

Path-Oblivious Entanglement Swapping for the Quantum Internet

Vincent Mutolo
Columbia University
New York, NY, USA

Rhea Parekh
Columbia University
New York, NY, USA

Dan Rubenstein
Columbia University
New York, NY, USA

Abstract

Proposed Bell pair swapping protocols, an essential component of the Quantum Internet, are planned-path: specific structured routing paths are reserved prior to the execution of the swapping process. This makes sense when one assumes the state used in the swapping process is expensive, fragile, and unstable. However, lessons from classical networking have shown that while reservations seem promising in concept, flexible, reservation-light or free approaches often outperform their more restrictive counterparts in well-provisioned networks. In this paper, we propose that a path-oblivious approach is more amenable to supporting swapping as quantum state evolves into a cheaper, more robust form. We formulate the swapping process as a linear program and present and evaluate a fairly naive baseline swapping protocol that tries to balance Bell pairs throughout the network. Preliminary results show that while naive balancing leaves room for improvement, investigating path-oblivious swapping is a promising direction.

CCS Concepts

• **Networks** → **Network protocol design**; • **Hardware** → **Quantum communication and cryptography**.

Keywords

Quantum networking, Path-oblivious swapping.

ACM Reference Format:

Vincent Mutolo, Rhea Parekh, and Dan Rubenstein. 2025. Path-Oblivious Entanglement Swapping for the Quantum Internet. In *The 24th ACM Workshop on Hot Topics in Networks (HotNets '25)*, November 17–18, 2025, College Park, MD, USA. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3772356.3772412>

November 17–18, 2025, College Park, MD, USA. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3772356.3772412>

1 Introduction

The Quantum Internet is expected to make extensive use of *teleportation* [6]: the transfer of an arbitrary quantum state (represented by a qubit) from one physical location (origin node) to another possibly distant location (destination node) via a sequence of local operations and transmission of only two bits of classical (non-quantum) information to complete the process, as summarized in Fig. 1. Aside from the qubit holding the state to be teleported, the process requires a Bell pair: two qubits that may be separated at an arbitrarily large distance, but remain *entangled*, meaning that the measurement of one qubit affects the outcome of the subsequent measurement of the other. The local operations include sequences of gate transformations (depicted as blue bolts) and measurement of qubits that collapses the qubit's state into a classical bit (1 or 0).

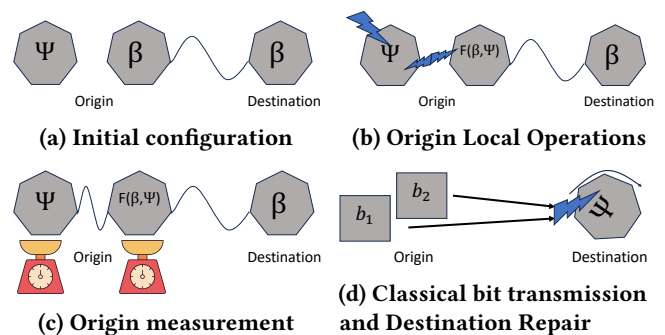


Figure 1: Teleportation process

In addition to teleportation [6], there are many additional proposed quantum applications that can benefit from having Bell pairs distributed throughout the network, such as entanglement-based quantum key distribution [16], distributed quantum computing [18], and entanglement-assisted sensing and metrology [24], as well as emerging novel applications [4]. While some pairs of nodes can directly *generate* these entangled pairs, others, often due to geographical distance, cannot, and it is challenging to physically move the qubits without corrupting them. Fortunately, there exists a *quantum swapping* process [40] as depicted in Fig. 2: if nodes

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org. *HotNets '25*, College Park, MD, USA
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-2280-6/25/11

<https://doi.org/10.1145/3772356.3772412>

A and C share a Bell pair and nodes B and C share a separate Bell pair, then C can implement a *swap operation*, which we write as $A \leftarrow C \rightarrow B$ whose steps bear resemblance to the teleportation process. The swap ends with A and B sharing a Bell pair, with both of C's formerly-entangled qubits collapsed via measurement, ending their entanglement, with C's 2-bit measurement result sent to either A or B who can make the necessary local corrections to complete the swap. In this paradigm, C is often referred to as a *repeater*, and, more generally, A and B can construct a shared Bell pair by performing a series of swaps across a *path* of repeaters [9], where each node N_i along the path shares a Bell pair with the next hop, N_{i+1} , in the path.

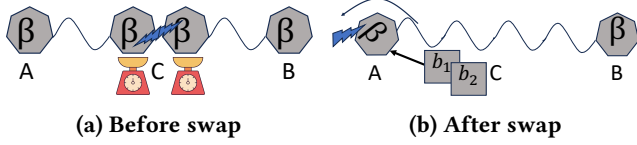


Figure 2: Swap operation

The swapping process along a path shares many characteristics to routing in a traditional network setting where data flows along the path instead of entanglement swaps. As in the traditional networking setting, there may exist multiple feasible paths over which swapping can connect an origin-destination pair, with differing paths having differing costs. The costs of concern in the swapping regime are also familiar to the networking community: the number of swaps needed grows with the number of hops in the path, such that longer paths will likely take more time and require more repeater resources. Also, differing paths may offer differing *fidelities* [23] that provide a measure of the intactness of Bell pairs, similar to how different classical networking paths can offer different reliabilities due to variations in corruption or packet loss along each path. And, just as classical networks have a means to compensate for different reliabilities (e.g., error correction, selective retransit), quantum swapping also has a means to compensate for varying fidelities (a combination of quantum error correction (QEC) and a process known as *distillation* or *purification* [7]).

