Abstract

The vision of a quantum internet is to enhance existing Internet technology by enabling quantum communication between any two points on Earth. To achieve this goal, a quantum network stack should be built from the ground up to account for the fundamentally new properties of quantum entanglement. The first quantum entanglement networks have been realised [Pompili21.1], but there is no practical proposal for how to organise, utilise, and manage such networks. In this draft, we attempt to lay down the framework and introduce some basic architectural principles for a quantum internet. This is intended for general guidance and general interest, but also to provide a foundation for discussion between physicists and network specialists. This document is a product of the Quantum Internet Research Group (QIRG).
This Internet-Draft will expire on 18 August 2022.

Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust’s Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction ........................................... 3
2. Quantum information ..................................... 4
   2.1. Quantum state ..................................... 4
   2.2. Qubit ............................................. 5
   2.3. Multiple qubits .................................... 6
3. Entanglement as the fundamental resource .................. 8
4. Achieving quantum connectivity ............................ 9
   4.1. Challenges ........................................ 9
       4.1.1. The measurement problem ...................... 9
       4.1.2. No-cloning theorem ........................... 10
       4.1.3. Fidelity ..................................... 10
       4.1.4. Inadequacy of direct transmission ............. 11
   4.2. Bell pairs ....................................... 11
   4.3. Teleportation ...................................... 12
   4.4. The life cycle of entanglement .................... 13
       4.4.1. Elementary link generation .................. 13
       4.4.2. Entanglement swapping ....................... 14
       4.4.3. Error Management ............................. 15
       4.4.4. Delivery ..................................... 19
5. Architecture of a quantum internet ....................... 19
   5.1. Challenges ........................................ 19
   5.2. Classical communication ............................ 21
   5.3. Abstract model of the network ..................... 22
       5.3.1. The control and data planes ................ 22
       5.3.2. Elements of a quantum network ................ 23
       5.3.3. Putting it all together ..................... 24
   5.4. Physical constraints ................................ 25
       5.4.1. Memory lifetimes ................................ 26
       5.4.2. Rates ......................................... 26
       5.4.3. Communication qubits ......................... 26
Quantum networks are distributed systems of quantum devices that utilise fundamental quantum mechanical phenomena such as superposition, entanglement, and quantum measurement to achieve capabilities beyond what is possible with non-quantum (classical) networks [Kimble08]. Depending on the stage of a quantum network [Wehner18] such devices may range from simple photonic devices capable of preparing and measuring only one quantum bit (qubit) at a time all the way to large-scale quantum computers of the future. A quantum network is not meant to replace classical networks, but rather form an overall hybrid classical-quantum network supporting new capabilities which are otherwise impossible to realise [VanMeterBook]. For example, the most well-known application of quantum communication, quantum key distribution (QKD), can create and distribute a pair of symmetric encryption keys in such a way that the security of the entire process relies on the laws of physics (and thus can be mathematically proven to be unbreakable) rather than the intractability of certain mathematical problems [Bennett14] [Ekert91]. Small networks capable of QKD have even already been deployed at short (roughly 100km) distances [Elliott03] [Peev09] [Aguado19] [Josh20].

The quantum networking paradigm also offers promise for a range of new applications beyond quantum cryptography, such as distributed quantum computation [Cirac99] [Crepeau02], secure quantum computing in the cloud [Fitzsimons17], quantum-enhanced measurement networks [Giovanetti04], or higher-precision, long-baseline telescopes [Gottesman12]. These applications are much more demanding than QKD and networks capable of executing them are in their infancy. The first fully quantum, multinode network capable of sending, receiving, and manipulating distributed quantum information has only recently been realized [Pompili21.1]

Whilst a lot of effort has gone into physically realising and connecting such devices, and making improvements to their speed and error tolerance, there are no worked out proposals for how to run
these networks. To draw an analogy with a classical network, we are at a stage where we can start to physically connect our devices and send data, but all sending, receiving, buffer management, connection synchronisation, and so on, must be managed by the application directly by using low-level, custom-built, and hardware-specific interfaces, rather than being managed by a network stack that exposes a convenient high-level interface, such as sockets. Only recently, was the first ever attempt at such a network stack experimentally demonstrated in a laboratory setting [Pompili21.2]. Furthermore, whilst physical mechanisms for transmitting quantum information exist, there are no robust protocols for managing such transmissions.

This document, produced by the Quantum Internet Research Group (QIRG), introduces quantum networks and presents general guidelines for the design and construction of such networks. Overall, it is intended as an introduction to the subject for network engineers and researchers. It should not be considered as a conclusive statement on how quantum network should or will be implemented. This document was discussed on the QIRG mailing list and several IETF meetings and represents the consensus of the QIRG members, both of experts in the subject matter (from the quantum as well networking domain) as well as newcomers who are the target audience.

2. Quantum information

In order to understand the framework for quantum networking, a basic understanding of quantum information theory is necessary. The following sections aim to introduce the minimum amount of knowledge necessary to understand the principles of operation of a quantum network. This exposition was written with a classical networking audience in mind. It is assumed that the reader has never before been exposed to any quantum physics. We refer the reader to [SutorBook] and [NielsenChuang] for an in-depth introduction to quantum information systems.

2.1. Quantum state

A quantum mechanical system is described by its quantum state. A quantum state is an abstract object that provides a complete description of the system at that particular moment. When combined with the rules of the system’s evolution in time, such as a quantum circuit, it also then provides a complete description of the system at all times. For the purposes of computing and networking, the classical equivalent of a quantum state would be a string or stream of logical bit values. These bits provide a complete description of what values we can read out from that string at that particular moment and when combined with its rules for evolution in time, such as a logical circuit, we will also know its value at any other time.
Just like a single classical bit, a quantum mechanical system can be simple and consist of a single particle, e.g. an atom or a photon of light. In this case, the quantum state provides the complete description of that one particle. Similarly, just like a string of bits consists of multiple bits, a single quantum state can be used to also describe an ensemble of many particles. However, because quantum states are governed by the laws of quantum mechanics their behaviour is significantly different to that of a string of bits. In this section we will summarise the key concepts to understand these differences and the we will explain their consequences for networking in the rest of the draft.

2.2. Qubit

The differences between quantum computation and classical computation begin at the bit-level. A classical computer operates on the binary alphabet \( \{0, 1\} \). A quantum bit, called a qubit, exists over the same binary space, but unlike the classical bit, its state can exist in a superposition of the two possibilities:

\[
|\text{qubit}\rangle = a|0\rangle + b|1\rangle,
\]

where \( |X\rangle \) is Dirac’s ket notation for a quantum state (the value that a qubit holds), here the binary 0 and 1, and the coefficients \( a \) and \( b \) are complex numbers called probability amplitudes. Physically, such a state can be realised using a variety of different technologies such as electron spin, photon polarisation, atomic energy levels, and so on.

Upon measurement, the qubit loses its superposition and irreversibly collapses into one of the two basis states, either \( |0\rangle \) or \( |1\rangle \). Which of the two states it ends up in may not be deterministic, but can be determined from the readout of the measurement. The measurement result is a classical bit, 0 or 1, corresponding to \( |0\rangle \) and \( |1\rangle \) respectively. The probability of measuring the state in the \( |0\rangle \) state is \( |a|^2 \) and similarly the probability of measuring the state in the \( |1\rangle \) state is \( |b|^2 \), where \( |a|^2 + |b|^2 = 1 \). This randomness is not due to our ignorance of the underlying mechanisms, but rather is a fundamental feature of a quantum mechanical system [Aspect81].

The superposition property plays an important role in fundamental gate operations on qubits. Since a qubit can exist in a superposition of its basis states, the elementary quantum gates are able to act on all states of the superposition at the same time. For example, consider the NOT gate:

\[
\text{NOT} (a|0\rangle + b|1\rangle) \rightarrow a|1\rangle + b|0\rangle.
\]
It is important to note that "qubit" can have two meanings. In the first meaning, "qubit" refers to a physical quantum *system* whose quantum state can be expressed as a superposition of two basis states, which we often label |0> and |1>. Here, "qubit" refers to a physical implementation akin to what a flip-flop, switch, voltage, or current would be for a classical bit. In the second meaning, "qubit" refers to the abstract quantum *state* of a quantum system with such two basis states. In this case, the meaning of "qubit" is akin to the logical value of a bit, from classical computing, i.e. "logical 0" or "logical 1". The two concepts are related, because a physical "qubit" (first meaning) can be used to store the abstract "qubit" (second meaning). Both meanings are used interchangeably in literature and the meaning is generally clear from the context.

2.3. Multiple qubits

When multiple qubits are combined in a single quantum state the space of possible states grows exponentially and all these states can coexist in a superposition. For example, the general form of a two-qubit register is

\[ a |00> + b |01> + c |10> + d |11> \]

where the coefficients have the same probability amplitude interpretation as for the single qubit state. Each state represents a possible outcome of a measurement of the two-qubit register. For example, |01> denotes a state in which the first qubit is in the state |0> and the second is in the state |1>.

Performing single qubit gates affects the relevant qubit in each of the superposition states. Similarly, two-qubit gates also act on all the relevant superposition states, but their outcome is far more interesting.

Consider a two-qubit register where the first qubit is in the superposed state \((|0> + |1>)/\sqrt{2}\) and the other is in the state |0>. This combined state can be written as:

\[ (|0> + |1>)/\sqrt{2} \times |0> = (|00> + |10>)/\sqrt{2}, \]

where \(\times\) denotes a tensor product (the mathematical mechanism for combining quantum states together).

The constant \(1/\sqrt{2}\) is called the normalisation factor and reflects the fact that the probabilities of measuring either a |0> or a |1> for the first qubit add up to one.
Let us now consider the two-qubit controlled-NOT, or CNOT, gate. The CNOT gate takes as input two qubits, a control and target, and applies the NOT gate to the target if the control qubit is set. The truth table looks like

<table>
<thead>
<tr>
<th>IN</th>
<th>OUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>00</td>
</tr>
<tr>
<td>01</td>
<td>01</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 1

Now, consider performing a CNOT gate on the state with the first qubit being the control. We apply a two-qubit gate on all the superposition states:

\[
\text{CNOT } (|00\rangle + |10\rangle)/\sqrt{2} \rightarrow (|00\rangle + |11\rangle)/\sqrt{2}.
\]

What is so interesting about this two-qubit gate operation? The final state is *entangled*. There is no possible way of representing that quantum state as a product of two individual qubits; they are no longer independent. That is, it is not possible to describe the quantum state of either of the individual qubits in a way that is independent of the other qubit. Only the quantum state of the system that consists of both qubits provides a physically complete description of the two-qubit system. The states of the two individual qubits are now correlated beyond what is possible to achieve classically. Neither qubit is in a definite \(|0\rangle\) or \(|1\rangle\) state, but if we perform a measurement on either one, the outcome of the partner qubit will *always* yield the exact same outcome. The final state, whether it’s \(|00\rangle\) or \(|11\rangle\), is fundamentally random as before, but the states of the two qubits following a measurement will always be identical. One can think of this as flipping two coins, but the coins always both land on "heads" or both land on "tails" together. Something that we know is impossible classically.

Once a measurement is performed, the two qubits are once again independent. The final state is either \(|00\rangle\) or \(|11\rangle\) and both of these states can be trivially decomposed into a product of two individual qubits. The entanglement has been consumed and the entangled state must be prepared again.
3. Entanglement as the fundamental resource

Entanglement is the fundamental building block of quantum networks. Consider the state from the previous section:

\[(|00\rangle + |11\rangle)/\sqrt{2}\].

Neither of the two qubits is in a definite \(|0\rangle\) or \(|1\rangle\) state and we need to know the state of the entire register to be able to fully describe the behaviour of the two qubits.

Entangled qubits have interesting non-local properties. Consider sending one of the qubits to another device. This device could in principle be anywhere: on the other side of the room, in a different country, or even on a different planet. Provided negligible noise has been introduced, the two qubits will forever remain in the entangled state until a measurement is performed. The physical distance does not matter at all for entanglement.

This lies at the heart of quantum networking, because it is possible to leverage the non-classical correlations provided by entanglement in order to design completely new types of application protocols that are not possible to achieve with just classical communication. Examples of such applications are quantum cryptography [Bennett14] [Ekert91], blind quantum computation [Fitzsimons17], or distributed quantum computation [Crepeau02].

