2025 International Conference on Quantum Communications, Networking, and Computing (QCNC) | 979-8-3315-3159-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/QCNC64685.2025.00061

Towards Stability and Utility Maximization in Adversarial Quantum Networks under Bandit Feedback (Invited)

Yu Chen
IIIS
Tsinghua University
Beijing, China
chenyu23@mails.tsinghua.edu.cn

Longbo Huang
IIIS
Tsinghua University
Beijing, China
longbohuang@tsinghua.edu.cn

John C.S. Lui
CSE
The Chinese University of Hong Kong
Shatin, N.T, Hong Kong
cslui@cse.cuhk.edu.hk

***Abstract*—This paper addresses the dual challenges of stabilizing entanglement scheduling and maximizing communication fidelity in multi-hop quantum networks under adversarial dynamics. Unlike prior works relying on stationary assumptions or perfect pre-decision knowledge, we consider scenarios where entanglement generation rates and task arrival patterns vary arbitrarily over time, with only bandit feedback available post-scheduling. We first propose `Q-NSO`, achieving stability for quantum networks under arbitrary time-varying entanglement request arrival rates and bandit feedback. Equipped with the key components of `Q-NSO`, we propose `Q-UMO`, a learning-based framework towards maximal utility by determining the entanglement request arrivals and quantum link allocation distribution. Integrating adversarial bandit convex optimization, an online learning algorithm, and Lyapunov drift-plus-penalty analysis, theoretical analysis demonstrates that `Q-UMO` ensures network stability with bounded queue lengths while achieving a sub-linear regret bound against the reference policy. This work provides a robust foundation for scalable quantum internet applications, balancing stability and performance in adversarial environments with limited feedback.**

***Index Terms*—Stability, Utility maximization, Quantum networks, Bandit Feedback, Adversarial settings.**

I. Introduction

Quantum networks are poised to revolutionize communication and computation by enabling the distribution of entanglement, a fundamental resource for quantum protocols such as teleportation, secure cryptography, and distributed quantum computing. A critical challenge in realizing scalable quantum networks lies in efficiently scheduling entanglement generation and routing under dynamic demand and stochastic link successes. While significant progress has been made in developing scheduling and routing algorithms for quantum networks, existing approaches predominantly rely on stationary assumptions about entanglement success probabilities and arrival rates. This paper addresses the pressing need for robust scheduling algorithms that adapt to adversarial dynamics in real-world quantum networks, where environmental fluctuations and unpredictable interference render stationary assumptions impractical.

Recent works have proposed a rich array of protocols, algorithms, and scheduling policies to enhance entanglement routing efficiency on diverse network architectures. For instance, prior work has investigated opportunistic routing mechanisms [1], and developed online approaches for entanglement routing [2]. Other studies analyze the stability aspect of quantum networks using the max-weight policy [3], build proactive entanglement generation schemes with overlay structures [4], or propose advanced frameworks for effectively increasing entanglement fidelity and fairness [5]. Additional methods incorporate end-to-end routing with purification [6], throughput-optimal memory allocation for bipartite requests [7], and reinforcement learning-based algorithms [8]. Attention has also been devoted to multi-entanglement routing [9], satellite-assisted entanglement distribution [10], capacity characterization under purification [11], and entanglement fusion techniques [12] that can simultaneously fuse multiple entangled states.

In parallel, research efforts have explored capacity analysis for quantum switches with or without decoherence [13], asynchronous routing and provisioning [14], transport layer protocols guiding quantum data [15], and novel mechanisms for integrating quantum swapping with optical switching [16]. There are also studies that seek to support multi-party entanglement [17], investigate the switching performance of quantum distribution policies [18], design routing algorithms suited for heterogeneous entanglement durations [19], apply Lyapunov drift optimization methods [20], offer entanglement routing protocols aimed at maximizing the number of successful quantum-user pairs [21], develop ranking-based distribution protocols [22], and propose scheduling strategies for multi-source entanglement distribution [23].

Despite these important advances, the vast majority of the existing literature primarily focuses on settings where the entanglement generation processes conform to stationary probabilistic models. Under such assumptions, entanglement success probabilities remain fixed or vary according to well-characterized stochastic distributions. In practice, however, the performance of quantum channels can fluctuate dramatically

979-8-3315-3159-1/25/$31.00 ©2025 IEEE
DOI 10.1109/QCNC64685.2025.00061

due to interference, hardware imperfections, or environmental uncertainties. These factors can render stationarity assumptions invalid, making many existing methods less robust or inapplicable to real-world scenarios where entanglement generation is subject to arbitrary and potentially adversarial dynamics. Accordingly, it becomes vital to develop scheduling and routing algorithms capable of adapting to unpredictable quantum channel conditions and dynamic user demands without relying on rigid statistical models.

In this work, we take a step toward realizing such robust quantum routing solutions by studying the problems of *stability* and *utility maximization* under adversarial variations in entanglement success probabilities and task arrival rates. Specifically, we assume that an oblivious adversary can choose the success probabilities and arrival processes arbitrarily over time. Moreover, the proposed algorithms are designed under *bandit feedback*, wherein after selecting a particular routing path for entanglement generation, the scheduler only observes the actual entanglement outcomes on that path, without access to any additional channel information. This feedback model accurately encapsulates scenarios where quantum network monitoring is prohibitively resource-heavy, or where inherent hardware constraints—such as in satellite-based systems—limit continuous surveillance of global link states.

Building on the novel adversarial learning-based framework recently developed in [24], we design two algorithms, `Q-NSO` and `Q-UMO`, that address the pressing need for both stability and utility guarantees in quantum multi-hop networks. The `Q-NSO` mechanism proves network stability by ensuring that queue backlogs remain bounded under adversarial entanglement requests and bandit feedback. To complement this, `Q-UMO` focuses on achieving strong utility performance by maximizing a well-defined network utility function, which captures key operational metrics related to entanglement distribution and user satisfaction. Both algorithms are lightweight, requiring only limited online observations, and are thus suitable for realistic quantum infrastructures where real-time data collection may be highly constrained.

