Efficient Routing Design Based on Entanglement Flow Loss Effect in Quantum Networks

***Abstract*—The fundamental technology for the construction of large-scale quantum networks lies in the efficient distribution of remote entanglement among quantum nodes. Entanglement routing tackles the challenge of path selection between two distant end nodes and realizes entanglement distribution through multi-hop entanglement swapping. Given the inherent probability of swapping, entanglement distribution flow always presents inevitable losses. The existing routing works adopt a parallel swapping mode, which only considers the final end-to-end distribution loss effect. This lag of failure detection and complete independence between swapping reduce distribution efficiency. We observe that the sequential swapping mode can improve this defect, but it is not thorough. Thus, in this paper, we explore the entanglement flow loss effect in the sequential entanglement swapping (Loss-SES) process in order to perceive failures earlier and save network resources. Subsequently, we introduce a novel metric to quantify path performance, namely the expected entanglement throughput (EET). Building upon EET, we propose the entanglement routing algorithm MaxEET adapted to the Loss-SES process to maximize EET. For efficient deployment, in the single-request scenario, we determine the optimal routing initiation direction by introducing a bidirectional routing mechanism. In the multi-request scenario, we design a greedy and contention-free iterative routing algorithm to exhaust resources. Based on extensive simulations, the results show the superiority of our proposed Loss-SES mode compared to other modes in terms of throughput, fairness, and utilization.**

***Index Terms*—Routing, Resource allocation, Sequential swapping, Entanglement flow loss, Quantum networks.**

I. INTRODUCTION

Recently, derived from the principles of quantum physics, quantum networks have gained considerable attention by serving as a novel network infrastructure. The critical function of quantum networks is to generate and distribute entanglement between two remote quantum nodes. End-to-end (E2E) entanglement between communication parties serves as the prerequisite implementation for a series of advanced quantum applications, such as secure quantum key distribution [1], clock synchronization [2], and distributed quantum computing [3], etc.

Due to the severe attenuation of direct long-distance entanglement generation, creating E2E entanglement requires several entanglement swapping operations to be performed on deployed repeaters [4]. To achieve this process between source-destination (SD) pairs, the challenge lies in selecting the optimal routing path and effectively utilizing network resources [5]. In most existing studies, entanglement routing is initially approached as finding the optimal path using metrics (entanglement distribution rate [6], entanglement fidelity [7], distribution delay [8], etc) suitable for quantum networks between arbitrary SD pairs. Furthermore, to address the performance limitation by a single path, entanglement routing algorithms leveraging multiple paths have been explored to enhance overall network throughput [9], [10] or improve metrics such as request service rate [11].

In recent research, we find that the running results of entanglement routing are closely related to entanglement swapping modes, which determine the operation method of multi-hop swapping on a given path between two distant nodes. There are two representative entanglement swapping modes: the parallel entanglement swapping (PES) [12], [13] and the sequential entanglement swapping (SES) [14], [15]. PES means all intermediate nodes simultaneously perform the Bell-state measurement (BSM) and report the BSM results to the destination node through the classical channel. SES means each intermediate node conducts BSM hop-by-hop and communicates BSM results to the next node along the path through a classical channel. The existing routing works which adopt PES can only consider the final E2E entanglement distribution loss. Lag caused by final end detection leads to the defect that any failed swapping operation destroys the complete distribution process, thereby reducing distribution efficiency. With the premise of tolerable delay, SES greatly improves the above-mentioned adverse effects, as the in-process swapping failure detection brought about by sequentiality saves downstream resources. However, merely changing the swapping mode along the existing paths only achieves partial improvement. Thus, we need to redesign the routing scheme to output a path set that better matches the SES mode because of the following two considerations.

Firstly, routing metrics are reflected as different quantifications in two modes, such as throughput, latency, etc [16]. Most of the existing routing results are based on quantitative metrics in PES mode, which is not suitable for SES mode. Secondly and more importantly, in SES mode, after multiple attempts by the source to distribute entanglement, the current entanglement flow experiences exponential loss with the increasing hops resembling the loss flow model, which is defined as Loss-SES. This is similar to the phenomenon of resource loss or signal attenuation during the transmission of signals through classical networks [17]. However, the existing works output the selected paths which consistently reserve consistent quantum memory units or entangled pairs on each link. A Loss-SES process is always reflected in deploying more upstream resources and fewer downstream resources along the path direction. Therefore, the routing algorithm should reduce the number of reserved resources on each link hop-by-hop based on the loss ratio (i.e., swapping probability). This aligns optimally with the actual entanglement flow, conserving entanglement resources in the network and alleviating competition on bottleneck links to accommodate more demands.

Hence, this work focuses on realizing the potential performance improvement mentioned above in the entanglement routing problem. In this paper, we model and quantify the routing metric EET (expected entanglement throughput) in Loss-SES mode, introducing the lossy allocation model manifested as a stepwise reduction in link capacity. EET reflects the expected E2E entanglement count along a given path before actual network operation which aims to differentiate the performance among alternative paths. Subsequently, we propose the MaxEET routing algorithm, which is capable of outputting an entanglement path with the highest EET for any SD pair. When deploying the MaxEET algorithm in a large-scale network, we address the proprietary problem of the best routing direction for a single request. Then, we also propose a contention-free routing scheme that balances performance and fairness in a multi-request scenario. Finally, we summarize the innovative contributions of our work as follows:

1) We introduce a novel routing metric EET to quantify the distribution performance along a path in Loss-SES mode rigorously. Based on dynamic programming, we propose the MaxEET routing algorithm to output the best path for any SD pair with the maximum EET.
2) For optimal deployment of the MaxEET algorithm, we propose a bidirectional routing mechanism to address the routing direction problem for a single request. Additionally, we propose a contention-free iterative routing scheme tailored for multiple requests, striking a balance between throughput and fairness.
3) We conducted simulations to implement our routing algorithm in Loss-SES mode and compared it with the other two modes. The extensive results verify that the routing results based on our Loss-SES mode are superior in terms of total throughput, fairness, and utilization.

