Quantum Internet Research Group
Internet-Draft
Intended status: Informational
Expires: May 6, 2020

W. Kozlowski
S. Wehner
QuTech
R. Van Meter
Keio Univeristy
B. Rijsman
Individual

November 3, 2019

Architectural Principles for a Quantum Internet
draft-irtf-qirg-principles-02

Abstract

The vision of a quantum internet is to fundamentally enhance 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 as the physical nature of the communication is fundamentally different. The first realisations of quantum networks are imminent, but there is no practical proposal for how to organise, utilise, and manage such networks. In this memo, we attempt 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.

Copyright Notice

Table of Contents

1. Introduction

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 classical networks. Depending on the stage of a quantum network [5] such devices may be 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. This new networking paradigm offers promise for a range of new applications such as secure communications [1], distributed quantum computation [2], or quantum sensor networks [3]. The field of quantum communication has been a subject of active research for many years and the most well-known application of quantum communication, quantum key distribution (QKD) for secure communications, has already been deployed at short (roughly 100km) distances.

Fully quantum networks capable of transmitting and managing entangled quantum states in order to send, receive, and manipulate distributed quantum information are now imminent [4] [5]. 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 itself at what is even lower than assembly level where no common interfaces yet exist. Furthermore, whilst physical mechanisms for forwarding quantum states exist, there are no robust protocols for managing such transmissions.

2. Model of communication

In order to understand the framework for quantum networking a basic understanding of quantum information is necessary. The following sections aim to introduce the bare minimum 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 to e.g. [10] for an in-depth introduction to quantum information.

2.1. 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, a qubit, exists over the same binary space, but unlike the classical bit, it can exist in a so-called superposition of the two possibilities:

a |0> + b |1>,

where |X> denotes a quantum state, 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> or |1>. Which of the two states it ends up in is not deterministic, but it can be determined from the readout of the measurement, a classical bit, 0 or 1 respectively. The probability of measuring the state in the |0> state is |a|^2 and similarly the probability of measuring the state in the |1> state is |b|^2, where |a|^2 + |b|^2 = 1. This randomness is not due to our ignorance of the underlying mechanisms, but rather it is a fundamental feature of a quantum mechanical system [6].

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:

NOT (a |0> + b |1>) -> a |1> + b |0>.

2.2. 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) x |0> = (|00> + |10>)/sqrt(2),

where x denotes a tensor product (the mathematical mechanism for combining quantum states together). Let us now consider the two-qubit 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

IN	OUT
00	00
01	01
10	11
11	10

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

CNOT (|00> + |10>)/sqrt(2) -> (|00> + |11>)/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 and their behaviour cannot be fully described without accounting for the other qubit. The states of the two individual qubits are now correlated beyond what is possible to achieve classically. Neither qubit is in a definite |0> or |1> 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> or |11>, is fundamentally random as

before, but the states of the two qubits following a measurement will always be identical.

Once a measurement is performed, the two qubits are once again independent. The final state is either |00> or |11> and both of these states can be trivially decomposed into a product of two individual qubits. The entanglement has been consumed and if the same measurement is to be repeated, the entangled state must be prepared again.

3. Entanglement as the fundamental service

Entanglement is the fundamental building block of quantum networks. To see this, consider the state from the previous section:

(|00> + |11>)/sqrt(2).

Neither of the two qubits is in a definite |0> or |1> 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, blind quantum computation, or distributed quantum computation.

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> + |11>)/sqrt(2) with each node holding one of the two qubits. Once any of the two nodes performs a measurement the state of the two qubits collapses to either |00> or |11> 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...> + |11...>)/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, than it is physically impossible for any other system to have any share of this entanglement. Hence, entanglement forms a sort of private and inherently untappable connection between two nodes once established.

It is impossible to entangle two qubits without ever having them directly interact with each other (e.g. by performing a local two-qubit gate, such as the CNOT). A local - or mediated - interaction is necessary to create entanglement and thus such states cannot be created between two quantum nodes that cannot transmit quantum states to each other. 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. [5]>

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. One cannot just send qubits like one can send bits over a wire. 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.

4.1.2. No-cloning theorem

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

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 checksums. Since we cannot read or copy a quantum state a similar approach is out of question for quantum networks.

To describe the quality of a quantum state a physical quantity called fidelity is used. 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 desire it to be in. It expresses the probability that one state will pass a test to identify as the other. 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 it 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.

4.2. Bell pairs

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 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 [7]. 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.