Entanglement has two very special features from which one can derive some intuition about the types of applications enabled by a quantum network.

The first stems from the fact that entanglement enables stronger than classical correlations, leading to opportunities for tasks that require coordination. As a trivial example, consider the problem of consensus between two nodes who want to agree on the value of a single bit. They can use the quantum network to prepare the state \((|00\rangle + |11\rangle)/\sqrt{2}\) with each node holding one of the two qubits. Once either of the two nodes performs a measurement, the state of the two qubits collapses to either \(|00\rangle\) or \(|11\rangle\), so whilst the outcome is random and does not exist before measurement, the two nodes will always measure the same value. We can also build the more general multi-qubit state \((|00...\rangle + |11...\rangle)/\sqrt{2}\) and perform the same algorithm between an arbitrary number of nodes. These stronger than classical correlations generalise to more complicated measurement schemes as well.
The second feature of entanglement is that it cannot be shared, in the sense that if two qubits are maximally entangled with each other, then it is physically impossible for these two qubits to also be entangled with a third qubit [Terhal04]. Hence, entanglement forms a sort of private and inherently untappable connection between two nodes once established.

Entanglement is created through local interactions between two qubits or as a product of the way the qubits were created (e.g. entangled photon pairs). To create a distributed entangled state, one can then physically send one of the qubits to a remote node. It is also possible to directly entangle qubits that are physically separated, but this still requires local interactions between some other qubits that the separated qubits are initially entangled with. Therefore, it is the transmission of qubits that draws the line between a genuine quantum network and a collection of quantum computers connected over a classical network.

A quantum network is defined as a collection of nodes that is able to exchange qubits and distribute entangled states amongst themselves. A quantum node that is able only to communicate classically with another quantum node is not a member of a quantum network.

More complex services and applications can be built on top of entangled states distributed by the network, see e.g. [ZOO]

4. Achieving quantum connectivity

This section explains the meaning of quantum connectivity and the necessary physical processes at an abstract level.

4.1. Challenges

A quantum network cannot be built by simply extrapolating all the classical models to their quantum analogues. Sending qubits over a wire like we send classical bits is simply not as easy to do. There are several technological as well as fundamental challenges that make classical approaches unsuitable in a quantum context.

4.1.1. The measurement problem

In classical computers and networks we can read out the bits stored in memory at any time. This is helpful for a variety of purposes such as copying, error detection and correction, and so on. This is not possible with qubits.
A measurement of a qubit’s state will destroy its superposition and with it any entanglement it may have been part of. Once a qubit is being processed, it cannot be read out until a suitable point in the computation, determined by the protocol handling the qubit, has been reached. Therefore, we cannot use the same methods known from classical computing for the purposes of error detection and correction. Nevertheless, quantum error detection and correction schemes exist that take this problem into account and how a network chooses to manage errors will have an impact on its architecture.

4.1.2. No-cloning theorem

Since directly reading the state of a qubit is not possible, one could ask if we can simply copy a qubit without looking at it. Unfortunately, this is fundamentally not possible in quantum mechanics [Park70] [Wootters82].

The no-cloning theorem states that it is impossible to create an identical copy of an arbitrary, unknown quantum state. Therefore, it is also impossible to use the same mechanisms that worked for classical networks for signal amplification, retransmission, and so on as they all rely on the ability to copy the underlying data. Since any physical channel will always be lossy, connecting nodes within a quantum network is a challenging endeavour and its architecture must at its core address this very issue.

4.1.3. Fidelity

In general, it is expected that a classical packet arrives at its destination without any errors introduced by hardware noise along the way. This is verified at various levels through a variety of error detection and correction mechanisms. Since we cannot read or copy a quantum state, error detection and correction is more involved.

To describe the quality of a quantum state, a physical quantity called fidelity is used [NielsenChuang]. Fidelity takes a value between 0 and 1 -- higher is better, and less than 0.5 means the state is unusable. It measures how close a quantum state is to the state we have tried to create. It expresses the probability that the state will behave exactly the same as our desired state. Fidelity is an important property of a quantum system that allows us to quantify how much a particular state has been affected by noise from various sources (gate errors, channel losses, environment noise).

Interestingly, quantum applications do not need perfect fidelity to be able to execute -- as long as the fidelity is above some application-specific threshold, they will simply operate at lower rates. Therefore, rather than trying to ensure that we always
deliver perfect states (a technologically challenging task) applications will specify a minimum threshold for the fidelity and the network will try its best to deliver it. A higher fidelity can be achieved by either having hardware produce states of better fidelity (sometimes one can sacrifice rate for higher fidelity) or by employing quantum error detection and correction mechanisms (see [Mural16] and [VanMeterBook] chapter 11).

4.1.4. Inadequacy of direct transmission

Conceptually, the most straightforward way to distribute an entangled state is to simply transmit one of the qubits directly to the other end across a series of nodes while performing sufficient forward quantum error correction (Section 4.4.3.2) to bring losses down to an acceptable level. Despite the no-cloning theorem and the inability to directly measure a quantum state, error-correcting mechanisms for quantum communication exist [Jiang09] [Fowler10] [Devitt13] [Mural16]. However, quantum error correction makes very high demands on both resources (physical qubits needed) and their initial fidelity. Implementation is very challenging and quantum error correction is not expected to be used until later generations of quantum networks are possible (see [Mural16] figure 2 and Section 4.4.3.3). Until then, quantum networks rely on entanglement swapping (Section 4.4.2) and teleportation (Section 4.3). This alternative relies on the observation that we do not need to be able to distribute any arbitrary entangled quantum state. We only need to be able to distribute any one of what are known as the Bell pair states [Briegel98].

4.2. Bell pairs

Bell pair states are the entangled two-qubit states:

\[ |00\rangle + |11\rangle, \quad |00\rangle - |11\rangle, \quad |01\rangle + |10\rangle, \quad |01\rangle - |10\rangle, \]

where the constant \(1/\sqrt{2}\) normalisation factor has been ignored for clarity. Any of the four Bell pair states above will do, as it is possible to transform any Bell pair into another Bell pair with local operations performed on only one of the qubits. When each qubit in a Bell pair is held by a separate node, either node can apply a series of single qubit gates to their qubit alone in order to transform the state between the different variants.

Distributing a Bell pair between two nodes is much easier than transmitting an arbitrary quantum state over a network. Since the state is known, handling errors becomes easier and small-scale error-correction (such as entanglement distillation discussed in a later section) combined with reattempts becomes a valid strategy.
The reason for using Bell pairs specifically as opposed to any other two-qubit state is that they are the maximally entangled two-qubit set of basis states. Maximal entanglement means that these states have the strongest non-classical correlations of all possible two-qubit states. Furthermore, since single-qubit local operations can never increase entanglement, less entangled states would impose some constraints on distributed quantum algorithms. This makes Bell pairs particularly useful as a generic building block for distributed quantum applications.

4.3. Teleportation

The observation that we only need to be able to distribute Bell pairs relies on the fact that this enables the distribution of any other arbitrary entangled state. This can be achieved via quantum state teleportation [Bennett93]. Quantum state teleportation consumes an unknown qubit state that we want to transmit and recreates it at the desired destination. This does not violate the no-cloning theorem as the original state is destroyed in the process.

To achieve this, an entangled pair needs to be distributed between the source and destination before teleportation commences. The source then entangles the transmission qubit with its end of the pair and performs a read out of the two qubits (the sum of these operations is called a Bell state measurement). This consumes the Bell pair’s entanglement, turning the source and destination qubits into independent states. The measurements yields two classical bits which the source sends to the destination over a classical channel.

Based on the value of the received two classical bits, the destination performs one of four possible corrections (called the Pauli corrections) on its end of the pair, which turns it into the unknown qubit state that we wanted to transmit. This requirement to communicate the measurement read out over a classical channel unfortunately means that entanglement cannot be used to transmit information faster than the speed of light.

The unknown quantum state that was transmitted was never fed into the network itself. Therefore, the network needs to only be able to reliably produce Bell pairs between any two nodes in the network. Thus, a key difference between a classical and quantum data planes is that a classical one carries user data, but a quantum data plane provides the resources for the user to transmit user data themselves without further involvement of the network.
4.4. The life cycle of entanglement

Reducing the problem of quantum connectivity to one of generating a Bell pair has facilitated the problem, but it has not solved it. In this section, we discuss how these entangled pairs are generated in the first place, and how their two qubits are delivered to the end-points.

4.4.1. Elementary link generation

In a quantum network, entanglement is always first generated locally (at a node or an auxiliary element) followed by a movement of one or both of the entangled qubits across the link through quantum channels. In this context, photons (particles of light) are the natural candidate for entanglement carriers, called flying qubits. The rationale for this choice is related to the advantages provided by photons such as moderate interaction with the environment leading to moderate decoherence, convenient control with standard optical components, and high-speed, low-loss transmissions. However, since photons are hard to store, a transducer must transfer the flying qubit’s state to a qubit suitable for information processing and/or storage (often referred to as a matter qubit).

Since this process may fail, in order to generate and store entanglement efficiently, we must be able to distinguish successful attempts from failures. Entanglement generation schemes that are able to announce successful generation are called heralded entanglement generation schemes.

There exist three basic schemes for heralded entanglement generation on a link through coordinated action of the two nodes at the two ends of the link [Cacciapuoti19]:

* "At mid-point": in this scheme an entangled photon pair source sitting midway between the two nodes with matter qubits sends an entangled photon through a quantum channel to each of the nodes. There, transducers are invoked to transfer the entanglement from the flying qubits to the matter qubits. In this scheme, the transducers know if the transfers succeeded and are able to herald successful entanglement generation via a message exchange over the classical channel.
"At source": in this scheme one of the two nodes sends a flying qubit that is entangled with one of its matter qubits. A transducer at the other end of the link will transfer the entanglement from the flying qubit to one of its matter qubits. Just like in the previous scheme, the transducer knows if its transfer succeeded and is able to herald successful entanglement generation with a classical message sent to the other node.

"At both end-points": in this scheme both nodes send a flying qubit that is entangled with one of their matter qubits. A detector somewhere in between the nodes performs a joint measurement on the two qubits, which stochastically projects the remote matter qubits into an entangled quantum state. The detector knows if the entanglement succeeded and is able to herald successful entanglement generation by sending a message to each node over the classical channel.

The "mid-point source" scheme is more robust to photon loss, but in the other schemes the nodes retain greater control over the entangled pair generation.

Note that whilst photons travel in a particular direction through the quantum channel the resulting entangled pair of qubits does not have a direction associated with it. Physically, there is no upstream or downstream end of the pair.

4.4.2. Entanglement swapping

The problem with generating entangled pairs directly across a link is that efficiency decreases with channel length. Beyond a few 10s of kilometres in optical fibre or 1000 kilometres in free space (via satellite) the rate is effectively zero and due to the no-cloning theorem we cannot simply amplify the signal. The solution is entanglement swapping [Briegel98].

A Bell pair between any two nodes in the network can be constructed by combining the pairs generated along each individual link on a path between the two end-points. Each node along the path can consume the two pairs on the two links that it is connected to in order to produce a new entangled pair between the two remote ends. This process is known as entanglement swapping. Pictorially it can be represented as follows:

```
+---------+      +---------+      +---------+
 |    A    |      |    B    |      |    C    |
 |         |------|         |------|         |
 |      X1˜˜˜˜˜˜˜˜˜˜X2   Y1˜˜˜˜˜˜˜˜˜˜Y2      |
 +---------+      +---------+      +---------+
```

where X1 and X2 are the qubits of the entangled pair X and Y1 and Y2 are the qubits of entangled pair Y. The entanglement is denoted with “~”. In the diagram above, nodes A and B share the pair X and nodes B and C share the pair Y, but we want entanglement between A and C.

To achieve this goal, we simply teleport the qubit X2 using the pair Y. This requires node B to perform a Bell state measurement on the qubits X2 and Y1 which result in the destruction of the entanglement between Y1 and Y2. However, X2 is recreated in Y2’s place, carrying with it its entanglement with X1. The end-result is shown below:

Depending on the needs of the network and/or application, a final Pauli correction at the recipient node may not be necessary since the result of this operation is also a Bell pair. However, the two classical bits that form the read out from the measurement at node B must still be communicated, because they carry information about which of the four Bell pairs was actually produced. If a correction is not performed, the recipient must be informed which Bell pair was received.