A. Notations

We use bold letters to denote vectors, e.g., $\boldsymbol{q}_t$, $\boldsymbol{\mu}_t$, $\boldsymbol{A}_t$, and denote their elements with corresponding normal letters, e.g., $q_{t,i}$, $\mu_{t,i}$, $A_{t,i}$. For an integer $n \geq 0$, $[n]$ stands for $\{1, 2, \ldots, n\}$. For a finite set S, $\Delta(S)$ is the simplex over S, i.e., $\{\boldsymbol{x} \in \mathbb{R}^{|S|} \mid \sum_{i=1}^{|S|} x_i = 1\}$, where every element $\boldsymbol{x} \in \Delta(S)$ is a discrete probability distribution over S. We use $\mathcal{O}$ to hide all absolute constants, and use $\tilde{\mathcal{O}}$ to additionally hide all logarithmic factors. For functions $f(T)$ and $g(T)$, we say $f(T) = \mathcal{O}_T(g(T))$ if $\limsup_{T\to\infty} \frac{f(T)}{g(T)} < \infty$, and $f(T) = o_T(g(T))$ if $\limsup_{T\to\infty} \frac{f(T)}{g(T)} = 0$.

II. Problem Formulation

A. System Model

We consider a discrete-time multi-hop quantum network composed of a finite set of quantum nodes $\mathcal{N} = \{1, 2, \ldots, N\}$,

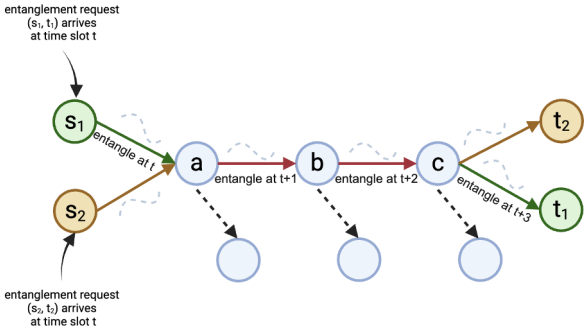

Fig. 1. An example of entanglement requests transiting through a multi-hop quantum network. Entanglement requests arrive at the nodes and end-to-end entanglements are established by link-level entanglements and swapping.

interconnected by directional quantum links $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$. A link $(n, m) \in \mathcal{L}$ indicates that there exists a physical or logical channel from node n to node m (e.g., a fiber optic link). At the beginning of each time slot t, a set of entanglement requests arrive at various source nodes and require end-to-end entanglements with designated destination nodes. The network then performs entanglement routing—activating link-level entanglements along a pre-specified path—until end-to-end entanglements are established.

Here we show an example of an entanglement request transiting through a multi-hop quantum network in Figure 1. At time slot t, we assume an end-to-end entanglement request (s_1, t_1) arrives at the quantum node s_1, which means that the scheduler needs to establish the entanglement between the qubits in node s_1 and t_1. Simultaneously, another entanglement request (s_2, t_2) arrives at node s_2. However, there may not exist a direct quantum link between s_1, t_1 and s_2, t_2. Therefore, we need to transmit this request through the path in this multi-hop quantum network. For example, to process the request (s_1, t_1), the scheduler first establishes the entanglement between the quantum nodes s_1 and a at time slot t, as shown in Figure 1. This can be viewed as "transmitting" the request from s_1 to a. Then, in the next time slot, by establishing an entanglement between a and b and performing entanglement swapping, we can establish an entanglement between s_1 and b, and the request can be similarly viewed as being "transmitted" to b now. This process goes on until the request is eventually delivered to its destination t_1. By then, an end-to-end entanglement is setup between s_1 and t_1 and we say that the request is served.

To model the procedure of scheduling the multi-hop quantum network with multiple requests, we consider a time horizon of T discrete slots, labeled $t = 1, 2, \ldots, T$. For each time slot t, the following notations will be used throughout the paper:

1) **Entanglement Request Arrivals:** Let $A_n^{(k)}(t)$ denote the number of new end-to-end entanglement requests arriving at node $n \in \mathcal{N}$ for destination node $k \in \mathcal{N}$. In other words, these arrivals represent requests for

establishing a quantum entanglement between node n and node k over multiple hops if necessary. We assume that $A_n^{(k)}(t)$ is bounded uniformly by $A_{\max}$ for all valid $n, k \in \mathcal{N}$ with $n \neq k$. Moreover, we denote $\mathcal{A}$ as the feasible request arrival set, such that $\boldsymbol{A}(t) \in \mathcal{A}$ for all $t \in [T]$.

2) **Entanglement Queues:** Each node $n \in \mathcal{N}$ maintains one or more logical queues $Q_n^{(k)}(t)$ to track the number of requests at node n that are still pending for a particular destination k. The queues evolve over time based on the requests served on outgoing links from n and any incoming requests from other nodes linking to n.

B. Entanglement Scheduling

Notice that in the example illustrated in Figure 1, two different requests (s_1, t_1) and (s_2, t_2) may meet at the same quantum node c at the same time. Constrained by the physical devices and environmental conditions, the scheduler needs to determine which quantum links $(n, m) \in \mathcal{L}$ to activate and with what capacity allocation. We assume that each activated link (n, m) has a capacity $C_{n,m}(t) \in [0, M]$, reflecting the probability of successfully established quantum entanglement qubits from node n to node m. The value of M is a known constant bounding link capacities.

To handle multi-hop flows of entanglement requests from different sources and destined for different nodes, we introduce an allocation plan $\boldsymbol{p}_{n,m}(t) \in \Delta(\mathcal{N})$ on each link (n, m), where $\Delta(\mathcal{N})$ is the probability simplex over destinations in $\mathcal{N}$. Specifically, $p_{n,m}^{(k)}(t)$ represents the fraction of link capacity $C_{n,m}(t)$ assigned to serve requests from queue $Q_n^{(k)}(t)$. Roughly speaking, approximately $p_{n,m}^{(k)}(t)\, C_{n,m}(t)$ requests can be transferred from $Q_n^{(k)}$ to $Q_m^{(k)}$ during slot t.

Formally, the number of entanglement requests of destination k successfully served from node n to m is a random variable $\mu_{n,m}^{(k)}(t) \in [0, M]$ with an expected value:

$$\mathbb{E}\big[\mu_{n,m}^{(k)}(t)\big] \;=\; C_{n,m}(t)\; p_{n,m}^{(k)}(t).$$

We operate under a *bandit feedback* model, meaning that at the end of time slot t, the scheduler only observes the random outcomes $\mu_{n,m}^{(k)}(t)$ for links actually chosen, as well as the realized link capacities $C_{n,m}(t)$, but gains no direct or prior knowledge about the capacities or outcomes on links not used. This setup contrasts with previous works such as [3], [13], where link rates are assumed to be fully known or can be accurately estimated before scheduling. Similar to [3], [4], we also adopt the simplifying assumption of *no entanglement decoherence*, i.e., entangled qubits remain valid as long as they have been successfully distributed.