An 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. Bell pair states are the entangled two-qubit states:

```
|00> + |11>,
|00> - |11>,
|01> + |10>,
|01> - |10>,
```

where the constant 1/sqrt(2) normalisation factor has been ignored for clarity. Any of the four Bell pair state 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. That is, either of the nodes that hold the two qubits of the Bell pair can apply a series of single qubit gates to just their qubit in order to transform the ensemble 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. Quantum state teleportation consumes an unknown quantum 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 on 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 quantum state that we wanted to transmit.

The unknown quantum state that was transmitted never entered the network itself. Therefore, the network needs to only be able to reliably produce Bell pairs between any two nodes in 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 its two qubits are delivered to the end-points.

4.4.1. Link generation

[waiting for contrib]

4.4.2. Entanglement swapping

The problem with generating entangled pairs directly across a link is that its efficiency decreases with its length. Beyond a few 10s of kms the rate is effectively zero and due to the no-cloning theorem we cannot simply amplify the signal. The solution is entanglement swapping.

A Bell pair between any two nodes in the network can be constructed by combining the pairs generated along each individual link on the 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 performs 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 transmitted and recreated in Y2's place carrying with it its entanglement with X1. The end-result is shown below:

```
+---------+      +---------+      +---------+
|    A    |      |    B    |      |    C    |
|         |------|         |------|         |
|      X1~~~~~~~~~~~~~~~~~~~~~~~~~~~X2      |
+---------+      +---------+      +---------+
```

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.

4.4.2.1. 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 or purification.

To purify 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 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. Purification makes low demands on fidelity and resources, but distributed protocols incur round-trip delays [11].

4.4.2.2. Delivery

The bare minimum requirements of an application for every Bell pair delivered to the two end-nodes are:

1. Information about which of the four Bell pairs was delivered. The network may choose to not perform Pauli corrections at all and simply notify the application of which state the delivered pair is in or it may perform the Pauli corrections and always deliver the same state.

2. An identifier that allows the applicatqion to unambiguously determine which qubits at the two end-points belong to which entangled pair.

3. An estimate of the fidelity of the delivered pair. This should be above the minimum threshold determined by the application. However, this will only be an estimate and not a guarantee. This has security implications for applications which will be discussed in the section on security.

There are several other features an application might want to be able to request (e.g. multiple pairs delivered together close in time, but doesn't matter when they are delivered), but they are beyond the scope of this memo.

4.4.3. Direct transmission vs. entanglement swapping

Direct state transmission whilst simpler conceptually is much more demanding to implement reliably in practice which means that any near-term practical realisation is more likely to succeed if it is based on the Bell pair and entanglement swapping architecture. All near-term experimental implementations of quantum repeaters are based on this approach. Therefore, this is the architecture that we will focus on in the rest of this memo.

Nevertheless, the direct transmission proposal may be relevant in the future as it has better fault-tolerance properties and much better scaling with transmission distance. It might even be beneficial to utilise a hybrid approach that combines the fault-tolerance of direct transmission with the generic nature of Bell pairs which lends itself to paralellisation and resource provisioning. That is, we still use Bell pairs for transmission of user data, but direct transmission may be used for some of hops for the purposes of Bell pair generation rather than just relying solely on entanglement swapping.

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. New challenges

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

1. There is no quantum equivalent of a payload carrying packet.

   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 pairs are handled individually -- they are not grouped into packets and they 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.

2. 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, 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.

3. Generating entanglement requires temporary state.

   Packet forwarding in a classical network is largely a stateless operation. When a packet is received, the router looks up 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 repeater 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.

4. Generating end-to-end entanglement is a parallelisable operation.

   Classical packets carry user data from source destination by performing a series of hops across the network. This process is necessarily sequential -- it is impossible to forward a packet ahead of time as the user data it carries cannot be known in

advance. A quantum network does not carry any user data. It is only responsible for generating entangled pairs in any of the generic Bell states. The process of creating an end-to-end Bell pair is by its nature parallelisable -- all of the individual link pairs can be generated independently of one another. Furthermore, there is no ordering requirement on the entanglement swapping operations either, they can happen in any order as long as the network can keep track of which pairs were swapped so that it can correctly identify the two ends of the final Bell pair. This parallelism must be exploited to make the most efficient use of the quantum network's resources.

5.2. Classical communication

In this memo 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 - this includes 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 limts). Therefore, quantum routers will need to manage two data planes in parallel, a classical one and a quantum one. Additionally, it must be able to correlate information between them 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. Elements of a quantum network

Collecting all the pieces described so far, a quantum network will consist of the following elements:

- Quantum repeaters - A quantum repeater is a node in the network that is capable of generating entangled pairs with its directly connected neighbours and performing entanglement swap operations on them.