This process of teleporting Bell pairs using other entangled pairs is called entanglement swapping. Quantum nodes that create long-distance entangled pairs via entanglement swapping are called quantum repeaters in academic literature [Briegel98] and we will use the same terminology in this draft.

4.4.3. Error Management

4.4.3.1. Distillation

Neither the generation of Bell pairs nor the swapping operations are noiseless operations. Therefore, with each link and each swap the fidelity of the state degrades. However, it is possible to create higher fidelity Bell pair states from two or more lower fidelity pairs through a process called distillation (sometimes also referred to as purification) [Dur07].

To distil a quantum state, a second (and sometimes third) quantum state is used as a "test tool" to test a proposition about the first state, e.g., "the parity of the two qubits in the first state is even." When the test succeeds, confidence in the state is improved, and thus the fidelity is improved. The test tool states are
destroyed in the process, so resource demands increase substantially when distillation is used. When the test fails, the tested state must also be discarded. Distillation makes low demands on fidelity and resources compared to quantum error correction, but distributed protocols incur round-trip delays due to classical communication [Bennett96].

4.4.3.2. Quantum Error Correction

Just like classical error correction, quantum error correction (QEC) encodes logical qubits using several physical (raw) qubits to protect them from errors described in Section 4.1.3 [Jiang09] [Fowler10] [Devitt13] [Mural16]. Furthermore, similarly to its classical counterpart, QEC can not only correct state errors but also account for lost qubits. Additionally, if all physical qubits which encode a logical qubit are located at the same node, the correction procedure can be executed locally, even if the logical qubit is entangled with remote qubits.

Although QEC was originally a scheme proposed to protect a qubit from noise, QEC can also be applied to entanglement distillation. Such QEC-applied distillation is cost-effective but requires a higher base fidelity.

4.4.3.3. Error management schemes

Quantum networks have been categorized into three "generations" based on the error management scheme they employ [Mural16]. Note that these "generations" are more like categories; they do not necessarily imply a time progression and do not obsolete each other, though the later generations do require more advanced technologies. Which generation is used depends on the hardware platform and network design choices.

Table 2 summarises the generations.
<table>
<thead>
<tr>
<th>Loss tolerance</th>
<th>First generation</th>
<th>Second generation</th>
<th>Third generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heralded entanglement</td>
<td>Heralded entanglement (bi-directional classical signaling)</td>
<td>Heralded entanglement generation (bi-directional classical signaling)</td>
<td>Quantum Error Correction (no classical signaling)</td>
</tr>
</tbody>
</table>

| Error tolerance             | Entanglement distillation (bi-directional classical signaling) | Entanglement distillation (uni-directional classical signaling) or Quantum Error Correction (no classical signaling) | Quantum Error Correction (no classical signaling) |

Table 2: Classical signaling and generations

Generations are defined by the directions of classical signalling required in their distributed protocols for loss tolerance and error tolerance. Classical signalling carries the classical bits and incurs round-trip delays described in Section 4.4.3.1, hence they affect the performance of quantum networks, especially as the distance between the communicating nodes increases.

Loss tolerance is about tolerating qubit transmission losses between nodes. Heralded entanglement generation, as described in Section 4.4.1, confirms the receipt of an entangled qubit using a heralding signal. A pair of directly connected quantum nodes repeatedly attempt to generate an entangled pair until the heralding signal is received. As described in Section 4.4.3.2, QEC can be applied to complement lost qubits eliminating the need for re-attempts. Furthermore, since the correction procedure is composed of local operations, it does not require a heralding signal. However, it is possible only when the photon loss rate from transmission to measurement is less than 50%.
Error tolerance is about tolerating quantum state errors. Entanglement distillation is the easiest mechanism for improved error tolerance to implement, but it incurs round-trip delays due to the requirement for bi-directional classical signalling. The alternative, QEC, is able to correct state errors locally so that it does not need any classical signalling between the quantum nodes. In between these two extremes, there is also QEC-applied distillation, which requires uni-directional classical signalling.

The three "generations" summarised:

1. First generation quantum networks use heralding for loss tolerance and entanglement distillation for error tolerance. These networks can be implemented even with a limited set of available quantum gates.

2. Second generation quantum networks improve upon the first generation with QEC codes for error tolerance (but not loss tolerance). At first, QEC will be applied to entanglement distillation only which requires uni-directional classical signalling. Later, QEC codes will be used to create logical Bell pairs which no longer require any classical signalling for the purposes of error tolerance. Heralding is still used to compensate for transmission losses.

3. Third generation quantum networks directly transmit QEC encoded qubits to adjacent nodes, as discussed in Section 4.1.4. Elementary link Bell pairs can now be created without heralding or any other classical signalling. Furthermore, this also enables direct transmission architectures in which qubits are forwarded end-to-end like classical packets rather than relying on Bell pairs and entanglement swapping.

Despite the fact that there are important distinctions in how errors will be managed in the different generations it is unlikely that all quantum networks will consistently use the same method. This is due to different hardware requirements of the different generations and the practical reality of network upgrades. Therefore, it is unavoidable that eventually boundaries between different error management schemes start forming. This will affect the content and semantics of messages that must cross those boundaries -- both for connection setup and real-time operation [Nagayama16].
4.4.4. Delivery

Eventually, the Bell pairs must be delivered to an application (or higher layer protocol) at the two end-nodes. A detailed list of such requirements is beyond the scope of this draft. At minimum, the end-nodes require information to map a particular Bell pair to the qubit in their local memory that is part of this entangled pair.

5. Architecture of a quantum internet

It is evident from the previous sections that the fundamental service provided by a quantum network significantly differs from that of a classical network. Therefore, it is not surprising that the architecture of a quantum internet will itself be very different from that of the classical Internet.

5.1. Challenges

This subsection covers the major fundamental challenges building quantum networks. Here, we only describe the fundamental differences. Technological limitations are described later.

1. Bell pairs are not equivalent to payload carrying packets.

In most classical networks, including Ethernet, Internet Protocol (IP), and Multi-Protocol Label Switching (MPLS) networks, user data is grouped into packets. In addition to the user data, each packet also contains a series of headers which contain the control information that lets routers and switches forward it towards its destination. Packets are the fundamental unit in a classical network.
In a quantum network, the entangled pairs of qubits are the basic unit of networking. These qubits themselves do not carry any headers. Therefore, quantum networks will have to send all control information via separate classical channels which the repeaters will have to correlate with the qubits stored in their memory. Furthermore, a Bell pair consists of two qubits distributed across two nodes which is unlike a classical packet which is located at a single node. This has a fundamental impact on how quantum networks will be managed and how protocols need to be designed. To make long-distance Bell pairs, the nodes may have to keep their qubits in their quantum memories and wait until control information is exchanged before proceeding with the next operation. This signalling will result in additional latency which will depend on the distance between the nodes holding the two ends of the Bell pair. Error management, such as entanglement distillation, is a typical example of such control information exchange [Nagayama21] (see also Section 4.4.3.3).

2. "Store and forward" vs "store and swap" quantum networks.

As described in Section 4.4.1, quantum links provide Bell pairs that are undirected network resources, in contrast to directed frames of classical networks. This phenomenological distinction leads to architectural differences between quantum networks and classical networks. Quantum networks combine multiple elementary link Bell pairs together to create one end-to-end Bell pair, whereas classical networks deliver messages from one end to the other end hop by hop.

Classical networks receive data on one interface, store it in local buffers, then forward the data to another appropriate interface. Quantum networks store Bell pairs and then execute entanglement swapping instead of forwarding in the data plane. Such quantum networks are "store and swap" networks. In "store and swap" networks, we do not need to care about the order in which the Bell pairs were generated since they are undirected. However, whilst the ordering does not matter, it is very important that the right entangled pairs get swapped, and that the intermediate measurement outcomes (see Section 4.4.2) are signalled to and correlated with the correct qubits at the other nodes. Otherwise, the final end-to-end entangled pair will not be created between the expected end-points or will be in a different quantum state than expected. For example, rather than Alice receiving a qubit that is entangled with Bob’s qubit, her qubit is entangled with Charlie’s qubit. This distinction makes control algorithms and optimisation of quantum networks different from classical ones, in the sense that swapping is stateful in contrast to stateless packet-by-packet forwarding. Note that
third generation quantum networks, as described in Section 4.4.1, will be able to support a "store and forward" architecture in addition to "store and swap".

3. An entangled pair is only useful if the locations of both qubits are known.

A classical network packet logically exists only at one location at any point in time. If a packet is modified in some way, whether headers or payload, this information does not need to be conveyed to anybody else in the network. The packet can be simply forwarded as before.

In contrast, entanglement is a phenomenon in which two or more qubits exist in a physically distributed state. Operations on one of the qubits change the mutual state of the pair. Since the owner of a particular qubit cannot just read out its state, it must coordinate all its actions with the owner of the pair’s other qubit. Therefore, the owner of any qubit that is part of an entangled pair must know the location of its counterpart. Location, in this context, need not be the explicit spatial location. A relevant pair identifier, a means of communication between the pair owners, and an association between the pair ID and the individual qubits is sufficient.

4. Generating entanglement requires temporary state.

Packet forwarding in a classical network is largely a stateless operation. When a packet is received, the router does a lookup in its forwarding table and sends the packet out of the appropriate output. There is no need to keep any memory of the packet any more.

A quantum node must be able to make decisions about qubits that it receives and is holding in its memory. Since qubits do not carry headers, the receipt of an entangled pair conveys no control information based on which the repeater can make a decision. The relevant control information will arrive separately over a classical channel. This implies that a repeater must store temporary state as the control information and the qubit it pertains to will, in general, not arrive at the same time.

5.2. Classical communication

In this draft we have already covered two different roles that classical communication must perform:
* communicate classical bits of information as part of distributed protocols such as entanglement swapping and teleportation,
* communicate control information within a network, including both background protocols such as routing as well as signalling protocols to set up end-to-end entanglement generation.

Classical communication is a crucial building block of any quantum network. All nodes in a quantum network are assumed to have classical connectivity with each other (within typical administrative domain limits). Therefore, quantum nodes will need to manage two data planes in parallel, a classical one and a quantum one. Additionally, a node must be able to correlate information between the two planes so that the control information received on a classical channel can be applied to the qubits managed by the quantum data plane.

5.3. Abstract model of the network

5.3.1. The control and data planes

Control plane protocols for quantum networks will have many responsibilities similar to their classical counterparts, namely discovering the network topology, resource management, populating data plane tables, etc. Most of these protocols do not require the manipulation of quantum data and can operate simply by exchanging classical messages only. There may also be some control plane functionality that does require the handling of quantum data, e.g. a quantum ping [I-D.irtf-qirg-quantum-internet-use-cases]. As it is not clear if there is much benefit in defining a separate quantum control plane given the significant overlap in responsibilities with its classical counterpart, the question of whether there should be a separate quantum control plane is beyond the scope of this document.

However, the data plane separation is much more distinct and there will be two data planes: a classical data plane and a quantum data plane. The classical data plane processes and forwards classical packets. The quantum data plane processes and swaps entangled pairs. Third generation quantum networks may also forward qubits in addition to swapping Bell pairs.
In addition to control plane messages, there will also be control information messages that operate at the granularity of individual entangled pairs, such as heralding messages used for elementary link generation (Section 4.4.1). In terms of functionality, these messages are closer to classical packet headers than control plane messages and thus we consider them to be part of the quantum data plane. Therefore, a quantum data plane also includes the exchange of classical control information at the granularity of individual qubits and entangled pairs.

5.3.2. Elements of a quantum network

We have identified quantum repeaters as the core building block of a quantum network. However, a quantum repeater will have to do more than just entanglement swapping in a functional quantum network. Its key responsibilities will include:

1. Creating link-local entanglement between neighbouring nodes.
2. Extending entanglement from link-local pairs to long-range pairs through entanglement swapping.
3. Performing distillation to manage the fidelity of the produced pairs.
4. Participating in the management of the network (routing, etc.).