C. Queue Dynamics

For each node n and destination k with $k \neq n$, the queue $Q_n^{(k)}(t)$ evolves according to

$$Q_n^{(k)}(t+1) = \left[Q_n^{(k)}(t) - \sum_{(n,m)\in\mathcal{L}} \mu_{n,m}^{(k)}(t) \right]_+ + \sum_{(o,n)\in\mathcal{L}} \mu_{o,n}^{(k)}(t) + A_n^{(k)}(t), \tag{1}$$

where $[x]_+ \triangleq \max\{x, 0\}$. Intuitively, $Q_n^{(k)}(t)$ first decreases by the number of requests successfully served on outgoing quantum links (n, m) for that destination k, and then increases by the number of newly arrived or incoming requests on links (o, n) and the arrivals $A_n^{(k)}(t)$. Requests with $k = n$ are considered self-destinations and immediately consumed, hence need not be queued.

D. Problem Objective

Each admission control or scheduling decision $\boldsymbol{A}(t) = \{A_n^{(k)}(t)\}_{n,k\in\mathcal{N}}$ yields a utility $g_t\big(\boldsymbol{A}(t)\big)$, capturing metrics such as the throughput of the network. We assume g_t is concave, nondecreasing in its arguments, and L-Lipschitz in some suitable norm, with values bounded in $[-G, G]$. The scheduler does not know the entire function g_t a priori and only observes the actual achieved value $g_t\big(\boldsymbol{A}(t)\big)$ once $\boldsymbol{A}(t)$ is chosen, which is again consistent with a full bandit feedback paradigm.

Our goal is twofold:

1) **Stability of Quantum Queues:** We require that the system's entanglement queues remain stable by choosing proper link allocation action $\boldsymbol{p}_{n,m}(t) \in \Delta(\mathcal{N})$ for each $(n, m) \in \mathcal{L}$. Formally, we require

$$\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{N}} Q_n^{(k)}(t)\right] = \mathcal{O}_T(1), \tag{2}$$

which means the time-averaged expected total queue size does not grow unbounded as T becomes large.

2) **Maximize Quantum Utility:** Subject to the stability constraint, we seek to maximize the average utility associated with entanglement distribution, i.e.,

$$\max \quad \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} g_t\big(\boldsymbol{A}(t)\big)\right] \qquad \text{subject to} \quad \text{the stability requirement in Eq. (2).} \tag{3}$$

In short, we seek an online scheduling and routing policy by choosing $\boldsymbol{A}(t)$ and $\boldsymbol{p}(t)$ for each time slot $t \in [T]$ that ensures provable stability of the quantum queues while simultaneously maximizing an unknown, time-varying adversarial utility function g_t. The subsequent sections outline the design of such algorithms and prove theoretical guarantees regarding their performance in multi-hop quantum networks.

Algorithm 1 Q-NSO: **Q**uantum **N**etwork **S**tability via Online Linear **O**ptimization (restated from [24, Algorithm 1])

1: Initialize AdaPFOL [24, Algorithm 2] instances for each quantum link $(n,m) \in \mathcal{L}$ with action set $\Delta(\mathcal{N})$ as $\texttt{AdaPFOL}_{n,m}$.
2: **for** $t = 1, 2, \ldots, T$ **do**
3: For each quantum link $(n,m) \in \mathcal{L}$, pass the maximum loss magnitude for this round $M\|\boldsymbol{Q}_m(t) - \boldsymbol{Q}_n(t)\|_\infty$ to $\texttt{AdaPFOL}_{n,m}$.
4: Pick quantum link allocation $\boldsymbol{p}_{n,m}(t) \in \Delta(\mathcal{N})$ as the output of $\texttt{AdaPFOL}_{n,m}$.
5: Observe entanglement request arrival rates $\boldsymbol{A}(t) \in \mathcal{A}$.
6: Observe quantum link capacities $\{C_{n,m}(t)\}_{(n,m)\in\mathcal{L}}$ and successfully establish entanglement counts $\{\mu^{(k)}_{n,m}(t)\}_{(n,m)\in\mathcal{L},k\in\mathcal{N}}$.
7: Calculate queue lengths for each quantum node $\boldsymbol{Q}(t+1)$ from $\boldsymbol{Q}(t)$ according to Eq. (1).
8: For each quantum link $(n,m) \in \mathcal{L}$, pass the loss vector $C_{n,m}(t)(Q_m(t) - Q_n(t))$ to $\texttt{AdaPFOL}_{n,m}$.
9: **end for**

III. Achieving Stability in Adversarial Multi-hop Quantum Networks

In this section, we present our Q-NSO algorithm (Algorithm 1), which ensures the stability of adversarial quantum multi-hop networks by provably bounding the average queue size. Formally, for large enough T and any time-varying arrival process $\boldsymbol{A}(t)$, Q-NSO guarantees

$$\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{n\in\mathcal{N},\,k\in\mathcal{N}} Q_n^{(k)}(t)\right] = \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\|\boldsymbol{Q}(t)\|_1\right] = \mathcal{O}_T(1).$$

Beyond establishing stability, Q-NSO will serve as a key component in our forthcoming Q-UMO (Algorithm 2) algorithm, dedicated to utility maximization under related adversarial conditions in multi-hop quantum networks. Our Q-NSO is a restatement of the NSO algorithm [24, Algorithm 1], adapted to the quantum setting. Throughout this section, the entanglement requests arrivals rate $\boldsymbol{A}(t)$ is arbitrary and time-varying vector which is obliviously decided

A. Algorithm for Quantum Network Stability Q-NSO

We propose the **Q**uantum **N**etwork **S**tability via Online Linear **O**ptimization (Q-NSO) algorithm to handle the network scheduling problem under adversarial request arrivals and bandit feedback. The overarching strategy is to recast the network stability objective in a form amenable to *Online Linear Optimization* (OLO), leveraging classical Lyapunov drift techniques. Because queue lengths in adversarial settings can grow large and induce correspondingly large loss magnitudes, we use a robust OLO subroutine named AdaPFOL [24, Algorithm 2]. This algorithm can effectively handle large and time-varying loss magnitudes, providing performance guarantees based on the geometric mean of observed losses.

