Intro to Quantum Networking and Fidelity-aware Entanglement Routing

Genya Ishigaki

Summer Seminar
June 2025

References

- Jessica Illiano, Marcello Caleffi, Antonio Manzalini, Angela Sara Cacciapuoti, "Quantum Internet protocol stack: A comprehensive survey," Computer Networks, Volume 213, 2022, 109092, ISSN 1389-1286. [box link]
- Qiaolun Zhang et al., "Link Configuration for Fidelity-Constrained Entanglement Routing in Quantum Networks," infocom 2025 [box link]

Quantum Internet protocol stack: A comprehensive survey

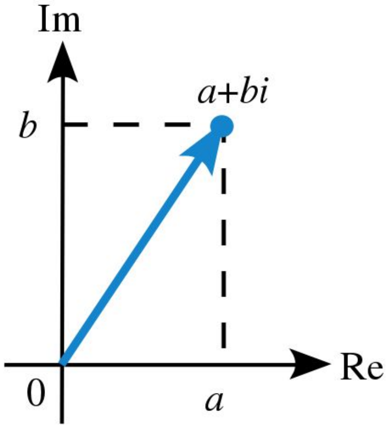
Jessica Illiano, Marcello Caleffi, Antonio Manzalini, Angela Sara Cacciapuoti

Computer Networks, 2022

Basics: Qubit

- A qubit can be in a superposition between “0” and “1”
- This means it is represented as a two-dimensional complex vector
- “bra-ket” notation

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

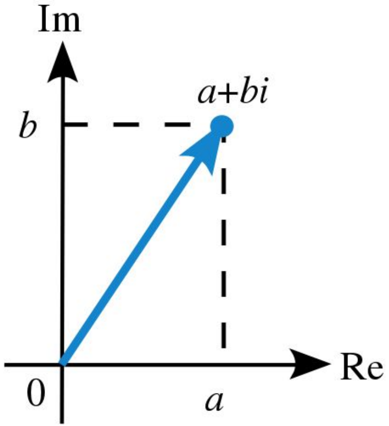


- Any one-qubit can be represented as a linear combination of an orthonormal basis of the vector space
- E.g., Standard basis

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$$

Basics: Qubit

- The bra is the Hermitian conjugate (complex conjugate transpose) of a ket

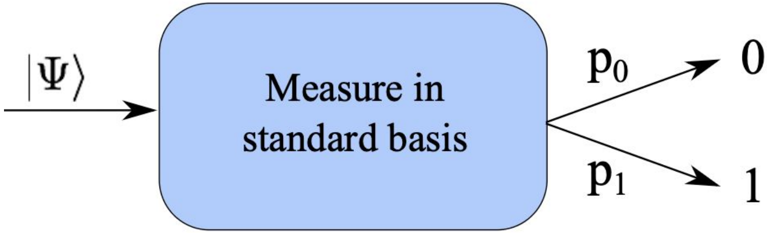


- A ket $$|\psi\rangle = \begin{bmatrix} 1 \\ i \end{bmatrix}$$

- Its bra $$\langle\psi| = \left(\begin{bmatrix} 1 \\ i \end{bmatrix}\right)^{\dagger} = \begin{bmatrix} 1 & -i \end{bmatrix}$$

Basics: Qubit Measurement (in the standard basis)

- Measurement ~ “Looking at how much ‘0’ is actually in our qubit vector”
- Quantum measurements can result in probabilistic outcomes
- We obtain different measurement outcomes corresponding to some probability distribution
- The probability of an outcome is quantified by the inner product between the qubit and the outcome



$$p_0 = |\langle\psi|0\rangle|^2 = \left|\begin{pmatrix}\alpha^* & \beta^*\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\right|^2 = |\alpha|^2,$$

$$p_1 = |\langle\psi|1\rangle|^2 = \left|\begin{pmatrix}\alpha^* & \beta^*\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}\right|^2 = |\beta|^2.$$

The Quantum Measurement Postulate

- (Roughly speaking,) we apply the inner product operation to a qubit (which is a physical entity)
- Measuring a quantum system generally changes the quantum state that describes that system
- The state of the system collapses (or is projected) onto one of the two orthogonal states used for the measurement
- This is one important constraint in quantum communication since the superposition of a qubit will be lost when measured (The quantum measurement postulate)

The No-cloning Theorem

- It is impossible to create an exact copy of an arbitrary unknown quantum state
- Why?
  - Quantum operations must be linear and unitary to preserve superposition and probability distributions
  - However, it is easily shown that there is no sequence of linear and unitary operations that can transfer a qubit state to another

Key Quantum Features

- The quantum measurement postulate
- The no-cloning theorem
- Collectively, they impose the impossibility of safely reading and copying quantum information without altering it
- This provides a mathematically-strong encryption
- However, this also becomes a constraint when designing communication paths