Not all quantum repeaters in the network will be the same; here we break them down further:

* Quantum routers (controllable quantum nodes) - A quantum router is a quantum repeater with a control plane that participates in the management of the network and will make decisions about which qubits to swap to generate the requested end-to-end pairs.

* Automated quantum nodes - An automated quantum node is a data plane only quantum repeater that does not participate in the network control plane. Since the no-cloning theorem precludes the use of amplification, long-range links will be established by chaining multiple such automated nodes together.

* End-nodes - End-nodes in a quantum network must be able to receive and handle an entangled pair, but they do not need to be able to perform an entanglement swap (and thus are not necessarily quantum repeaters). End-nodes are also not required to have any quantum memory as certain quantum applications can be realised by having the end-node measure its qubit as soon as it is received.
* Non-quantum nodes - Not all nodes in a quantum network need to have a quantum data plane. A non-quantum node is any device that can handle classical network traffic.

Additionally, we need to identify two kinds of links that will be used in a quantum network:

* Quantum links - A quantum link is a link which can be used to generate an entangled pair between two directly connected quantum repeaters. This may include additional mid-point elements described in Section 4.4.1. It may also include a dedicated classical channel that is to be used solely for the purpose of coordinating the entanglement generation on this quantum link.

* Classical links - A classical link is a link between any node in the network that is capable of carrying classical network traffic.

Note that passive elements, such as optical switches, do not destroy the quantum state. Therefore, it is possible to connect multiple quantum nodes with each other over an optical network and perform optical switching rather than routing via entanglement swapping at quantum routers. This does require coordination with the elementary link entanglement generation process and it still requires repeaters to overcome the short-distance limitations. However, this is a potentially feasible architecture for local area networks.

5.3.3. Putting it all together

A two-hop path in a generic quantum network can be represented as:

```
+-----+                                        +-----+
| App |- - - - - - - - - -CC- - - - - - - - | App |
+-----+                +------+                +-----+
| EN  |------ CL ------|  QR  |------ CL ------| EN  |
|     |------ QL ------|      |------ QL ------|     |
+-----+                +------+                +-----+
```

App - user-level application  
EN - end-node  
QL - quantum link  
CL - classical link  
CC - classical channel (traverses one or more CLs)  
QR - quantum repeater

An application (App) running on two end-nodes (ENs) attached to a network will at some point need the network to generate entangled pairs for its use. This may require negotiation between the end-nodes (possibly ahead of time), because they must both open a
communication end-point which the network can use to identify the two ends of the connection. The two end-nodes use a classical channel (CC) available in the network to achieve this goal.

When the network receives a request to generate end-to-end entangled pairs it uses the classical communication links (CLs) to coordinate and claim the resources necessary to fulfill this request. This may be some combination of prior control information (e.g. routing tables) and signalling protocols, but the details of how this is achieved are an active research question. A thought experiment on what this might look like be can be found later in this draft in Section 7

During or after the distribution of control information, the network performs the necessary quantum operations such as generating entanglement over individual quantum links (QLs), performing entanglement swaps at quantum repeaters (QRs), and further signalling to transmit the swap outcomes and other control information. Since Bell pairs do not carry any user data, some of these operations can be performed before the request is received in anticipation of the demand.

The entangled pair is delivered to the application once it is ready, together with the relevant pair identifier. However, being ready does not necessarily mean that all link pairs and entanglement swaps are complete, as some applications can start executing on an incomplete pair. In this case the remaining entanglement swaps will propagate the actions across the network to the other end, sometimes necessitating fixup operations at the end node.

5.4. Physical constraints

The model above has effectively abstracted away the particulars of the hardware implementation. However, certain physical constraints need to be considered in order to build a practical network. Some of these are fundamental constraints and no matter how much the technology improves, they will always need to be addressed. Others are artifacts of the early stages of a new technology. Here, we consider a highly abstract scenario and refer to [Wehner18] for pointers to the physics literature.
5.4.1. Memory lifetimes

In addition to discrete operations being imperfect, storing a qubit in memory is also highly non-trivial. The main difficulty in achieving persistent storage is that it is extremely challenging to isolate a quantum system from the environment. The environment introduces an uncontrollable source of noise into the system which affects the fidelity of the state. This process is known as decoherence. Eventually, the state has to be discarded once its fidelity degrades too much.

The memory lifetime depends on the particular physical setup, but the highest achievable values in quantum network hardware currently are on the order of seconds [Abobeih18] although a lifetime of a minute has also been demonstrated for qubits not connected to a quantum network [Bradley19] (as of 2020). These values have increased tremendously over the lifetime of the different technologies and are bound to keep increasing. However, if quantum networks are to be realised in the near future, they need to be able to handle short memory lifetimes, for example by reducing latency on critical paths.

5.4.2. Rates

Entanglement generation on a link between two connected nodes is not a very efficient process and it requires many attempts to succeed [Hensen15] [Dahlberg19]. For example, the highest achievable rates of success between nitrogen-vacancy center nodes, which in addition to entanglement generation are also capable of storing and processing the resulting qubits, are on the order of 10 Hz. Combined with short memory lifetimes this leads to very tight timing windows to build up network-wide connectivity.

Other platforms have shown higher entanglement rates, but this usually comes at the cost of other hardware capabilities, such as no quantum memory and/or limited processing capabilities [Wei22]. Nevertheless, the current rates are not sufficient for practical applications beyond simple experimental proofs of concept. However, they are expected to improve over time as quantum network technology evolves [Wei22].

5.4.3. Communication qubits

Most physical architectures capable of storing qubits are only able to generate entanglement using only a subset of available qubits called communication qubits [Dahlberg19]. Once a Bell pair has been generated using a communication qubit, its state can be transferred into memory. This may impose additional limitations on the network. In particular, if a given node has only one communication qubit it
cannot simultaneously generate Bell pairs over two links. It must generate entanglement over the links one at a time.

5.4.4. Homogeneity

Currently all existing quantum network implementations are homogeneous and they do not interface with each other. In general, it is very challenging to combine different quantum information processing technologies.

There are many different physical hardware platforms for implementing quantum networking hardware. The different technologies differ in how they store and manipulate qubits in memory and how they generate entanglement across a link with their neighbours. For example, hardware based on optical elements and atomic ensembles [Sangouard11] is very efficient at generating entanglement at high rates, but provides limited processing capabilities once the entanglement is generated. On the other hand, nitrogen-vacancy based [Hensen15] or trapped ion [Mohring07] platforms offer a much greater degree of control over the qubits, but have a harder time generating entanglement at high rates.

In order to overcome the weaknesses of the different platforms, coupling the different technologies will help to build fully functional networks. For example, end-nodes may be implemented using technology with good qubit processing capabilities to enable complex applications, but automated quantum nodes that that serve only to "repeat" along a linear chain, where the processing logic is much simpler, can be implemented with technologies that sacrifice processing capabilities for higher entanglement rates at long distances [Askarani21].

This point is further exacerbated by the fact that quantum computers (i.e. end-nodes in a quantum network) are often based on different hardware platforms than quantum repeaters thus requiring a coupling (transduction) between the two. This is especially true for quantum computers based on superconducting technology which are challenging to connect to optical networks. However, even trapped ion quantum computers, which is a platform that has shown promise for quantum networking, will still need to connect to other platforms that are better at creating entanglement at high rates over long distances (hundreds of kms).
6. Architectural principles

Given that the most practical way of realising quantum network connectivity is using Bell pair and entanglement swapping repeater technology, what sort of principles should guide us in assembling such networks such that they are functional, robust, efficient, and most importantly, do they work? Furthermore, how do we design networks so that they work under the constraints imposed by the hardware available today, but do not impose unnecessary burdens on future technology?

As quantum networking is a completely new technology that is likely to see many iterations over its lifetime, this draft must not serve as a definitive set of rules, but merely as a general set of recommended guidelines for the first generations of quantum networks based on principles and observations made by the community. The benefit of having a community built document at this early stage is that expertise in both quantum information and network architecture is needed in order to successfully build a quantum internet.

6.1. Goals of a quantum internet

When outlining any set of principles we must ask ourselves what goals do we want to achieve as inevitably trade-offs must be made. So what sort of goals should drive a quantum network architecture? The following list has been inspired by the history of computer networking and thus it is inevitably very similar to one that could be produced for the classical Internet [Clark88]. However, whilst the goals may be similar the challenges involved are often fundamentally different. The list will also most likely evolve with time and the needs of its users.

1. Support distributed quantum applications

This goal seems trivially obvious, but makes a subtle, but important point which highlights a key difference between quantum and classical networks. Ultimately, quantum data transmission is not the goal of a quantum network – it is only one possible component of more advanced quantum application protocols [Wehner18]. Whilst transmission certainly could be used as a building block for all quantum applications, it is not the most basic one possible. For example, entanglement-based QKD, the most well known quantum application protocol, only relies on the stronger-than-classical correlations and inherent secrecy of entangled Bell pairs and does not have to transmit arbitrary quantum states [Ekert91].
The primary purpose of a quantum internet is to support distributed quantum application protocols and it is of utmost importance that they can run well and efficiently. Thus, it is important to develop performance metrics meaningful to application to drive the development of quantum network protocols. For example, the Bell pair generation rate is meaningless if one does not also consider their fidelity. It is generally much easier to generate pairs of lower fidelity, but quantum applications may have to make multiple re-attempts or even abort if the fidelity is too low. A review of the requirements for different known quantum applications can be found in [Wehner18] and an overview of use-cases can be found in [I-D.irtf-qirg-quantum-internet-use-cases].

2. Support tomorrow’s distributed quantum applications

The only principle of the Internet that should survive indefinitely is the principle of constant change [RFC1958]. Technical change is continuous and the size and capabilities of the quantum internet will change by orders of magnitude. Therefore, it is an explicit goal that a quantum internet architecture be able to embrace this change. We have the benefit of having been witness to the evolution of the classical Internet over several decades and seen what worked and what did not. It is vital for a quantum internet to avoid the need for flag days (e.g. NCP to TCP/IP) or upgrades that take decades to roll out (e.g. IPv4 to IPv6).

Therefore, it is important that any proposed architecture for general purpose quantum repeater networks can integrate new devices and solutions as they become available. The architecture should not be constrained due to considerations for early-stage hardware and applications. For example, it is already possible to run QKD efficiently on metropolitan scales and such networks are already commercially available. However, they are not based on quantum repeaters and thus will not be able to easily transition to more sophisticated applications.

3. Support heterogeneity

There are multiple proposals for realising practical quantum repeater hardware and they all have their advantages and disadvantages. Some may offer higher Bell pair generation rates on individual links at the cost of more difficult entanglement swap operations. Other platforms may be good all around, but are more difficult to build.
In addition to physical boundaries, there may be distinctions in how errors are managed (Section 4.4.3.3). These differences will affect the content and semantics of messages that cross these boundaries -- both for connection setup and real-time operation.

The optimal network configuration will likely leverage the advantages of multiple platforms to optimise the provided service. Therefore, it is an explicit goal to incorporate varied hardware and technology support from the beginning.

4. Ensure security at the network level

The question of security in quantum networks is just as critical as it is in the classical Internet, especially since enhanced security offered by quantum entanglement is one of the key driving factors.

Fortunately, from an application’s point of view, as long as the underlying implementation corresponds to (or sufficiently approximates) theoretical models of quantum cryptography, quantum cryptographic protocols do not need the network to provide any guarantees about the confidentiality or integrity of the transmitted qubits or the generated entanglement (though they may impose requirements on the classical channel, e.g. to be authenticated [Wang21]). Instead, applications will leverage the classical networks to establish the end-to-end security of the results obtained from the processing of entangled qubits. However, it is important to note that whilst classical networks are necessary to establish these end-to-end guarantees, the security relies on the properties of quantum entanglement. For example, QKD uses classical information reconciliation [Tang19] for error correction and privacy amplification [Elkouss11] for generating the final secure key, but the raw bits that are fed into these protocols must come from measuring entangled qubits [Ekert91]. In another application, secure delegated quantum computing, the client hides its computation from the server by sending qubits to the server and then requesting it (in a classical message) to measure them in an encoded basis. The client then decodes the results it receives from the server to obtain the result of the computation [Broadbent10]. Once again, whilst a classical network is used to achieve the goal of secure computation, the remote computation is strictly quantum.