Not surprisingly, the swapping process has noticeable differences from classical network formulations. The first is the magic of teleportation: that a complex quantum state can be moved by the combination of local operations and 2-bit classical transmission. Second, the Bell pairs that are swapped carry no explicit payload and, aside from where the pair of qubits reside, are *interchangeable*: a Bell pair used at any point in the swapping process to support a teleportation between nodes A and B could, if needed, be diverted to the swapping process to support teleportation between nodes C and D. Thus, each Bell pair whose qubits reside at nodes

N_1 and N_2 are effectively identical and for our purposes we refer to any such Bell pair as $[N_1, N_2]$. In contrast, a data packet has an explicit payload which rarely can be diverted to support another application.

The interchangeable nature of Bell pairs leads to a third noticeable distinction from traditional networking: unlike classical routing where each node performing its routing action must await the data item from the node one hop closer to the origin, the order in which repeaters perform their swapping action is arbitrary - any shuffle of the order in which a node on the path executes its swap will succeed, with the repeater performing the swap effectively extracting itself from the chain of entanglements that connects origin to destination: it replaces its two entanglements that connect it to its neighbors with a single entanglement directly entangling those neighbors. This scenario is depicted in Fig. 3, in which repeater R_3 performs $R_2 \leftarrow R_3 \rightarrow R_4$, collapsing a $[R_2, R_3]$ and a $[R_3, R_4]$ to form a $[R_2, R_4]$. In our example, this swap is performed on the path even before R_1 and R_2 have yet to even establish a Bell pair: a phenomenon that cannot happen in classical communication.

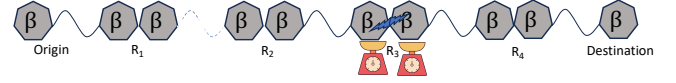


Figure 3: A swapping path example

Planned-path swapping: To our knowledge, all prior art that explores the distribution via swapping of Bell pairs through a quantum network uses what we refer to as a *planned-path* approach [11, 19, 26, 29, 38]: When a pair of nodes desires a Bell pair between them (i.e., for teleportation), a set (perhaps of size 1) of possible paths are identified, one is selected, and swaps are subsequently performed along this path. Such approaches can be classified into *connection-oriented* [26] and *connectionless* [38], where the former immediately maps Bell pairs to a specific swap path while the latter permits multiple paths that might both have interest in the same swap configuration (i.e., $A \leftarrow C \rightarrow B$) to compete over the Bell pairs that can implement this swap). However, both approaches require the idea of identifying the path prior to implementing the swapping process.

Planned-path approaches overlook an important evolution that took place in the classical networking community for applications where there is a distributed demand for the same content [28, 31, 33, 34]. Restricting communication along pre-defined paths can overlook potential efficiencies of scale, even when attempts are made to converge toward well performing paths [3, 13]. In short, flexible strategies that pre-position content in a way that makes it more easily accessible when the demand arises or increases are preferred

to only trying to find an optimal path from static points at the immediate instance of demand.

In this paper, we consider the benefits and challenges toward constructing a Quantum Internet using path-oblivious approaches where swapping is used to balance Bell pairs holistically toward needed locations. We begin in §2 with early assumptions about cost that targeted early designs toward planned-path approaches, and why current evolution in the field supports the design of path-oblivious approaches. In §3, we show how to formulate the path-oblivious approach as a linear program. We present a simple distributed balancing algorithm in §4, compare its performance to planned-path approaches in §5, describe some preliminary thoughts on future directions in §6, and end with related work and conclusion respectively in §7 and §8.

2 Planned v. Oblivious

In this section, we justify why the quantum networking community fixates on planned-path approaches, and motivate why we think path-oblivious approaches have a future.

Planned-path approaches are (as we experienced in classical networking [17, 39]) initially appealing because if well organized, they can make optimal use of networking resources [8]. They are needed in high-cost, under-utilized or limited capacity networks. But as the network scales and costs come down, the complexity of reserving access on multiple paths can bottleneck the supposed gain in efficiency. So let us first present arguments why over the short term, planned-path approaches make sense.

Decoherence: Even if left alone, Bell pairs naturally decohere, i.e., the entangled qubits naturally disentangle [9]. Planned-path advocates note that current and near-term coherence times are so short, that the entire swapping process must be completed in milliseconds before the Bell pairs to be used along the path collapse [37]. Applying a collapsed Bell pair anywhere in the swapping/teleportation process corrupts the measurement.

Fidelity: A further complication is the stochastic nature of quantum information. Since measurement collapses the entanglement, it is impossible to validate via measurement that the entangled quantum state is in its desired form. Fidelity can be thought of as a confidence measure of how likely a given quantum state is actually in its desired form [23]. One approach to increasing fidelity is a process called *distillation* a.k.a. *purification* [7]: a predictive process that uses (and destroys) one Bell pair to assess the correctness of another. Because of decoherence, the fidelity of a Bell pair must be continuously reaffirmed. More recently, quantum error correction (QEC) approaches have been considered as an alternate and likely less costly approach to maintaining high

fidelity [22], and there is ongoing research that considers hybrid approaches [30].