Key Quantum Features

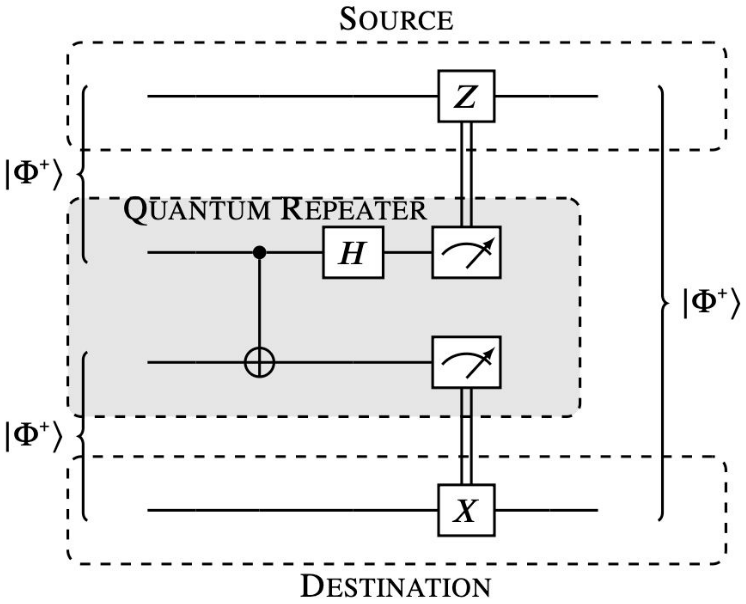
- **Coherence time**: a measurement of how long a qubit can maintain its complex quantum state — essentially, a qubit's lifespan
- **Gate fidelity**: how many gates you can run in the first place, and that determines the size of the algorithms you can run
    - Fidelity, which is the inverse of the error rate — a 1% gate error rate equals 99% fidelity

Quantum Repeaters

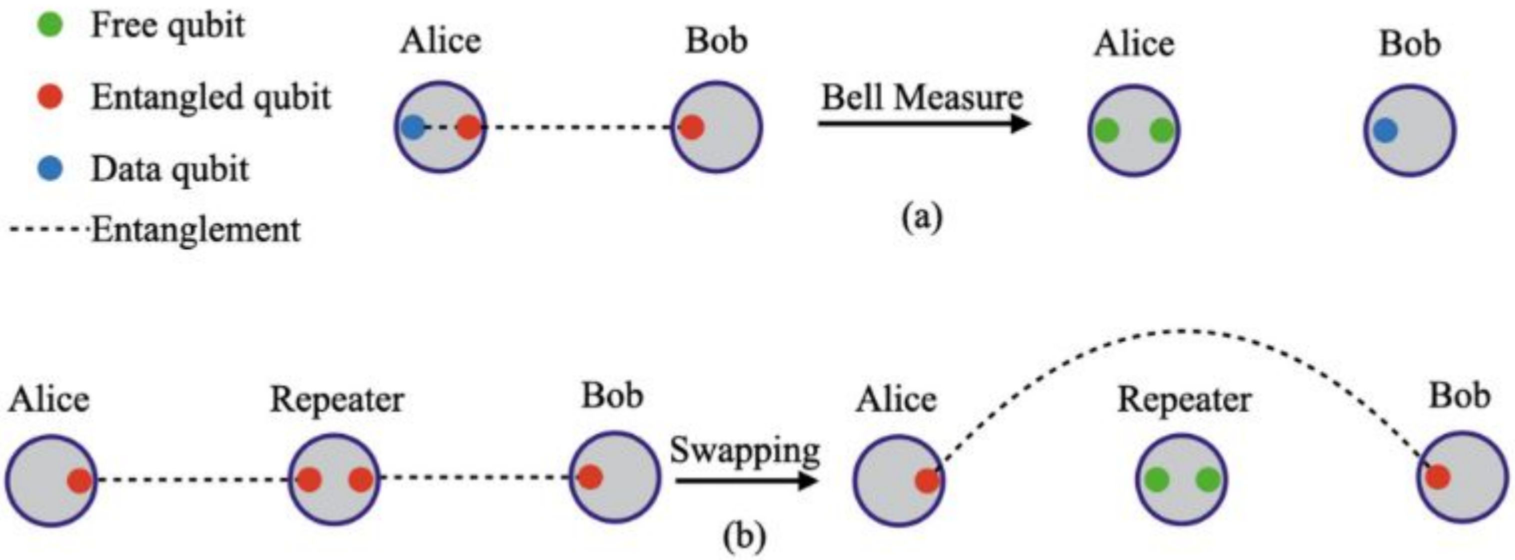
- Decoherence affects entangled states, also during the distribution to remote nodes.
- As a consequence, the distribution of entangled pairs suffers from distance limitations.
- Entanglement distribution over longer distances can be achieved through ***quantum repeaters*** (devices implementing the physical process called entanglement swapping.)
- A quantum repeater acts as intermediate node between source and destination, splitting so the total distance into two smaller sub-links.

Quantum Repeaters

- Source-repeater and repeater-destination entanglement couples
- Through the Bell state measurement (BSM) operations at the repeater, the entanglement is eventually distributed between source and destination
- (Then, due to the measurement, the original entanglements will be consumed.)



Quantum Repeaters



Quantum Repeaters

- Can be extended to multiple repeaters
- The only constraint on the swapping operations is a time constraint arising from the decoherence effects on the entangled pair
- As long as such operations are performed within the coherence time of the entangled pairs over the sub-links, the swapping ordering can be ignored (***deferred measurement principle***)