B. Lyapunov Drift Analysis

Q-NSO uses classic Lyapunov drift arguments [25] to establish stability. Define the quadratic Lyapunov function

$$L_t \triangleq \frac{1}{2}\sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{N}}\left(Q_n^{(k)}(t)\right)^2, \tag{4}$$

and the one-step Lyapunov drift

$$\Delta(\boldsymbol{Q}(t)) \triangleq \mathbb{E}\left[L_{t+1} - L_t \mid \boldsymbol{Q}(t)\right]. \tag{5}$$

Using standard drift inequalities, one can show:

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}\Delta(\boldsymbol{Q}(t))\right] \le \mathbb{E}\Bigg[\sum_{t=1}^{T}\sum_{n,k\in\mathcal{N}} Q_n^{(k)}(t) \\ \cdot\left(\sum_{(o,n)\in\mathcal{L}}\mu_{o,n}^{(k)}(t) + A_n^{(k)}(t) - \sum_{(n,m)\in\mathcal{L}}\mu_{n,m}^{(k)}(t)\right)\Bigg] \\ + \frac{1}{2}N^2((NM)^2 + 2(NM)^2 + 2R^2)T,\end{aligned} \tag{6}$$

where R is an upper bound on arrival requests magnitudes and M is the upper bound of capacity for each quantum link. As $\boldsymbol{A}(t)$ is chosen obliviously, our main objective is to use online learning to minimize the term corresponding to $\boldsymbol{\mu}(t)$:

$$f(\boldsymbol{\mu}) \triangleq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{(n,m)\in\mathcal{L}}\langle\boldsymbol{\mu}_{n,m}(t), \boldsymbol{Q}_m(t) - \boldsymbol{Q}_n(t)\rangle\right]. \tag{7}$$

Since $\boldsymbol{\mu}_{n,m}(t) = C_{n,m}(t)\,\boldsymbol{p}_{n,m}(t)$, each link $(n,m) \in \mathcal{L}$ faces an *Online Linear Optimization* problem with loss vectors $C_{n,m}(t)(\boldsymbol{Q}_m(t) - \boldsymbol{Q}_n(t))$. Notice that we need to decide the allocation vector $\boldsymbol{p}_{n,m}(t)$, we can write:

$$\begin{aligned}&\sum_{t=1}^{T}\langle\boldsymbol{\mu}_{n,m}(t), \boldsymbol{Q}_m(t) - \boldsymbol{Q}_n(t)\rangle \\ &= \sum_{t=1}^{T}\langle C_{n,m}(t)\left(\boldsymbol{Q}_m(t) - \boldsymbol{Q}_n(t)\right), \boldsymbol{p}_{n,m}(t)\rangle.\end{aligned}$$

C. Plug-In Online Learning Optimization Framework

To solve each link's OLO subproblem, we use AdaPFOL [24, Algorithm 2], a robust online learning method capable of handling large, dynamically changing loss magnitudes. Let $\{\texttt{AdaPFOL}_{n,m}\}$ denote the set of AdaPFOL instances for each link (n,m), which is detailed in Line 3 of Algorithm 1. We measure the efficiency of AdaPFOL via a reference link allocation sequence $\{\bar{\boldsymbol{p}}(t)\}$, satisfying *multi-hop piecewise stability* conditions [24, Assumption 1], an extension of assumptions from [26, Assumption 1]. Intuitively, $\{\bar{\boldsymbol{p}}(t)\}$ describes a "baseline" policy that stabilizes the network in a piecewise manner over certain intervals. Formally, we write

Assumption III.1 (Extension Piecewise Stability [24, Assumption 1])**.** There exists a reference policy $\{\bar{\boldsymbol{p}}(t)\}_{t=1}^{T}$ with link allocations $\bar{\boldsymbol{p}}_{n,m}(t) \in \triangle(\mathcal{N})$ and constants $\epsilon_W > 0$, $C_W \ge 0$, such that for some partitioned intervals $\{W_j\}_{j=1}^{J}$ of $[T]$:

1) The partition of $[T]$ satisfies that

$$\sum_{j=1}^{J}(|W_j|-1)^2 \le C_W T$$

2) For $\forall j \in [J], n \in \mathcal{N}, k \in \mathcal{N}$:

$$\frac{1}{|W_j|}\sum_{t\in W_j}\sum_{(n,m)\in\mathcal{L}} C_{n,m}(t)\bar{p}^{(k)}_{n,m}(t) \ge \epsilon_W + \frac{1}{|W_j|}\sum_{t\in W_j}\left(A^{(k)}_n(t) + \sum_{(o,n)\in\mathcal{L}} C_{o,n}(t)\bar{p}^{(k)}_{o,n}(t)\right),$$

where $A^{(k)}_n(t)$ is the obliviously decided arrival rates that we assume in this section.

If $\{\bar{\boldsymbol{p}}_{n,m}(t)\}$ is such a reference policy, define $\bar{\mu}^{(k)}_{n,m}(t) = C_{n,m}(t)\,\bar{p}^{(k)}_{n,m}(t)$ and let $\mu^{(k)}_{n,m}(t) = C_{n,m}(t)\,p^{(k)}_{n,m}(t)$ be the allocations chosen by Q-NSO. By [24, Theorem 3.5], one obtains regret-like bounds on the total cost difference

$$f(\boldsymbol{\mu}) - f(\bar{\boldsymbol{\mu}}) = \mathcal{O}\left(M\sqrt{1+P^{\bar{\boldsymbol{p}}}_T}\mathbb{E}\left[\left(\sum_{t=1}^{T}\|\boldsymbol{Q}(t)\|_2^2\right)^{1/2}\cdot\log T \cdot\log\left(\max_{t\in[T],(n,m)\in\mathcal{L}} M\|\boldsymbol{Q}_m(t)-\boldsymbol{Q}_n(t)\|_\infty\right)\right]\right),$$

where $P^{\bar{\boldsymbol{p}}}_T$ is the path length of $\{\bar{\boldsymbol{p}}(t)\}_{t=1}^T$, defined by

$$P^{\bar{\boldsymbol{p}}}_T \triangleq \sum_{t=1}^{T-1}\sum_{(n,m)\in\mathcal{L}}\|\bar{\boldsymbol{p}}_{n,m}(t)-\bar{\boldsymbol{p}}_{n,m}(t+1)\|_1. \tag{8}$$

Restricting $\{\bar{\boldsymbol{p}}(t)\}$ to have controlled path length is necessary in adversarial settings, since without such constraints, performance guarantees would be unattainable, as shown in [27].