Waste: It should be clear that Bell pairs that are used for teleportation are in fact the product of the consumption of numerous Bell pairs that were used to form its entanglement (via swapping) and ensure high fidelity (via distillation). It is wasteful to manufacture such a Bell pair and not consume it before it decoheres, or to delay its use and unnecessarily incur additional costs of continual distillation or QEC.

Classical overheads: Finally, swapping, teleportation, and distillation all require transmission of classical information (albeit only a few bits) across the network between points that wish to maintain entanglements. Both planned-path and path-oblivious approaches will need to account for this classical transmission, as well as any additional classical coordination to (in the case of planned-path) establish and execute swapping routes, and (in the case of path-oblivious) learn about the status of the distributed state of Bell pairs such that a path-oblivious algorithm can make informed decisions about how to proceed.

2.1 A Push Analogy: Water systems

While in the near-term each Bell pair should be considered an expensive, short-lived entity, based on current progress trends [27], we optimistically anticipate that as physicists and engineers put their heads together, the cost of generation will decline to a point where each Bell pair can simply be viewed as a drop of water in a very large bucket. This water analogy is intentional: path-planned approaches are analogous to a water park that only turns on the water for a ride when a person (valuable commodity) actively engages the ride and turns the water off immediately on completion. Water, while not free, is in many places sufficiently prevalent such that approaches that “push” water to the faucet are preferable, even if water is wasted (e.g., pipes or taps drip, samples are used for testing, etc.) As costs associated with Bell pairs drop, we speculate that a similar “push” architecture will in the long-term be preferable to path-based approaches, in that the small loss of efficiency will be offset by the significantly simpler control mechanism that reduces the need for planning and reservation.

Anticipating that over the next decades, Bell pair generation costs will drop and coherence times will grow, one can expect such networks to experience the same provisioning boom that classical networks have experienced over the last few decades. This will presumably lead to a similar realization within the Quantum Internetworking community that scalability and flexibility should trump optimality.

3 Path-oblivious LP formulation

In contrast to a planned-path formulation that establishes paths for swapping to support a specific teleportation, a path-oblivious formulation can be thought of in its crudest form as “swap gossip”: a repeater that holds two Bell pairs, each entangled with a different node, can implement a swap to directly entangle those neighbors without necessarily supporting any explicit teleportation goal: it is simply trying to help “spread” the availability of different pairs of nodes connected by Bell pairs. If this spreading is well-designed, then as teleportation needs appear, the Bell pairs supporting them may already be present, or may simply be a small number of swaps away.

We can best understand the asymptotic capabilities of a path-oblivious approach by formulating the generation and consumption of entangled state as a linear flow program. Our formulation presumes an ideal environment where each repeater has a limitless buffer to store its qubit components of Bell pairs, and generated Bell pairs do not decohere (for analysis purposes).

The problem is formulated over a set of nodes $\mathcal{N}$. Each pair of nodes $x, y \in \mathcal{N}$ has an average rate $g(x, y) = g(y, x) \geq 0$ at which they can, from scratch, *generate* Bell pairs $[x, y]$, where some (perhaps many) pairs of x, y are unable to generate pairwise, such that $g(x, y) = 0$ (e.g., perhaps they are physically too far from one another to form Bell pairs without the assistance of swapping). Similarly there are pairs x, y that *consume* Bell pairs $[x, y]$ (i.e., to implement teleportation) at an average rate of $c(x, y) \geq 0$, where many pairs may have no need for consumption, such that $c(x, y) = 0$. We can define the *generation graph*, G , consisting of undirected edges (x, y) for which $g(x, y) > 0$. Observing that a swap never increases the number of Bell pairs held at a node, any node x that wishes to be part of the quantum Internet must satisfy $g(x, y) > 0$ for some y , and that $\sum_y c(x, y) \leq \sum_y g(x, y)$ if it is to maintain its consumption rate. Further, if a generation graph G is not fully connected, then it is impossible to build (via generation and swaps) a Bell pair that entangles x with y when they lie in distinct connected components. Last, we are only interested in the generation and consumption of Bell pairs that are distributed between nodes, such that we assume $g(x, x) = c(x, x) = 0$.

We can observe the asymptotic rate at which node i implements the swap $x \leftarrow i \rightarrow y$ and write this rate as $\sigma_i(x, y) = \sigma_i(y, x)$ where $\sigma_i(x, i) = \sigma_i(x, x) = 0$ since these result in Bell pairs entangled at a single node.

3.1 Steady State Analysis

For each pair x, y , we define $r_+(x, y) = r_+(y, x)$ to be the asymptotic rate at which entanglements arrive between x

and y and $r_-(x, y) = r_-(y, x)$ the rate at which they depart:

$$r_+(x, y) = g(x, y) + \sum_i \sigma_i(x, y) \quad (1)$$

$$r_-(x, y) = c(x, y) + \sum_i (\sigma_x(i, y) + \sigma_y(i, x)) \quad (2)$$

i.e., a Bell pair arrives between x and y when it is explicitly generated or when some third node i implements $x \leftarrow i \rightarrow y$, and a Bell-pair departs between x and y when it is explicitly consumed or when either $i \leftarrow x \rightarrow y$ or $i \leftarrow y \rightarrow x$.