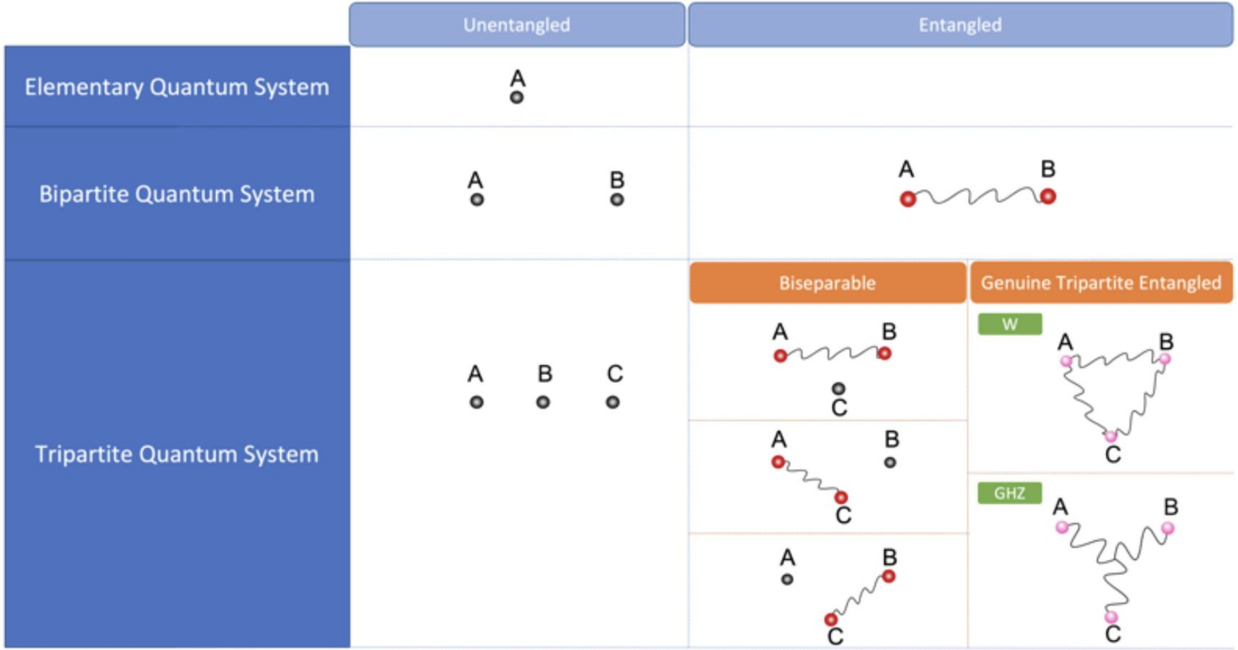
Classical bits vs Qubits

	Bit	Qubit	Entanglement
Temporal Constraints	**no**: can be stored indefinitely	**yes**: irreversibly degrades over time as a consequence of the decoherence process	
Duplication Constraints	**no**	**yes**: due to the no-cloning theorem	**no**: entangled states exploited in the network are in a known state, so they can be prepared repeatedly
Singleton	**yes**: self-contained entities		**no**: a single entangled qubit is useless in the network without the awareness of the remaining entangled qubits
Scope	**local**: any processing affects only the information available locally at the node		**non-local**: any processing of a single entangled qubit has an instantaneous effect on the remaining entangled qubits

Classical bits vs Qubits

	Bit	Qubit	Entanglement
State	**nearly stateless**: the node storing the bit does not need to retain any additional information	**stateful**: the node storing the qubit needs to retain at least temporal information	**profoundly stateful**: the node storing the entangled qubit needs to retain temporal information and the identities of the entangled nodes
Value	**local** and **pre-determined**: the encoded information is valuable only for the destination and not for the intermediate nodes		**global** and **dynamic**: the entangled state represents a valuable resource for any set of nodes sharing it
Order of Operations & Flow Direction	**yes**, with a strict ordering: source, intermediate nodes, destination	**flexible** the order: among the communication channels traversed by a quantum information carrier, can be indefinite	**flexible**: the swapping operation can happen simultaneously or without any particular order