- Quantum routers - A quantum router is a quantum repeater that is connected to more than two quantum repeaters as neighbours. This distinguishes it from quantum repeaters composed into a linear

chain to connect two quantum routers (since no-cloning prohibits quantum signal amplification).

- 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.

- Quantum links - A quantum link is a link which can be used to generate an entangled pair between two directly connected quantum repeaters. It may 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.

5.3.2. Putting it all together

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

```
| App |-------------------CC-------------------| App |
   ||                                             ||
 ------                 ------                 ------
|  EN  |----QC & CC----|  QR  |----QC & CC----|  EN  |
 ------                 ------                 ------
```

App - user-level application
QR - quantum repeater
EN - end-node
QC - quantum channel
CC - classical channel

An application running on two end-nodes attached to a network will at some point need the network to generate entangled pairs for its use. This will require negotiation between the end-nodes, because they must both open a communication end-point (a quantum socket) which the network can use to identify the two ends of the connection. The two end-nodes use the classical connectivity 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 channels to coordinate and claim the resources necessary to fulfil 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 and thus beyond the scope of this memo.

During or after the control information is distributed the network performs the necessary quantum operations such as generating entangled over individual links, performing entanglement swaps, and further signalling to transmit the swap outcomes and other control information. Since none of the entangled pairs 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 once 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.

5.4. Network boundaries

Just like classical network, there will various boundaries will exist in quantum networks.

5.4.1. Boundaries between different physical architectures

There are many different physical architectures for implementing quantum repeater technology. The different technologies differ in how they store and manipulate qubits in memory and how they generate entanglement across a link with their neighbours. Different architectures come with different trade-offs and thus a functional network will likely consist of a mixture of different types of quantum repeaters.

For example, architectures based on optical elements and atomic ensembles are very efficient at generating entanglement, but provide little control over the qubits once the pair is generated. On the other hand nitrogen-vacancy architectures offer a much greater degree of control over qubits, but have a harder time generating the entanglement across a link.

It is an open research question where exactly the boundary will lie. It could be that a single quantum repeater node provides some backplane connection between the architectures, but it also could be that special quantum links delineate the boundary.

5.4.2. Boundaries between different administrative regions

Just like in classical networks, multiple quantum networks will connect into a global quantum internet. This necessarily implies the existence of borders between different administrative regions. How these boundaries will be handled is also an open question and thus beyond the scope of this memo.

5.5. 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 artefacts of the early stages of a new technology. We here consider a highly abstract scenario and refer to [5] for pointers to the physics literature.

5.5.1. 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 currently are on the order of seconds. 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. An architecture that handles short lifetimes may also be more cost-efficient in the future.

5.5.2. Rates

Entanglement generation on a link between two connected nodes is not a very efficient process and it requires many attempts to succeed. A fast repetition rate for Bell Pair generation is achievable, but only one in a few thousands will succeed. Currently, the highest achievable rates of success between nodes capable of storing the resulting qubits are of the order of 10 Hz. Combined with short memory lifetimes this leads to very tight timing windows to build up

network-wide connectivity. Achievable rates are likely to increase with time, but just like with quantum memories, it may be more cost-efficient in the future to provide low-rate links in some parts of the network.

5.5.3. Communication qubit

Most physical architectures capable of storing qubits are only able to generate entanglement using only a subset of its available qubits called communication qubits. 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.

5.5.4. Homogeneity

Currently all hardware implementations are homogeneous and they do not interface with each other. In general, it is very challenging to combine different quantum information processing technologies at present. Coupling different technologies with each other is of great interest as it may help overcome the weaknesses of the different implementations, but this may take a long time to be realised with high reliability and thus is not a near-term goal.

5.6. 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: 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 burden on future technology. Redeploying network technology is a non-trivial process.

As this is a completely new technology that is likely to see many iterations over its lifetime, this memo must not serve as a definitive set of rules, but merely as a general set of recommended guidelines 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.

5.6.1. 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 the classical Internet, but it will inevitably evolve with time and the needs of its users. The goals are listed in order of priority which in itself may also evolve as the community learns more about the technology.

1. Support distributed quantum applications

   The primary purpose of a quantum internet is to run distributed quantum protocols and it is of utmost importance that they can run well and efficiently. Therefore, the needs of quantum applications should always be considered first. The requirements for different applications can be found in [5].

   If a network is able to distribute entanglement it is officially quantum. However, if it is unable to distribute these states with a sufficiently high fidelity at a reasonable rate for a majority of potential applications it is not practical.