Nevertheless, whilst applications can ensure their own end-to-end security, network protocols themselves should be security aware in order to protect the network itself and limit disruption. Whilst the applications remain secure they are not necessarily operational or as efficient in the presence of an attacker. For
example, if an attacker can measure every qubit between two
parties trying to establish a key using QKD, no secret key can be
generated. Security concerns in quantum networks are described
in more detail in [Satoh17] [Satoh20].

5. Make them easy to monitor

In order to manage, evaluate the performance of, or debug a
network it is necessary to have the ability to monitor the
network while ensuring there will be mechanisms in place to
protect the confidentiality and integrity of the devices
connected to it. Quantum networks bring new challenges in this
area so it should be a goal of a quantum network architecture to
make this task easy.

The fundamental unit of quantum information, the qubit, cannot be
actively monitored as any readout irreversibly destroys its
contents. One of the implications of this fact is that measuring
an individual pair’s fidelity is impossible. Fidelity is
meaningful only as a statistical quantity which requires the
constant monitoring and the sacrifice of generated Bell pairs for
tomography or other methods.

Furthermore, given one end of an entangled pair, it is impossible
to tell where the other qubit is without any additional classical
metadata. It is impossible to extract this information from the
qubits themselves. This implies that tracking entangled pairs
necessitates some exchange of classical information. This
information might include (i) a reference to the entangled pair
that allows distributed applications to coordinate actions on
qubits of the same pair, and (ii) the two bits from each
entanglement swap necessary to identify the final state of the
Bell pair (Section 4.4.2).

6. Ensure availability and resilience

Any practical and usable network, classical or quantum, must be
able to continue to operate despite losses and failures, and be
robust to malicious actors trying to disable connectivity. What
differs in quantum networks as compared to classical networks in
this regard is that we now have two data planes and two types of
channels to worry about: a quantum and a classical one.
Therefore, availability and resilience will most likely require a
more advanced treatment than they do in classical networks.
6.2. The principles of a quantum internet

The principles support the goals, but are not goals themselves. The goals define what we want to build and the principles provide a guideline in how we might achieve this. The goals will also be the foundation for defining any metric of success for a network architecture, whereas the principles in themselves do not distinguish between success and failure. For more information about design considerations for quantum networks see [VanMeter13.1] [Dahlberg19].

1. Entanglement is the fundamental service

The key service that a quantum network provides is the distribution of entanglement between the nodes in a network. All distributed quantum applications are built on top of this key resource. Applications such as clustered quantum computing, distributed quantum computing, distributed quantum sensing networks, and certain kinds of quantum secure networks all consume quantum entanglement as a resource. Some applications (e.g. quantum key distribution) simply measure the entangled qubits to obtain a shared secret key [QKD]. Other applications (e.g. distributed quantum computing) build more complex abstractions and operations on the entangled qubits, e.g., distributed CNOT gates [DistCNOT] or teleportation of arbitrary qubit states [Teleportation].

A quantum network may also distribute multipartite entangled states (entangled states of three or more qubits) [Meignant19] which are useful for applications such as conference key agreement [Murta20], distributed quantum computing [Cirac99], secret sharing [Qin17], and clock synchronisation [Komar14]. Though it was worth noting that multipartite entangled states can also be constructed from multiple entangled pairs distributed between the end-nodes.

2. Bell Pairs are indistinguishable

Any two Bell Pairs between the same two nodes are indistinguishable for the purposes of an application provided they both satisfy its required fidelity threshold. This observation is likely to be key in enabling a more optimal allocation of resources in a network, e.g. for the purposes of provisioning resources to meet application demand. However, the qubits that make up the pair themselves are not indistinguishable and the two nodes operating on a pair must coordinate to make sure they are operating on qubits that belong to the same Bell pair.
3. Fidelity is part of the service

In addition to being able to deliver Bell pairs to the communication end-points, the Bell Pairs must be of sufficient fidelity. Unlike in classical networks where most errors are effectively eliminated before reaching the application, many quantum applications only need imperfect entanglement to function. However, quantum applications will generally have a threshold for Bell pair fidelity below which they are no longer able to operate. Different applications will have different requirements for what fidelity they can work with. It is the network’s responsibility to balance the resource usage with respect to the applications’ requirements. It may be that it is cheaper for the network to provide lower fidelity pairs that are just above the threshold required by the application than it is to guarantee high fidelity pairs to all applications regardless of their requirements.

4. Time is an expensive resource

Time is not the only resource that is in short supply (memory, and communication qubits are as well), but ultimately it is the lifetime of quantum memories that imposes some of the most difficult conditions for operating an extended network of quantum nodes. Current hardware has low rates of Bell pair generation, short memory lifetimes, and access to a limited number of communication qubits. All these factors combined mean that even a short waiting queue at some node could be enough for a Bell pair to decohere or result in an end-to-end pair below an application’s fidelity threshold. Therefore, managing the idle time of qubits holding live quantum states should be done carefully. Ideally by minimising the idle time, but potentially also by moving the quantum state for temporary storage to a quantum memory with a longer lifetime.

5. Be flexible with regards to capabilities and limitations

This goal encompasses two important points. First, the architecture should be able to function under the physical constraints imposed by the current generation hardware. Near-future hardware will have low entanglement generation rates, quantum memories able to hold a handful of qubits at best, and decoherence rates that will render many generated pairs unusable.

Second, the architecture should not make it difficult to run the network over any hardware that may come along in the future. The physical capabilities of repeaters will improve and redepolying a technology is extremely challenging.
7. A thought experiment inspired by classical networks

To conclude, we discuss a plausible quantum network architecture inspired by MPLS. This is not an architecture proposal, but rather a thought experiment to give the reader an idea of what components are necessary for a functional quantum network. We use classical MPLS as a basis as it is well known and understood in the networking community.

Creating end-to-end Bell pairs between remote end-points is a stateful distributed task that requires a lot of a-priori coordination. Therefore, a connection-oriented approach seems the most natural for quantum networks. In connection-oriented quantum networks, when two quantum application end-points wish to start creating end-to-end Bell pairs, they must first create a quantum virtual circuit (QVC). As an analogy, in MPLS networks end-points must establish a label switched path (LSP) before exchanging traffic. Connection-oriented quantum networks may also support virtual circuits with multiple end-points for creating multipartite entanglement. As an analogy, MPLS networks have the concept of multi-point LSPs for multicast.

When a quantum application creates a quantum virtual circuit, it can indicate quality of service (QoS) parameters such as the required capacity in end-to-end Bell pairs per second (BPPS) and the required fidelity of the Bell pairs. As an analogy, in MPLS networks applications specify the required bandwidth in bits per second (BPS) and other constraints when they create a new LSP.

Different applications will have different QoS requirements. For example, applications such as QKD, that don’t need to process the entangled qubits and only need measure them and store the resulting outcome, may require a large volume of entanglement, but will be tolerant of delay and jitter for individual pairs. On the other hand, distributed/cloud quantum computing applications may need fewer entangled pairs, but instead, may need all of them to be generated in one go so that they can be processed all together before any of them decohere.
Quantum networks need a routing function to compute the optimal path (i.e., the best sequence of routers and links) for each new quantum virtual circuit. The routing function may be centralized or distributed. In the latter case, the quantum network needs a distributed routing protocol. As an analogy, classical networks use routing protocols such as open shortest path first (OSPF) and intermediate-system to intermediate system (IS-IS). However, note that the definition of "shortest-path"/"least-cost" may be different in a quantum network to account for its non-classical features, such as fidelity [VanMeter13.2].

Given the very scarce availability of resources in early quantum networks, a traffic engineering function is likely to be beneficial. Without traffic engineering, quantum virtual circuits always use the shortest path. In this case, the quantum network cannot guarantee that each quantum end-point will get its Bell pairs at the required rate or fidelity. This is analogous to "best effort" service in classical networks.

With traffic engineering, quantum virtual circuits choose a path that is guaranteed to have the requested resources (e.g., bandwidth in BPPS) available, taking into account the capacity of the routers and links and taking into account the resources already consumed by other virtual circuits. As an analogy, both OSPF and IS-IS have traffic engineering (TE) extensions to keep track of used and available resources, and can use constrained shortest path first (CSPF) to take resource availability and other constraints into account when computing the optimal path.

The use of traffic engineering implies the use of call admission control (CAC): the network denies any virtual circuits for which it cannot guarantee the requested quality of service a-priori. Or alternatively, the network pre-empts lower priority circuits to make room for the new one.

Quantum networks need a signaling function: once the path for a quantum virtual circuit has been computed, signaling is used to install the "forwarding rules" into the data plane of each quantum router on the path. The signaling may be distributed, analogous to the resource reservation protocol (RSVP) in MPLS. Or the signaling may be centralized, similar to OpenFlow.

Quantum networks need an abstraction of the hardware for specifying the forwarding rules. This allows us to de-couple the control plane (routing and signaling) from the data plane (actual creation of Bell pairs). The forwarding rules are specified using abstract building blocks such as "creating local Bell pairs", "swapping Bell pairs", "distillation of Bell pairs". As an analogy, classical networks use
abstractions that are based on match conditions (e.g. looking up header fields in tables) and actions (e.g. modifying fields or forwarding a packet to a specific interface). The data-plane abstractions in quantum networks will be very different from those in classical networks due to the fundamental differences in technology and the stateful nature of quantum networks. In fact, choosing the right abstractions will be one of the biggest challenges when designing interoperable quantum network protocols.

In quantum networks, control plane traffic (routing and signaling messages) is exchanged over a classical channel, whereas data plane traffic (the actual Bell pair qubits) is exchanged over a separate quantum channel. This is in contrast to most classical networks, where control plane traffic and data plane traffic share the same channel and where a single packet contains both user fields and header fields. There is, however, a classical analogy to the way quantum networks work. Generalized MPLS (GMPLS) networks use separate channels for control plane traffic and data plane traffic. Furthermore, GMPLS networks support data planes where there is no such thing as data plane headers (e.g. DWDM or TDM networks).

8. Security Considerations

Security is listed as an explicit goal for the architecture and this issue is addressed in the section on goals. However, as this is an informational draft it does not propose any concrete mechanisms to achieve these goals.

9. IANA Considerations

This draft includes no request to IANA.

10. Acknowledgements

The authors want to thank Carlo Delle Donne, Matthew Skrzypczyk, Axel Dahlberg, Mathias van den Bossche, Patrick Gelard, Chonggang Wang, Scott Fluhrer, Joey Salazar, Joseph Touch, and the rest of the QIRG community as a whole for their very useful reviews and comments to the document.

11. Informative References


[Bradley19]

[Briegel98]

[Broadbent10]

[Cacciapuoti19]

[Cirac99]

[Clark88]

[Crepeau02]

[Dahlberg19]
Dahlberg, A., Skrzypczyk, M., Coopmans, T., Wubben, L., Rozpedek, F., Pompili, M., Stolk, A., Pawelczak, P.,


Authors’ Addresses

Wojciech Kozlowski
QuTech
Building 22
Lorentzweg 1
2628 CJ Delft
Netherlands

Email: w.kozlowski@tudelft.nl
Stephanie Wehner
QuTech
Building 22
Lorentzweg 1
2628 CJ Delft
Netherlands
Email: s.d.c.wehner@tudelft.nl

Rodney Van Meter
Keio University
5322 Endo, Kanagawa
252-0882
Japan
Email: rdv@sfc.wide.ad.jp

Bruno Rijksman
Individual
Email: brunorijsman@gmail.com

Angela Sara Cacciapuoti
University of Naples Federico II
Department of Electrical Engineering and Information Technologies
Claudio 21
80125 Naples
Italy
Email: angelasara.cacciapuoti@unina.it

Marcello Caleffi
University of Naples Federico II
Department of Electrical Engineering and Information Technologies
Claudio 21
80125 Naples
Italy
Email: marcello.caleffi@unina.it

Shota Nagayama
Mercari, Inc.
Roppongi Hills Mori Tower 18F
6-10-1 Roppongi, Minato-ku,
106-6118
Japan

Email: shota.nagayama@mercari.com
Abstract

The Quantum Internet has the potential to improve application functionality by incorporating quantum information technology into the infrastructure of the overall Internet. This document provides an overview of some applications expected to be used on the Quantum Internet and categorizes them. Some general requirements for the Quantum Internet are also discussed. The intent of this document is to describe a framework for applications, and describe a few selected application scenarios for the Quantum Internet.
1. Introduction

The Classical Internet has been constantly growing since it first became commercially popular in the early 1990's. It essentially consists of a large number of end nodes (e.g., laptops, smart phones, network servers) connected by routers and clustered in Autonomous Systems. The end nodes may run applications that provide service for the end users such as processing and transmission of voice, video or data. The connections between the various nodes in the Internet include backbone links (e.g., fiber optics) and access links (e.g., fiber optics, WiFi, cellular wireless, Digital Subscriber Lines (DSLs)). Bits are transmitted across the Classical Internet in packets.
Research and experiments have picked up over the last few years for developing the Quantum Internet [Wehner]. End nodes will also be part of the Quantum Internet, in that case called quantum end nodes that may be connected by quantum repeaters/routers. These quantum end nodes will also run value-added applications which will be discussed later.