Basic entanglement concepts for quantum systems

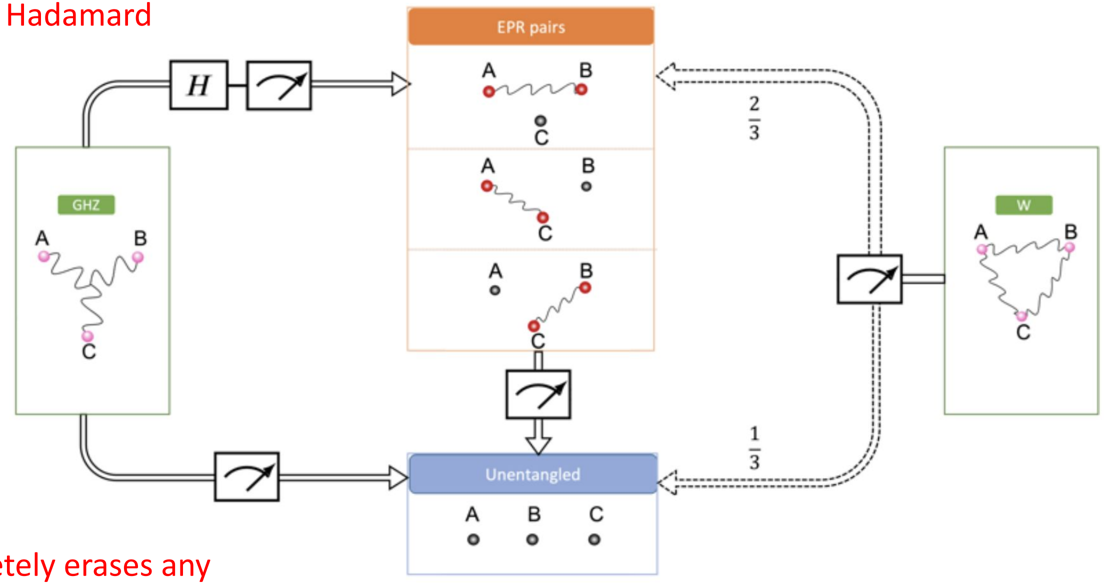


Bell states or EPR pairs

W states or GHZ states

Distillation of EPR pairs from GHZ and W states

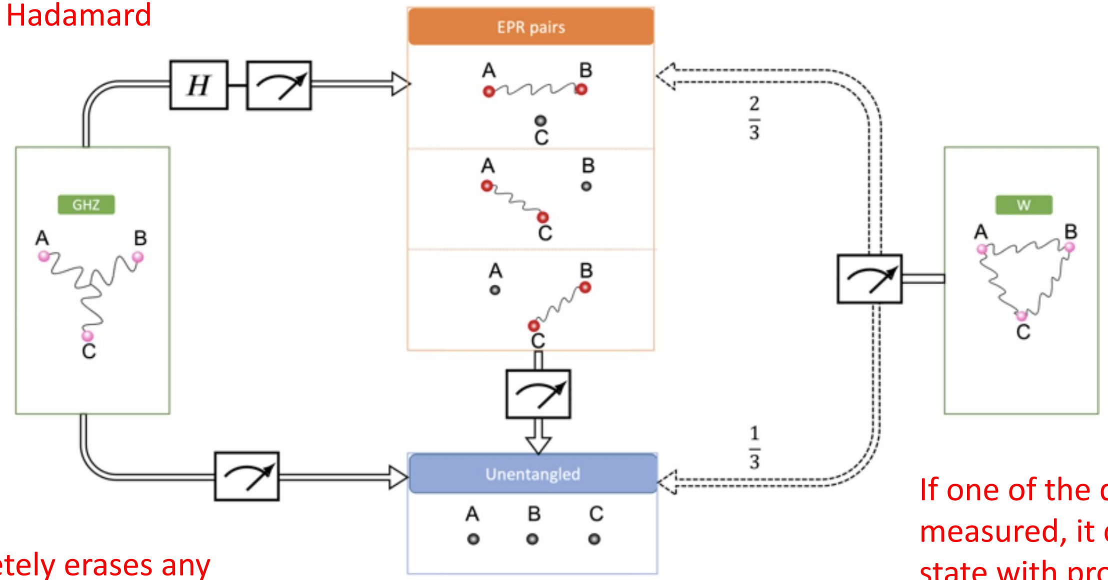
An EPR pair between two parties can be deterministically extracted from a 3-qubit GHZ by applying a Hadamard gate on the residual qubit



Any measurement completely erases any entanglement within a GHZ state, which collapses into a fully separable state.

Distillation of EPR pairs from GHZ and W states

An EPR pair between two parties can be deterministically extracted from a 3-qubit GHZ by applying a Hadamard gate on the residual qubit



Any measurement completely erases any entanglement within a GHZ state, which collapses into a fully separable state.

If one of the qubits in a W state is measured, it collapses in an unentangled state with probability equal to 1/3, while preserving maximal entanglement with probability equal to 2/3.

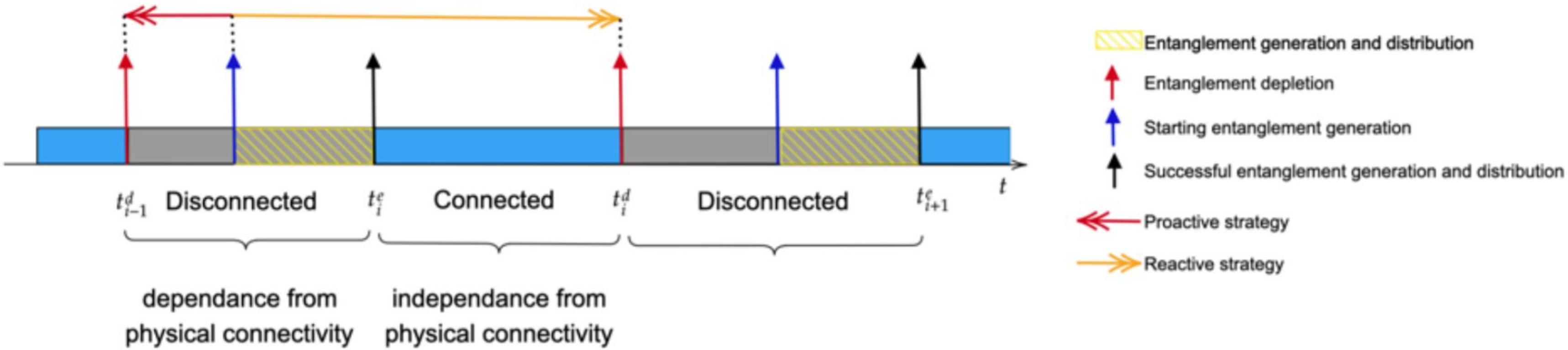
Entanglement Generation and Distribution

- The rate for direct transmission of qubits – including entangled states – over quantum links decays exponentially with the distance
  - This is why we need the entanglement swapping to keep the pairs
- Noise within generation and/or distribution process contributes to the generation of imperfect entangled states, with the imperfection usually reflecting in a non-maximal entangled state that jeopardizes the performance of the overlying communication protocol
  - We need entanglement distillation (purification) or quantum error correction

1. Virtual Connectivity

- In classical networks, a single concept of connectivity arises, referred to as physical connectivity (because successful transmission of information solely depends on the physical connections)
- Quantum teleportation enables the transmission of one qubit without any use of a quantum link as long as an entangled state – say an EPR pair for the sake of simplicity – is shared between two nodes
    - The qubit transmission is still possible even when the quantum link(s) go down after sharing an EPR pair! (This has a lot of implications in intermittent connectivity scenarios. <- this is my personal view)
    - Virtual quantum connectivity

