Designing Quantum Repeater Networks

Rodney Van Meter and Joe Touch

ABSTRACT

Quantum networks generate distributed entangled state or relocate quantum state, uniquely ensuring eavesdropper detection or reaching agreement more quickly than their classical counterparts. These capabilities rely on the composition of link and multihop mechanisms into a coherent system, with particular attention to managing errors in and loss of delicate quantum states. This article explores quantum networking in terms of fundamental network architecture principles, and explains where and how it diverges from its classical counterparts. It discusses engineering principles that ensure robust and interoperable communication by introducing new protocol layers to support quantum sessions, and considers how these layers interact with quantum link mechanisms to support user-level quantum-enabled applications.

INTRODUCTION

Quantum networks are distributed systems that utilize entanglement and teleportation. They uniquely enable tamper-evident communication, distributed quantum computation, and quantum sensor networks that support tests of the foundations of quantum mechanics itself [1, 2]. These applications are driving substantial amounts of research into quantum networks, especially a form known as quantum repeater networks.

Quantum networks appear impossible. Classical networks depend on relaying, in which a message is copied from one link to another to compose links into an end-to-end path. Quantum information — encoded in single photons or groups of photons — is extremely fragile. Reception of photons transmitted either through free space or over an optical fiber typically succeeds only with low probability, and the signal’s quantum nature can be destroyed with the loss of even a small fraction of the signal. Quantum information cannot be copied, a restriction known as the no-cloning theorem [3]. This restriction prohibits us from using classical repeaters and amplification, and even prevents us from keeping copies to retransmit.

It would seem to be impossible to transport quantum data in any but a limited area. Paradoxically, the same quantum properties that limit a classical solution enable a uniquely quantum approach — we can create distributed quantum entanglement independent of the quantum state we want to transmit, then use that entanglement to teleport data rather than transmit it. The discovery of quantum teleportation and quantum error management led to the development of quantum repeaters. These quantum repeaters are the core of a quantum network architecture because they relay data between potentially different link technologies, just as Internet routers do in classical networks. Other quantum router functions, such as path determination and routing exchange, use classical methods adapted with quantum-related metrics and constraints.

Very little work has been done to date on quantum repeater network architectures, but a half dozen approaches to repeater communication sessions have been proposed [4–9]. Each of these approaches can be organized into a protocol stack in the spirit of the OSI seven-layer model. We focus on the relationship between the session architecture and the restrictions imposed by the physical technologies, and the resulting impact on network architectures.

A network architecture must operate within technological limits to provide a complete description of how nodes and links are composed and managed to allow long-distance end-to-end communication across a series of hops known as a path. Although classical design principles can be applied to quantum networking, the resulting architectures can be quite different due to the radical restrictions and unique capabilities of the quantum domain.

In this article, we discuss the issues in combining these technologies into large-scale long-lived networks. We review basic networking concepts, including group communication and quantum communication, as well as principles of quantum repeaters. We present our quantum network architecture, including its vertically layered and horizontally distributed aspects and our layered recursive quantum repeaters. We compare our approach to other quantum repeater network architectures and evaluate aspects of their feasibility, and we summarize our conclusions.

BACKGROUND

Typical network architecture descriptions span two dimensions: vertically, as a layered protocol stack, and horizontally to describe using this stack across different nodes for group communication. This section presents our view of the defining properties of a network architecture and reviews some of the key quantum communication concepts needed to explain our quantum network architecture in this context.
GROUP COMMUNICATION

Networking describes the way in which a group communicates with each other. Such communication is trivial given full connectivity, such as when \( N \) parties are interconnected with \( N(N - 1) \) direct links. The more interesting case is where these \( N \) parties are connected with far fewer direct links, and pairwise communication requires the transitive closure of a sequence of links to create a path. This case describes the horizontal component of network architecture, specifying the components of a group communication system and how they interact. Group communication falls into two approaches:

- Message transfer — relocating information between physical locations
- Distributed state coordination — establishing simultaneous information at a set of physical locations

Although a message transfer system can be used to establish distributed state, and a distributed state system can be used to transfer messages, typically a given architecture focuses primarily on only one of these capabilities. Note that this is the typical “message passing vs. shared memory” duality, although we use the terms transfer and state to more accurately handle our quantum case.