The physical layer quantum channels between the various nodes in the Quantum Internet can be either waveguides such as optical fibers or free space. Photonic channels are particularly useful because light (photons) is very suitable for physically realizing qubits. The Quantum Internet will operate according to quantum physical principles such as quantum superposition and entanglement [I-D.irtf-qirg-principles].

The Quantum Internet is not anticipated to replace, but rather to enhance the Classical Internet and/or provide breakthrough applications. For instance, quantum key distribution can improve the security of the Classical Internet; quantum computing can expedite and optimize computation-intensive tasks in the Classical Internet. The Quantum Internet will run in conjunction with the Classical Internet. The process of integrating the Quantum Internet with the Classical Internet is similar to, but with more profound implications, the process of introducing any new communication and networking paradigm into the existing Internet. The intent of this document is to provide a common understanding and framework of applications and application scenarios for the Quantum Internet.

This document represents the consensus of the Quantum Internet Research Group (QIRG). It has been reviewed extensively by Research Group (RG) members with expertise in both quantum physics and Classical Internet operation.

2. Terms and Acronyms List

This document assumes that the reader is familiar with the quantum information technology related terms and concepts that are described in [I-D.irtf-qirg-principles]. In addition, the following terms and acronyms are defined herein for clarity:

* Bell Pairs - A special type of two-qubits quantum state. The two qubits show a correlation that cannot be observed in classical information theory. We refer to such correlation as quantum entanglement. Bell pairs exhibit the maximal quantum entanglement. One example of a Bell pair is (|00> + |11>)/sqrt(2). The Bell pairs are a fundamental resource for quantum communication.
* Bit - Binary Digit (i.e., fundamental unit of information in classical communications and classical computing).

* Classical Internet - The existing, deployed Internet (circa 2020) where bits are transmitted in packets between nodes to convey information. The Classical Internet supports applications which may be enhanced by the Quantum Internet. For example, the end-to-end security of a Classical Internet application may be improved by secure communication setup using a quantum application.

* Entanglement Swapping: It is a process of sharing an entanglement between two distant parties via some intermediate nodes. For example, suppose there are three parties A, B, C, and each of the parties (A, B) and (B, C) share Bell pairs. B can use the qubits it shares with A and C to perform entanglement swapping operations, and as a result, A and C share Bell pairs.

* Fast Byzantine Negotiation - A Quantum-based method for fast agreement in Byzantine negotiations [Ben-Or] [Taherkhani].

* Local Operations and Classical Communication (LOCC) - A method where nodes communicate in rounds, in which (1) they can send any classical information to each other; (2) they can perform local quantum operations individually; and (3) the actions performed in each round can depend on the results from previous rounds.

* Noisy Intermediate-Scale Quantum (NISQ) - NISQ was defined in [Preskill] to represent a near-term era in quantum technology. According to this definition, NISQ computers have two salient features: (1) The size of NISQ computers range from 50 to a few hundred physical qubits (i.e., intermediate-scale); and (2) Qubits in NISQ computers have inherent errors and the control over them is imperfect (i.e., noisy).

* Packet - A self-identified message with in-band addresses or other information that can be used for forwarding the message. The message contains an ordered set of bits of determinate number. The bits contained in a packet are classical bits.

* Prepare-and-Measure - A set of Quantum Internet scenarios where quantum nodes only support simple quantum functionalities (i.e., prepare qubits and measure qubits). For example, BB84 [BB84] is a prepare-and-measure quantum key distribution protocol.

* Quantum Computer (QC) - A quantum end node that also has quantum memory and quantum computing capabilities is regarded as a full-fledged quantum computer.
* Quantum End Node - An end node that hosts user applications and interfaces with the rest of the Internet. Typically, an end node may serve in a client, server, or peer-to-peer role as part of the application. A quantum end node must also be able to interface to the Classical Internet for control purposes and thus also be able to receive, process, and transmit classical bits/packets.

* Quantum Internet - A network of Quantum Networks. The Quantum Internet is expected to be merged into the Classical Internet. The Quantum Internet may either improve classical applications or may enable new quantum applications.

* Quantum Key Distribution (QKD) - A method that leverages quantum mechanics such as no-cloning theorem to let two parties create the same arbitrary classical key.

* Quantum Network - A new type of network enabled by quantum information technology where quantum resources such as qubits and entanglement are transferred and utilized between quantum nodes. The Quantum Network will use both quantum channels, and classical channels provided by the Classical Internet, referred to as a hybrid implementation.

* Quantum Teleportation - A technique for transferring quantum information via local operations and classical communication (LOCC). If two parties share a Bell pair, then using quantum teleportation a sender can transfer a quantum data bit to a receiver without sending it physically via a quantum channel.

* Qubit - Quantum Bit (i.e., fundamental unit of information in quantum communication and quantum computing). It is similar to a classic bit in that the state of a qubit is either "0" or "1" after it is measured, and is denoted as its basis state vector |0⟩ or |1⟩ using Dirac’s ket notation. However, the qubit is different than a classic bit in that the qubit can be in a linear combination of both states before it is measured and termed to be in superposition. Any of several Degrees of Freedom (DOF) of a photon (e.g., polarization, time bib, and/or frequency) or an electron (e.g., spin) can be used to encode a qubit.

* Transmit a Qubit - An operation of encoding a qubit into a mobile carrier (i.e., typically photon) and passing it through a quantum channel from a sender (a transmitter) to a receiver.

* Teleport a Qubit - An operation on two or more carriers in succession to move a qubit from a sender to a receiver using quantum teleportation.
Transfer a Qubit - An operation to move a qubit from a sender to a receiver without specifying the means of moving the qubit, which could be "transmit" or "teleport".

3. Quantum Internet Applications

3.1. Overview

The Quantum Internet is expected to be beneficial for a subset of existing and new applications. The expected applications for the Quantum Internet are still being developed as we are in the formative stages of the Quantum Internet [Castelvecchi] [Wehner]. However, an initial (and non-exhaustive) list of the applications to be supported on the Quantum Internet can be identified and classified using two different schemes. Note, this document does not include quantum computing applications that are purely local to a given node.

3.2. Classification by Application Usage

Applications may be grouped by the usage that they serve. Specifically, applications may be grouped according to the following categories:

* Quantum cryptography applications - Refer to the use of quantum information technology for cryptographic tasks (e.g., quantum key distribution [Renner]).

* Quantum sensors applications - Refer to the use of quantum information technology for supporting distributed sensors (e.g., clock synchronization [Jozsa2000] [Komar] [Guo]).

* Quantum computing applications - Refer to the use of quantum information technology for supporting remote quantum computing facilities (e.g., distributed quantum computing [Denchev]).

This scheme can be easily understood by both a technical and non-technical audience. The next sections describe the scheme in more detail.

3.2.1. Quantum Cryptography Applications

Examples of quantum cryptography applications include quantum-based secure communication setup and fast Byzantine negotiation.

1. Secure communication setup - Refers to secure cryptographic key distribution between two or more end nodes. The most well-known method is referred to as Quantum Key Distribution (QKD) [Renner], which has been mathematically proven to be unbreakable.
2. Fast Byzantine negotiation - Refers to a Quantum-based method for fast agreement in Byzantine negotiations [Ben-Or], for example, to reduce the number of expected communication rounds and in turn achieve faster agreement, in contrast to classical Byzantine negotiations. A quantum aided Byzantine agreement on quantum repeater networks as proposed in [Taherkhani] includes optimization techniques to greatly reduce the quantum circuit depth and the number of qubits in each node. Quantum-based methods for fast agreement in Byzantine negotiations can be used for improving consensus protocols such as practical Byzantine Fault Tolerance (pBFT), as well as other distributed computing features which use Byzantine negotiations.

3. Quantum money - The main security requirement of money is unforgeability. A quantum money scheme aims to fulfill by exploiting the no-cloning property of the unknown quantum states. Though the original idea of quantum money dates back to 1970, these early protocols allow only the issuing bank to verify a quantum banknote. However, the recent protocols such as public-key quantum money [Zhandry] allow anyone to verify the banknotes locally.

3.2.2 Quantum Sensing/Metrology Applications

The entanglement, superposition, interference, squeezing properties can enhance the sensitivity of the quantum sensors and eventually can outperform the classical strategies. Examples of quantum sensor applications include network clock synchronization, high sensitivity sensing, etc. These applications mainly leverage a network of entangled quantum sensors (i.e. quantum sensor networks) for high-precision multi-parameter estimation [Proctor].

1. Network clock synchronization - Refers to a world wide set of high-precision clocks connected by the Quantum Internet to achieve an ultra precise clock signal [Komar] with fundamental precision limits set by quantum theory.

2. High sensitivity sensing - Refers to applications that leverage quantum phenomena to achieve reliable nanoscale sensing of physical magnitudes. For example, [Guo] uses an entangled quantum network for measuring the average phase shift among multiple distributed nodes.

3. Interferometric Telescopes using Quantum Information - Interferometric techniques are used to combine signals from two or more telescopes to obtain measurements with higher resolution than what could be obtained with either telescope individually. It can make measurements of very small astronomical objects if
the telescopes are spread out over a wide area. However, the phase fluctuations and photon loss introduced by the communication channel between the telescopes put a limitation on the baseline lengths of the optical interferometers. This limitation can be potentially avoided using quantum teleportation. In general, by sharing EPR-pairs using quantum repeaters, the optical interferometers can communicate photons over long distances, providing arbitrarily long baselines [Gottesman2012].

3.2.3. Quantum Computing Applications

In this section, we include the applications for the quantum computing. Note that, for the next couple of years we will have quantum computers as a cloud service. Sometimes to run such applications in the cloud while preserving the privacy, a client and a server need to exchange qubits. Therefore, such privacy preserving quantum computing applications require a Quantum Internet to execute.

Examples of quantum computing include distributed quantum computing and secure quantum computing with privacy preservation, which can enable new types of cloud computing.

1. Distributed quantum computing – Refers to a collection of remote small-capacity quantum computers (i.e., each supporting a relatively small number of qubits) that are connected and work together in a coordinated fashion so as to simulate a virtual large capacity quantum computer [Wehner].

2. Secure quantum computing with privacy preservation – Refers to private, or blind, quantum computation, which provides a way for a client to delegate a computation task to one or more remote quantum computers without disclosing the source data to be computed over [Fitzsimons].

4. Selected Quantum Internet Application Scenarios

The Quantum Internet will support a variety of applications and deployment configurations. This section details a few key application scenarios which illustrates the benefits of the Quantum Internet. In system engineering, an application scenario is typically made up of a set of possible sequences of interactions between nodes and users in a particular environment and related to a particular goal. This will be the definition that we use in this section.
4.1. Secure Communication Setup

In this scenario, two banks (i.e., Bank #1 and Bank #2) need to have secure communications for transmitting important financial transaction records (see Figure 1). For this purpose, they first need to securely share a classic secret cryptographic key (i.e., a sequence of classical bits), which is triggered by an end-user banker at Bank #1. This results in a source quantum node A at Bank #1 to securely establish a classical secret key with a destination quantum node B at Bank #2. This is referred to as a secure communication setup. Note that the quantum node A and B may be either a bare-bone quantum end node or a full-fledged quantum computer. This application scenario shows that the Quantum Internet can be leveraged to improve the security of Classical Internet applications of which the financial application shown in Figure 1 is an example.