1. Virtual Connectivity

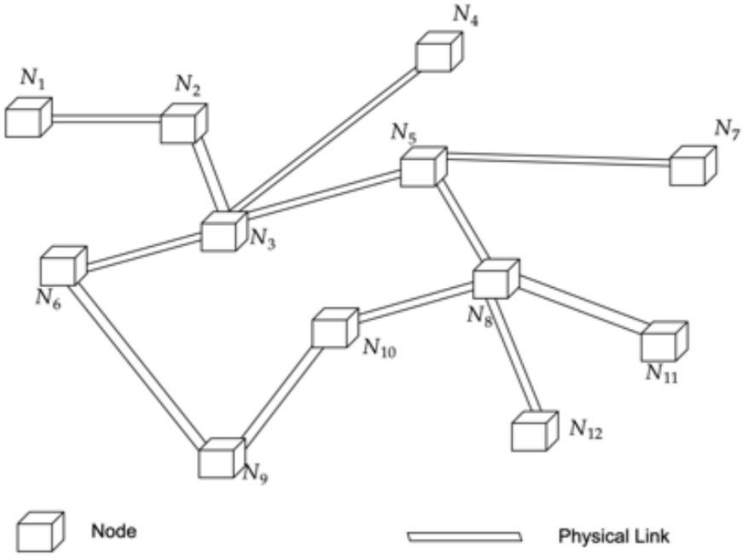


The virtual link is connected after the successful distribution of the entanglement, and it remains in such a state until entanglement is consumed.

1. When to prepare the virtual connectivity via entanglements

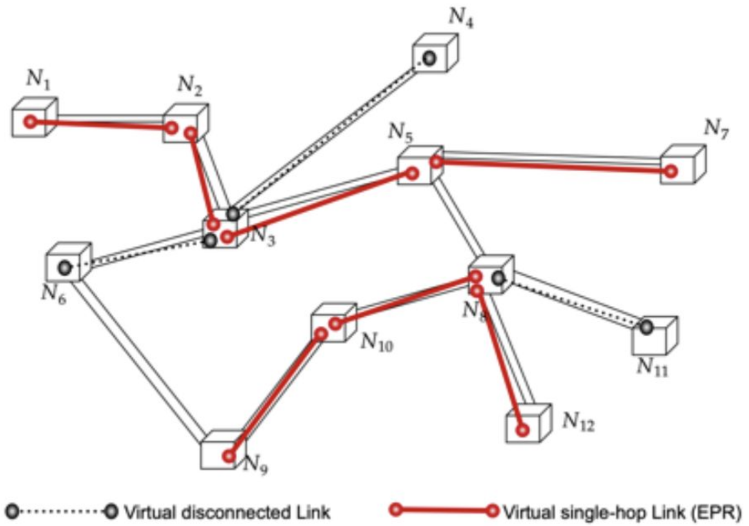
- There exist two different strategies for the entanglement distribution from a network engineering prospective: **proactive** or **reactive**
- **Proactive** strategies aim at early distribution of entanglement resources – ideally, with a new generation process starting as soon as the entanglement resource is depleted
- **Reactive** strategies aim at on-the-fly distribution of entanglement, with a new generation process starting on demand, when needed

2. Augmented Connectivity



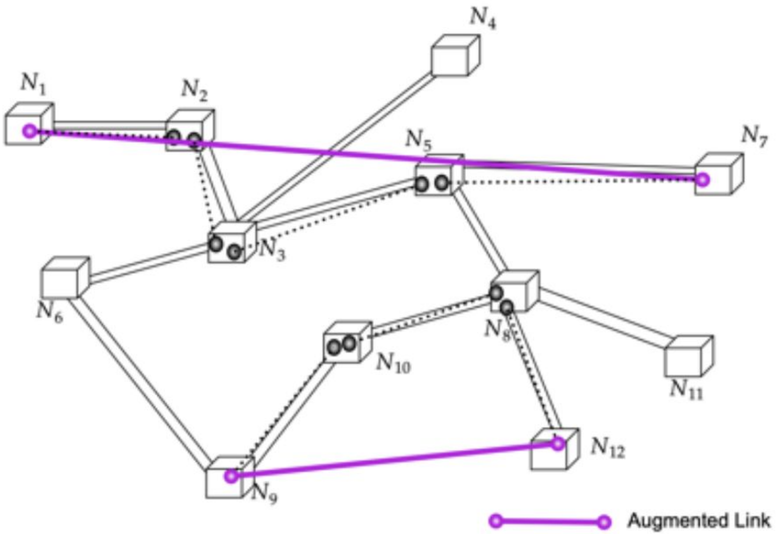
(a) Physical Graph, representing the physical connectivity within the network. In the example, Nodes N_3 and N_6 are connected by a physical communication link.

2. Augmented Connectivity



(b) Virtual Graph, representing the virtual connectivity within the network. In the example, although node N_6 is physically connected to N_3, it does not belong to the virtual graph as it does not share an EPR pair. Hence, the virtual link between N_3 and N_6 is disconnected.

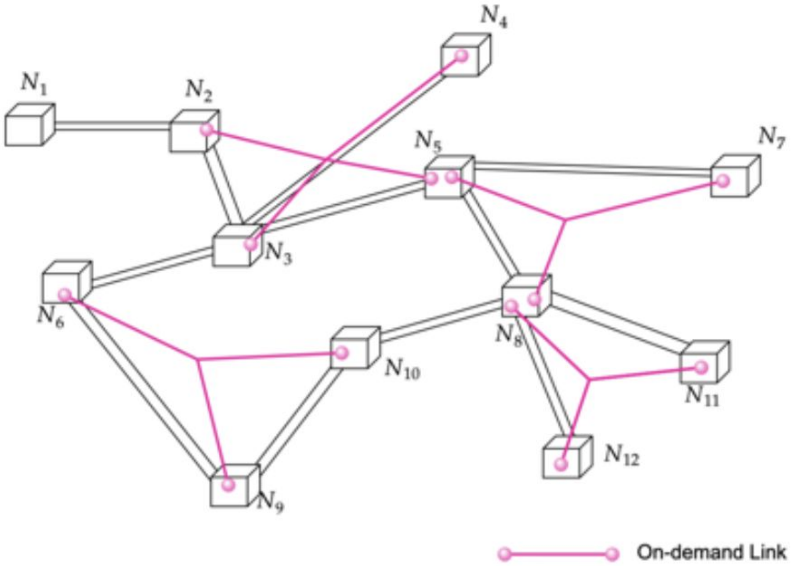
2. Augmented Connectivity



(c) Augmented Graph obtained through entanglement swapping. In the example, nodes N_1 and N_7 are directly connected by an augmented virtual link, although they are not connected in the physical graph nor in the virtual graph.

