A Link Layer Protocol for Quantum Networks

Axel Dahlberg , Matthew Skrzypczyk , Tim Coopmans , Leon Wubben Filip Rozpędek, Matteo Pompili , Arian Stolk , Przemysław Pawełczak , Robert Knegjens , Julio de Oliveira Filho , Ronald Hanson, Stephanie Wehner

QuTech, Delft University of Technology and TNO
Kavli Institute of Nanoscience, Delft University of Technology

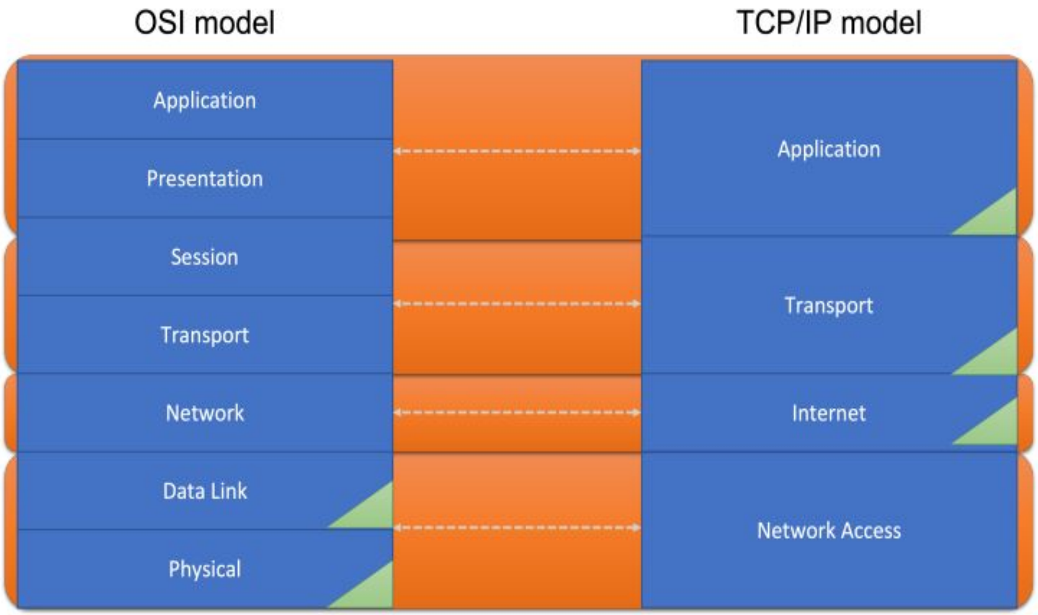
Diego Abreu (Federal University of Pará) - Brazil
Advisor: Prof. Antônio Abelém

Presentation Outline

1. Quantum Internet Stack
2. Design Considerations for the Link Layer
3. The Physical Layer Protocol
4. The Link Layer Protocol
5. Experiments and Results
6. Conclusion and Main Outcomes