One requirement for this secure communication setup process is that it should not be vulnerable to any classical or quantum computing attack. This can be realized using QKD which is unbreakable in principle. QKD can securely establish a secret key between two quantum nodes, using a classical authentication channel and insecure quantum channel without physically transmitting the key through the network and thus achieving the required security. However, care must be taken to ensure that the QKD system is safe against physical side channel attacks which can compromise the system. An example of a physical side channel attack is to surreptitiously inject additional light into the optical devices used in QKD to learn side information about the system such as the polarization. Other specialized physical attacks against QKD also use a classical authentication channel and insecure quantum channel such as the phase-remapping attack, photon number splitting attack, and decoy state attack [Zhao2018]. QKD can be used for many other cryptographic communications, such as IPSec and Transport Layer Security (TLS) where involved parties need to establish a shared security key, although it usually introduces a high latency.

QKD is the most mature feature of the quantum information technology, and has been commercially released in small-scale and short-distance deployments. More QKD use cases are described in ETSI documents [ETSI-QKD-UseCases]; in addition, the ETSI document [ETSI-QKD-Interfaces] specifies interfaces between QKD users and QKD devices.

In general, the prepare and measure QKD protocols (e.g., [BB84]) without using entanglement work as follows:
1. The source quantum node A encodes classical bits to qubits. Basically, the source node A generates two random classical bit strings X, Y. Among them, it uses the bit string X to choose the basis and uses Y to choose the state corresponding to the chosen basis. For example, if X=0 then in case of BB84 protocol Alice prepares the state in \(|0\rangle, |1\rangle\)-basis; otherwise she prepares the state in \(|+\rangle, |->\rangle\)-basis. Similarly, if Y=0 then Alice prepares the qubit either \(|0\rangle\) or \(|+\rangle\) (depending on the value of X), and if Y =1, then Alice prepares the qubit either \(|1\rangle\) or \(|->\rangle\).

2. The source quantum node A sends qubits to the destination quantum node B via quantum channel.

3. The destination quantum node receives qubits and measures each of them in one of the two basis at random.

4. The destination quantum node informs the source node of its choice of basis for each qubit.

5. The source quantum node informs the destination node which random quantum basis is correct.

6. Both nodes discard any measurement bit under different quantum basis and remaining bits could be used as the secret key. Before generating the final secret key, there is a post-processing procedure over authenticated classical channels. The classical post-processing part can be subdivided into three steps, namely parameter estimation, error-correction, and privacy amplification. In the parameter estimation phase, both Alice and Bob use some of the bits to estimate the channel error. If it is larger than some threshold value, they abort the protocol otherwise move to the error-correction phase. Basically, if an eavesdropper tries to intercept and read qubits sent from node A to node B, the eavesdropper will be detected due to the entropic uncertainty relation property theorem of quantum mechanics. As a part of the post-processing procedure, both nodes usually also perform information reconciliation [Elkouss] for efficient error correction and/or conduct privacy amplification [Tang] for generating the final information-theoretical secure keys.
7. The post-processing procedure needs to be performed over an authenticated classical channel. In other words, the source quantum node and the destination quantum node need to authenticate the classical channel to make sure there is no eavesdroppers or man-in-the-middle attacks, according to certain authentication protocols such as [Kiktenko]. In [Kiktenko], the authenticity of the classical channel is checked at the very end of the post-processing procedure instead of doing it for each classical message exchanged between the quantum source node and the quantum destination node.

It is worth noting that:

1. There are some entanglement-based QKD protocols, such as [Treiber][E91][BBM92], which work differently than the above steps. The entanglement-based schemes, where entangled states are prepared externally to the source quantum node and the destination quantum node, are not normally considered "prepare-and-measure" as defined in [Wehner]; other entanglement-based schemes, where entanglement is generated within the source quantum node can still be considered "prepare-and-measure"; send-and-return schemes can still be "prepare-and-measure", if the information content, from which keys will be derived, is prepared within the source quantum node before being sent to the destination quantum node for measurement.

2. There are many enhanced QKD protocols based on [BB84]. For example, a series of loopholes have been identified due to the imperfections of measurement devices; there are several solutions to take into account these attacks such as measurement-device-independent QKD [Zhang2019]. These enhanced QKD protocols can work differently than the steps of BB84 protocol [BB84].

3. For large-scale QKD, QKD Networks (QKDN) are required, which can be regarded as a subset of a Quantum Internet. A QKDN may consist of a QKD application layer, a QKD network layer, and a QKD link layer [Qin]. One or multiple trusted QKD relays [Zhang2018] may exist between the source quantum node A and the destination quantum node B, which are connected by a QKDN. Alternatively, a QKDN may rely on entanglement distribution and entanglement-based QKD protocols; as a result, quantum-repeaters/routers instead of trusted QKD relays are needed for large-scale QKD.

4. QKD provides an information-theoretical way to share secret keys between two parties in the presence of Eve. However, this is true in theory, and there is a significant gap between theory and practice. By exploiting the imperfection of the detectors Eve
can gain information about the shared key [Xu]. To avoid such side-channel attacks in [Lo], the researchers provide a QKD protocol called Measurement Device-Independent (MDI) QKD that allows two users (a transmitter "Alice" and a receiver "Bob") to communicate with perfect security, even if the (measurement) hardware they are using has been tampered with (e.g., by an eavesdropper) and thus is not trusted. It is achieved by measuring correlations between signals from Alice and Bob rather than the actual signals themselves.

5. QKD protocols based on Continuous Variable (CV-QKD) have recently seen plenty of interest as they only require telecommunications equipment that is readily available and is also in common use industry-wide. This kind of technology is a potentially high-performance technique for secure key distribution over limited distances. The recent demonstration of CV-QKD shows compatibility with classical coherent detection schemes that are widely used for high bandwidth classical communication systems [Grosshans]. Note that we still do not have a quantum repeater for the continuous variable systems; hence, this kind of QKD technologies can be used for the short distance communications or trusted relay-based QKD networks.

6. Secret sharing can be used to distribute a secret key among multiple nodes by letting each node know a share or a part of the secret key, while no single node can know the entire secret key. The secret key can only be re-constructed via collaboration from a sufficient number of nodes. Quantum Secret Sharing (QSS) typically refers to the scenario: The secret key to be shared is based on quantum states instead of classical bits. QSS enables to split and share such quantum states among multiple nodes.

As a result, the Quantum Internet in Figure 1 contains quantum channels. And in order to support secure communication setup especially in large-scale deployment, it also requires entanglement generation and entanglement distribution [I-D.van-meter-qirg-quantum-connection-setup], quantum repeaters/routers, and/or trusted QKD relays.
4.2. Secure Quantum Computing with Privacy Preservation

Secure computation with privacy preservation refers to the following scenario:

1. A client node with source data delegates the computation of the source data to a remote computation node (i.e., a server).

2. Furthermore, the client node does not want to disclose any source data to the remote computation node, which preserves the source data privacy.

3. Note that there is no assumption or guarantee that the remote computation node is a trusted entity from the source data privacy perspective.

As an example illustrated in Figure 2, a terminal node can be a small quantum computer with limited computation capability compared to a remote quantum computation node (e.g., a remote mainframe quantum computer), but the terminal node needs to run a computation-intensive task (e.g., Shor’s factoring algorithm). The terminal node can create individual qubits and send them to the remote quantum computation node. Then, the remote quantum computation node can entangle the qubits, calculate on them, measure them, generate measurement results in classical bits, and return the measurement results to the terminal node. It is noted that those measurement
results will look like purely random data to the remote quantum computation node because the initial states of the qubits were chosen in a cryptographically secure fashion.

As a new client/server computation model, BQC generally enables: 1) The client delegates a computation function to the server; 2) The client does not send original qubits to the server, but send transformed qubits to the server; 3) The computation function is performed at the server on the transformed qubits to generate temporary result qubits, which could be quantum-circuit-based computation or measurement-based quantum computation. The server sends the temporary result qubits to the client; 4) The client receives the temporary result qubits and transforms them to the final result qubits. During this process, the server can not figure out the original qubits from the transformed qubits. Also, it will not take too much efforts on the client side to transform the original qubits to the transformed qubits, or transform the temporary result qubits to the final result qubits. One of the very first BQC protocols such as [Childs] follows this process, although the client needs some basic quantum features such as quantum memory, qubit preparation and measurement, and qubit transmission. Measurement-based quantum computation is out of the scope of this document and more details about it can be found in [Jozsa2005].

It is worth noting that:

1. The BQC protocol in [Childs] is a circuit-based BQC model, where the client only performs simple quantum circuit for qubit transformation, while the server performs a sequence of quantum logic gates. Qubits are transmitted back and forth between the client and the server.

2. Universal BQC in [Broadbent] is a measurement-based BQC model, which is based on measurement-based quantum computing leveraging entangled states. The principle in UBQC is based on the fact the quantum teleportation plus a rotated Bell measurement realizes a quantum computation, which can be repeated multiple times to realize a sequence of quantum computation. In this approach, the client first prepares transformed qubits and sends them to the server and the server needs first to prepare entangled states from all received qubits. Then, multiple interaction and measurement rounds happen between the client and the server. For each round, the client computes and sends new measurement instructions or measurement adaptations to the server; then, the server performs the measurement according to the received measurement instructions to generate measurement results (qubits or in classic bits); the client receives the measurement results and transforms them to the final results.
3. A hybrid universal BQC is proposed in [Zhang2009], where the server performs both quantum circuits like [Childs] and quantum measurements like [Broadbent] to reduce the number of required entangled states in [Broadbent]. Also, the client is much simpler than the client in [Childs]. This hybrid BQC is a combination of circuit-based BQC model and measurement-based BQC model.

4. It will be ideal if the client in BQC is a purely classical client, which only needs to interact with the server using classical channel and communications. [Huang] demonstrates such an approach, where a classical client leverages two entangled servers to perform BQC, with the assumption that both servers cannot communicate with each other; otherwise, the blindness or privacy of the client cannot be guaranteed. The scenario as demonstrated in [Huang] is essentially an example of BQC with multiple servers.

5. How to verify that the server will perform what the client requests or expects is an important issue in many BQC protocols, referred to as verifiable BQC. [Fitzsimons] discusses this issue and compares it in various BQC protocols.

In Figure 2, the Quantum Internet contains quantum channels and quantum repeaters/routers for long-distance qubits transmission [I-D.irtf-qirg-principles].

4.3. Distributed Quantum Computing

There can be two types of distributed quantum computing [Denchev]:

Figure 2: Secure Quantum Computing with Privacy Preservation
1. Leverage quantum mechanics to enhance classical distributed computing. For example, entangled quantum states can be exploited to improve leader election in classical distributed computing, by simply measuring the entangled quantum states at each party (e.g., a node or a device) without introducing any classical communications among distributed parties [Pal]. Normally, pre-shared entanglement needs first be established among distributed parties, followed by LOCC operations at each party. And it generally does not need to transmit qubits among distributed parties.

2. Distribute quantum computing functions to distributed quantum computers. A quantum computing task or function (e.g., quantum gates) is split and distributed to multiple physically separate quantum computers. And it may or may not need to transmit qubits (either inputs or outputs) among those distributed quantum computers. Entangled states will be needed and actually consumed to support such distributed quantum computing tasks. It is worth noting that: 1) Entangled states can be created beforehand and stored or buffered; 2) The rate of entanglement creation will limit the performance of practical quantum internet applications including distributed quantum computing, although entangled states could be buffered. For example, [Gottesman1999] and [Eisert] have proved that a CNOT gate can be realized jointly by and distributed to multiple quantum computers. The rest of this section focuses on this type of distributed quantum computing.

As a scenario for the second type of distributed quantum computing, Noisy Intermediate-Scale Quantum (NISQ) computers distributed in different locations are available for sharing. According to the definition in [Preskill], a NISQ computer can only realize a small number of qubits and has limited quantum error correction. This scenario is referred to as distributed quantum computing [Caleffi] [Cacciapuoti2020] [Cacciapuoti2019]. This application scenario reflects the vastly increased computing power which quantum computers as a part of the Quantum Internet can bring, in contrast to classical computers in the Classical Internet, in the context of distributed quantum computing ecosystem [Cuomo]. According to [Cuomo], quantum teleportation enables a new communication paradigm, referred to as teledata [VanMeter2006-01], which moves quantum states among qubits to distributed quantum computers. In addition, distributed quantum computation also needs the capability of remotely performing quantum computation on qubits on distributed quantum computers, which can be enabled by the technique called telegate [VanMeter2006-02].