The rest of this paper is organized as follows. We list and discuss the related work in Section II. We introduce the system model and swapping modes in Section III. Then, we formalize the performance of different swapping modes and demonstrate our research motivation in Section IV. Section V presents the routing algorithm in detail. Then, in Section VI, we perform extensive simulations to demonstrate the effectiveness of our routing algorithm compared with the existing works. Finally, we conclude the work and give prospects in Section VII.

II. Related work

In this section, we analyze and compare routing strategies available for two entanglement swapping modes: PES and SES. In addition, we discuss the potential improvement opportunities with our proposed Loss-SES mode.

For PES mode, initially, Van Meter *et al.* [18] explored the application of the Dijkstra algorithm and defined optimal link costs considering the physical properties, to find a single path for each SD pair in the quantum network. Pant *et al.* [12] introduced a greedy-based algorithm in grid quantum networks, utilizing paths with the fewest hops to establish entangled pairs based on information about entanglement generation results. While this multi-path routing scheme improves performance compared to a single-path scheme, decisions based solely on the fewest hops have a limitation of low network resource efficiency. Considering more comprehensive quantum properties, Shi *et al.* [13] proposed the Q-CAST scheme, utilizing the extended Dijkstra's algorithm to find the most reliable paths by incorporating the successful probability of entanglement generation and swapping. Q-CAST utilizes the recovery paths in the residual graph to recover link failures of the major paths to maximize the network throughput. However, Q-CAST compensates too late for failures and lacks flexibility in utilizing successfully created entanglements to connect these sub-connections. In addition, waste of network resources is expected due to redundant path selection for some segments with higher generation probability. Overall, the paths output by these routing schemes all have the same characteristics, namely the consistent reserved resources of each link. These solutions which are suitable for PES mode cause any failed swapping to destroy one complete distribution attempt and lower E2E distribution efficiency.

For SES mode, Li *et al.* [19] designed a connection-oriented entanglement distribution protocol that focuses on guaranteeing the distribution rate of entangled pairs between any two quantum nodes in a large-scale quantum network. However, this work still reserves the same share of resources in the pre-selected path, resulting in some resources not being effectively utilized. Analogous to packet forwarding in classical networks, Chen *et al.* [14] proposed a full spontaneous routing algorithm to adaptively evaluate the congestion on adjacent nodes to avoid potential congestion. We found that this work implements sequential swapping through a best-effort mechanism, which cannot strictly guarantee distribution performance. Xiao *et al.* [15] designed a connectionless remote entanglement distribution protocol to let SD pairs compete for entanglement resources effectively and simultaneously. This work focuses on solving resource allocation problems in distributed settings after path selection, resulting in its overall network performance being limited by the initial path set. Although these works are inspired by the SES mode, the entanglement flow loss effect is rarely considered in path selection, which leads to the objective potential for performance improvement.

III. System Model

In this section, we first introduce the quantum network model based on graph theory. Then, we describe two swapping modes on a path connecting an SD pair.

A. Network Model

We consider a quantum network which consists of the multiple quantum nodes and the quantum links between nodes,

$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{C})$. Here, $\mathcal{V} = \{v_i\}_{i=1}^{N}$ denotes the set of N nodes with the capability of quantum operations, $\mathcal{E} = \{(v_i, v_j) \mid v_i, v_j \in \mathcal{V}, i \neq j\}$ represents the set of edges and the existence of an edge between two nodes implies that these nodes share one or more link-level entanglements. The set $\mathcal{C} = \{C_{i,j}\}^{(v_i,v_j)\in\mathcal{E}}$ encompasses all link-level entanglements, each identified by its two end nodes. The maximum number of link-level entanglements (i.e., memory units) on an edge is referred to as the link capacity $C_{i,j}$.

B. Time Slot and Process Model

To achieve batch processing and better scheduling of resources for optimal routing decisions, the appropriate synchronization duration among all nodes is necessary, which can be achieved by existing synchronization protocols via the Internet [20] [21] [22]. In a time slot similar to existing methodologies, our study systematically executes the following four phases for E2E entanglement distribution, ensuring generality and precision.

- **Receives SD pairs**: The centralized controller collects the SD pairs sponsored by user nodes to form an SD pair set $\mathcal{S} = \{(s_i, d_i)\}$.
- **Entanglement routing**: Paths are found for each SD pair according to our proposed routing algorithm. Each node then binds its memory units to channels belonging to a specific SD pair.
- **Link-level entanglement generation**: Each node attempts to generate entanglements with neighbors on the bound channels and notify the centralized controller of the global generation results.
- **Distribute E2E entangled pairs**: According to the entangled pairs allocated in the previous phase, the relevant repeaters perform entanglement swapping to distribute E2E entangled pairs next.