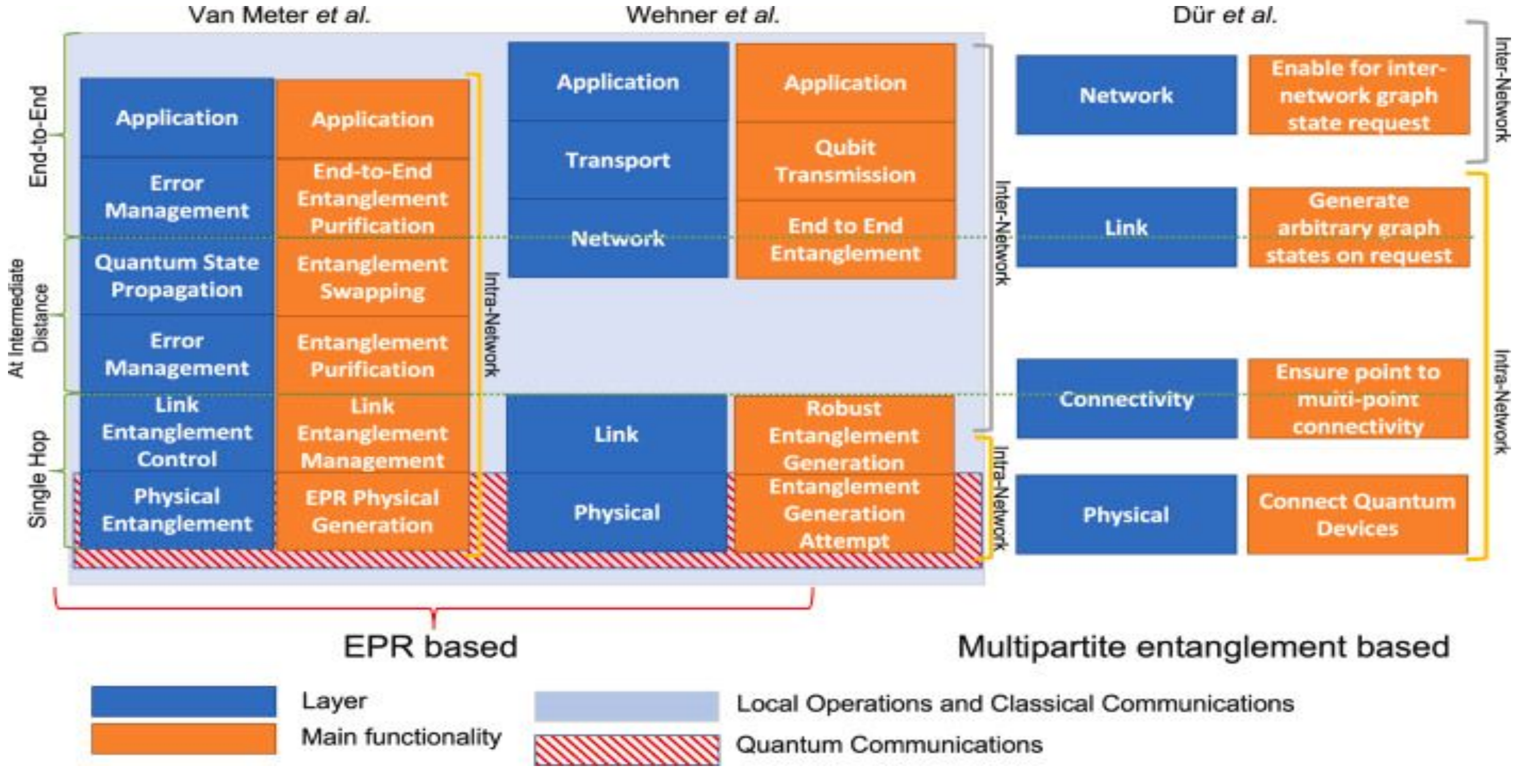
Why do we need a Internet Stack

- Helps develop specific applications
- Each layer serves the other one
- Modularity, Scalability
- Standardization
- Works well in the classical Internet



Illiano, Jessica, et al. "Quantum internet protocol stack: A comprehensive survey." *Computer Networks* 213 (2022): 109092.

Quantum Internet Stack



Illiano, Jessica, et al. "Quantum internet protocol stack: A comprehensive survey." *Computer Networks* 213 (2022): 109092.

Quantum Internet Stack

Wehner *et al.* Quantum Internet Stack

- Application Layer : Quantum Application Protocols
- Transport: End-to-end Qubit Delivery
- Network: Long-distance entanglement generation
- Link: Entanglement Generation on a link
- Physical: Quantum Device Layer

Application	
Transport	Qubit transmission
Network	Long distance entanglement
Link	Robust entanglement generation
Physical	Attempt entanglement generation

Dahlberg, Axel, et al. "A link layer protocol for quantum networks." *Proceedings of the ACM special interest group on data communication*. 2019. 159-173.

Wehner *et al.* Quantum Internet Stack

1. Physical Layer:

- Main Task: **synchronization**
- Actual quantum hardware devices and physical connections such as fibers.
- No decision making elements, keep no state about the production of entanglement

2. Link layer

- Main Task : robust entanglement generation service.
- Turn the physical layer
- Requests can be made by higher layers to the link layer to produce entanglement
- Request are either fulfilled or result in a time-out .

