Quantum Data Networking for Distributed Quantum Computing: Opportunities and Challenges

Chunming Qiao*, Yangming Zhao†, Gongming Zhao†, and Hongli Xu†
*Department of Computer Science and Engineering, University at Buffalo
†School of Computer Science and Technology, University of Science and Technology of China

***Abstract*—Quantum Data Networking can significantly transform the landscape of quantum computing by enabling several small quantum computers (QCs) to form a distributed system to achieve the same computing power as a large quantum computer which is infeasible to build. However, this requires quantum state information, in the form of qubits, to be exchanged among multiple geographically distributed QCs, and there are many challenges associated with reliably transferring qubits from one QC to another efficiently. In this paper, we discuss various QDN design options, present main challenges and describe promising solutions to tackle the challenges.**

I. INTRODUCTION

Quantum computing holds a great potential to solve certain types of problems more efficiently than classic computers [1]. For example, it can solve the integer factorization problem, which is NP-hard with classic computers, in polynomial time [2]. However, in a foreseeable future, it is expected that each quantum computer (QC) can have only a small number of quantum bits (called *qubits*). To overcome such a limitation, we propose to use a quantum data network (QDN) to network many small QCs to form a distributed processing system [3, 4]. Ideally, this would enable for example two QCs with a and b qubits respectively to work together to achieve the same power of one larger QC with $a + b$ qubits (in reality, due to overhead involved in distributed computing, the power of such a distributed system may be on the same as a larger QC with $a + b - c$ qubits, where c is a small number).

With such a QDN, a large quantum circuit can be properly designed and then partitioned into multiple sub-circuits in order to exploit locality in quantum gate operations among the qubits. In other words, we can map each subset of all the qubits needed for the large quantum circuit, which often interact with each other via local (quantum gate) operations in the corresponding sub-circuit to one of the QCs. Occasionally, (and when needed), we can send the quantum state information carried by a few qubits mapped to one QC to another QC, in order to perform non-local operations among the qubits initially mapped to these two QCs. In this paper, we focus on how to design QDNs and leave additional discussions related to quantum circuit designs and partitioning to another work.

While the concept of using QDNs to support distributed quantum computing is similar to distributed computing in the classical world, there are several unqiue challenges and opportunities in QDNs. For example, the no–cloning theorem [5] prevents one from copying a qubit for later restranmissions in case of transmission errors. Moreover, a data qubit carrying quantum state information from one source QC (Alice) must not be "converted" into a classical bit (via measurements) before reaching the other destination QC (Bob), as such a conversion would result in the collapse of the carried quantum state. In other words, the qubits must be transferred in an end-to-end (E2E) fashion, so none of the traditional methods for data exchanges based on *e.g.*, TCP/IP, would be sufficient. On the other hand, QDNs offer unique opportunities such as teleportation of qubits over entanglement connections.

It is also worth noting that today's QKD networks [6] are fundamentally different from the envisioned QDNs, since the former is used only to establish a shared encryption key between two (classical) computers. Specifically, with QKD, data generated by one computer is sent as classical bits to another computer. In addition, all existing QKD protocols (such as BB84 [7] and E91 [8]) are designed for sending "non-data" qubits which can tolerate a high loss and/or error rate.

In general, a QDN consists of source and destination QCs and other quantum nodes. These nodes are interconnected with quantum links, which can be fibers or free space optical links. There are two types of QDNs. We call the first type Tell-n-Go (TAG) QDNs, where Alice directly sends her data qubits (in the form of *e.g.*, photons) to Bob. A TAQ-QDN consists of QCs, optical (quantum) links, and all-optical switches. It is relatively simple but due to the fact that a photon carrying a data qubit has typically a very low energy and thus can be easily lost during transmission, such a TAQ-QDN is useful to connect QCs in a small geographical area. On the other hand, in the second type which we call Teleportation (TELE) QDN, Alice and Bob first establishes an E2E entanglement connection, and then Alice teleports a data qubit to Bob. We will describe in more details a few variations in which a TELE-QDN operates later, but in general, a TELE-QDN needs, in addition to quantum repeaters, entanglement photon sources (EPS) that generate entangled photon pairs (often called Bell pairs), and Bell-State Measurement (BSM) devices. The Bell pairs carry non-data qubits which are used to establish an entanglement link between two adjacent quantum nodes (QCs or repeaters), and as a result, each quantum node holds one of the Bell pair quantum in its quantum memory. Two quantum links, both adjacent to repeater R, can then be stitched together by using a BSM device to perform a joint measurement of the two non-data qubits at R, corresponding to the two entanglement links. An E2E entanglment connection can be established after we successfully establish and stitch all