Since Bell pairs cannot depart at a rate faster than they arrive, we must always have $r_-(x, y) \leq r_+(x, y)$, and in steady-state, we expect $r_-(x, y) = r_+(x, y)$ for all pairs x, y .

3.2 Distillation and QEC

Our simple LP can easily be extended to handle constant rates of distillation per node-pair. We assume that the pair x and y only utilize Bell pairs for consumption and swapping after their fidelity has achieved a sufficient level which is accomplished with an expected number $\mathcal{D}_{x,y}$ of distillations. We can reformulate r_- as

$$r_-(x, y) = \mathcal{D}_{x,y}(c(x, y) + \sum_i (\sigma_x(i, y) + \sigma_y(i, x))) \quad (3)$$

Alternatively, QEC can be added simply by assuming that QEC is applied to generated Bell pairs, after which QEC is the form of correction. If the overhead of the QEC (i.e., the number of physical qubits per logical qubit) is $\mathcal{R}$, we can simply thin the generation rate by $g(x, y)$ to be $g(x, y)/\mathcal{R}$.

3.3 Optimization Objectives

It is worth noting that in our formulation, one can think of the external inputs being a set of maximum generation rates $\{\gamma(x, y)\}$ between pairs of nodes x and y such that deployed generation rates must satisfy $\forall x \forall y, g(x, y) \leq \gamma(x, y)$ and desired consumption rates $\{\kappa_{x,y}\}$ that bounds the maximum needed consumption rate between all pairs $(x, y) : c(x, y) \leq \kappa(x, y)$. With these constraints, the $\{g(x, y)\}$, $\{c(x, y)\}$ and $\{\sigma_i(x, y)\}$ being the set of variables that need to be solved. The optimal setting of the $\{\sigma_i(x, y)\}$ depend of course on the desired optimization criterion. We list a few possibilities:

- **Conserve generation:** When generation is sufficient to handle consumption (i.e., there is a swapping strategy for which generation rates $\{\gamma(x, y)\}$ can support all desired consumption rates $\{\kappa_{x,y}\}$, then conserve generation: fix $\forall x \forall y, c(x, y) = \kappa(x, y)$ and determine $\{\sigma_i(x, y)\}$ and $\{g(x, y)\}$ that minimizes $\sum_{x < y} g(x, y)$ or alternatively, minimizes the maximum $g(x, y)$.
- **“Fairly” reduce consumption:** When generation is insufficient to handle consumption, then reduce consumption amongst pairs according to some fairness criterion. More formally, solve for $\{g(x, y)\}$, $\{c(x, y)\}$ and $\{\sigma_i(x, y)\}$ that

maximizes $\sum_{x < y} c(x, y)$, or alternatively maximizes the minimum $c(x, y) < \kappa(x, y)$, or alternatively finds the largest α such that $c(x, y) = \alpha \kappa_{x, y}$.

3.4 Observations about LP Optimality

We make three important observations about the LP solution.

LP follows shortest paths in unconstrained environments. First, let us look at the environment where there is no distillation, and there is no network congestion along paths such that the ideal swapping path for a consumption between (x, y) could always be supported along the shortest generation path between (x, y) . Then the LP solution to minimizing the aggregate generation rate must construct swaps that would follow this shortest path, the argument being that if a longer path were used, this longer path would require additional generations, and the LP would have failed to minimize the aggregate generation rate!

LP finds the optimal "mix" of paths in constrained environments. When congestion does exist, such that the network cannot support all consumption demand along shortest paths, the fact that the LP is minimizing the aggregate generation rate means that it is identifying the combination of most efficient paths that could support the needed consumption, given the available generation rates.

LP uses optimal distillation ordering. While in general, for a given swap path, the order of swaps along the path can be arbitrary, when distillation is required (i.e. $\mathcal{D} > 1$, the order in which swaps occur is known to affect the number of swaps (and hence number of generated Bell pairs) needed to render the Bell pair to be used for consumption. For instance, in uncongested settings, a *nested* ordering that recursively performs the middle swap last, is known to be optimal [9, 15]. The swap rates discovered by the LP must support such nested swapping, since again this minimizes generation rates! We note that while the LP determines the swap rates to support these swapping approaches when there is distillation, and it follows that swapping at these rates will automatically enforce a nested ordering without ever needing to make it explicit! However, note that distillation processes generally require the pair of Bell pairs used in distillation have equivalent fidelities

4 Max-min distributed swapping

While our LP formulation assists with understanding asymptotic behavior of a path-oblivious swapping system, it does not address the practicalities of how and when to swap given the stochastic nature of generation, consumption, distillation and decoherence processes. We present a simple *balancing* approach that, as an initial baseline, appears to work well.

We suppose for now that each node x maintains a count $C_x(y)$ of the number of Bell pairs it stores that are shared with each y in the network (note $C_x(y) = C_y(x)$). For any two neighbors y and y' , we say that the swap $y' \leftarrow x \rightarrow y$ is *preferable* when $C_y(y') + 1 \leq \min(C_x(y) - \mathcal{D}_{x, y}, C_x(y') - \mathcal{D}_{x, y'})$, such that x would only reduce its own counts if doing so aids a pair whose count would be no larger after the swap. If x has more than one preferable swap candidate, it performs a swap with minimal $C_y(y')$.