Wehner *et al.* Quantum Internet Stack

3. Network Layer:

- **Main Task** : Produce Long-distance entanglement
- Keeps track of entanglement in the network, and which may choose to pregenerate entanglement to service later requests from higher layers

4. Transport layer:

- **Main Task:** Transmitting qubits deterministically (e.g. using teleportation)
- Use of a dedicated layer allows two nodes to preshare entanglement that is used as applications of the system demand it

5. Application Layer:

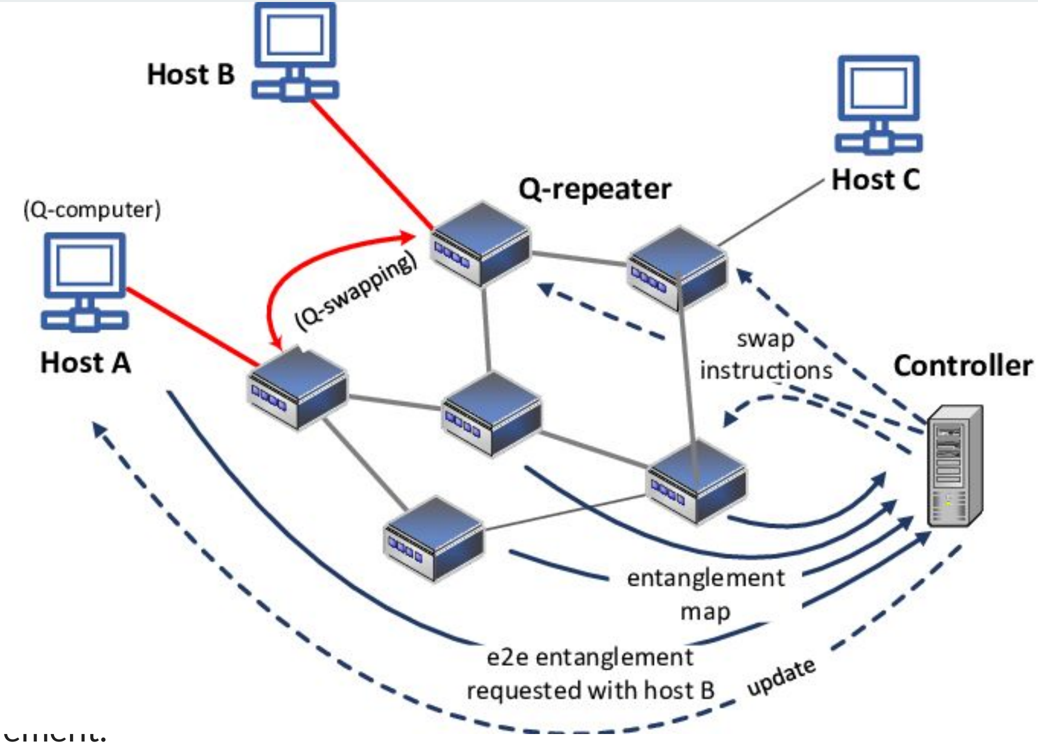
- **Main Task:** Generate Requests. Multiple services

Design Considerations

- **Quantum Network Devices**
- Use Cases
- Desired Service
- Physical Platform

- **Controllable Node**
  - **Full stack, decisions**
- **Repeter Nodes/automated nodes**
  - Devices triggered at a given time instant

responsible of the actual attempt to generate entanglement.



Nguyen, Tu & Ambarani, Kashyab & Le, Linh & Djordjevic, Ivan & Zhang, Zhi-Li. (2022). A Multiple-Entanglement Routing Framework for Quantum Networks.

Use Cases

- **Quantum Network Devices**
- **Use Cases**
- Desired Service
- Physical Platform

1. Measure Directly (MD)

-Both qubits are immediately measured to produce classical correlations.

-no quantum memory is needed to store the entanglement and it is not necessary to produce all entangled pairs at the same time.

- QKD , secure identification

2. Create and Keep (CK)

-require genuine entanglement,even multiple entangled pairs to exist simultaneously

3. Send Qubit (SQ)

-ask for the transmission of (unknown) qubits,using teleportation.