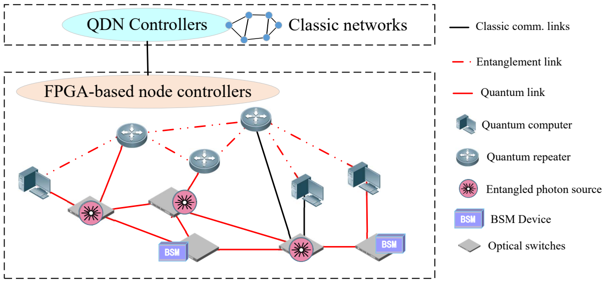


Fig. 1. The architecture of a typical QDN.

entanglement links along a path. Accordingly, a TELE-QDN is more complex than a TAG-QDN but the advantage is that it can be used to connect QCs far apart from each other. We will focus only on TELE-QDNs (and omit its prefix ”TELE”) in the remainder of the paper.

There are many challenges stemming from quantum physics, and related quantum communication technologies. In this paper, we will focus on E2E entanglement establishment in QDNs and discuss various network architecture and protocol level design considerations.

The rest of the paper is organized as follows. In Section II, we introduce the background on teleportation, and then, we discuss the challenges associated with establishing E2E entanglement connections for reliable teleportation of qubits in Section III. After that, promising approaches to overcome the challenges and opportunities for future research are presented in Section III, followed by concluding remarks in Section V.

II. BACKGROUND

In this section, we present basic concepts on teleportation in QDNs. Fig. 1 shows a typical QDN, which as shown, uses an overlay classical control or signaling network for sending auxiliary information. To simply the illustration, only quantum nodes (either QCs or repeaters) are shown, while other quantum resources such as quantum memory, EPSes and BSM devices are omitted from Fig. 1.

The basic operations involved in establishing an E2E entanglement connection between Alice and Bob are as follows:

Path Determination. Similarly to routing in a classical network, a path specifying intermediate quantum nodes (repeaters) is first determined. For simplicity, one may assume that there are physical quantum links connecting two quantum nodes, although in a typical QDN, two quantum nodes are likely connected through a BSM device instead.

Entangled Photon Generation and Entanglement Link. An EPS can generate a pair of entangled photons, often called a Bell pair. Since they carry no data and are used to establish an entanglement, we will call them e-bits. If one quantum node (either a QC or a repeater) possesses one e-bit of a Bell pair in its quantum memory, and the other quantum node possesses the other e-bit of the same Bell pair in its memory, then we consider that these two nodes have an entanglement link.

Quantum Swapping and Entanglement Segment. Assume that a repeater R has one entanglement link with quantum node A by sharing one Bell pair, and in addition, has another entanglement link with quantum node B by sharing the second Bell pair. If R performs a quantum swapping, which basically involves a joint BSM operation on these two e-bits, then nodes A and B will have a two-hop entanglement segment. In addition, both entanglement links will cease to exist and the two units of quantum memory at R can be freed.

E2E Entanglement Connection. Similar to how an entanglement segment is established, an E2E entanglement connection can be established by first establishing each and every entanglement link along a chosen path, and then having each repeater along the path perform a joint BSM operation.

Teleportation. Once an E2E connection is set up between Alice and Bob, Alice can teleport her data qubit by performing a joint BSM operation between the data qubit and the e-bit she used to entangle with Bob.