2. Support tomorrow's distributed quantum applications

   There are many applications already proposed to run over a quantum internet. However, more algorithms will be invented as the community grows as well as the robustness and the reliability of the technology. Any proposed architecture should not constrain the capabilities of the network for short-term benefit.

3. Hardware heterogeneity

   There are multiple proposals for realising practical quantum repeaters and they all have their advantages and disadvantages. It is also very likely that the most optimal technologies in the future will be hybrid combinations of the many different solutions currently under development. It should be an explicit goal of the architecture to allow for a large variety of hardware implementations.

4. Be flexible with regards to hardware 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. Second, it 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 redeploying a technology is extremely challenging.

5. Security

Whilst the priority for the first quantum networks should be to simply work, we cannot forget that ultimately they have to also be secure. This has implications for the physical realisations (do they satisfy the idealised theoretical models) and also the design of the control stack.

It is actually difficult to guarantee security at the network level and even if the network did provide such guarantees, the application would still need to perform its own verification similarly to how one ensures end-to-end security in classical networks.

It turns out that 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 authenticity, confidentiality, or integrity of the transmitted qubits or the generated entanglement. Instead, applications such as QKD establish such guarantees using the classical network in conjunction with he quantum one. This is much easier than demanding that the network deliver secure entanglement, which indeed is not needed for quantum applications.

Nevertheless, control protocols themselves should be security aware in order to protect the operation of the network itself and limit disruption.

6. Availability and resilience

A practical and usable network is able to continue to operate despite losses and failures, and will be robust to malicious actors trying to disable connectivity. These may be simply considered different aspects of security, but it is worthwhile to address them explicitly at the architectural level already.

7. Easy to manage and monitor

Quantum networks rely on complex physical phenomena and require hardware that is challenging to build. Furthermore, the quantum resources will at first be very scarce and potentially very expensive. This entails a need for a robust management solution. It is important that a good management solution needs to come with adequate monitoring capabilities.

Good management solutions may also be key to optimising the networks which in turn may be crucial in making them economically feasible. Unlike user data that is transmitted over classical networks, quantum networks only need to generate generic Bell Pairs. This leaves a lot of room for pre-allocating resources in an efficient manner.

5.6.2. 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 [8] [9] .

1. Bell Pairs are the fundamental building block

   The key service that a quantum network provides is the distribution of entanglement between the nodes in a network. This point additionally specifies that the entanglement is primarily distributed in the form of the entangled Bell Pair states which should be used as a building block in providing other services, including more complex entangled states.

2. 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 errors should essentially be eliminated for most application protocols, many quantum applications only need imperfect entanglement to function. However, 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 application's 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.

3. 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 point is crucial in enabling the reuse of resources of a network and 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.

4. Time as an expensive resource

With the current technology, time is the most expensive resource. It 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 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 the Bell Pairs to decohere.

However, time is only expensive once quantum operations are underway. If no quantum operations are currently being processed then the network can use this time to prepare and provision resources.

As hardware improves, the need for carefully timing quantum operations may become smaller. It is currently unknown what the cost of these improvements will be, but it is conceivable that there is value in having relatively cheap and undemanding links connected at the edges of a network which will have very short memory lifetimes and low rates of Bell Pair generation.

5. Limit classical communication

This point offers a practical guideline to the issue of timing. A bottleneck in many quantum networked algorithms is the classical communication needed between quantum operations to synchronise state. Ideally, classical control mechanisms that require increased memory lifetimes should be avoided.

For example, some quantum protocols may need to perform a correction for the random outcome of a quantum measurement. For this, they will block the state from further operations until a classical message is received with the information necessary to perform the correction. The time during which the quantum state is blocked is effectively wasted. It reduces the time available for subsequent operations possibly rendering the state useless for an application.

Trade-offs that allow a protocol to limit the number of blocking classical communication rounds once quantum operations have commenced will in general be worth considering.

6. Parallelise quantum operations

A further point to address the issue of timing constraints in the network. The Bell Pairs on the individual links need not be generated one after another along the path between the communication end-points. The order does not matter at all. Furthermore, the order of the swap operations is flexible as long as they don't reduce the fidelity too much. Parallelising these operations is key to optimising quantum protocols.

7. Avoid time-based coordination when possible

A solution to timing constraints is to synchronise clocks and agree on the timing of events. However, such solutions have several downsides. Whilst network clock synchronisation may be accurate enough for certain purposes it introduces an additional element of complexity, especially when multiple nodes in different networks must be synchronised. Furthermore, clock synchronisation will never be perfect and it is conceivable that hardware capabilities advance so much that time-based mechanisms under-utilise resources in the more efficient parts of the network.