Were generation and consumption to both cease, this swapping process would balance bell pairs in the network such that the minimum count would be maximized, i.e., when no swap candidates are preferable, a max-min fair allocation has been established, in that no buffer count can be increased without reducing another that was already smaller [21]. Intuitively, as Bell pairs are generated, swaps will flow from the points of generation, and as Bell pairs are consumed or used in swaps, swaps will flow toward these points of drainage.

Note that in this initial formulation of the algorithm, we are not concerning ourselves with classical costs, which would include not only the 2 bits required to complete the swap, but also the sharing of information between all nodes of the $\binom{|N|}{2}$ edge counts. This may very well be feasible given existing and emerging high speed classical networks. Still, we discuss possible ways to reduce classical communication overheads in §6.

5 Performance Analysis

To evaluate the efficacy of our baseline algorithm, we simulate a Quantum Internet on a variety of topologies, where a subset of repeater pairs (that form a connected graph) act as generators, and a (possibly overlapping) subset of repeater pairs act as consumers. The generators generate Bell pairs and these Bell pairs are swapped using the baseline algorithm, after which consumers may consume them.

We consider two generation graph topologies in this paper: a *cycle graph* containing $|N|$ numbered 0 through $|N| - 1$ where $g(x, y) > 0 \iff y = x \pm 1 \pmod{|N|}$, and an embedding on a wraparound $\sqrt{|N|} \times \sqrt{|N|}$ grid, in which each node x 's position can be described by coordinates (x_i, x_j) , $0 \leq i, j < \sqrt{|N|}$ where $g(x, y) > 0$ can only occur when either $x_i = y_i$ and $x_j = y_j \pm 1 \pmod{\sqrt{|N|}}$ or $x_j = y_j$ and $x_i = y_i \pm 1 \pmod{\sqrt{|N|}}$ (i.e., 1-hop neighbors on the grid). Generation edges are added uniformly at random on the grid until the underlying generation graph connects all nodes. In these preliminary simulations, $g(x, y) = 1$ for all edges where $g(x, y) > 0$.

We generate 35 pairs from the set of all possible $\binom{|N|}{2}$ pairs as our consumers, and construct a sequence of consumption requests from these pairs that must be satisfied in the order

of the sequence.¹ All nodes perform the swapping process at an identical rate. We found that varying this rate did not significantly alter the results presented below.

To measure the efficiency of our swapping process, we compare the number of swaps our algorithm performs in comparison to the minimum number of swaps needed where each consumption event is satisfied by swaps along the shortest path in the underlying generation graph. Specifically, if the consumption pair is (x, y) , we identify the length of the shortest path along the generation graph and compute the number of swaps needed to form the Bell pair that entangles x and y . When the length of the shortest path between x and y is n hops and all distillation overheads $\mathcal{D}_{x,y}$ are set to an identical value of $\mathcal{D}$, the optimal swapping process (i.e., nested swapping [9, 15]) requires $s(n)$ swaps where $s(1) = 0$ and $s(n) = \mathcal{D}(1 + s(\lfloor n/2 \rfloor) + s(\lceil n/2 \rceil))$ for $n > 1$ (build the (already-distilled) entanglement between the first half and second half of the path recursively, perform the swap to entangle the desired endpoints, then distill). The *swap overhead* of our distributed algorithm is defined as the number of swaps performed in simulation divided by $\sum_c s(\ell(c))$ where the sum over c covers all consumption events that were satisfied in simulation, and $\ell(c)$ is the number of hops in the shortest path within the generation graph for consumption event c . Note this measure can be no less than 1, since the denominator is the minimum number of swaps that could satisfy all consumption events. However, there are several reasons why this scoring is conservative, in that our algorithm probably does better against practical planned-path approaches:

- Practical planned-path approaches need not always take the shortest swapping path (i.e., in attempts to load balance or avoid congestion). For similar reasons, the optimal nested swapping order that minimizes distillation costs may not always be practical.
- When the simulation ends, there are multiple swaps that occurred within the simulation that were not used for consumption (i.e., they would still be available for future consumptions). These are extraneous counts in the numerator that are unaccounted for in the denominator.

Figures 4 and 5 plot the swap overhead on the y -axis as $\mathcal{D}$ and $|\mathcal{N}|$ are respectively varied along the x -axis with $|\mathcal{N}| = 50$ and $\mathcal{D} = 1, 2$ respectively in the graphs. Our preliminary results show that varying $|\mathcal{N}|$ has an unpredictable impact (no clear monotonic trend) on the swap overhead of our baseline swapping protocol. We suspect the observed variation on the grid topologies as we vary the number of nodes is most likely due to the variance in the process constructing the topology on the grid (i.e., the process of selecting

¹This was done to prevent biasing the cost toward easy-to-satisfy pair requests - a point addressed in §6.