BSM Results and Unitary Operations. To teleport, Alice (and every repeater along the path) will send their BSM results using the classical network to Bob, who will perform a unitary operation based on these BSM results on its own e-bit, in order to “receive” the quantum state information from Alice.

III. CHALLENGES

While teleportation enables two remote QCs to exchange quantum state information, establishment of an E2E entanglement connection between the two QCs could be quite challenging. In this section, we start with quantum communications technology related challenges, and then move on to QDN architecture and protocol design challenges.

In addition to the no-cloning theorem mentioned above, a QDN, unlike any other classical communications networks studied earlier, has the following unique set of low-level constraints: (i). non-negligible quantum channel loss and interference [9]; (ii). non-deterministic process of entanglement link establishment [10] due to interference during entangled photon generation and photon transmission processes, which will lead to a low fidelity of an entanglement link; (iii). fast decoherence of entangled quantum states [11] (which limits the duration of an established entanglement link); (iv). limited (amount of) quantum memory and other quantum resources such as EPSes, in addition to the limited amount of time a quantum state can be stored in a quantum memory [12]; and (v). low efficiency in single photon detection [13] as a part of BSM operations, which impacts the ability to connect multiple entanglement links into an E2E entanglement connection.

While it is highly desirable to overcome these and other challenges through technological advances, in order to achieve some overall QDN optimization objectives such as maximizing the throughout (or more precisely good put) in QDNs, we aim to explore QDN architecture and protocol level approaches that can establish a maximal number of (high fidelity) entanglement connections given the low-level constraints. To this end, we first discuss how the low-level constraints introduce the challenges at the architecture and protocol level.

Limited quantum memory or entanglement bits (e-bits). One unit of quantum memory is needed to store a qubit, which can be either a data qubit used by Alice and Bob to carry out quantum gate operation, or a non-data qubit used by Alice and Bob as well as the repeaters to establish entanglement links or connections, hereafter referred as entanglement bits or e-bits. In particular, only a few e-bits are available at each repeater, and each intermediate repeater needs at least to use two e-bits in order to be able to stitch together two adjacent entanglement links as a part of the process of establishing an E2E entanglement connection. This means that the number of entanglement links and connections that can be concurrently established is limited. On the other hand, we note that one can reuse the two units of quantum memory associated with the two e-bits once the two entanglement links have been stitched together, without having to wait till the E2E connection is fully established, let alone the completion of the teleportation of a data qubit over the E2E connection. Accordingly, one of the design challenges is to determine how to efficiently utilize (and schedule) the use of these limited number of e-bits (or equivalently quantum memory).

Difficulties in Bell pair generation and limited number of EPSes. It is not easy to generate (highly quality) Bell pairs which is needed to establish even an entanglement link. Often, it takes an EPS multiple (*e.g.*, hundreds or more) tries to generate a Bell pair. In addition, EPSes are expensive, and their numbers are limited. The limited rate at which Bell pairs can be generated also limits the rate at which entanglement links can be established. Therefore, one of the challenges related to the QDN designs is to determine how many EPSes one should deploy in a QDN to maximize the cost-effectiveness, and where they should be placed, assuming some knowledge about the set of locations for QCs, repeaters and other quantum resources such as BSM devices, as well as the the knowledge of the (potential) quantum links. Another related design challenge is once the number and locations of these EPSes (and their Bell pair generation rate) are known, determine which selected entanglement links should we establish, in order to achieve the overall QDN optimization objectives. There are also several other open problems related to the topological design and resource provisioning in QDNs.