C. Entanglement between Adjacent Nodes and Remote Nodes

The extension of entanglement occurs through the creation of elementary link-level entangled pairs between adjacent nodes and subsequent swapping at each repeater along the selected path [23] [24]. However, the imperfect quantum devices introduce uncertainty in quantum gates operations [25]. Consequently, the generation and swapping of entanglement are non-deterministic processes, with success probabilities denoted as p and q, respectively. In existing research involving entanglement distribution, there are primarily two deployable swapping modes, as illustrated in Fig. 1:

Parallel entanglement swapping (PES): In this mode, all intermediate nodes concurrently conduct BSM and tell the results to the destination node through classical channels. While this mode reduces waiting time, it lacks the timely detection of failed swapping. Failed swapping events can disrupt the E2E distribution process and are only detected at the destination node.

Sequential entanglement swapping (SES): Sequentiality means that each intermediate node performs BSM hop-by-hop and forwards the BSM result to the next node along the path via classical channels. Successful entanglement swapping operations along the path enable remote nodes to share an entangled pair for subsequent applications. However, most attempts fail midway in the hop-by-hop process, resembling a transmission model akin to flow loss. With the development of more advanced physical devices, the effective lifetime of entanglement is up to the minute level in existing research [26]–[28], so the SES swapping mode is compatible with quantum networks.

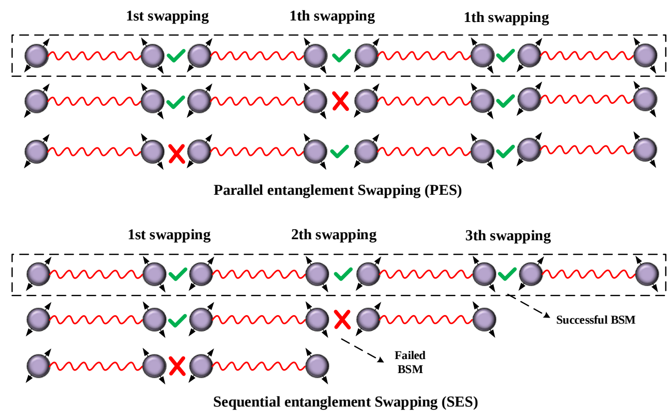


Fig. 1. Comparison: parallel entanglement swapping and sequential entanglement swapping

IV. Formulation and Moviation

In this section, we formalize the performance of different swapping modes and provide their numerical calculation results to visually demonstrate our research motivation. Firstly, in PES mode, Shi *et al.* [13] have defined a routing metric to quantify EET. To evaluate a $(C_1, ..., C_L)$-path, they denote the probability of the k-th hop with reserved capacity $W_k \leq C_k$ on the path having exactly i successful link-level entanglements as Q_k^i, which conforms to the binomial distribution.

$$Q_k^i = \binom{W_k}{i} p^i (1-p)^{W_k - i}. \tag{1}$$

They define the probability of each of the first k hops of has i successful links as $\hat{P}_k^i$. Then we get the recursive formula set, for $i \in \{1, 2, \cdots, W = W_k = min\{C_1, ..., C_L\}\}$, and $k \in \{1, 2, \cdots, L\}$. Further, considering the success probability q of each entanglement swapping, they get EET as:

$$\begin{aligned} \hat{P}_1^i &= Q_1^i, \\ \hat{P}_k^i &= \hat{P}_{k-1}^i \cdot \sum_{l=i}^{W} Q_k^l + Q_k^i \cdot \sum_{l=i+1}^{W} \hat{P}_{k-1}^l, \\ EET &= q^L \cdot \sum_{i=1}^{W} i \cdot \hat{P}_L^i. \end{aligned} \tag{2}$$

As a comparison, with the same resource supply, we formalize the quantification of EET in SES mode. If we define $\overline{P}_k^i$ as the probability of establishing exactly i entangled pairs that cross the first k hops. Specifically, there exists $\overline{P}_1^i = Q_1^i$.

TABLE I
PERFORMANCE WHEN $q = p = 0.8$

Nodes	Scheme	4	5	6	7	8	9	10
EET (pairs)	PES	9.265	7.232	5.679	4.476	3.537	2.800	2.219
	SES	**9.538**	**7.623**	**6.098**	**4.878**	**3.902**	**3.122**	**2.497**
	Loss-SES	9.400	7.434	5.894	4.695	3.745	2.990	2.388
Cost (units/pair)	PES	6.475	11.060	17.605	26.804	39.577	57.142	81.104
	SES	6.290	10.493	16.398	24.597	35.871	51.245	72.063
	Loss-SES	**6.169**	**9.953**	**14.929**	**21.511**	**30.166**	**41.463**	**56.111**

Based on the forward recursive calculation, we obtain EET that can be represented as:

$$\begin{aligned}
\overline{P}_k^i = &\sum_{n=i}^{W}(\overline{P}_{k-1}^n \binom{n}{i} q^i(1-q)^{n-i} \cdot \sum_{l=n}^{W} Q_k^l) + \\
&\sum_{n=i+1}^{W}(\overline{P}_{k-1}^n \cdot \sum_{l=i}^{n-1} Q_k^l \binom{l}{i} q^i(1-q)^{l-i}), \\
EET = &\sum_{i=1}^{W} i \cdot \overline{P}_L^i.
\end{aligned} \tag{3}$$

