Service Manag.*, early access, May 12, 2023, doi: 10.1109/TNSM.2023.3275815.

[52] B. M. Waxman, "Routing of multipoint connections," *IEEE J. Sel. Areas Commun.*, vol. 6, no. 9, pp. 1617–1622, Dec. 1988.

[53] H. F. Salama, *Multicast Routing for Real-Time Communication of High-Speed Networks*. Raleigh, NC, USA: North Carolina State Univ., 1996.

[54] A. Dahlberg et al., "A link layer protocol for quantum networks," in *Proc. Conf. ACM Special Interest Group Data Commun. (SIGCOMM)*, 2019, pp. 159–173.

Zhonghui Li (Graduate Student Member, IEEE) received the bachelor's degree from the School of Information and Software Engineering, University of Electronic Science and Technology of China in 2018, and the Doctoral degree in information security from the School of Cyber Science and Technology, University of Science and Technology of China in 2023, where he is currently a Postdoctoral Researcher with the School of Cyber Science and Technology. His current research interests include quantum Internet architecture design, quantum networking, and network security.

Jian Li (Member, IEEE) received the bachelor's degree from the Department of Electronics and Information Engineering, Anhui University in 2015, and the Doctoral degree from the Department of Electronic Engineering and Information Science, University of Science and Technology of China (USTC) in 2020. From November 2019 to November 2020, he was a Visiting Scholar with the Department of Electronic and Computer Engineering, University of Florida. From December 2020 to December 2022, he was a Postdoctoral Researcher with the School of Cyber Science and Technology, USTC. He is currently a Research Associate with the School of Cyber Scie.nce and Technology, USTC. His research interests include wireless networks, next-generation Internet, and quantum networks. He serves as the Session Chair for IWQoS 2021 and a TPC Member for conferences, including ICCCN 2023 and ICNC 2024. He also serves as an Editor for *China Communications*.

Kaiping Xue (Senior Member, IEEE) received the bachelor's degree from the Department of Information Security, University of Science and Technology of China (USTC) in 2003 and the Doctoral degree from the Department of Electronic Engineering and Information Science, USTC in 2007. From May 2012 to May 2013, he was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, University of Florida. He is currently a Professor with the School of Cyber Science and Technology, USTC, where he is also the Director of Network and Information Center. His research interests include next-generation Internet architecture design, transmission optimization, and network security. His work won the Best Paper Awards in IEEE MSN 2017 and IEEE HotICN 2019, the Best Paper Honorable Mention in ACM CCS 2022, the Best Paper Runner-Up Award in IEEE MASS 2018, and the Best Track Paper in MSN 2020. He serves on the editorial board for several journals, including the IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT. He has also served as a (Lead) Guest Editor for many reputed journals/magazines, including IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, *IEEE Communications Magazine*, and IEEE NETWORK. He is an IET Fellow.

David S. L. Wei (Life Senior Member, IEEE) received the Ph.D. degree in computer and information science from the University of Pennsylvania in 1991. He is currently a Full Professor with the Computer and Information Science Department, Fordham University. He has published over 140 technical papers in various archival journals and conference proceedings. He has served as a Program Committee Member and the Session Chair for several well-known international conferences, including Infocom. Moreover, he was a Lead Guest Editor or a Guest Editor for several special issues in the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, the IEEE TRANSACTIONS ON CLOUD COMPUTING, and the IEEE TRANSACTIONS ON BIG DATA. Additionally, he served as an Associate Editor for IEEE TRANSACTIONS ON CLOUD COMPUTING from 2014 to 2018, an Editor of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS for the Series on Network Softwarization and Enablers from 2018 to 2020, and an Associate Editor of the *Journal of Circuits, Systems, and Computers* from 2013 to 2018. His contributions to information security in wireless and satellite communications and cyber–physical systems were recognized with the IEEE Region 1 Technological Innovation Award (Academic) in 2020. He is a member of ACM and AAAS and holds a life senior memberships in IEEE Computer Society and IEEE Communications Society. His current research is focused on cloud and edge computing, cybersecurity, and quantum computing and communications.

Nenghai Yu received the B.S. degree from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1987, the M.E. degree from Tsinghua University, Beijing, China, in 1992, and the Ph.D. degree from the Department of Electronic Engineering and Information Science, University of Science and Technology of China (USTC), Hefei, China, in 2004. He is currently a Professor with the School of Cyber Science and Technology and the School of Information Science and Technology, USTC, where he is also the Executive Dean of the School of Cyber Security and the Director of the Information Processing Center. He has authored or coauthored more than 130 papers in journals and international conferences. His research interests include multimedia security, multimedia information retrieval, video processing, and information hiding.

Qibin Sun (Fellow, IEEE) received the Ph.D. degree from the Department of Electronic Engineering and Information Science, University of Science and Technology of China in 1997, where he is currently a Professor with the School of Cyber Security. He has published more than 120 papers in international journals and conferences. His research interests include multimedia security and network intelligence and security.

Jun Lu received the bachelor's degree from Southeast University in 1985, and the master's degree from the Department of Electronic Engineering and Information Science, University of Science and Technology of China in 1988, where he is currently a Professor. He is an Academician of the Chinese Academy of Engineering. His research interests include theoretical research and system development in the field of integrated electronic information systems.