Error-prone Bell pair distribution. This is closely related to the previous constraint. In particular, wherever (and however) a Bell pair of e-bits is generated, at least one of them needs to be sent to a remote quantum node to establish an entanglement link, but due to the extremely low energy of a photon carrying such an e-bit, its transmission can be easily lost. In particular, the transmission error on a fiber-optical link increases exponentially with the transmission distance. This implies that where to place EPSes is important as mentioned above. In addition, there is a non-negligible failure probability associated with each attempt to establish an entanglement link (as to be discussed next, a related challenge due to fast decoherence as well is that one can hardly establish any entanglement link with a 100% fidelity). This implies that different links may be associated with different failure probabilities, Accordingly, a design challenge is to determine which set of entanglement links to create (and then stitched together) to establish a set of E2E connections. Note that this notion is similar to routing or path selection in a classical network, but here we must take into consideration the limited quantum memory and other constraints mentioned earlier. In particular, given that transmitting e-bits is error-prone, but quantum swapping needed to stitch two entanglement links together is also not 100% reliable, besides the fact quantum swapping consumes additional quantum memory, one interesting question is whether there is room for a hybrid approach, whereby instead of establishing many entanglement links each having a short span (in terms of geographical distance), one could establish a few longer-span entanglement "segments" with help from all-optical switches, and then stitch these segments together to form an E2E connection.

Fast decoherence, low fidelity, and other errors. Every Bell pair of e-bits will suffer from (fast) decoherence which affects how long the entanglement will last and the fidelity of such entanglement. Current technology cannot maintain entanglement for more than a few seconds after which the entanglement may cease to exist (due to environmental interference). Even within a short period of time, the fidelity of an entanglement link will be low, reducing the fidelity of the E2E entanglement connection established by stitching such entanglement links and the reliability of the corresponding teleportation operation. Also, in addition to transmission, quantum memory read/write operations on e-bits or data qubits are also error prone. Even BSM operations needed for quantum swapping may result in errors. These error further exacerbate the problem of being able to establish high-quality (or usable) entanglement links and connections. One of the challenges is thus how to speed up the process of establishing E2E connections before decoherence causes serious damages. For example, instead of processing all requests in a batch mode which affords a global optimality, how we can process requests in an online fashion while still achieving a good performance. A related challenge is whether one can come up with a distributed control approach that is not only more scalable, but also at least as effective as a central control approach. Another challenge is how to utilize limited resources effectively when considering trade-offs between either using some resources to increase the redundancy (or fidelity) for one or a few E2E entanglement connections or using the resources to establish additional E2E connections. A typical example is the use of extra Bell pair of e-bits as sacrificial pairs to improve fidelity, in a scheme called purification.

The above challenges naturally present opportunities for us to pursue some interesting open research problems.

IV. PROMISING APPROACHES AND OPPORTUNITIES

In this section, we present several promising approaches to E2E entanglement establishment in QDNs, taking into consideration the constraints and challenges described earlier. The overall objective is to maximize the number of E2E entanglement connections (and thus throughput), while fairness among multiple SD pairs could also be a secondary objective.

Entanglement routing to maximize network throughput. Most of the current QDN networking research addressed the challenge of limited quantum memory. Since one E2E entanglement connection can be used to teleport only one data qubit, it is desirable to establish as many entanglement connections as possible. The problem of how to figure a best path for each SD pair (such as Alice and Bob), along which we will create entanglement links and then form an E2E entanglement connection from Alice to Bob is referred to as *entanglement routing* problem.

Early works have focused on the entanglement routing problem on some specific types of topology, such as diamond [14], ring or sphere [15], star [16], and chain [17]. The works in [18] and [19]are perhaps the first works considering entanglement routing problem on a general topology. However, both of them assumed that the entanglement links have been successfully created, and only focused on how to connect the existing entanglement links to form E2E entanglement connections. In other words, they ignored the issue that an entanglement link is difficult to be generated, and hence in order to fully utilize the limited quantum memory, we need to discuss how to selectively generate entanglement links, and then how to connect the successfully created entanglement links to establish E2E entanglement connections.