Similar to the loss flow model, for our Loss-SES mode, the link width is not consistent but rather satisfies the loss ratio. We design $W_k = \lceil W_{k-1} \cdot q \rceil + 1$ for the following reasons. Although $\overline{P}_k^i$ is a complex probability distribution function, through numerical calculations, we find that when i equals an integer close to $W_{k-1} \cdot q$, $\overline{P}_k^i$ always dominates over $\overline{P}_k^j$, when $j \ll i$ or $j \gg i$. The swapping result that approximates the reduced allocation quantity always occurs with the maximum probability, instead of some small probability events that can be ignored. The width of the first two links needs to be consistent because only the first swapping occurs there ($W_2 = W_1$). Similarly, we can deduce EET as follows:

$$\begin{aligned}
&W_2 = W_1, W_k = \lceil W_{k-1} \cdot q \rceil + 1, k > 2, \\
&\overline{P}_k^i = \sum_{n=i}^{W_k}(\overline{P}_{k-1}^n \binom{n}{i} q^i(1-q)^{n-i} \cdot \sum_{l=n}^{W_k} Q_k^l) + \\
&\quad \sum_{n=i+1}^{W_{k-1}}(\overline{P}_{k-1}^n \cdot \sum_{l=i}^{min\{n-1, W_k\}} Q_k^l \binom{l}{i} q^i(1-q)^{l-i}), \\
&EET = \sum_{i=1}^{W_L} i \cdot \overline{P}_L^i.
\end{aligned} \tag{4}$$

We present numerical results of EET of three modes on paths with varying hop counts, ranging from 4 to 10. Additionally, we define the intuitive "cost" metric to reflect the proportion of reserved memory units and EET, given by $cost = \frac{\sum_{k=1}^{L} W_k}{EET}$. Each link's capacity is set to 20 units, and we set typical values $q = p = 0.8$. In both PES and SES modes, the consumed resources are identical, indicating that all path memory units are utilized to complete entanglement distribution. However, in Loss-SES mode, the consumption of memory units gradually decreases during forward sequential swapping, resulting in the least amount of resources consumed.

Table I elucidates the following two observations. First, when the resource consumption along the path is identical, the SES mode enhances the EET compared to the PES mode, which conforms to the previous analysis. Second, deploying the Loss-SES mode allows for the reduction of resource consumption along the path without significantly diminishing the EET (approaching optimal), as evidenced by the lowest cost. Therefore, by amalgamating the entanglement flow loss effect and SES mode, we markedly alleviate resource contention on core bottleneck links in the network. The preserved entanglement resources can accommodate more requests, thereby enhancing the overall performance.

V. ROUTING DESIGN

In this section, we design the entanglement routing algorithm based on the previous routing metric EET and solve the key problems of applying it in scenarios of single request and multiple requests.

A. MaxEET Routing Algorithm

Motivated by the original Dijkstra's algorithm for finding the shortest path, we aim to construct an optimal spanning tree with the source node s as the root to determine the optimal path with the maximum EET defined in Eq. (4) within a quantum network. While Dijkstra's classical algorithm is effective for finding the shortest path in cases where the path cost is additive, the proposed EET metric involves complex iterative computations. Fortunately, the evaluation function EET for a given path P exhibits monotonic decreasing behavior as P is extended to a longer path by appending another link at the end. As P expands, the link width aligns with the loss ratio, at least, does not increase. In other words, the new edge W_k is narrower than W_{k-1}. Additionally, the addition of another hop implies more swapping operations, neither of them can elevate EET. This monotonic decreasing property enables us to utilize the dynamic programming approach to find the best path with maximum EET. The detailed algorithm process is provided in Algorithm 1.

Now, assuming the source node is denoted as s and the destination node as d, we establish two sets of nodes: the determined node set $\mathcal{D}$ and the undetermined node set $\mathcal{U}$. Initially, the determined set only includes s. Meanwhile, we introduce the set $\mathcal{T}$ which records the maximum EET from the s to each node in $\mathcal{V}$. The evaluation value from s to an unvisited node n is set as $-\infty$, or the evaluation value $EET(s,n)$ of the edge (s,n) if s and n are neighbors.

At each step, we traverse from the just-determined node u ($\mathcal{D}$.dequeue()) to its adjacent nodes and calculate the next link width W_k based on the loss ratio and the previous link width W_{k-1}. Specifically, $Loss(.)$ is a piecewise function and cannot exceed the maximum link capacity C_k, as shown in Eq. (4). The evaluation values from s to any other node v are updated if v and u are neighbors and the newly obtained value is greater than the existing recorded value to ensure that the current evaluation value is optimal. Subsequently, the current node $\hat{u} \in \mathcal{U}$ with the maximum evaluation value $\mathcal{T}[\hat{u}]$ is added to the set $\mathcal{D}$ and deleted from $\mathcal{U}$. The cycle stops when the destination $d \in \mathcal{D}$.

Then, we need to perform a reverse correction on the path to release more redundant resources. For example, $W_k = Loss(W_{k-1}) = min\{\lceil W_{k-1} \cdot q \rceil + 1, C_k\}$, if $C_k < \lceil W_{k-1} \cdot q \rceil + 1$, i.e., demand exceeds supply, so the traversed links need to reduce the allocated capacity $(W_1, ..., W_{k-1})$. Reverse correction matches the uneven link capacity with the decreasing allocation amount per hop. Ultimately, the algorithm can output the optimal path from s to d, i.e., with the maximum EET. Next, we introduce the application of this routing algorithm in the demand scenarios of single and multiple requests in quantum networks.