D. Theoretical Results of Q-NSO

Now we are able to present the theoretical average queue length bound (and hence stability) result of Q-NSO (Algorithm 1), as the following Theorem III.2 which is a direct corollary of Theorem 3.6 of [24].

Theorem III.2 (Main Theorem Q-NSO (Algorithm 1))**.** *Suppose $\{\bar{\boldsymbol{p}}_{n,m}(t)\}$ is the reference policy satisfying Assumption III.1, and the path length $P^{\bar{\boldsymbol{p}}}_t$ satisfies*

$$P^{\bar{\boldsymbol{p}}}_t \triangleq \sum_{s=1}^{t-1}\|\bar{\boldsymbol{p}}(s)-\bar{\boldsymbol{p}}(s+1)\|_1 \le C^p t^{1/2-\delta_p}, \tag{9}$$

for every $t \in [T]$ and some known constants C and δ_p. Then, we have

$$\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\|\boldsymbol{Q}(t)\|_1\right] = \mathcal{O}\Big(\epsilon_W^{-1}\cdot\Big((N^2(2NM+R)^2+\epsilon_W N^2(2NM+R))C_W + (N^4M^2+N^2R^2)\Big)\Big) + o_T(1).$$

That is, when $T \gg 0$, we have $\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\|\boldsymbol{Q}(t)\|_1\right] = \mathcal{O}_T(1)$, i.e., Eq. (2) holds and the system is stable.

Remark III.3. In other words, Q-NSO extends the robust online stability results of [24] to quantum multi-hop settings under limited feedback. The theorem guarantees that queues remain stable, addressing the fundamental goal of preventing queue explosion in adversarial multi-hop quantum networks, laying the groundwork for additional objectives such as utility maximization and service quality within the same adversarial framework. In the subsequent section, the algorithmic framework of Q-NSO in Algorithm 1 will act as the crucial component for the utility maximization task to maintain the stability of the quantum network.

IV. Maximizing Utility in Adversarial Multi-Hop Quantum Networks

With the above stability foundation, we now turn to the utility maximization task. In this case, the entanglement request rates are also decided by the scheduler with the objective of maximizing the unknown and time-varying utility function while maintaining the stability of the multi-hop quantum network. We have already introduced the algorithm Q-NSO for quantum networks, which achieves stability under arbitrary request arrival rates $\boldsymbol{A}(t)$. Equipped with the key intuition of Q-NSO, we need to determine the arrival rates for maximizing the average utility it gains (as formally stated in Eq. (3)).

A. Algorithm for Utility Maximization Q-UMO

We summarize our approach with the pseudo-code in Algorithm 2, a quantum adaptation of Algorithm 3 from [24]. This method, called Q-UMO, preserves key elements of Q-NSO (Algorithm 1), in that each link allocation is determined by the plug-in learning subroutine AdaPFOL, which ensures stability under adversarial conditions. On top of that, Q-UMO now selects the entanglement arrival rates $\boldsymbol{A}(t)$ each slot in a manner designed to maximize cumulative utility $\sum_t g_t(\boldsymbol{A}(t))$. To balance these dual goals, Q-UMO combines the Lyapunov drift-plus-penalty analysis [25, Theorem 4.2] with a bandit convex optimization approach known as AdaBGD [24, Algorithm 4], which adapts to possible infinite queue lengths and time-varying loss magnitudes.

B. Lyapunov Drift-Plus-Penalty Analysis

Under the drift-plus-penalty framework [25, Theorem 4.2], we modify the classical Lyapunov analysis by including a penalty term $-V\,\mathbb{E}[g_t(\boldsymbol{A}(t)) \mid \boldsymbol{Q}(t)]$ in the drift expression. Recall the Lyapunov function $L_t = \frac{1}{2}\|\boldsymbol{Q}(t)\|_2^2$ and drift $\Delta(\boldsymbol{Q}(t)) = \mathbb{E}[L_{t+1} - L_t \mid \boldsymbol{Q}(t)]$ from before. With penalty, the Drift-Plus-Penalty (DPP) function is:

$$\Delta(\boldsymbol{Q}(t)) - V\mathbb{E}[g_t(\boldsymbol{A}(t)) \mid \boldsymbol{Q}(t)],$$

where $\Delta(\boldsymbol{Q}(t))$ is defined in Eq. (5) and V is arbitrarily determined for our purpose. Following similar steps as in Section III-B, one obtains two subproblems: one for controlling $\boldsymbol{\mu}(t)$ via AdaPFOL, and one for managing $\boldsymbol{A}(t)$ to explore and

Algorithm 2 Q−UMO: **U**tility **M**aximization via Online Linear **O**ptimization in **Q**uantum Networks (restated from [24, Algorithm 3])

1: Initialize AdaPFOL [24, Algorithm 2] instances for each quantum link $(n,m) \in \mathcal{L}$ with action set $\Delta(\mathcal{N})$ as $\texttt{AdaPFOL}_{n,m}$.
2: Initialize a bandit convex optimization algorithm AdaBGD for learning rates defined in Eq. (12) and feasible action set $\mathcal{A}$.
3: **for** $t = 1, 2, \ldots, T$ **do**
4: For each quantum link $(n,m) \in \mathcal{L}$, pass the maximum loss magnitude for this round $M\|\boldsymbol{Q}_m(t) - \boldsymbol{Q}_n(t)\|_\infty$ to $\texttt{AdaPFOL}_{n,m}$.
5: Pick quantum link allocation $\boldsymbol{p}_{n,m}(t) \in \Delta(\mathcal{N})$ as the output of $\texttt{AdaPFOL}_{n,m}$.
6: Pick entanglement request arrival rates $\boldsymbol{A}(t)$ as the output of AdaBGD under learning rates $\{\eta_s, \delta_s, \alpha_s\}_{s=1}^t$ given in Eq. (12).
7: Observe quantum link capacities $\{C_{n,m}(t)\}_{(n,m)\in\mathcal{L}}$ and successfully establish entanglement counts $\{\mu_{n,m}^{(k)}(t)\}_{(n,m)\in\mathcal{L},k\in\mathcal{N}}$.
8: Observe the collected utility $g_t(\boldsymbol{A}(t))$.
9: Calculate queue lengths for each quantum node $\boldsymbol{Q}(t+1)$ from $\boldsymbol{Q}(t)$ according to Eq. (1).
10: For each quantum link $(n,m) \in \mathcal{L}$, pass the loss vector $C_{n,m}(t)(\boldsymbol{Q}_m(t) - \boldsymbol{Q}_n(t))$ to $\texttt{AdaPFOL}_{n,m}$, and the loss $\langle \boldsymbol{Q}(t), \boldsymbol{A}(t)\rangle - V g_t(\boldsymbol{A}(t))$ to AdaBGD.
11: **end for**