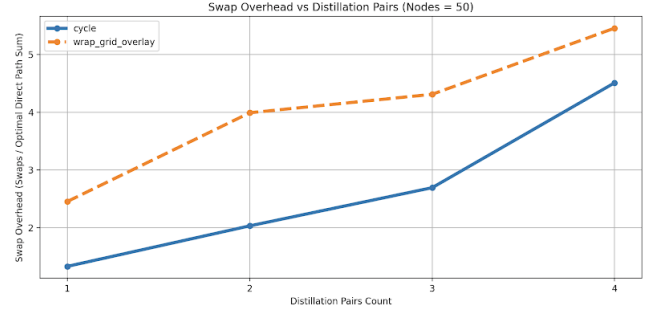


Figure 4: Simulation results $|\mathcal{N}| = 50$, varying $\mathcal{D}$

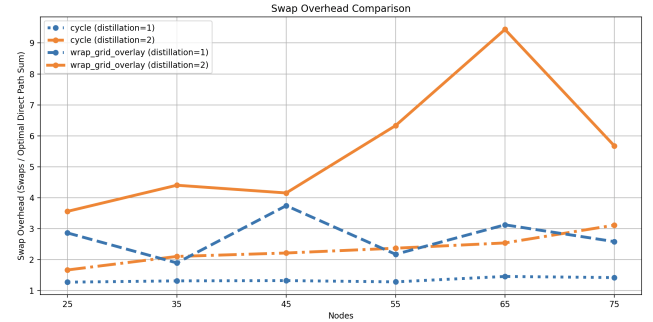


Figure 5: Simulation results, $\mathcal{D} = 1, 2$, grid, varying $|\mathcal{N}|$

generating edges via a random process). In contrast, growth appears (almost) linear as $\mathcal{D}$ increases. The growth as a function of $\mathcal{D}$ is not surprising as our current approach makes no explicit effort to ensure that the swapping ordering is nested, which is needed for efficiency when $\mathcal{D} > 1$ [9, 15].

In summary, as quantum memories, coherence times, and fidelity preserving mechanisms improve, even a simple baseline approach is reasonably efficient in comparison to more stringent planned-path approaches, and there are plenty of dimensions over which path-oblivious protocols can be further tuned.

6 Discussion

Baseline Our initial attempt at a path-oblivious swapping algorithm makes no attempts to optimize the problem beyond a simple max-min balancing scheme: how best to refine the path-oblivious swapping algorithm to further reduce overhead is an ongoing trajectory of our research. If costs in high-distillation regimes seem excessively high, there are several things to do. For instance, reducing the likelihood that node i , very distant from both x and y in the generation graph, implements a swap between x and y when they are close to one another in the generation graph.

Realistic coherence, QEC and distillation overheads Admittedly, our models of coherence time and distillation

processes are oversimplified, and our use has been simply to ballpark how these additional phenomena might affect performance of path-oblivious approaches. Future study should better incorporate the actual effects of such processes, and consider protocol adjustments that might result (e.g., avoiding combining Bell pairs with short expected remaining coherence times with those that have longer times).

Classical overheads If classical communication overheads are of concern, the immediate global knowledge of all buffers can likely be relaxed as well. For instance, a BitTorrent-like approach [14] with a similar choke/unchoke mechanism, where each node knows only the status of a rotating but small number of neighbors, would intuitively scale well.

Toward a layered Quantum Internet Stack Path-oblivious approaches also give rise to a cleaner separation within a layered approach. For instance, the max-min algorithm we describe can be thought of as a network-layer protocol, whereas decisions about distillation processes and rejection of aged Bell pairs can be thought of as a transport layer functionality that “cleanses” the Bell pairs passed through the system. While planned-path approaches generally need a transport layer that fixes problems all along the swapping path, a path-oblivious transport layer need only evaluate the quality of Bell pairs without co-mingling the evaluation with their explicit future trajectory.

Hybrid oblivious with minimal planning Path-oblivious and planned-path approaches can in fact work together, with the oblivious component providing a distributed base of Bell pairs. For any consumption pairs for which a Bell pair is not immediately available upon consumption request, a planned-path search can identify a shortest path among the existing Bell pairs (which could be much shorter than their shortest path on the underlying generation graph). From a path-oblivious perspective, doing this could help mitigate the starvation effect we saw in our simulations, where consumption requests between nodes who are close on the generation graph would usurp the Bell pairs needed to form the longer paths. From a planned-path perspective, the amount of resources that require reservation at any point in time could significantly be reduced.

7 Related Work

Prior work in quantum network resource allocation often assumes short coherence times, motivating reactive, planned-path approaches that generate Bell pairs on-demand and rely on explicit path selection [11, 19, 26, 29] with some such as [38] proposing a connectionless variant where the path is reserved but Bell pairs are not, such that criss-crossing paths compete for available Bell pairs at shared links. Our LP formulation is similar in flavor to that of [10]. However, that formulation is significantly more complex in that it

attempts to identify specific paths that prioritize high fidelity without taking into account the potential use of distillation and quantum error correction processes.

Proactive strategies [11] propose pre-preparing virtual links but assumes short lifetimes, requiring frequent regeneration. [36] studies continuous distribution in regular topologies under fidelity decay but offers no optimization framework or support for distillation. [20] analyzes cutoffs in chain topologies and references purification, though operational integration remains limited. Percolation-based studies such as [2] explore probabilistic global connectivity but do not account for distillation, memory dynamics, or network-wide optimization. Surveys and architecture proposals [1, 5, 35] offer taxonomies or protocol-level mechanisms but do not model fidelity-aware resource distribution.