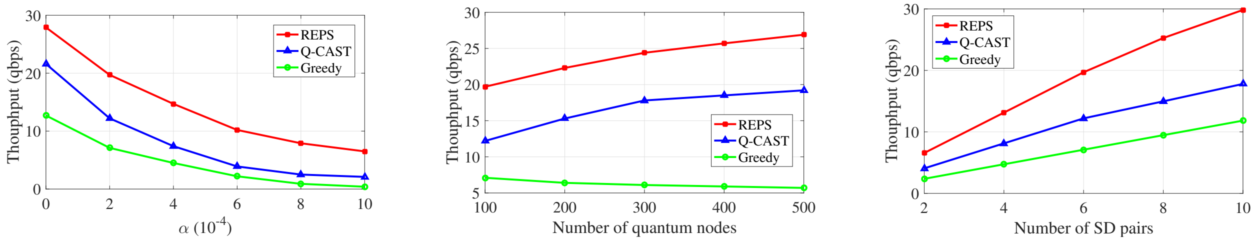
Following the above idea, a two-step approach, called the Q-CAST [20] algorithm, was proposed. Q-CAST used the extended-Dijkstra's algorithm to first find most reliable paths (that have the best chance to successfully establish entanglement connections) for the given set of SD pairs. It then tries to create entanglement links along the chosen paths. If some of them failed to be created, one would end up (or be stuck) with what we call "sub-connections" (*i.e.*, each being a part of an E2E entanglement connection). As a remedy, in the second step of Q-CAST, each quantum node will leverage the information gathered from neighboring nodes in order to to try to establish as many E2E entanglement connections as possible, by salvaging these sub-connections.

Note hat step 2 of Q-CAST may be considered as a best-effort reaction to having failed to establish entanglement links in the first step. Since such a failure probability could be high, it'd be better to take a more proactive approach. More specifically, it would be better to create some redundant entanglement links in the first step, especially since multiple entanglement links (belonging to different entanglement paths) need to be created which creates an opportunity to "share" the redundancy. When some entanglement paths fail to be created, the redundant entanglement links can be used selectively to establish as many E2E entanglement connections possible to achieve a maximal utilization of resources allocated in the first step. Following this idea, we proposed REPS in [21].

Fig. 2 compares the performance of Q-CAST and REPS under different settings with that of the greedy algorithm (which identifies a minimum hop path for a SD pair and will fail to establish an E2E connection if at least one entanglement link along the path fails), The α in Fig. 2(a) is the parameter that affects the success probability to create an entanglement. More specifically, an entanglement link can be created over a quantum link (u, v) following the probability $p_{uv} = e^{-\alpha l(u,v)}$ [18], where $l(u, v)$ is the physical distance between node u and v and α is a system specific parameter. The larger α is, the smaller probability there will be to create an entanglement link. I tis not surprising that from the above figure, we can observe that REPS performs best due to its proactively provisioned redundancy when creating entanglement links which affords the flexibility when trying to connect successfully created entanglement links to form E2E entanglement connections. The greedy algorithm performs worst since it chooses paths with least number of hops but does not take the success probability to create entanglement links into consideration.

There are several other opportunities to further improve QDN throughput via advanced entanglement approaches. For example, since a QDN usually has optical switches connecting EPSes and BSM devices to all quantum nodes (computers or repeaters) using a switched network, we can deliver e-bits generated by an EPS to quantum nodes (quantum computers or quantum repeaters) that are not geographically close using the switching network (consisting of switches and quantum links, along with EPSes and BSM devices), in order to create an entanglement segment, which would otherwise have to be formed by stitching two or more entanglement links together. Doing so can not only save the precious quantum memory at some repeaters, but also avoid performing quantum swapping which may fail. Of course, the downside is that such e-bit distributions through multiple optical switches and quantum links could result in a higher failure probability for the entanglement segment than that for each individual entanglement link. Accordingly, we can explore tradeoffs involved in leveraging both switching and quantum swapping to establish E2E entanglement connections.