exploit the best utility. We derive the optimization objective corresponding to $\boldsymbol{\mu}(t)$ and $\boldsymbol{A}(t)$, respectively:

$$f(\boldsymbol{\mu}) \triangleq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{(n,m)\in\mathcal{L}} \langle \boldsymbol{\mu}_{n,m}(t), \boldsymbol{Q}_m(t) - \boldsymbol{Q}_n(t)\rangle\right], \quad (10)$$

$$h(\boldsymbol{A}) \triangleq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{n\in\mathcal{N}} \langle \boldsymbol{Q}_n(t), \boldsymbol{A}_n(t)\rangle\right] - V\mathbb{E}\left[\sum_{t=1}^{T}(g_t(\boldsymbol{A}(t)))\right]. \quad (11)$$

Section III-B has already discuss the optimization on $f(\boldsymbol{\mu})$ (Eq. (10)). Here we focus on the optimization problem on $h(\boldsymbol{A})$ (Eq. (11)).

Concretely, the second subproblem translates into a *bandit convex optimization* (BCO) objective with only partial feedback on g_t, i.e., only $g_t(\boldsymbol{A}(t))$ is observable once $\boldsymbol{A}(t)$ is chosen. Because high queue backlogs can inflate both losses and Lipschitz constants, an adaptive BCO approach is required. Hence, Q−UMO applies the AdaBGD [24, Algorithm 4] method, designed for such adversarial bandit settings, to track the optimal arrival rates despite large, time-varying losses.

C. Plug-In Bandit Convex Optimization Algorithm

The *Bandit Convex Optimization* (BCO) problem (Eq. (11)) impose the loss vector ℓ_t for each round t

$$\ell_t(\boldsymbol{A}) \triangleq \langle \boldsymbol{Q}(t), \boldsymbol{A}\rangle - V\, g_t(\boldsymbol{A}).$$

To handle this BCO problem under bandit feedback, Q−UMO calls AdaBGD (Algorithm 4 in [28]) each round. This subroutine adaptively adjusts its learning rate (detailed in Eq. (12)) based on the magnitude of $\|\boldsymbol{Q}(t)\|_\infty$, thus coping with substantial and unpredictable fluctuations to handle the case when the loss magnitude $\|\boldsymbol{Q}(t)\|_\infty + VG$ and the Lipschitzness $\|\boldsymbol{Q}(t)\|_2 + VL$ are both large when $\|\boldsymbol{Q}(t)\|$ is large. This approach also assumes only that $g_t(\boldsymbol{A}(t))$ (i.e., the utility realized by the chosen entanglement request arrival rates $\boldsymbol{A}(t)$) is observed each time slot $t \in [T]$, rather than having knowledge of g_t over the entire domain $\mathcal{A}$. This partial-information setting closely matches real-world quantum network scenarios in which measuring the complete set of channel utilities is prohibitively expensive. Hence only $\ell_t(\boldsymbol{A}(t))$, the actual loss associated with our action, can be accurately calculated.

To show the theoretical guarantees of AdaBGD, we first introduce the following assumption on the reference policy including the self-determined entanglement request arrival rates $\boldsymbol{A}$, which is similar to Assumption III.1. We assume a reference control sequence $\left(\bar{\boldsymbol{p}}(t), \bar{\boldsymbol{A}}(t)\right)$ that piecewise stabilizes the system, similar to Assumption III.1 ([24, Assumption 2]), but now incorporating dynamic arrival rates. The performance of Q−UMO is then benchmarked against this reference, yielding both queue stability and near-optimal long-term utility.

Assumption IV.1 (Extension Piecewise Stability for Utility Maximization [24, Assumption 2])**.** There exists a reference policy $\{(\bar{\boldsymbol{p}}(t), \bar{\boldsymbol{A}}(t))\}_{t\in[T]}$ and constants $\epsilon_W > 0$, $C_W \geq 0$, such that for some partitioned intervals $\{W_j\}_{j=1}^J$ of $[T]$:

1) The partition of $[T]$ satisfies that

$$\sum_{j=1}^{J}(|W_j| - 1)^2 \leq C_W T$$

2) For $\forall j \in [J], n \in \mathcal{N}, k \in \mathcal{N}$:

$$\frac{1}{|W_j|}\sum_{t\in W_j}\sum_{(n,m)\in\mathcal{L}} C_{n,m}(t)\bar{p}_{n,m}^{(k)}(t) \geq \epsilon_W + \frac{1}{|W_j|}\sum_{t\in W_j}\left(\bar{A}_n^{(k)}(t) + \sum_{(o,n)\in\mathcal{L}} C_{o,n}(t)\bar{p}_{o,n}^{(k)}(t)\right).$$

[24, Theorem 4.4] shows that For the reference arrival rates $\{\bar{\boldsymbol{A}}(t)\}_{t\in[T]}$ defined in Assumption IV.1, suppose that its path length ensures for every $t \in [T]$:

$$P_t^{\bar{\boldsymbol{A}}} \triangleq \sum_{t=1}^{T-1}\|\bar{\boldsymbol{A}}(t+1) - \bar{\boldsymbol{A}}(t)\|_1 \leq C^A t^{1/2-\delta_A},$$

where C^A and δ_A are known constants but the precise $P_t^{\bar{\boldsymbol{A}}}$ or $\{\bar{\boldsymbol{A}}(t)\}_{t\in[T]}$ both remain unknown.