A group communication architecture describes how these goals are achieved by defining:

- Information type — communication content, typically bits (qubits for quantum communication)
- Group communication type — message transfer or distributed state coordination
- Link — the communications medium that connects parties or intermediates
- Identifiers — to differentiate the communicating parties, messages, and/or states
- Paths — a way to support group communication among \( N \) parties with far fewer than \( N^2 \) links, and must define:
  - The time at which the path is decided
  - A means for composing links into a path
  - A way to relay information between adjacent links

The architectural elements above are tightly coupled, such that a change in any one would result in significant changes to the others. Additional mechanisms enable communications systems to support reliable in-order delivery, avoid resource competition, and increase capacity. Those enabling mechanisms are very important additions but typically independent, and thus not considered part of the architecture itself.

One of the most common examples of group communication is the Internet, which communicates information as bits. The Internet transfers messages of variable-length byte sequences that can be reordered, lost, or duplicated. Links are either two-party simplex or multiparty broadcast channels. The Internet uses globally unique party identifiers. Paths are constructed as the result of messages traversing the network at the time of that traversal (i.e., on the fly). The destination identifier within the message is matched to rules at a node that attaches to a set of links. Messages are relayed between adjacent links by classical means (i.e., by making a classical copy of the message).

Additional capabilities critical to using the architecture are outside the scope of the architecture, but are established by enabling mechanisms. For the Internet, these enabling mechanisms support:

- In-order reliable delivery that adjusts to congestion (TCP)
- End-to-end state (HTTP, FTP, etc.)
- Distributed translation of symbolic names to Internet addresses (Domain Name Service [DNS])
- Distributed computation of local rules that achieve local and global paths (Open Shortest Path First [OSPF], Intermediate System to Intermediate System [IS-IS], and Border Gateway Protocol [BGP])
- Distributed translation between Internet addresses and link addresses (Address Resolution Protocol [ARP] or its equivalent)

These mechanisms can be — and often are — replaced without impact on the architecture. Internet in-order reliable delivery is provided by several different mechanisms (TCP, Stream Control Transport Protocol [SCTP]), and various lower physical layers (wireless LAN or copper cable) and the upper application layers (HTTP, FTP) coexist within the Internet. This is why these mechanisms are not considered a part of the architecture.

QUANTUM COMMUNICATION CONCEPTS

The following is a brief summary of the key aspects of quantum communication that impact network architecture.

Qubits — Quantum information is most often discussed in terms of qubits. A qubit, like a classical bit, is something with two possible values that we can label zero and one. Unlike a classical bit, a qubit can occupy both values simultaneously, known as superposition.

Superposition and Measurement — A qubit can represent multiple values in different proportions at the same time (e.g., two-thirds of a one and one-third of a zero). This superposition determines the relative probability of finding each value when we measure the state. When we measure the qubit, we get only a single classical bit of information (the one or zero) with 100 percent probability, and the superposition collapses.

Entanglement and Bell Pairs — Some groups of qubits exhibit strong correlation between the qubits that cannot be explained by independent probabilities for individual qubits. Instead, the group must be considered as a whole, with interdependent probabilities. This phenomenon is known as quantum entanglement. A special entangled state known as a Bell pair or EPR pair, consisting of two qubits, figures prominently in quantum communication. Each qubit in the pair has a 50 percent probability of having a value of 1 and a 50 percent probability of having a value of 0 when we measure it. Although we cannot predict which will be found, when we measure one member of the pair, the value of the other is...
immediately determined. This happens independent of the distance between the two members of the Bell pair.

**No Cloning** — As mentioned above, a key restriction of quantum systems is that we cannot make independent copies of an unknown state [3]. This makes error correction exceedingly difficult.

**Fidelity** — The quality of a quantum state is described by its fidelity, which is, roughly, the probability that we correctly understand the state — if we ran the same experiment many times and measured the results, how close do our desired statistics would we be? Unfortunately, any physical operation results in a loss of fidelity, gradually degrading the state as we manipulate or even store it. We can counter this by using a form of error correction or detection.