Algorithm 1: Maximum EET Entanglement Routing

```
Input: Graph G = (V, E, C), an SD pair (s, d);
Output: The best path P = [W_1, W_2, ..., W_L].
Determined set D = {s};
Undetermined set U = {V/s};
Set T of maximum EET, record value of each node is
  initialized to -∞;
while d not in D do
    u = D.dequeue();
    for each node v ∈ neighbors of u do
        Calculate the next link width
          W_k = Loss(W_{k-1});
        Calculate EET(s, v) via node u;
        if EET(s, v) > T[v] then
            T[v] = EET(s, v);
        end
    end
    Find the node û ∈ U which has maximum T[û];
    Push node û to D;
    Delete node û from U;
end
Reverse path correction and update T[d];
Return T[d] and path P = [W_1, W_2, ..., W_L];
```

B. Bidirectional Routing for Single SD Pair

In contrast to existing routing schemes where the initiator of routing calculations consistently yields the same result regardless of whether it is the source or destination, the Loss-SES mode introduces differentiated resource reservation for each link on the path. In a network with uneven resource distribution, the EET calculated from the source or destination may differ. For instance, in a path with capacity [5, 4, 4, 3], after source routing, the path may be [4, 4, 3, 2], while after destination routing, it could be [3, 3, 2, 1]. Therefore, our proposed Loss-SES mode raises the problem of optimal path selection in bidirectional routing.

Within the same network topology, the optimal routing direction can be evaluated by comparing the EET from the source to the destination and the EET from the destination to the source. Subsequently, the larger routing result between the two alternative calculations can be chosen to output the path, update the topology, and iterate the above process repeatedly to achieve multi-path output for a single SD pair. It is important to emphasize that the routing direction signifies the direction of subsequent sequential entanglement swapping.

C. Multi-round Iterative Routing for Multiple SD Pairs

When managing multiple requests, the primary objective of entanglement routing is to establish multiple paths based on the knowledge of SD pairs. These paths must be contention-free in resource allocation while maximizing the overall EET and ensuring equitable treatment among requests.

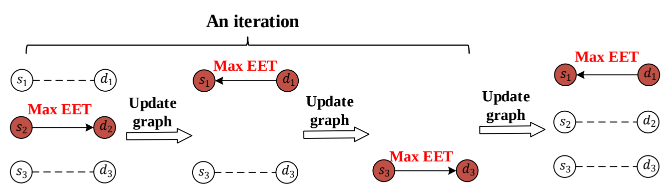


Fig. 2. A multi-round iterative routing algorithm for multiple requests

We propose to search multiple contention-free paths for online SD pairs using a greedy algorithm, as shown in Algorithm 2. In each iteration, we initialize an identical SD pairs set $\mathcal{S}' = \mathcal{S}$ and there are three cyclic steps involved.

1) For every SD pair, we use the Algorithm 1 to find the best path while considering the routing direction.
2) Among the best paths $\mathcal{P}$ for all SD pairs, it further selects the path with the highest EET and reserves the resources (memory units and channels) of this path.
3) The network topology is updated to the residual graph by removing the reserved resources. Remove the corresponding SD pair from $\mathcal{S}'$.

When the set $\mathcal{S}'$ is empty or no path can be found for any remaining requests in $\mathcal{S}'$, this iteration ends (line 12). Then a new iteration is initiated, and the algorithm terminates when it finds that no available path can be found for any SD pair, as shown in Fig. 2. Meanwhile, for any request, the path selection follows our proposed bidirectional routing strategy, which is reflected in the arrow direction between SD pairs. We can merge the output of the path from all iterations for each request to obtain the final routing result $\mathbb{P} = \{P_1, ..., P_n\}$ for each SD pair. In summary, in each iteration, the algorithm strives to

Algorithm 2: Multi-round Iterative Routing

Input: Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{C})$, SD pairs set $\mathcal{S}$;
Output: Path set $\mathbb{P} = \{P_1, ..., P_n\}$ for each SD pair;
while *True* **do**
  Temporary SD pairs set $\mathcal{S}' = \mathcal{S}$;
  `/* Once iterative routing; */`
  while $\mathcal{S}'$ **do**
    for *each SD pair i in* $\mathcal{S}'$ **do**
      Get $EET(s_i, d_i)$ and $EET(d_i, s_i)$;
      Preserve maximum routing results;
    end
    Get path set $\mathcal{P}$ in this round ;
    if $\mathcal{P}$ ==*NULL and* $\mathcal{S}' = \mathcal{S}$ **then**
      Return path set $\mathbb{P}$ for each SD pair;
    end
    if $\mathcal{P}$ ==*NULL* **then**
      Break;
    end
    Add path with the maximum EET in $\mathcal{P}$ to $\mathbb{P}$;
    Remove the corresponding SD pair from $\mathcal{S}'$;
    Update network topology $\mathcal{G}$;
  end
end

output one optimal path for each request, improving fairness. The exhaustive iteration also maximizes the utilization of the entire network.

VI. Simulation

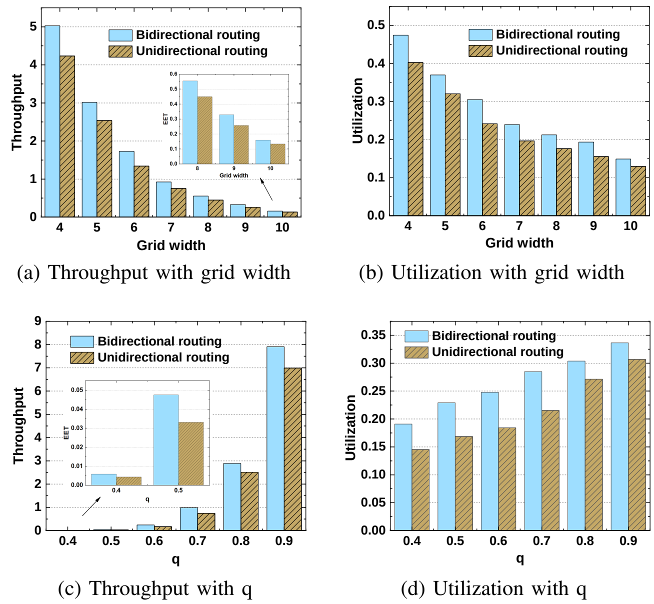
In this section, we provide a detailed methodology for simulation and introduce relevant comparison schemes as well as corresponding evaluation results.