Ensuring high fidelity E2E entanglement. To teleport a data qubit from Alice to Bob over an entanglement connection, Alice needs to perform a joint BSM on her data qubit and one e-bit, and Bob has to perform a unitary operation on his e-bit, based on the BSM results from Alice and all other repeaters involved in quantum swapping in the process of establishing the E2E connection. Accordingly, the fidelity of the entanglement connection used for teleportation determines the reliability of the teleportation. [22] is perhaps the first work that took entanglement link fidelity into consideration, albert indirectly. The main idea in [22] is to tackle the time-induced decoherence issue by reducing the duration that every entanglement link needs to be maintained before an quantum swapping operation is performed, so that these links are likely to stay in the expected state when establishing the desired E2E entanglement connection. This work did not however quantify the fidelity of each entanglement link when considering which entanglement links to create or stitch together. To improve the entanglement fidelity, purification is required. The main idea of purification of a Bell pair (or its corresponding entanglement link) is to use another (called sacrificial) Bell pair to test if the (first) entanglement link is in the expected state or not. If not, we will get rid of the entanglement link and generate a new one. [23] discussed several purification schemes but when it comes to entanglement routing, the main idea is to assign



(a) Entanglement success probability vs. throughput.

(b) Network scale vs. throughput.

(c) Number of SD pairs vs. throughput.

Fig. 2. Performance of REPS and Q-CAST.

each link a cost which is inversely proportional to the total number of sacrificial pairs supported by the link, and then use the Dijkstra's algorithm to find a shortest path, hoping that the path will have a high enough E2E fidelity. The work didn't quantify, let alone guarantee, the E2E fidelity. [24] proposed an elaborate entanglement routing solution based on the idea of first purifying as many entanglement links as possible, and then use only the links whose fidelity can be purified above a given threshold in the routing step. However, it still didn't quantify the E2E fidelity. In addition, purifying many links so their fidelity is above a threshold will not only waste resources, but also fail to guarantee a high enough E2E fidelity.

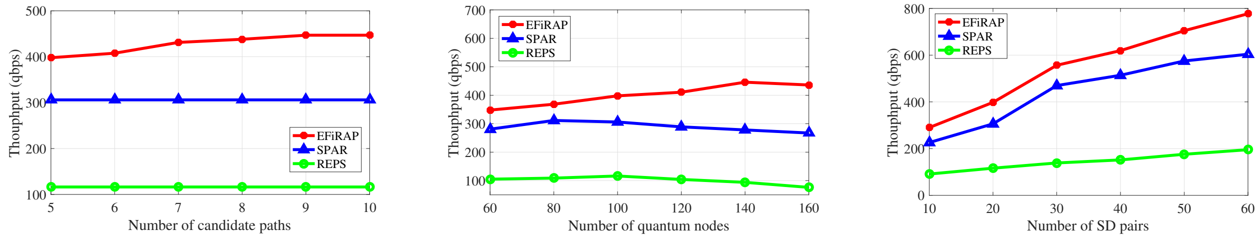
To guarantee a high E2E fidelity, we should first calculate it. However this is not trivial even if the fidelity of all the entanglement links is known. [25] proposed an approach to calculating the E2E fidelity of such an entanglement connection by deriving the relationship between the E2E fidelity and the fidelity of each entanglement link, assuming that there are only bit flip errors. Under the bit-flip error assumption, a counter-intuitive finding is that sometimes, over-purifying an entanglement link will reduce the overall E2E fidelity. Based on derived relationship between the E2E fidelity and the fidelity of each entanglement link, [25] also proposed an efficient way to determine which entanglement links to purify and how much purification should be performed, in addition to finding a path. In particular, the proposed EFiRAP approach, one first prepares a Candidate Entanglement Path Set (CEPS) whose elements consist of not only the entanglement paths but also the corresponding purification schemes to ensure that the E2E fidelity will be satisfied. CEPS may produce multiple candidates for each SD pair. Then, EFiRAP maximizes the network throughput by selecting one candidate (*i.e.*, a combination of an entanglement path and the corresponding purification schemes along the path).

Fig. 3 shows the performance of EFiRAP, assuming the threshold requirement is 0.8. As comparison baselines, the performances of SPAR and REPS are also shown. In SPAR (which is similar to the approach taken in [24]), each entanglement link will be purified to have a fidelity larger than 0.94 as a heuristic to ensure that the E2E fidelity would be higher than the required threshold, while REPS does not perform any purification. In this figure, only those entanglement connections with E2E fidelity larger than 0.8 will be counted. From this figure, we can see that EFiRAP outperforms the other two while REPS performs the worst.