Nevertheless, it may not be possible to avoid clocks, but such solutions should be adequately justified.

8. Pre-allocate resources

Regardless of what application is running over the network it will have the same needs as any other application: a number of Bell Pairs of sufficient fidelity. Whilst the fidelity is a variable number, the indistinguishability of Bell Pairs means that there is lots of flexibility in how a network may provision resources to meet demand. The additional timing constraints mean that pre-allocation of resources will be central to a usable quantum network.

6. Security Considerations

Even though no user data enters a quantum network security is listed as an explicit goal for the architecture and this issue is addressed in the section on goals. Even though user data doesn't enter the network, it is still possible to attack the control protocols and violate the authenticity, confidentiality, and integrity of

communication. However, as this is an informational memo it does not propose any concrete mechanisms to achieve these goals.

In summary:

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 authenticity, confidentiality, or integrity of the transmitted qubits or the generated entanglement. Instead, applications such as QKD establish such guarantees using the classical network in conjunction with he quantum one. This is much easier than demanding that the network deliver secure entanglement.

7. IANA Considerations

This memo includes no request to IANA.

8. Acknowledgements

The authors of this memo acknowledge funding received from the EU Flagship on Quantum Technologies through Quantum Internet Alliance project.

The authors would further like to acknowledge Carlo Delle Donne, Matthew Skrzypczyk, and Axel Dahlberg for useful discussions on this topic prior to the submission of this memo.

9. Informative References

[1] Bennett, C. and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing", Theoretical Computer Science 560, 7-11, 2014, <http://www.sciepub.com/reference/53249>.

[2] Crepeau, C., Gottesman, D., and A. Smith, "Secure multi-party quantum computation. Proceedings of Symposium on Theory of Computing", Proceedings of Symposium on Theory of Computing , 2002, <https://arxiv.org/abs/quant-ph/0206138>.

[3] Giovanetti, V., Lloyd, S., and L. Maccone, "Quantum-enhanced measurements: beating the standard quantum limit", Science 306(5700), 1330-1336, 2004, <https://arxiv.org/abs/quant-ph/0412078>.

[4] Castelvecchi, D., "The Quantum Internet has arrived (and it hasn't)", Nature 554, 289-292, 2018, <https://www.nature.com/articles/d41586-018-01835-3>.

[5] Wehner, S., Elkouss, D., and R. Hanson, "Quantum internet: A vision for the road ahead", Science 362, 6412, 2018, <http://science.sciencemag.org/content/362/6412/eaam9288.full>.

[6] Aspect, A., Grangier, P., and G. Roger, "Experimental Tests of Realistic Local Theories via Bell's Theorem", Phys. Rev. Lett. 47 (7): 460-463, 1981, <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.47.460>.

[7] Muralidharan, S., Kim, J., Lutkenhaus, N., Lukin, M., and L. Jiang, "Ultrafast and Fault-Tolerant Quantum Communication across Long Distances", Phys. Rev. Lett. 112 (25-27), 250501, 2014, <https://arxiv.org/abs/1310.5291>.

[8] Meter, R. and J. Touch, "Designing quantum repeater networks", IEEE Communications Magazine 51, 64-71, 2013, <https://ieeexplore.ieee.org/document/6576340>.

[9] Dahlberg, A., Skrzypczyk, M., Coopmans, T., Wubben, L., Rozpedek, F., Pompili, M., Stolk, A., Pawelczak, P., Knegjens, R., de Oliveira Filho, J., Hanson, R., and S. Wehner, "A Link Layer Protocol for Quantum Networks", arXiv 1903.09778, 2019, <https://arxiv.org/abs/1903.09778>.

[10] Nielsen, M. and I. Chuang, "Quantum Computation and Quantum Information", Cambridge University Press , 2011.

[11] Bennett, C., DiVincenzo, D., Smolin, J., and W. Wootters, "Mixed State Entanglement and Quantum Error Correction", Phys. Rev. A Vol. 54, Iss. 5, 1996, <https://arxiv.org/abs/quant-ph/9604024>.

Authors' Addresses

Wojciech Kozlowski
QuTech
Building 22
Lorentzweg 1
Delft 2628 CJ
Netherlands

Email: w.kozlowski@tudelft.nl

Stephanie Wehner
QuTech
Building 22
Lorentzweg 1
Delft 2628 CJ
Netherlands

Email: S.D.C.Wehner@tudelft.nl

Rodney Van Meter
Keio Univeristy
5322 Endo
Fujisawa, Kanagawa 252-0882
Japan

Email: rdv@sfc.wide.ad.jp

Bruno Rijsman
Individual

Email: brunorijsman@gmail.com