**Purification** — The form of error detection historically favored in quantum repeater networks is purification, which uses minimal resources [4]. It sacrifices some quantum states to test the fidelity of others. There are various purification mechanisms, with different purification algorithms and different methods for determining which states are sacrificed, each with particular trade-offs.

**Quantum Error Correction** — QEC may be based on classical codes or purely quantum concepts. The primary difficulties are extraction of errors without damaging quantum state, avoiding error propagation, and the increased resources required [1, 5, 6; references therein].

**Teleportation** — Teleportation destroys the state of a qubit at the sender and recreates that state at the destination, teleporting information rather than matter, as explained in Fig. 1 [10]. The process uses a Bell pair’s long-distance correlation, followed by transmission of a pair of classical bits.

With these basic concepts, we can begin to construct networks. Bell pairs are consumed by teleportation, so one way to organize a network is to create a continuous stream of Bell pairs between source and destination — as long as we identify those sources and destinations, choose paths to get there, and manage the resources along the way.

**PURIFY AND SWAP: EARLY QUANTUM Repeaters**

In the late 1990s, researchers recognized that teleportation can be used to extend entanglement, and that purification could be used to detect errors introduced in the process [4]. This first architecture we call purify and swap, although the originators called it nested purification.

The process of entanglement swapping uses teleportation to splice two Bell pairs spanning adjacent short distances into one pair over the corresponding longer distance. If node $A$ shares a Bell pair with node $B$, and node $B$ shares another Bell pair with node $C$, node $B$ can teleport its member of the $A \leftrightarrow B$ pair to node $C$ using the $B \leftrightarrow C$ Bell pair. In the process, the $B \leftrightarrow C$ pair is consumed, and at the end we have a single $A \leftrightarrow C$ Bell pair.

Entanglement swapping is independent of the distances between $A$ and $B$, and between $B$ and $C$. Only local quantum operations are required, supported by classical communication. We combine one-hop Bell pairs into two-hop Bell pairs, then combine two-hop pairs into four-hop pairs, and so on, doubling the length of the remaining pair at each step, as shown in Fig. 2.

To compensate for the errors introduced,
purification is used, as shown in Fig. 3: local quantum operations are performed at both nodes on two Bell pairs, then one of the Bell pairs is measured. The measurement results are exchanged and compared. If they agree, the pair’s fidelity has improved, and it is kept for reuse. If the measurement results disagree, the pair is discarded.

Purify-and-swap is the combination of these two concepts, interleaving purification with entanglement swapping. Purification is performed over one hop, then two hops, then four, resulting in a recursive, interleaved power-of-two approach called simply nested purification. The principles have proved to be flexible, so we refer to the entire group of specific designs as purify and swap session architectures.

**A Quantum Network Architecture**

The design philosophy of our quantum network architecture is inspired by the Internet architecture, leveraging it as much as possible, except where modifications are absolutely necessary to distribute quantum state. As noted earlier, there are two dimensions of an architecture: vertical layered communication and horizontal distributed group communication. Here, we describe layering in terms of the model we have developed [11], and group communication in terms of our Quantum Recursive Network Architecture (QRNA) [12]; these are presented in the following two subsections.

**Vertical: Layered Quantum Communication**

Layered communication describes how protocol functions are vertically composed within a communications node to provide increasingly complex capabilities. Layered quantum communication relies on five key vertical layer functions that are uniquely quantum.