Suppose that the action set $\mathcal{A}$ is bounded by $[r, R]$ (*i.e.*, $r\mathbb{B} \subseteq \mathcal{A} \subseteq R\mathbb{B}$, where $\mathbb{B}$ is the unit ball). [24] shows that, if we execute AdaBGD over $\mathcal{A}$ with loss functions $\ell_t(\boldsymbol{A}) =$

$\langle \boldsymbol{Q}(t), \boldsymbol{A} \rangle - V g_t(\boldsymbol{A})$ and parameters $\eta_t, \delta_t, \alpha_t$ defined in Eq. (12):

$$\eta_t = \left(C^A T^{\frac{1}{2}-\delta_A} \Big/ \begin{matrix} \left(C^A T^{1/2-\delta_A}\right)^{\frac{7}{3}} \left(4r^{-3}d^2\right)^{\frac{28}{9}} (M+R)^{\frac{4}{3}} + \\ C^A T^{\frac{1}{2}-\delta_A} (r^{-3}d^2 V G^2/L)^{\frac{4}{3}} + \\ \sum_{s=1}^{t} \left((\|\boldsymbol{Q}_s\|_\infty + VG)^2 (\|\boldsymbol{Q}_s\|_2 + VL)^2 \right)^{\frac{1}{3}} \end{matrix} \right)^{\frac{3}{4}},$$
$$\delta_t = \left(\eta_t d^2 \frac{(\|\boldsymbol{Q}(t)\|_\infty + VG)^2}{(\|\boldsymbol{Q}(t)\|_2 + VL)} \right)^{\frac{1}{3}}, \quad \alpha_t = \frac{\delta_t}{r}, \tag{12}$$

then the outputs $\boldsymbol{A}(1), \boldsymbol{A}(2), \ldots, \boldsymbol{A}(T) \in \mathcal{A}$ of AdaBGD ensure

$$h(\boldsymbol{A}) - h(\bar{\boldsymbol{A}}) = \mathcal{O}\left(\frac{R(2NM+R)}{r^7} d^{14/3} (C^A T^{1/2-\delta_A})^2 \right) + \mathcal{O}\left(\mathbb{E}\left[\left(\frac{R}{r} d^{2/3} + R \right) (C^A T^{1/2-\delta_A})^{1/4} \cdot \left(\sum_{t=1}^{T} \left(\|\boldsymbol{Q}(t)\|_2 + V(L+G) \right)^{4/3} \right)^{3/4} \right] \right).$$

D. Theoretical Results of Q-UMO

Finally, we state the main outcome for Q-UMO in Theorem IV.2, as the following Theorem IV.2, which is a direct corollary of Theorem 4.5 of [24]. Roughly speaking, if there is a piecewise-stabilizing reference policy with bounded path lengths for both link allocations and arrivals, Q-UMO can sustain bounded queues while achieving an average utility that approaches that of the reference policy on the order of $\mathcal{O}_T(V^{-1})$ as T grows. By carefully tuning V as a sub-polynomial in T, the algorithm can reconcile the competing goals of fast utility convergence and stable queue operation in large-scale adversarial quantum networks.

Theorem IV.2 (Main Theorem Q-UMO (Algorithm 2))**.** *Suppose that the feasible set of arrival rates vector $\mathcal{A}$ is bounded by $[r, R]$. Assume all (unknown) utility functions g_t to be concave, L-Lipschitz, and $[-G, G]$-bounded. Consider a reference action sequence $\{(\bar{\boldsymbol{p}}(t), \bar{\boldsymbol{A}}(t))\}_{t\in[T]}$ satisfying Assumption IV.1, such that their path lengths satisfy*

$$P_t^{\bar{\boldsymbol{p}}} \triangleq \sum_{s=1}^{t-1} \|\bar{\boldsymbol{p}}(s) - \bar{\boldsymbol{p}}(s+1)\|_1 \le C^p t^{1/2-\delta_p},$$
$$P_t^{\bar{\boldsymbol{A}}} \triangleq \sum_{s=1}^{t-1} \|\bar{\boldsymbol{A}}(s) - \bar{\boldsymbol{A}}(s+1)\|_1 \le C^A t^{1/2-\delta_A}, \quad \forall t \in [T].$$

Here, $M, R, r, L, G, C^p, \delta_p, C^A, \delta_A$ are known constants, whereas the specific $\{(\bar{\boldsymbol{p}}(t), \bar{\boldsymbol{A}}(t))\}_{t\in[T]}$ remains unknown. If we execute the Q-UMO in Algorithm 2 with the plug-in online learning optimization algorithm AdaPFOL and the bandit convex optimization algorithm AdaBGD given in [24], when T is large enough such that the constant $V = o_T(\min\{T^{2\delta_p/3}, T^{2\delta_A/7}\})$, the following inequalities hold simultaneously:

$$\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} \|\boldsymbol{Q}(t)\|_1 \right] = \mathcal{O}\Big(\epsilon_W^{-1} \cdot \Big(N^2(2NM+R)^2 + \epsilon_W N^2 (2NM+R)) C_W + (N^4M^2 + N^2R^2) \Big) \Big) + o_T(1),$$
$$\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} \left(g_t(\bar{\boldsymbol{A}}(t)) - g_t(\boldsymbol{A}(t)) \right) \right] = \mathcal{O}\Big(V^{-1} \Big((N^2(2NM+R)^2 + \epsilon_W N^2(2NM+R)) C_W + (N^4M^2 + N^2R^2) \Big) \Big) + o_T(V^{-1}).$$

That is, when $T \gg 0$, our algorithm not only stabilizes the system so that $\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} \|\boldsymbol{Q}(t)\|_1\right] = \mathcal{O}_T(1)$, but also enjoys an average utility approaching that of the reference policy polynomially fast, i.e., $\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} \left(g_t(\bar{\boldsymbol{A}}(t)) - g_t(\boldsymbol{A}(t))\right)\right] = \mathcal{O}_T(V^{-1})$ – the utility maximization objective Eq. (3) is ensured.

Remark IV.3. As shown in Theorem IV.2, Q-UMO jointly ensures bounded average queue length and guarantees an $\mathcal{O}(V^{-1})$ gap in utility with respect to the reference policy. By appropriately selecting V (e.g., as a sub-polynomial function in T), the scheduler can maintain stability and simultaneously approach the best attainable utility in an adversarial multi-hop quantum network. This result demonstrates that Q-UMO can achieve both $\mathcal{O}_T(1)$ average queue length and $\mathcal{O}(V^{-1})$ utility gap simultaneously by carefully handling the drift-plus-penalty framework in adversarial quantum networks.