the augmented connectivity redefines the same concept of "neighborhood".Two nodes can be "neighbors" in the augmented graph whenever they are directly connected by an augmented link, even though they are physically remote located

3. On-Demand Connectivity



(d) On-Demand Graph obtained through multipartite entanglement. In the example, an EPR pair can be obtained on-demand between any node pair among nodes N_8, N_{11} and N_{12}.

Open Issues

Note: This survey was published in 2022. Some problems may be obsolete already.

- The decoherence process imposes strong temporal constraints on quantum information and quantum entanglement.
  - The synchronization of two end nodes should be coordinated strictly with classical communication
  - Decoherence should be modeled to consider the virtual, augmented, and on-demand links

Open Issues

Note: This survey was published in 2022. Some problems may be obsolete already.

- The number of communication qubits [10, 108] available at the network nodes.
    - Entanglement distribution among network nodes requires that at least one qubit at each processor, referred to as communication qubit, must be reserved for the generation of the entangled state
    - The more communication qubits are available within a network node, the more entanglement resource is available at that node (High entanglement rate achievable)
    - But the more communication qubits are available, the less resources – i.e., data qubits – are available for quantum computing
    - This is a clear tradeoff (especially qubits are precious resource at this point.)

Open Issues

Note: This survey was published in 2022. Some problems may be obsolete already.

- Neighbor discovery
  - Virtual connectivity makes it difficult to define it
  - Temporal nature of the connectivity and additional state to store at each node
  - With tripartite on-demand connectivity, “is a certain node one of my neighbors?” is a complex question.

Link Configuration for Fidelity-Constrained Entanglement Routing in Quantum Networks

Qiaolun Zhang, Nicola Di Cicco and Memedhe Ibrahimi (Politecnico di Milano, Italy); Raul Almeida Júnior (Federal University of Pernambuco (UFPE), Brazil); Alberto Gatto (Politecnico di Milano, Italy); Raouf Boutaba (University of Waterloo, Canada); Massimo Tornatore (Politecnico di Milano, Italy)

infocom 2025

Entanglement Routing (ER)

- To establish entanglement between non-adjacent nodes, intermediate nodes along the path may perform an operation called entanglement swapping on the link-level entanglements
- Since the distributed entanglements may be
  - (1) consumed by applications or
  - (2) lost due to decoherence,
- quantum networks must continuously generate entanglements between adjacent nodes and distribute them across non-adjacent nodes via entanglement swapping

Entanglement Routing (ER)

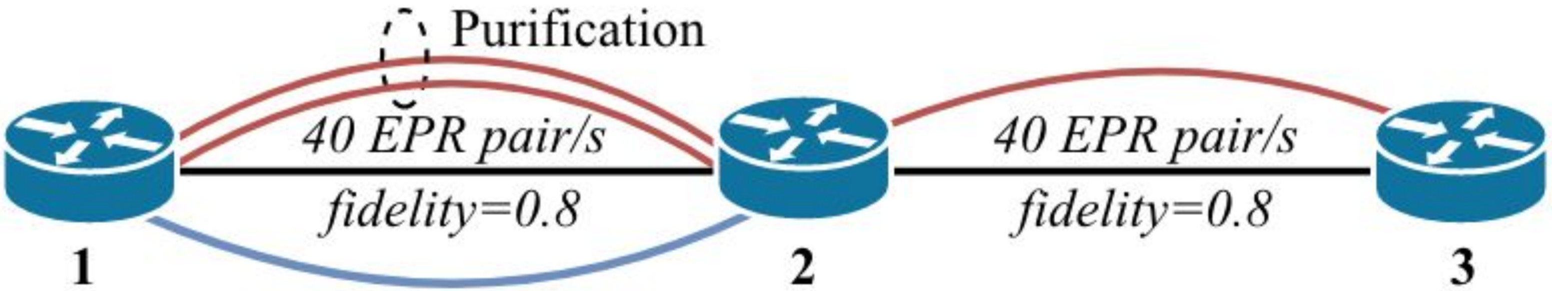
- One of the most fundamental questions in ER is how to distribute entanglements <u>with the guaranteed quality required by quantum applications</u>
- The quality of entanglements can be measured with the <u>**fidelity**</u> metric, which is defined as the probability that the set of qubits (two qubits in both end nodes) is in the desired state (entangled state) [15]

Link Configuration

- Most entanglement routing algorithms assume that quantum links can only create entanglements with a fixed fidelity, and
- address the fidelity constraint through entanglement purification, which can distill high-fidelity entanglement starting from low-fidelity entanglements
- However, recent pioneering research has demonstrated that quantum links can **create higher fidelity entanglements by trading off the entanglement generation rate**, and vice versa [13, 17]
  - This is what they call link configuration
  - I.e., fidelity/entanglement-rate combination