In contrast, our work assumes long-lived quantum memories (motivated by advances such as [25, 32]) and explicitly incorporates distillation costs. We present both a centralized linear program and a distributed protocol for proactive, path-oblivious, and connectionless Bell pair distribution. This enables efficient, fidelity-aware resource balancing across arbitrary topologies, supporting asynchronous many-to-many traffic without relying on precomputed paths.

A very interesting, very recent idea in [12] propose a graph-complement approach that achieves path-obliviousness by manipulating multipartite graph states (much more complex to build and further off from realistic deployment than the Bell pairs our work is limited to) to directly rewire the entanglement topology between remote nodes, eliminating the need for explicit pathfinding. In contrast, our work achieves path-obliviousness through a lighter-weight, operational framework that balances Bell-pair generation and swapping rates without requiring complex multipartite-state preparation or graph-level transformations.

8 Conclusion

As coherence times and fidelities of Bell pairs increase, swapping mechanisms will become less dependent on real-time construction of the end-to-end path, such that we anticipate approaches migrating away from planned-path approaches. Specifically, the interchangeability of Bell pairs makes the formulation of quantum internetworking extremely amenable to path-oblivious methods that distribute the availability of such state in a manner that is divorced from the underlying topological constraints. While the interplay between distillation and QEC remains an important issue to resolve, the classical networking community’s experience with massive distribution of content emphasizing stateless approaches will become more relevant and the classical networking community’s expertise could have positive impact on this important genre.