A. Methodology

Setting: We construct grid topologies of different scales for simulation, with the capacity of each link set to a random value of 10 to 20. When simulating single request routing results, we deploy an SD pair at the two endpoints of the diagonal of the grid. In a multi-request scenario, SD pairs are randomly deployed but not duplicated, ensuring that at least one entanglement swapping operation is required between any requests. In any scenario, we set a typical value for the probability of entanglement generation $p = 0.75$. For each metric, we repeat the simulation 100 times to reduce accidental errors. Our simulation code is developed based on Python version 3.10.

Evaluation metrics: We adopt the following metrics to comprehensively evaluate the performance of routing schemes in different swapping modes:

1) Total throughput: This metric provides an intuitive reflection of the overall serviceability available for all network requests, representing the sum of throughput values for all requests.



(a) Throughput with grid width
(b) Utilization with grid width
(c) Throughput with q
(d) Utilization with q

Fig. 3. Simulation results of single SD pair.

2) Minimum throughput: It assesses fairness among requests and indicates the ability to provide performance guarantees for non-dominant requests.
3) Cost: We define "cost" as the ratio of resource (i.e., reserved memory units) consumption to final throughput, i.e., $cost = \frac{\sum^{\mathbb{P}} \sum_{k=1}^{L} W_k}{total\ throughput}$. A lower cost signifies achieving equivalent entanglement distribution performance with less resource overhead.
4) Utilization: This metric represents the percentage of network resources reserved by all requests in the network. A better routing algorithm maximizes resource utilization to enhance overall network performance.

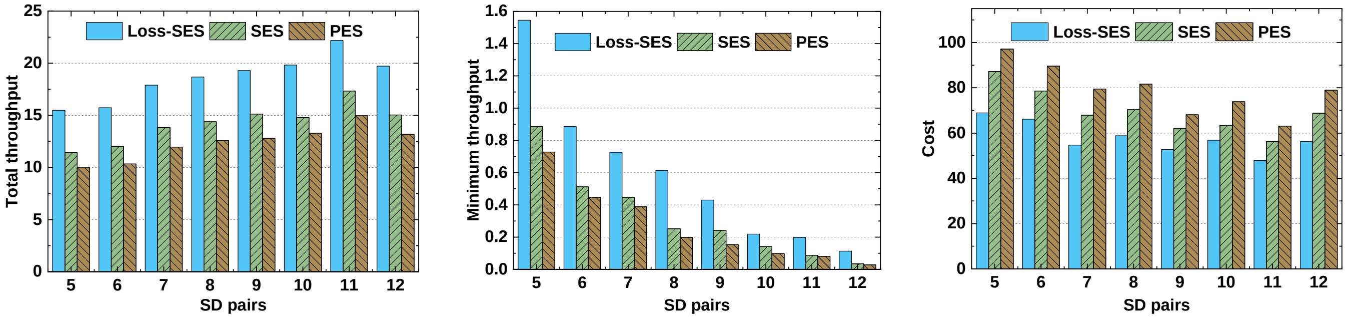
B. Comparison Schemes

Single request: We simulate the iterative routing results of bidirectional routing and unidirectional routing for the same request until no available paths remain in the network.

Multiple requests: Our comparison scheme is inspired by the other two EET calculation methods mentioned in Section IV. We implement Algorithm 1 based on Eq. (2), which reserves the same resources for each link on the path and adopts the PES mode. Additionally, Algorithm 1 is implemented based on Eq. (3), where path resources are also lossless, but entanglement swapping is achieved using the SES mode. These two comparison schemes still adhere to our proposed contention-free iteration routing strategy when exhausting network resources for multiple requests.

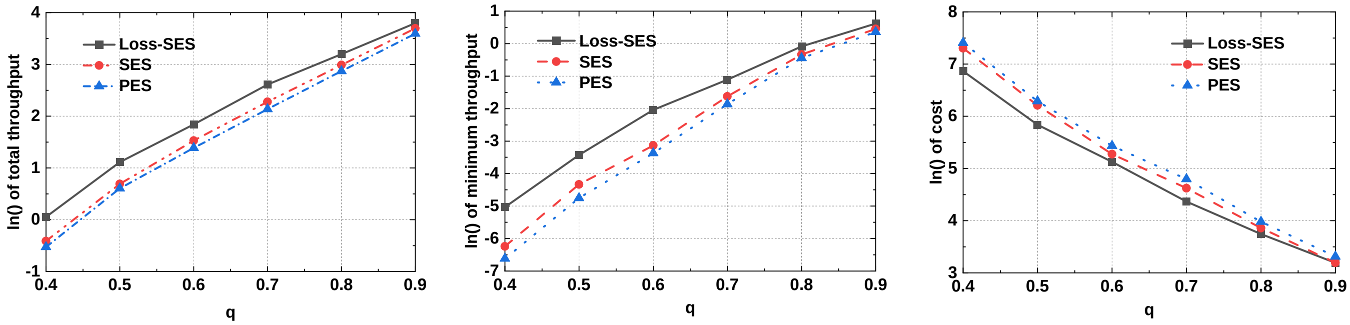
C. Evaluation Results

Firstly, we verify the importance of routing direction. When $q = 0.75$, as the grid size gradually expands, it is observed that with the increase in path length, both throughput and utilization gradually decrease, as shown in Fig. 3 (a)(b). However, the selection of the optimal routing direction



(a) Total throughput with SD pairs (b) Minimum throughput with SD pairs (c) Cost with SD pairs

Fig. 4. Performance comparison results as the number of SD pairs increases for multiple requests.