4. Network Layer (NL)

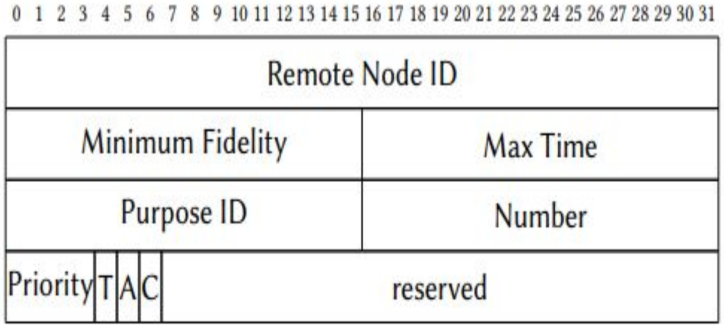
-producing entanglement between neighboring nodes

Desired Service

- **Quantum Network Devices**
- **Use Cases**
- **Desired Service**
- Physical Platform

Performance Metrics

- Throughput (entangled pairs/s)
- **Latency**
  - Latency per request
  - Latency per pair
  - **scaled latency** (Latency per request per number of requested pairs)



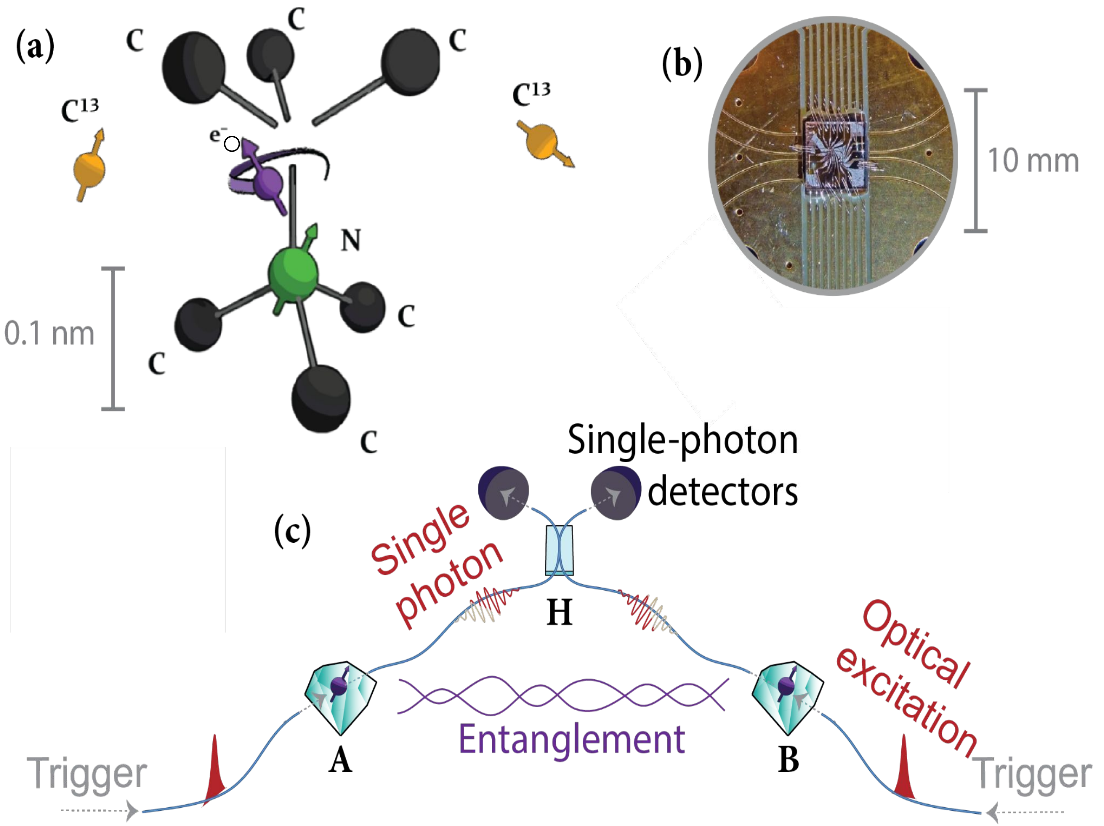
Dahlberg, Axel, et al. "A link layer protocol for quantum networks." *Proceedings of the ACM special interest group on data communication*. 2019. 159-173.