References

- [1] Amar Abane, Michael Cubeddu, Van Sy Mai, and Abdella Battou. Entanglement routing in quantum networks: A comprehensive survey. *IEEE Transactions on Quantum Engineering*, 2025.
- [2] Antonio Acín, J Ignacio Cirac, and Maciej Lewenstein. Entanglement percolation in quantum networks. *Nature Physics*, 3(4):256–259, 2007.
- [3] David Andersen, Hari Balakrishnan, Frans Kaashoek, and Robert Morris. Resilient overlay networks. In *Proceedings of the eighteenth ACM symposium on Operating systems principles*, pages 131–145, 2001.
- [4] Venkat Arun, Vijay Chidambaram, and Scott Aaronson. Faster-than-light coordination for networked systems with quantum non-local games. In *Proceedings of the 24th ACM Workshop on Hot Topics in Networks (HotNets '25)*, November 2025.
- [5] Leonardo Bacciotini, Matheus Guedes De Andrade, Shahrooz Pouryousof, Emily A Van Milligen, Aparimit Chandra, Nitish K Panigrahy, Nageswara SV Rao, Gayane Vardoyan, and Don Towsley. Leveraging internet principles to build a quantum network. *IEEE Network*, 2025.
- [6] Charles H Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Physical review letters*, 70(13):1895, 1993.
- [7] Charles H Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A Smolin, and William K Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Physical review letters*, 76(5):722, 1996.
- [8] Lee Breslau and Scott Shenker. Best-effort versus reservations: A simple comparative analysis. In *Proceedings of the ACM SIGCOMM'98 conference on Applications, technologies, architectures, and protocols for computer communication*, pages 3–16, 1998.
- [9] H-J Briegel, Wolfgang Dür, Juan I Cirac, and Peter Zoller. Quantum repeaters: the role of imperfect local operations in quantum communication. *Physical Review Letters*, 81(26):5932, 1998.
- [10] Kaushik Chakraborty, David Elkouss, Bruno Rijsman, and Stephanie Wehner. Entanglement distribution in a quantum network: A multicommodity flow-based approach. *IEEE Transactions on Quantum Engineering*, 1:1–21, 2020.
- [11] Kaushik Chakraborty, Filip Rozpedek, Axel Dahlberg, and Stephanie Wehner. Distributed routing in a quantum internet. *arXiv preprint arXiv:1907.11630*, 2019.
- [12] Si-Yi Chen, Angela Sara Cacciapuoti, and Marcello Caleffi. Beyond traditional quantum routing, 2025.
- [13] Yang-hua Chu, Sanjay G Rao, and Hui Zhang. A case for end system multicast (keynote address). *ACM SIGMETRICS Performance Evaluation Review*, 28(1):1–12, 2000.
- [14] Bram Cohen. Incentives build robustness in bittorrent. In *Workshop on Economics of Peer-to-Peer systems*, volume 6, pages 68–72. Berkeley, CA, USA, 2003.
- [15] Wolfgang Dür, H-J Briegel, Juan Ignacio Cirac, and Peter Zoller. Quantum repeaters based on entanglement purification. *Physical Review A*, 59(1):169, 1999.
- [16] Artur K. Ekert. Quantum cryptography based on bell's theorem. *Physical Review Letters*, 67(6):661–663, 1991.
- [17] G Gallassi, G Rigolio, and Luigi Fratta. Atm: Bandwidth assignment and bandwidth enforcement policies. In *1989 IEEE Global Telecommunications Conference and Exhibition 'Communications Technology for the 1990s and Beyond'*, pages 1788–1793. IEEE, 1989.
- [18] Daniel Gottesman and Isaac L. Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, 402(6760):390–393, 1999.
- [19] Joy Halder, Akhmadjon Rajabov, Riccardo Bassoli, Frank H. P. Fitzek, and Gerhard P. Fettweis. Optimal routing and end-to-end entanglement distribution in quantum networks. *Scientific Reports*, 14(1):19262, 2024.
- [20] Álvaro G Iñesta, Gayane Vardoyan, Lara Scavuzzo, and Stephanie Wehner. Optimal entanglement distribution policies in homogeneous repeater chains with cutoffs. *npj Quantum Information*, 9(1):46, 2023.
- [21] Jeffrey Jaffe. Bottleneck flow control. *IEEE Transactions on Communications*, 29(7):954–962, 1981.
- [22] Liang Jiang, Jacob M Taylor, Kae Nemoto, William J Munro, Rodney Van Meter, and Mikhail D Lukin. Quantum repeater with encoding. *Physical Review A—Atomic, Molecular, and Optical Physics*, 79(3):032325, 2009.
- [23] Richard Jozsa. Fidelity for mixed quantum states. *Journal of modern optics*, 41(12):2315–2323, 1994.
- [24] P. Kómár, E. M. Kessler, M. Bishof, L. Jiang, A. S. Sørensen, J. Ye, and M. D. Lukin. A quantum network of clocks. *Nature Physics*, 10:582–587, 2014.
- [25] Yisheng Lei, Faezeh Kimiaee Asadi, Tian Zhong, Alex Kuzmich, Christoph Simon, and Mahdi Hosseini. Quantum optical memory for entanglement distribution. *Optica*, 10(11):1511–1528, 2023.
- [26] Zhonghui Li, Kaiping Xue, Jian Li, Nenghai Yu, David SL Wei, and Ruidong Li. Connection-oriented and connectionless remote entanglement distribution strategies in quantum networks. *IEEE Network*, 36(6):150–156, 2022.
- [27] Jianqing Liu, Thinh Le, Tingxiang Ji, Ruozhou Yu, Demitry Farfurnik, Greg Bryd, and Daniel Stancil. The road to quantum internet: Progress in quantum network testbeds and major demonstrations. *Progress in Quantum Electronics*, page 100551, 2024.
- [28] Andy Myers, Peter Dinda, and Hui Zhang. Performance characteristics of mirror servers on the internet. In *IEEE INFOCOM'99*, volume 1, pages 304–312. IEEE, 1999.
- [29] Mihir Pant, Hari Krovi, Don Towsley, Leandros Tassioulas, Liang Jiang, Prithwish Basu, Dirk Englund, and Saikat Guha. Routing entanglement in the quantum internet. *npj Quantum Information*, 5(1):25, 2019.
- [30] Christopher A Pattison, Gefen Baranes, J Ataide, Mikhail D Lukin, and Hengyun Zhou. Fast quantum interconnects via constant-rate entanglement distillation. *arXiv preprint arXiv:2408.15936*, 2024.
- [31] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp, and Scott Shenker. A scalable content-addressable network. In *Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 161–172, 2001.
- [32] Matthew Reagor, Wolfgang Pfaff, Christopher Axline, Reinier W. Heeres, Nissim Ofek, Katrina Sliwa, Eric Holland, Chen Wang, Jacob Blumoff, Kevin Chou, Michael J. Hatridge, Luigi Frunzio, Michel H. Devoret, Liang Jiang, and Robert J. Schoelkopf. Quantum memory with millisecond coherence in circuit qed. *Physical Review B*, 94(1):014506, July 2016.
- [33] Rüdiger Schollmeier. A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. In *Proceedings first international conference on peer-to-peer computing*, pages 101–102. IEEE, 2001.
- [34] Ion Stoica, Robert Morris, David Liben-Nowell, David R Karger, M Frans Kaashoek, Frank Dabek, and Hari Balakrishnan. Chord: a scalable peer-to-peer lookup protocol for internet applications. *IEEE/ACM Transactions on networking*, 11(1):17–32, 2003.
- [35] Ze-Zhou Sun, Yuan-Bin Cheng, Yu-Chen Liu, Dong Ruan, Dong Pan, and Gui-Lu Long. Message-oriented entanglement distribution network. *IEEE Internet of Things Journal*, 2024.
- [36] Lars Talsma, Álvaro G Iñesta, and Stephanie Wehner. Continuously distributing entanglement in quantum networks with regular topologies. *Physical Review A*, 110(2):022429, 2024.
- [37] Michelle Victora, Stefan Krastanov, Alexander Sanchez de la Cerda, Steven Willis, and Prineha Narang. Purification and entanglement routing on quantum networks. *arXiv preprint arXiv:2011.11644*, 2020.

- [38] Zirui Xiao, Jian Li, Kaiping Xue, Zhonghui Li, Nenghai Yu, Qibin Sun, and Jun Lu. A connectionless entanglement distribution protocol design in quantum networks. *IEEE Network*, 38(1):131–139, 2023.
- [39] Lixia Zhang, Steve Deering, Deborah Estrin, Scott Shenker, and Daniel Zappala. Rsvp: A new resource reservation protocol. *IEEE network*, 7(5):8–18, 1993.
- [40] Marek Zukowski, Anton Zeilinger, M Horne, and Artur Ekert. "event-ready-detectors" bell experiment via entanglement swapping. *Physical review letters*, 71(26), 1993.