- **Physical layer:** We rely on a quantum physical layer using light to encode quantum state. Many technologies for this layer are under development.
- **Link-level entanglement:** We rely on existing techniques to support entanglement across a link. Because most physical entanglement mechanisms are probabilistic, the link layer will include an acknowledgment to the sender indicating which attempts succeeded.
- **Remote state composition:** In the Internet, links are composed by copying packets from one link to the next. In a quantum network, links are less readily composed due to the no-cloning theorem. Quantum paths thus either establish end-to-end entanglement from entangled links, or use that entanglement to teleport quantum state from one end to the other. This layer is very sensitive to the link-layer capabilities, as well as the error management mechanism.
- **Error management:** In the classical Internet, errors are managed using redundancy (e.g., forward error correction) or error detection and retransmission. As noted earlier, the no-cloning theorem prevents straightforward use of either of these mechanisms. The fidelity of quantum states is critical in reducing the need for error management.
- **Application:** The application may be a sensor network, or a numeric computation or decision algorithm based on shared state [13]. The application will determine if end-to-end entanglement is required, or if our quantum states can be measured on a pay-as-you-go basis. Some applications may also desire quantum states other than Bell pairs, including any of several common forms of three-party or larger states. Of course, the application is driven by a classical program, presumably using a socket-like data structure.
Composing Quantum Links: Purify and Swap — To compose links, we use recursive (nested) purify and swap along a path. In classical networks, composition is a matter of copying and separately applying error correction and control. In purify and swap, composition and error management are sometimes viewed as an integrated operation, but in our layering they are natively distinct operations.

With this architectural background, let us return to the canonical purify and swap approach. Figure 4 shows a five-node example. The physical and link layers are the two layers at the bottom, labeled physical entanglement (PE) and entanglement control (EC), respectively. The key feature in the communication session architecture is the recursive nature of the error management and remote state composition layers, which in purify and swap we call purification control (PC) and entanglement swapping control (ESC), respectively. In this example, purification is run over individual links, two-hop entanglement, and finally four-hop entanglement. Entanglement swapping is run at all intermediate nodes, first at B and D to create two-hop entanglement, then after purification at C to create four-hop entanglement. One characteristic of this nested approach is that the end nodes of an n-node path must communicate with \( \log_2 n \) other nodes along the path, which has implications for the path selection and composition mechanisms.

We can observe that entanglement swapping can be thought of as the middle node taking a Bell pair qubit from its left and teleporting it to the right using another Bell pair. As an example, node B in the figure in theory need only communicate with C. A’s role in the process is entirely passive. However, as the goal is to create end-to-end entanglement, A must participate; after entanglement swapping, the next operation may be purification, another swapping operation, or transfer of control to the application, any of which requires A’s involvement.

Other Aspects of Quantum Layering — Organizing the layering is the first step in developing the key purify-and-swap insight described earlier into a functional, robust, distributed implementation. To be practically implementable, details of the management of requests must be defined. Our approach to doing so is to use protocol state machines to govern the memories themselves [11]. One important facet of this problem is management of the Bell pairs to maximize the end-to-end success rate of purification and swapping, which in turn affects the overall system throughput. We call this the purification scheduling problem. Minor extensions are also required when a path is not a power of two hops long.

HORIZONTAL: DISTRIBUTED QUANTUM COMMUNICATION

Distributed group communication describes how protocol functions are horizontally composed across different communication nodes. Distributed quantum communication extends this principle to quantum communication by explicitly managing distributed state through the use of recursive composition.

As with the Internet, our architecture composes links into paths, manages state and errors, and supports applications. This section introduces the key differences that result in quantum networks having different architectures from classical: type of information, state management, path composition, and identifiers.

Our Quantum Recursive Network Architecture (QRNA) [12] provides a general-purpose request mechanism abstracted from underlying layers, to accommodate any of the models presented above. Rather than explicit state transfer, it supports requests for creation of distributed states (including both two-party and multi-party states) and operations on those states. Requests may be recursively decomposed and distributed throughout the network in order to build the end-to-end state requested by an application. A link in QRNA may be a physical link or a recursively organized network. QRNA uses globally unique identifiers that represent the locations where the shared quantum state will be established; the structure of these identifiers affects how paths are determined, but is outside the scope of the architecture. Paths are constructed by classical means prior to communication.

Table 1 compares QRNA with the Internet architecture.

Information Type and Group Communication — A quantum network architecture can be organized to present either the generation of distributed entangled state or the relocation of quantum state as the fundamental communication semantics. Relocation, using either direct transmission or simple teleportation, may seem easier and more natural, but distributed state generation natively supports a broader range of applications.
State relocation across a network would be sufficient for some applications. One-way teleportation from a client to a server is sufficient for universal blind quantum computation, in which the server is oblivious to the computation it performs for the client. State relocation also appears to extend smoothly from unentangled networks. Applications that need simultaneous long-distance entangled states must build them, because state relocation does not provide entangled states. State relocation does not demand long-lived memory unless the session architecture itself does, but it also cannot easily take advantage of resources in the middle of the network to operate more efficiently.