There are still many open research questions related to the E2E fidelity aware entanglement routing in QDNs. For example, how can we calculate the E2E fidelity when taking the phase flip error into consideration.

Dealing with fast decoherence. Most of the above mentioned approaches assumed centralized control and relied on complex and time-consuming algorithms to optimize the process of establishing E2E entanglement connections. In particular, a typical entanglement routing solution would perform the following steps for a set of SD pairs in a time-slotted fashion: (i) calculate all entanglement paths; (ii) create all entanglement links; (iii) perform quantum swapping; and (iv) teleportation. To deal with fast decoherence, the duration of a time slot should be less than the time an entanglement can be maintained, which is about 1 second. This means that there is a very little time to carry out each of these steps, especially steps (i), (ii) and (iii), which could be relatively more time consuming than step (iv). Indeed, in a large network with many QCs and SD pairs, we may not be able to calculate the entanglement paths for all SD pairs in a timely manner. To speed up this step (i), one can use an online entanglement routing method which processes each request as soon as it arrives. Typical, this results in an less efficient utilization of the quantum resources as a trade-off due to the lack of more global knowledge about all the requests.

In order to speed up steps (ii) and (iii), we may explore approaches that use centralized control to determine routes but distributed signaling to reserve quantum resources(and establish the E2E connection link-by-link) in a synchronous fashion. For example, when an entanglement path from Alice to Bob (with H hops) is identified, a basic approach is for Alice to forward a request to establish up to N entanglement connections as possible through this path. At the i-th intermediate node R_i (where $1 \leq i < H$), it determines (over all previous $i < H$ hops) the maximal number of concurrent entanglement links (from Alice to R) that can be established on each of the previous $i < H$ quantum links. Denote this number by W_i (whose upper bound is limited by the availability of the quantum resources such as the quantum memory at previous i hops). If R has a total $2W_{i+1}$ units of free quantum memory, then it will be establish at most W_{i+1} entanglement links in the upstream (towards Alice) and downstream (towards Bob) direction. Thus, it can create $W = \min\{W_i, W_{i+1}\}$ entanglement links with its upstream node. If $W < W_i$, then up to $W_i - W$ entanglement links over each of the previous i hops



(a) Number of candidate paths vs. throughput. (b) Network scale vs. throughput. (c) Number of SD pairs vs. throughput.

Fig. 3. Performance of EFiRAP.

may be released (or used by another SD pair). In addition, R should notify its downstream that at most W links over each hop can be established. This helps improve the utilization of the quantum memory. Since it may take a while (multiple tries) to successfully establish an entanglement link at each hop, a more aggressive approach is to allow a request to move to the next intermediate node before an entanglement link on the current hop is established. This allows multiple nodes to try to establish their entanglement links in a pipelined fashion. An even more advanced approach is for a central controller to send multiple requests, one for each link along the path, so as to be able to establish multiple entanglement links concurrently.

In addition, in either the basic, pipelined, or the concurrent signaling approach, as soon as an intermediate node R has created entanglement links with both of its upstream and downstream nodes, it will perform quantum swapping without waiting for all the entanglement links along the entanglement path to be established. This will not only reduce the time needed to maintain entanglement status (and thus help improve the fidelity), but also more importantly can help free up the quantum memory at R quickly for other links adjacent to R.

V. CONCLUSIONS

In this paper, we have outlined a vision for supporting distributed quantum computing using quantum data networks (QDNs). We have presented major low-level constraints imposed by quantum physics and quantum communication technologies, and discussed how they impact high-level QDN architecture and (routing) protocol designs in terms of both research challenges and opportunities. Promising approaches to addressing the challenges along with preliminary research results have also been presented. Much future work on QDNs, including quantum mechanisms for FEC, need to be pursued.

ACKNOWLEDGMENT

The work by Hongli Xu is supported in part by the Anhui Initiative in Quantum Information Technologies under Grant AHY150300.