(a) Total throughput with q (b) Minimum throughput with q (c) Cost with q

Fig. 5. Performance comparison results as the swapping probability q increases for multiple requests.

from bidirectional routing consistently outperforms fixed unidirectional routing. This phenomenon can be explained by the fact that unidirectional routing always exhausts all entanglement resources around the source (or destination) earlier, thereby hindering further performance improvement.

As the swapping probability q gradually increases, it becomes easier to construct an E2E entangled pair. Consequently, the throughput gradually increases, and the corresponding loss ratio also gradually increases, resembling more closely the SES mode, thus occupying more entanglement resources and leading to increased utilization. In SES mode, there is no issue of selecting the routing direction. Therefore, it is observed that the relative advantage of bidirectional routing diminishes, and the utilization gap also gradually narrows as the probability q increases, as shown in Fig. 3 (c)(d).

Secondly, in the multi-request scenario, performance simulation comparisons are conducted when changing the number of SD pairs and swapping probability q. Firstly, in a 64-node grid topology, $q = 0.75$, the number of SD pairs is increased from 5 to 12 without any positional duplicates, as shown in Fig.4. Overall, in the same routing framework, the performance achieved based on Loss-SES mode is superior to the other two modes (SES mode and PES mode). As the number of SD pairs increases, network resources are utilized more efficiently, leading to a gradual increase in the total throughput, eventually approaching a saturated allocation, as shown in Fig.4 (a). However, this heightened resource competition intensifies, resulting in a continuous decrease in the minimum throughput, as shown in Fig.4 (b). Our proposed routing algorithm based on Loss-SES mode consistently exhibits the highest total throughput and the maximum minimum throughput, demonstrating its advantages in both throughput and fairness. Additionally, in Fig.4 (c), due to the uncertainty in path length and width, the cost of resource consumption fluctuates. Nonetheless, the resource allocation results based on the loss ratio ensure more efficient entanglement distribution, resulting in our approach achieving the lowest cost.

Subsequently, simulations are conducted on the same 64-node grid topology, with the swapping probability q varying from 0.4 to 0.9 to reflect an increasingly higher quality of swapping operations. In each round of simulation, 8 SD pairs are selected, with their positions randomly selected. Due to the wide range of result values, the corresponding natural logarithms are calculated and used as the vertical axis. Overall, it is observed that all performance metrics exhibit exponential variations as q increases. With the probability of successful swapping operations rising, the limited resources along the paths have more potential to form E2E entangled pairs. Consequently, the total throughput and minimum throughput of the entire network continue to increase, as shown in Fig.5 (a)(b). Due to higher-quality swapping operations, the cost measuring resource consumption steadily decreases in Fig.5 (c). Nonetheless, the routing results based on the Loss-SES mode consistently outperform the other two modes, underscoring the inherent advantage of our approach in conserving resources to serve more SD pairs.

Another important observation is that this advantage for both three metrics becomes more significant as the probability decreases, even by several times. This is because a lower probability q implies a lower loss ratio, allowing more resources along the paths to be freed up without significantly affecting the distribution performance. As q approaches 1, the Loss-SES mode gradually approaches the conventional SES mode, leading to a diminishing performance gap between them. Therefore, it can be concluded that our proposed routing scheme is better suited for early-stage quantum networks with lower swapping probability.

VII. Conclusion and Prospect

Considering the entanglement flow loss effect raised by probabilistic swapping, we introduced a novel routing metric EET to quantify the distribution performance for a path in Loss-SES mode. Based on dynamic programming, we proposed a routing algorithm MaxEET to output the best path with the maximum EET. For optimal deployment, we presented a bidirectional routing mechanism to address the routing direction problem for a single request. Additionally, we designed a contention-free iterative routing scheme tailored for multiple requests, striking a balance between throughput and fairness. The extensive simulation verified that the routing results based on our Loss-SES mode are superior in terms of total throughput, fairness, and utilization.

Furthermore, our work can be improved in the following aspects: it is evident that although network resource utilization has improved, there are some sub-paths (entanglement fragments) in the network. Although these cannot form complete paths, they can be utilized as auxiliary connections to some segments of the established path (even a single link), thereby further enhancing the performance of that path. We will explore more advanced solutions in subsequent work to enhance the effect of this study.

References

[1] H.-K. Lo, M. Curty, and K. Tamaki, "Secure quantum key distribution," *Nature Photonics*, vol. 8, no. 8, pp. 595–604, 2014.

[2] E. O. Ilo-Okeke, L. Tessler, J. P. Dowling, and T. Byrnes, "Remote quantum clock synchronization without synchronized clocks," *npj Quantum Information*, vol. 4, no. 1, p. 40, 2018.

[3] J. F. Fitzsimons and E. Kashefi, "Unconditionally verifiable blind quantum computation," *Physical Review A*, vol. 96, no. 1, p. 012303, 2017.