Distributed state generation supports a more general distributed computation model. It works well with both two-party and multiparty entangled states. However, in the basic form it requires long-lived memory. Asynchronous distributed state generation is actually the most general model, subsuming both of the above. This model, which QRNA adopts, provides the most direct match to applications such as entanglement-based quantum key distribution, in which long-distance Bell pairs are measured at each end soon after creation.

Identifiers — Networks naturally require names for the nodes or communication endpoints. Unlike the Internet, purify-and-swap end nodes communicate directly with nodes along the path.

On the Internet, a packet is directed to transit a particular subnet (Internet autonomous system), rather than given a complete hop-by-hop source route. QRNA’s recursive naming allows an operation, such as Bell pair creation or entanglement swapping, to be similarly directed to a subnetwork rather than to a specific node. Paths then can be transparently relocated within the subnetwork. This partially relaxes the path constraint, simplifying end node knowledge of network components and returning local operation decisions to the local neighborhood.

The entangled states built within the network also must be named, to facilitate their management and delivery to applications. On the Internet, packets are mapped to a connection using a tuple consisting of node addresses, a connection identifier (port numbers), and possibly an application-level identifier. In quantum networks, such a tuple may not yet exist because a distributed state, such as a Bell pair in the middle of the network, might not yet be assigned to serve a particular end-to-end session. QRNA is designed to accommodate this delayed association (a type of late binding) and to reassign state identifiers when necessary.

**Figure 4.** Protocol layers and their interaction in purify-and-swap repeaters in a five-node four-hop chain. The left labels indicate the model layer represented, and the boxed labels and right labels indicate the protocol name for purify-and-swap repeaters. Double-headed arrows indicate that bidirectional classical communication is required. Only the physical layer is quantum, shown propagating left to right.
Purify and swap was developed because a perfect physical quantum link cannot exist. Purify and swap’s demand for round-trip end-to-end communication limits throughput and demands long memory lifetimes. The quest to better match available technological capabilities and improve performance has driven the development of several new approaches to the vertical layering and horizontal distributed communication interaction, as summarized in Table 2.

<table>
<thead>
<tr>
<th>Comm type</th>
<th>Internet</th>
<th>QRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>Message transfer</td>
<td>Distributed state</td>
</tr>
<tr>
<td>Link</td>
<td>Byte sequences of varying length</td>
<td>Entangled state</td>
</tr>
<tr>
<td>Identifier</td>
<td>Simplex 2-party or broadcast N-party</td>
<td>Two-party with entangled state supporting quantum teleportation</td>
</tr>
<tr>
<td>Path</td>
<td>Global fixed-length</td>
<td>Global, any that supports the path composition algorithm</td>
</tr>
</tbody>
</table>

We can compensate for low memory lifetime by using quantum error correction in the repeater nodes, or reengineering the protocol stack to avoid round-trip delays. The encoded link [5] and quasi-asynchronous [7] approaches each require an individual memory lifetime longer than the link round-trip time (RTT), but for \( n \) hops require this for \( n \) separate memories, in which the total time a state is stored in memory is proportional to the end-to-end latency. The surface code [6] and memoryless [8] approaches can tolerate short memory lifetimes, but at the expense of needing a high probability of entanglement success.

The availability of sufficient buffer memory is also a problem. The earliest purify-and-swap proposals required a few tens of qubits per node, proportional to the log of the number of repeater hops in a network’s longest path. Although this suggests a scalable solution, it exceeds current experimental capabilities. An adapted version uses only two qubits per node [15]. Encoded link and surface code, which depend on QEC, require orders of magnitude more memory than purify and swap. The memoryless approach takes advantage of clever encoding to avoid storing qubits in memory.

If the generated entanglement is already of high fidelity, all of these schemes will work well. Purify-and-swap and measurement-based schemes operate with low-fidelity entanglement, but can reduce the round-trip purification delays when entanglement fidelity is high.