REFERENCES

[1] L. K. Grover, “A fast quantum mechanical algorithm for database search,” in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, 1996, p. 212–219.

[2] P. W. Shor, “Algorithms for quantum computation: discrete logarithms and factoring,” in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, 1994, pp. 124–134.

[3] D. Cuomo, M. Caleffi, and A. S. Cacciapuoti, “Towards a distributed quantum computing ecosystem,” *IET Quantum Communication*, vol. 1, no. 1, p. 3–8, Jul 2020.

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

[5] J. L. Park, “The concept of transition in quantum mechanics,” *Foundations of Physics*, vol. 1, p. 23–33, 1970.

[6] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, “The security of practical quantum key distribution,” *Rev. Mod. Phys.*, vol. 81, pp. 1301–1350, 2009.

[7] C. H. Bennett and G. Brassard, “Quantum cryptography: Public key distribution and coin tossing,” in *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing*, 1984.

[8] A. K. Ekert, “Quantum cryptography based on bell’s theorem,” *Physical review letters*, vol. 67, no. 6, pp. 661–663, 1991.

[9] M. M. Weston, S. Slussarenko, H. M. Chrzanowski *et al.*, “Heralded quantum steering over a high-loss channel,” *Science Advances*, vol. 4, no. 1, 2018.

[10] J. M. Auger, H. Anwar, M. Gimeno-Segovia, T. M. Stace, and D. E. Browne, “Fault-tolerant quantum computation with nondeterministic entangling gates,” *Phys. Rev. A*, vol. 97, p. 030301, Mar 2018.

[11] S. Haroche, “Entanglement, decoherence and the quantum/classical boundary,” *Physics Today*, vol. 51, no. 7, pp. 36–42, 1998.

[12] A. I. Lvovsky, B. C. Sanders, and W. Tittel, “Optical quantum memory,” *Nature Photonics*, vol. 3, pp. 706–714, 2009.

[13] R. H. Hadfield, “Single-photon detectors for optical quantum information applications,” *Nature Photonics*, vol. 3, pp. 696–705, 2009.

[14] S. Pirandola, “End-to-end capacities of a quantum communication network,” *Communications Physics*, vol. 3, pp. 1—-10, 2019.

[15] E. Schoute, L. Mancinska, T. Islam, I. Kerenidis, and S. Wehner, “Shortcuts to quantum network routing,” *CoRR*, vol. abs/1610.05238, 2016. [Online]. Available: http://arxiv.org/abs/1610.05238

[16] G. Vardoyan, S. Guha, P. Nain, and D. Towsley, “On the stochastic analysis of a quantum entanglement switch,” *SIGMETRICS Perform. Eval. Rev.*, vol. 47, no. 2, pp. 27—-29, dec 2019.

[17] M. Caleffi, “Optimal routing for quantum networks,” *IEEE Access*, vol. 5, pp. 22 299–22 312, 2017.

[18] 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, Dec. 2019.

[19] R. V. Meter, T. Satoh, T. D. Ladd, W. J. Munro, and K. Nemoto, “Path selection for quantum repeater networks,” *Networking Science*, vol. 3, pp. 82–95, 2013.

[20] S. Shi and C. Qian, “Concurrent entanglement routing for quantum networks: Model and designs,” in *Proceedings of the ACM SIGCOMM*, 2020.

[21] Y. Zhao and C. Qiao, “Redundant entanglement provisioning and selection for throughput maximization in quantum networks,” in *Proceedings of the IEEE INFOCOM*, 2021.

[22] W. Kozlowski, A. Dahlberg, and S. Wehner, “Designing a quantum network protocol,” in *ACM CoNEXT*, 2020, pp. 1 – 16.

[23] M. Victora, S. Krastanov, A. S. de la Cerda, S. Willis, and P. Narang, “Purification and entanglement routing on quantum networks,” 2020.

[24] 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, 01 2021.

[25] Y. Zhao, G. Zhao, and C. Qiao, “E2e fidelity aware routing and purification for throughput maximization in quantum networks,” in *Proceedings of the IEEE INFOCOM*, 2022.