An Example

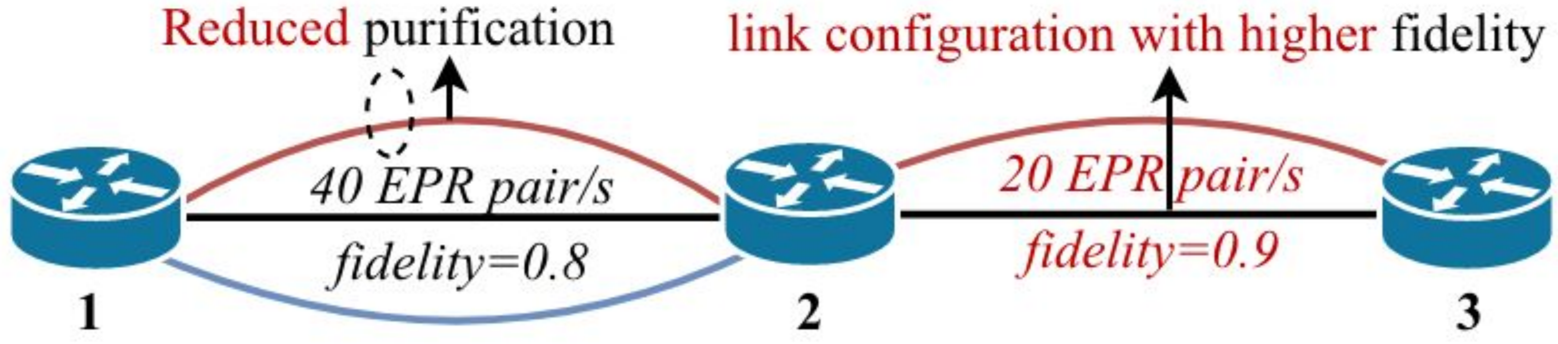
- Each link can only create entanglements with a fixed fidelity of 0.8 and a generation rate of 40 EPR pair/s
- E2E request (1,3) requires a higher fidelity
- → purification is performed on link (1,2), consuming some EPR pairs



(a) Conventional approach with purification

An Example

- Link (2, 3) is configured to generate entanglements with a fidelity of 0.9 and a generation rate of 20 EPR pairs/s
    - E2E fidelity is positively correlated with the diodelity involved in swapping
- The fidelity constraint of request (1, 3) might be satisfied without purification in link (1, 2), in which more entanglements can be used to serve



(b) Link configuration for entanglement routing

The Fidelity-aware Entanglement Routing

- **Given**
  - a quantum network topology
  - node memory capacity
  - a set of possible link configurations
  - a set of time-slots
  - the relation between fidelity and the number of required entanglements per purification round
  - set of requests with the desired generation rate and fidelity, along with swapping success probability,
- **Decide** the link configuration, routing and purification in each edge for each request
- **Constrained by** the generation rate of each link, node memory capacity constraint, and the fidelity constraint
- With the objective of maximizing the served generation rate

The Fidelity-aware Entanglement Routing

In this paper, they cite many physics papers to provide reasonable assumptions on the tradeoff and configuration benefits

- **Given**
  - a quantum network topology
  - node memory capacity
  - a set of possible link configurations
  - a set of time-slots
  - the relation between fidelity and the number of required entanglements per purification round
  - set of requests with the desired generation rate and fidelity, along with swapping success probability,
- **Decide** the link configuration, routing and purification in each edge for each request
- **Constrained by** the generation rate of each link, node memory capacity constraint, and the fidelity constraint
- With the objective of maximizing the served generation rate

Proposed Solution

- They formulate the problem formally in MILP
- This is reduced to the multi-commodity flow problem, which is NP-hard
- A greedy-heuristic to accept as many requests as possible
- Also, the algorithm computes the average request arrival rate at each link in the past and greedily determines the link configurations based on the demand
- The initial value of the fidelity setting is given by the initial algorithm
- Subsequently, Bayesian Optimization (BO) is used to improve the link configuration for ER
    - Iteratively searches link configurations

Experiments

- CPLEX solver for the MILP
- Two topologies, a small topology (the German topology in Ref. [36] with 7 nodes and 11 links) and a large topology (the USnet topology in Ref. [37] with 24 nodes and 43 links).
- Link setups: a fidelity of 0.8 with a generation rate uniformly distributed in [250, 500] EPR pair/s similar to Ref. [11]
- The memory storage is set to 12,000 as in Ref. [11]
  - 12,000 qubits exists at each quantum node for potential entanglement
  - Current experiments have demonstrated a memory capacity of 225 qubits [27]
  - This work envisions technological advances and assumes a larger capacity
- The time duration of each time slot is 10 seconds. The number of considered time slots for German and USnet topology is set to 2 and 8, respectively

Comparison

- We evaluate the improvement brought by link configuration on the performance of three baseline ER algorithms
  - **Progressive Filling (PF) [21]**: PF performs purification on each link to achieve the same pre-defined fidelity threshold
  - **Low-complexity Routing (Qleap) [10]**: Qleap performs purification on links to achieve a fidelity threshold according to the fidelity required by each request
  - **Purification-enabled Iterative Routing (Qpath) [10]**: Qpath is an advanced version of Qleap, which performs purification on critical links that have the highest improvement in fidelity with the least entanglements

Metrics

- the acceptance ratio (defined as the percentage of served generation rate over the total requested generation rate)
- The resource utilization (probably, the used qubits in nodes over all qubits available)
- End-to-end fidelity (next slide)

E2E Fidelity vs Link Fidelity

$$\rho_w = w|\Psi^+\rangle\langle\Psi^+| + (1-w)\frac{\mathbb{I}_4}{4}$$

A Bell state　　　　A completely mixed state (noisy!)