As an example, a user can leverage these connected NISQ computers to solve highly complex scientific computation problems, such as analysis of chemical interactions for medical drug development [Cao]
In this case, qubits will be transmitted among connected quantum computers via quantum channels, while the user's execution requests are transmitted to these quantum computers via classical channels for coordination and control purpose. Another example of distributed quantum computing is secure Multi-Party Quantum Computation (MPQC) [Crepeau], which can be regarded as a quantum version of classical secure Multi-Party Computation (MPC). In a secure MPQC protocol, multiple participants jointly perform quantum computation on a set of input quantum states, which are prepared and provided by different participants. One of the primary aims of the secure MPQC is to guarantee that each participant will not know input quantum states provided by other participants. Secure MPQC relies on verifiable quantum secret sharing [Lipinska].

For the example shown in Figure 3, we want to move qubits from one NISQ computer to another NISQ computer. For this purpose, quantum teleportation can be leveraged to teleport sensitive data qubits from one quantum computer A to another quantum computer B. Note that Figure 3 does not cover measurement-based distributed quantum computing, where quantum teleportation may not be required. When quantum teleportation is employed, the following steps happen between A and B. In fact, LOCC [Chitambar] operations are conducted at the quantum computers A and B in order to achieve quantum teleportation as illustrated in Figure 3.

1. The quantum computer A locally generates some sensitive data qubits to be teleported to the quantum computer B.

2. A shared entanglement is established between the quantum computer A and the quantum computer B (i.e., there are two entangled qubits: q1 at A and q2 at B). For example, the quantum computer A can generate two entangled qubits (i.e., q1 and q2) and sends q2 to the quantum computer B via quantum communications.

3. Then, the quantum computer A performs a Bell measurement of the entangled qubit q1 and the sensitive data qubit.

4. The result from this Bell measurement will be encoded in two classical bits, which will be physically transmitted via a classical channel to the quantum computer B.

5. Based on the received two classical bits, the quantum computer B modifies the state of the entangled qubit q2 in the way to generate a new qubit identical to the sensitive data qubit at the quantum computer A.
In Figure 3, the Quantum Internet contains quantum channels and quantum repeaters/routers [I-D.irtf-qirg-principles]. This application scenario needs to support entanglement generation and entanglement distribution (or quantum connection) setup [I-D.van-meter-qirg-quantum-connection-setup] in order to support quantum teleportation.

![Diagram of Distributed Quantum Computing](image)

Figure 3: Distributed Quantum Computing

5. General Requirements

5.1. Background

Quantum technologies are steadily evolving and improving. Therefore, it is hard to predict the timeline and future milestones of quantum technologies as pointed out in [Grumbling] for quantum computing. Currently, a NISQ computer can achieve fifty to hundreds of qubits with some given error rate.

On the network level, six stages of Quantum Internet development are described in [Wehner] as follows:
1. Trusted repeater networks (Stage-1)
2. Prepare and measure networks (Stage-2)
3. Entanglement distribution networks (Stage-3)
4. Quantum memory networks (Stage-4)
5. Fault-tolerant few qubit networks (Stage-5)
6. Quantum computing networks (Stage-6)

The first stage is simple trusted repeater networks, while the final stage is the quantum computing networks where the full-blown Quantum Internet will be achieved. Each intermediate stage brings with it new functionality, new applications, and new characteristics. Figure 4 illustrates Quantum Internet application scenarios as described in this document mapped to the Quantum Internet stages described in [Wehner]. For example, secure communication setup can be supported in Stage-1, Stage-2, or Stage-3, but with different QKD solutions. More specifically:

In Stage-1, basic QKD is possible and can be leveraged to support secure communication setup but trusted nodes are required to provide end-to-end security. The primary requirement is the trusted nodes.

In Stage-2, the end users can prepare and measure the qubits. In this stage, the users can verify classical passwords without revealing it.

In Stage-3, end-to-end security can be enabled based on quantum repeaters and entanglement distribution, to support the same secure communication setup application. The primary requirement is entanglement distribution to enable long-distance QKD.

In Stage-4, the quantum repeaters gain the capability of storing and manipulating entangled qubits in the quantum memories. Using these kind of quantum networks, one can run sophisticated applications like blind quantum computing, leader election, quantum secret sharing.

In Stage-5, quantum repeaters can perform error correction; hence they can perform fault-tolerant quantum computations on the received data. With the help of these repeaters, it is possible to run distributed quantum computing and quantum sensor applications over a smaller number of qubits.

Finally, in Stage-6, distributed quantum computing relying on more qubits can be supported.
Figure 4: Example Application Scenarios in Different Quantum Internet Stages

5.2. Requirements

Some general and functional requirements on the Quantum Internet from the networking perspective, based on the above application scenarios, are identified as follows:

1. Methods for facilitating quantum applications to interact efficiently with entangled qubits are necessary in order for them to trigger distribution of designated entangled qubits to potentially any other quantum node residing in the Quantum Internet. To accomplish this, specific operations must be performed on entangled qubits (e.g., entanglement swapping, entanglement distillation). Quantum nodes may be quantum end nodes, quantum repeaters/routers, and/or quantum computers.

2. Quantum repeaters/routers should support robust and efficient entanglement distribution in order to extend and establish high-fidelity entanglement connection between two quantum nodes. For
achieving this, it is required to first generate an entangled pair on each hop of the path between these two nodes, and then perform entanglement swapping operations at each of the intermediate nodes.

3. Quantum end nodes must send additional information on classical channels to aid in transferring qubits across quantum repeaters/receivers. This is because qubits are transferred individually and do not have any associated packet header which can help in transferring the qubit. Any extra information to aid in routing, identification, etc., of the qubit(s) must be sent via classical channels.

4. Methods for managing and controlling the Quantum Internet including quantum nodes and their quantum resources are necessary. The resources of a quantum node may include quantum memory, quantum channels, qubits, established quantum connections, etc. Such management methods can be used to monitor network status of the Quantum Internet, diagnose and identify potential issues (e.g. quantum connections), and configure quantum nodes with new actions and/or policies (e.g. to perform a new entanglement swapping operation). New management information model for the Quantum Internet may need to be developed.

6. Conclusion

This document provides an overview of some expected application categories for the Quantum Internet, and then details selected application scenarios. The applications are first grouped by their usage which is a natural and easy to understand classification scheme. This set of applications may, of course, naturally expand over time as the Quantum Internet matures. Finally, some general requirements for the Quantum Internet are also provided.

This document can also serve as an introductory text to readers interested in learning about the practical uses of the Quantum Internet. Finally, it is hoped that this document will help guide further research and development of the Quantum Internet functionality required to implement the application scenarios described herein.

7. IANA Considerations

This document requests no IANA actions.
8. Security Considerations

This document does not define an architecture nor a specific protocol for the Quantum Internet. It focuses instead on detailing application scenarios, requirements, and describing typical Quantum Internet applications. However, some salient observations can be made regarding security of the Quantum Internet as follows.

It has been identified in [NISTIR8240] that once large-scale quantum computing becomes reality that it will be able to break many of the public-key (i.e., asymmetric) cryptosystems currently in use. This is because of the increase in computing ability with quantum computers for certain classes of problems (e.g., prime factorization, optimizations). This would negatively affect many of the security mechanisms currently in use on the Classical Internet which are based on public-key (Diffie-Hellman) encryption. This has given strong impetus for starting development of new cryptographic systems that are secure against quantum computing attacks [NISTIR8240].

Interestingly, development of the Quantum Internet will also mitigate the threats posed by quantum computing attacks against Diffie-Hellman based public-key cryptosystems. Specifically, the secure communication setup feature of the Quantum Internet as described in Section 4.1 will be strongly resistant to both classical and quantum computing attacks against Diffie-Hellman based public-key cryptosystems.

A key additional threat consideration for the Quantum Internet is pointed to by [RFC7258], which warns of the dangers of pervasive monitoring as a widespread attack on privacy. Pervasive monitoring is defined as a widespread, and usually covert, surveillance through intrusive gathering of application content or protocol metadata such as headers. This can be accomplished through active or passive wiretaps, traffic analysis, or subverting the cryptographic keys used to secure communications.

The secure communication setup feature of the Quantum Internet as described in Section 4.1 will be strongly resistant to pervasive monitoring based on directly attacking (Diffie-Hellman) encryption keys. Also, Section 4.2 describes a method to perform remote quantum computing while preserving the privacy of the source data. Finally, the intrinsic property of qubits to decohere if they are observed, albeit covertly, will theoretically allow detection of unwanted monitoring in some future solutions.

Modern networks are implemented with zero trust principles where classical cryptography is used for confidentiality, integrity protection, and authentication on many of the logical layers of the
network stack, often all the way from device to software in the cloud [NISTSP800-207]. The cryptographic solutions in use today are based on well-understood primitives, provably secure protocols and state-of-the-art implementations that are secure against a variety of side-channel attacks.

In contrast to conventional cryptography and Post-Quantum Cryptography (PQC), the security of QKD is inherently tied to the physical layer, which makes the threat surfaces of QKD and conventional cryptography quite different. QKD implementations have already been subjected to publicized attacks [Zhao2008] and the National Security Agency (NSA) notes that the risk profile of conventional cryptography is better understood [NSA]. The fact that conventional cryptography and PQC are implemented at a higher layer than the physical one means PQC can be used to securely send protected information through untrusted relays. This is in stark contrast with QKD, which relies on hop-by-hop security between intermediate trusted nodes. The PQC approach is better aligned with the modern technology environment, in which more applications are moving toward end-to-end security and zero-trust principles. It is also important to note that while PQC can be deployed as a software update, QKD requires new hardware.

Regarding QKD implementation details, the NSA states that communication needs and security requirements physically conflict in QKD and that the engineering required to balance them has extremely low tolerance for error. While conventional cryptography can be implemented in hardware in some cases for performance or other reasons, QKD is inherently tied to hardware. The NSA points out that this makes QKD less flexible with regard to upgrades or security patches. As QKD is fundamentally a point-to-point protocol, the NSA also notes that QKD networks often require the use of trusted relays, which increases the security risk from insider threats.

The UK’s National Cyber Security Centre cautions against reliance on QKD, especially in critical national infrastructure sectors, and suggests that PQC as standardized by the NIST is a better solution [NCSC]. Meanwhile, the National Cybersecurity Agency of France has decided that QKD could be considered as a defense-in-depth measure complementing conventional cryptography, as long as the cost incurred does not adversely affect the mitigation of current threats to IT systems [ANNSI].
9. Acknowledgments

The authors want to thank Michele Amoretti, Mathias Van Den Bossche, Xavier de Foy, Patrick Gelard, Alvaro Gomez Inesta, Wojciech Kozlowski, John Mattsson, Rodney Van Meter, Joey Salazar, and Joseph Touch, and the rest of the QIRG community as a whole for their very useful reviews and comments to the document.

10. Informative References


[ETSI-QKD-Interfaces] ETSI GR QKD 003 V2.1.1, "Quantum Key Distribution (QKD); Components and Internal Interfaces", 2018, <https://www.etsi.org/deliver/etsi_gr/QKD/001_099/003/02.01.01_60/gr_QKD003v020101p.pdf>.


Authors’ Addresses

Chonggang Wang
InterDigital Communications, LLC
1001 E Hector St
Conshohocken, 19428
United States of America
Email: Chonggang.Wang@InterDigital.com

Akbar Rahman
InterDigital Communications, LLC
1000 Sherbrooke Street West
Montreal  H3A 3G4
Canada
Email: rahmansakbar@yahoo.com

Ruidong Li
Kanazawa University
Kakuma-machi,
Ishikawa Prefecture 920-1192
Japan
Email: lrd@se.kanazawa-u.ac.jp

Melchior Aelmans
Leidseweg 144
Voorschoten
Email: melchior@aelmans.eu