The measurement-based scheme can be considered a new implementation of purify and swap, and a carefully engineered protocol stack would allow it as a drop-in replacement for individual nodes. Conversely, memoryless is a new link architecture whose benefits are realized only when the entire protocol stack is optimized. Encoded link, surface code, and quasi-asynchronous are not specific to a particular link layer, and may as a group be able to support the same upper layer protocols, including ESC.

In this article, we have introduced both the vertical and horizontal aspects of designing quantum repeater networks. The vertical direction, which we describe in terms of layered communication, has recently seen a flood of new proposals. Each new proposal typically focuses on a new mechanism for managing errors, the primary constraint on creating quantum networks. The horizontal aspect, involving group communication via paths through complex arrangements of links and nodes, as we propose in QRNA, will be critical for deploying functional, robust, efficient networks.

With such a proliferation of layered communication schemes and a large experimental community working on many mechanisms for creating long-distance entanglement, there is no reason yet to assume that we have exhausted the possibilities.
range of possible link-level constructions or approaches to communication sessions. The true test of a quantum Internet architecture will be its resilience in the face of continuing innovation. QRNA provides a framework in which relaying between physically heterogeneous links or even converting from one layering scheme to another can be described, but how to actually implement such mechanisms remains an open problem. Engineering long-lived quantum repeater networks will require an explicit focus on internetworking, to connect heterogeneous technologies and allow older nodes to remain active as the network expands over time.

ACKNOWLEDGMENTS

The development and testing of the concepts presented here leans heavily on the work of all of our collaborators, not all of whom can be cited and acknowledged here in the depth they deserve. This work was supported by JSPS KAKENHI 24012706. This project has been made possible in part by a gift from the Cisco University Research Program Fund, a corporate advised fund of the Silicon Valley Community Foundation. This work is partially supported by the NSF Center for Integrated Access Networks (CIAN) under contract 0812072.

REFERENCES


Table 2. Comparison of several quantum repeater communication session architectures. RTT is round trip time, E2E is end to end.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Memory lifetime</th>
<th>Local operation fidelity</th>
<th>Entanglement success probability</th>
<th>Entanglement fidelity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hop-by-hop teleportation</td>
<td>E2E RTT</td>
<td>Very high</td>
<td>Low</td>
<td>Very high</td>
</tr>
<tr>
<td>Purify+swap [4]</td>
<td>Multiple E2E RTTs</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Encoded link [5]</td>
<td>E2E RTT</td>
<td>High enough for QEC</td>
<td>Low</td>
<td>Fairly high</td>
</tr>
<tr>
<td>Surface code [6]</td>
<td>Local QEC cycle time</td>
<td>High enough for QEC</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Quasi-asynchronous [7]</td>
<td>E2E RTT/2</td>
<td>Very high</td>
<td>Low</td>
<td>Fairly high</td>
</tr>
<tr>
<td>Memoryless [8]</td>
<td>Very low</td>
<td>Very high</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Measurement-based [9]</td>
<td>Multiple E2E RTTs</td>
<td>Fairly high</td>
<td>N/A</td>
<td>Low</td>
</tr>
</tbody>
</table>


BIographies

RODNEY VAN METER (rdv@sfc.wide.ad.jp) holds degrees from the California Institute of Technology, the University of Southern California (USC), and Keio University. His research interests include storage systems, networking, and post-Moore’s Law computer architecture. He has held positions in both industry and academia in the United States and Japan. He is now an associate professor of environment and information studies at Keio University’s Shonan Fujiwara Campus.

JOE TOUCH (SM) is the Postel Center Director in USC’s Information Sciences Institute and a research associate professor in USC’s Computer Science and EE/Systems Departments. He received his Ph.D. from the University of Pennsylvania and joined ISI in 1992, and his current projects involve virtual networks, optical Internets, and network security. He is in Sigma Xi, an ACM Distinguished Scientist, and a ComSoc Distinguished Lecturer. He serves on various committees and the editorial boards of IEEE Network and Elsevier’s Journal of Computer and Systems Sciences, and is active in the IETF.

IEEE Communications Magazine • August 2013