[4] J.-W. Pan, D. Bouwmeester, H. Weinfurter, and A. Zeilinger, "Experimental entanglement swapping: entangling photons that never interacted," *Physical review letters*, vol. 80, no. 18, p. 3891, 1998.

[5] F. Dupuy, C. Goursaud, and F. Guillemin, "A survey of quantum entanglement routing protocols—challenges for wide-area networks," *Advanced Quantum Technologies*, p. 2200180, 2023.

[6] W. Dai, T. Peng, and M. Z. Win, "Optimal remote entanglement distribution," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 3, pp. 540–556, 2020.

[7] J. Li, M. Wang, K. Xue, R. Li, N. Yu, Q. Sun, and J. Lu, "Fidelity-guaranteed entanglement routing in quantum networks," *IEEE Transactions on Communications*, vol. 70, no. 10, pp. 6748–6763, 2022.

[8] M. Ghaderibaneh, C. Zhan, H. Gupta, and C. Ramakrishnan, "Efficient quantum network communication using optimized entanglement swapping trees," *IEEE Transactions on Quantum Engineering*, vol. 3, pp. 1–20, 2022.

[9] Y. Zhao and C. Qiao, "Redundant entanglement provisioning and selection for throughput maximization in quantum networks," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2021, pp. 1–10.

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

[11] Y. Zeng, J. Zhang, J. Liu, Z. Liu, and Y. Yang, "Multi-entanglement routing design over quantum networks," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2022, pp. 510–519.

[12] M. Pant, H. Krovi, D. Towsley, L. Tassiulas, L. Jiang, P. Basu, D. Englund, and S. Guha, "Routing entanglement in the quantum internet," *npj Quantum Information*, vol. 5, no. 1, pp. 1–9, 2019.

[13] S. Shi and C. Qian, "Concurrent entanglement routing for quantum networks: Model and designs," in *Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM)*, 2020, pp. 62–75.

[14] L. Chen, K. Xue, J. Li, R. Li, N. Yu, Q. Sun, and J. Lu, "Q-DDCA: Decentralized dynamic congestion avoid routing in large-scale quantum networks," *IEEE/ACM Transactions on Networking*, vol. 32, no. 1, pp. 368–381, 2024.

[15] Z. Xiao, J. Li, K. Xue, Z. Li, N. Yu, Q. Sun, and J. Lu, "A connectionless entanglement distribution protocol design in quantum networks," *IEEE Network*, 2023.

[16] A. Chang and G. Xue, "Order matters: On the impact of swapping order on an entanglement path in a quantum network," in *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2022, pp. 1–6.

[17] A. S. Avestimehr, S. N. Diggavi, and N. David, "Wireless network information flow: A deterministic approach," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 1872–1905, 2011.

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

[19] J. Li, Q. Jia, K. Xue, D. S. Wei, and N. Yu, "A connection-oriented entanglement distribution design in quantum networks," *IEEE Transactions on Quantum Engineering*, vol. 3, pp. 1–13, 2022.

[20] L. Chen, K. Xue, J. Li, N. Yu, R. Li, J. Liu, Q. Sun, and J. Lu, "A heuristic remote entanglement distribution algorithm on memory-limited quantum paths," *IEEE Transactions on Communications*, vol. 70, no. 11, pp. 7491–7504, 2022.

[21] C. Cicconetti, M. Conti, and A. Passarella, "Request scheduling in quantum networks," *IEEE Transactions on Quantum Engineering*, vol. 2, pp. 2–17, 2021.

[22] K. Chakraborty, D. Elkouss, B. Rijsman, and S. Wehner, "Entanglement distribution in a quantum network: A multicommodity flow-based approach," *IEEE Transactions on Quantum Engineering*, vol. 1, pp. 1–21, 2020.

[23] A. Dahlberg, M. Skrzypczyk, T. Coopmans, L. Wubben, F. Rozpędek, M. Pompili, A. Stolk, P. Pawełczak, R. Knegjens, J. de Oliveira Filho *et al.*, "A link layer protocol for quantum networks," in *Proceedings of the ACM Special Interest Group on Data Communication (SIGCOMM)*, 2019, pp. 159–173.

[24] A. S. Cacciapuoti, M. Caleffi, R. Van Meter, and L. Hanzo, "When entanglement meets classical communications: Quantum teleportation for the quantum internet," *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3808–3833, 2020.

[25] W. Kozlowski, A. Dahlberg, and S. Wehner, "Designing a quantum network protocol," in *Proceedings of the ACM International Conference on Emerging Networking Experiments and Technologies (CoNEXT)*, 2020, pp. 1–16.

[26] W. Kozlowski and S. Wehner, "Towards large-scale quantum networks," in *Proceedings of the 2019 ACM International Conference on Nanoscale Computing and Communication (NANOCOM)*. ACM, 2019, pp. 1–7.

[27] C. E. Bradley, J. Randall, M. H. Abobeih, R. Berrevoets, M. Degen, M. A. Bakker, M. Markham, D. Twitchen, and T. H. Taminiau, "A ten-qubit solid-state spin register with quantum memory up to one minute," *Physical Review X*, vol. 9, no. 3, p. 031045, 2019.

[28] I. Inlek, C. Crocker, M. Lichtman, K. Sosnova, and C. Monroe, "Multispecies trapped-ion node for quantum networking," *Physical Review Letters*, vol. 118, no. 25, p. 250502, 2017.