V. Conclusion

In this paper, we proposed two scheduling protocols, Q-NSO and Q-UMO, to tackle adversarial quantum network control with bandit feedback. Our methods are tightly built upon the algorithmic framework and analytical tools in [24]. Specifically, Q-NSO ensures the quantum network's stability. Meanwhile, Q-UMO extends Q-NSO to incorporate utility maximization, combining Lyapunov drift-plus-penalty analysis with adaptive bandit convex optimization. This approach jointly stabilizes quantum queues and optimizes the adversarial utility function under bandit feedback, enabling the network to handle dynamic arrival rates while continuously improving overall utility.

References

[1] A. Farahbakhsh and C. Feng, "Opportunistic routing in quantum networks," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 490–499.

[2] L. Yang, Y. Zhao, H. Xu, and C. Qiao, "Online entanglement routing in quantum networks," in *2022 IEEE/ACM 30th International Symposium on Quality of Service (IWQoS)*. IEEE, 2022, pp. 1–10.

[3] T. Vasantam and D. Towsley, "Stability analysis of a quantum network with max-weight scheduling," *arXiv preprint arXiv:2106.00831*, 2021.

[4] S. Pouryousef, N. K. Panigrahy, and D. Towsley, "A quantum overlay network for efficient entanglement distribution," in *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*. IEEE, 2023, pp. 1–10.

[5] M. Chehimi, M. Elhattab, W. Saad, G. Vardoyan, N. K. Panigrahy, C. Assi, and D. Towsley, "Reconfigurable intelligent surface (ris)-assisted entanglement distribution in fso quantum networks," *IEEE Transactions on Wireless Communications*, 2025.

[6] C. Qiao, Y. Zhao, G. Zhao, and H. Xu, "Quantum data networking for distributed quantum computing: Opportunities and challenges," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2022, pp. 1–6.

[7] P. Promponas, V. Valls, S. Guha, and L. Tassiulas, "Maximizing entanglement rates via efficient memory management in flexible quantum switches," *IEEE Journal on Selected Areas in Communications*, 2024.

[8] L. Le and T. N. Nguyen, "Dqra: Deep quantum routing agent for entanglement routing in quantum networks," *IEEE Transactions on Quantum Engineering*, vol. 3, pp. 1–12, 2022.

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

[10] N. K. Panigrahy, P. Dhara, D. Towsley, S. Guha, and L. Tassiulas, "Optimal entanglement distribution using satellite based quantum networks," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2022, pp. 1–6.

[11] N. K. Panigrahy, T. Vasantam, D. Towsley, and L. Tassiulas, "On the capacity region of a quantum switch with entanglement purification," in *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*. IEEE, 2023, pp. 1–10.

[12] Y. Zeng, J. Zhang, J. Liu, Z. Liu, and Y. Yang, "Entanglement routing over quantum networks using greenberger-horne-zeilinger measurements," in *2023 IEEE 43rd International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2023, pp. 350–360.

[13] V. Valls, P. Promponas, and L. Tassiulas, "On the capacity of the quantum switch with and without entanglement decoherence," *IEEE Communications Letters*, 2023.

[14] L. Yang, Y. Zhao, L. Huang, and C. Qiao, "Asynchronous entanglement provisioning and routing for distributed quantum computing," in *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*. IEEE, 2023, pp. 1–10.

[15] Y. Zhao and C. Qiao, "Distributed transport protocols for quantum data networks," *IEEE/ACM Transactions on Networking*, vol. 31, no. 6, pp. 2777–2792, 2023.

[16] G. Zhao, J. Wang, Y. Zhao, H. Xu, L. Huang, and C. Qiao, "Segmented entanglement establishment with all-optical switching in quantum networks," *IEEE/ACM Transactions on Networking*, vol. 32, no. 1, pp. 268–282, 2023.

[17] N. K. Panigrahy, M. G. De Andrade, S. Pouryousef, D. Towsley, and L. Tassiulas, "Scalable multipartite entanglement distribution in quantum networks," in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 2. IEEE, 2023, pp. 391–392.

[18] G. Vardoyan, P. Nain, S. Guha, and D. Towsley, "On the capacity region of bipartite and tripartite entanglement switching," *ACM Transactions on Modeling and Performance Evaluation of Computing Systems*, vol. 8, no. 1-2, pp. 1–18, 2023.

[19] Y. Gan, X. Zhang, R. Zhou, Y. Liu, and C. Qian, "A routing framework for quantum entanglements with heterogeneous duration," in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, vol. 1. IEEE, 2023, pp. 1132–1142.

[20] P. Fittipaldi, A. Giovanidis, and F. Grosshans, "A linear algebraic framework for dynamic scheduling over memory-equipped quantum networks," *IEEE Transactions on Quantum Engineering*, 2023.

[21] Y. Zeng, J. Zhang, J. Liu, Z. Liu, and Y. Yang, "Entanglement routing design over quantum networks," *IEEE/ACM Transactions on Networking*, vol. 32, no. 1, pp. 352–367, 2023.

[22] L. Bacciottini, L. Lenzini, E. Mingozzi, and G. Anastasi, "Redip: Ranked entanglement distribution protocol for the quantum internet," *IEEE Open Journal of the Communications Society*, 2023.

[23] H. Gu, R. Yu, Z. Li, X. Wang, and F. Zhou, "Esdi: Entanglement scheduling and distribution in the quantum internet," in *2023 32nd International Conference on Computer Communications and Networks (ICCCN)*. IEEE, 2023, pp. 1–10.

[24] Y. Dai and L. Huang, "Adversarial network optimization under bandit feedback: Maximizing utility in non-stationary multi-hop networks," *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, vol. 8, no. 3, pp. 1–48, 2024.

[25] M. Neely, *Stochastic network optimization with application to communication and queueing systems*. Morgan & Claypool Publishers, 2010.

[26] J. Huang, L. Golubchik, and L. Huang, "When lyapunov drift based queue scheduling meets adversarial bandit learning," *IEEE/ACM Transactions on Networking*, 2024.

[27] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proceedings of the 20th international conference on machine learning (icml-03)*, 2003, pp. 928–936.

[28] W. Dai, A. Rinaldi, and D. Towsley, "The capacity region of entanglement switching: Stability and zero latency," in *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 2022, pp. 389–399.