- w: the Werner parameter, which denotes the extent to which the Werner state retains the entangled state |Ψ+⟩ versus being mixed state I4
- Ideal quantum channel, w = 1
- Partially noisy channel, w = 0.9, 0.8, …
- Maximally noisy, w = 0.5
  - E.g., Quantum key distribution (QKD): security may be compromised below F ≈ 0.7

E2E Fidelity vs Link Fidelity

- After performing entanglement swapping, the fidelity of a newly created entanglement between non-adjacent nodes can be calculated through the Werner parameters
- Assume that we have a path φ and the Werner parameter of the entanglement in link e traversed by path φ is w_e

$$w_{\phi}^{e2e} = \prod_{e \in k} w_e$$

Linear transformation for LP

$$\log(w^{req}) \leq \sum_{e \in k} \log(w_e)$$

Results

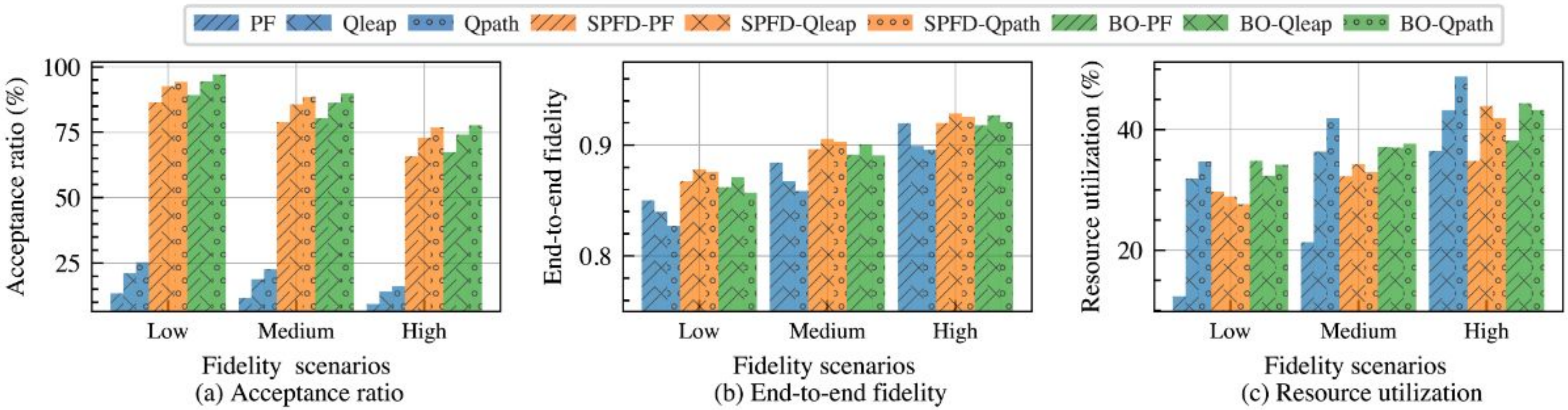


Fig. 6: Performance of link configuration for different fidelity scenarios in the USnet topology.

low fidelity scenario, medium fidelity scenario, and high fidelity scenario, with average requested fidelities of 0.80, 0.85, and 0.90, respectively

Results

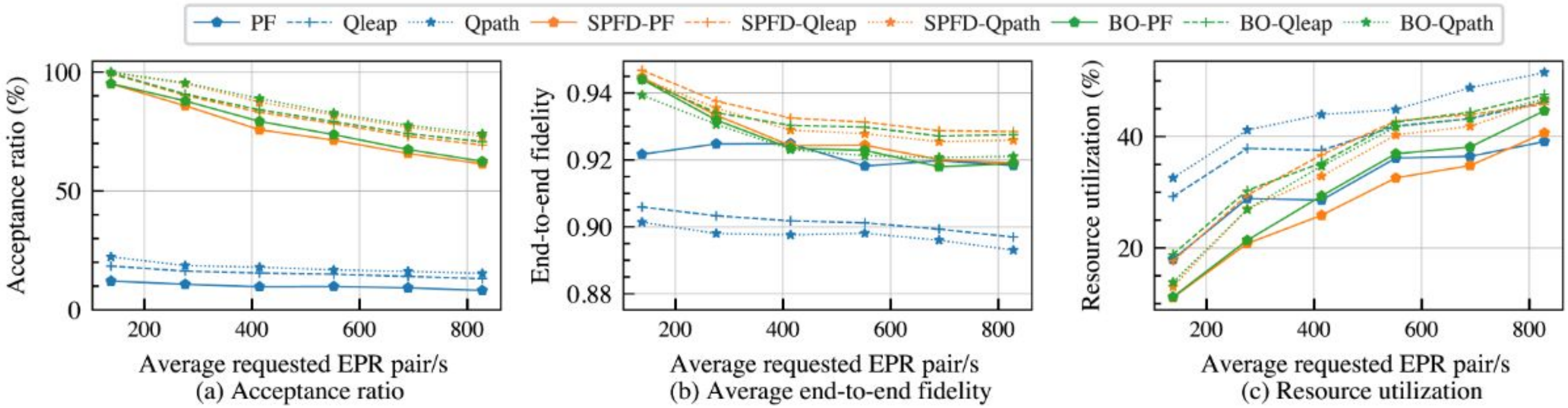


Fig. 5: Performance of link configuration for different load scenarios in the USnet topology.

Notes

"Recently it has been discovered that the order among the communication channels traversed by a quantum information carrier can be indefinite"

1 EPR pair = 1 ebit

I was not sure how much decoherence is considered