Physical Platform

- Quantum Network Devices
- Use Cases
- Desired Service
- Physical Platform

- Nodes A,B and a Heralding Station (H)
- Two Types of Qubits
  - Memory Qubits
  - Communication Qubits

- Similar implementations with
  - Ion Traps
  - Neutral Atoms

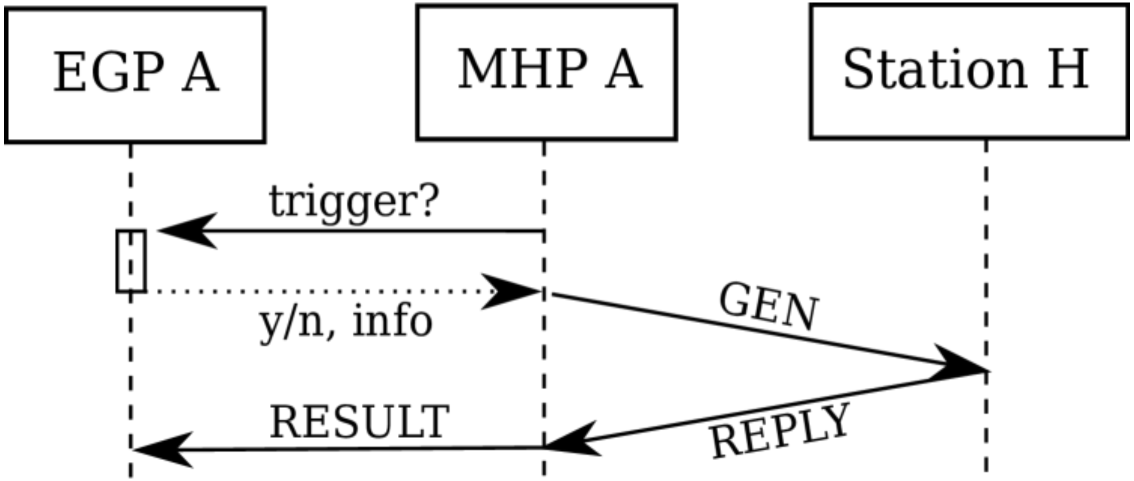


Nitrogen-Vacancy (NV) center platform

Dahlberg, Axel, et al. "A link layer protocol for quantum networks." *Proceedings of the ACM special interest group on data communication*. 2019. 159-173.

Physical Layer Protocol : Midpoint Heralding Protocol (MHP)

- Heralded Entanglement
  - Confirm entanglement generation by performing heralded entanglement generation
- On top of physical implementations
  - **Additional control information**
- Can be adapted to other forms of heralded entanglement

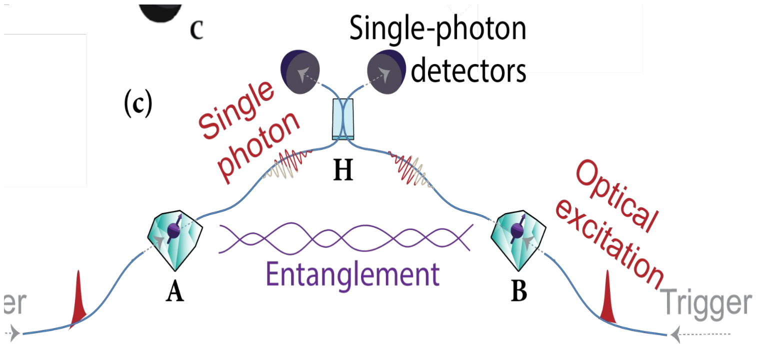


EGP : Entanglement Generation

Dahlberg, Axel, et al. "A link layer protocol for quantum networks." *Proceedings of the ACM special interest group on data communication*. 2019. 159-173.

Physical Layer Protocol : Midpoint Heralding Protocol (MHP)

1. A microwave pulse prepares the communication qubit depending on a parameter **α**
2. Laser pulse trigger the photon emission (total duration 5.5μs)
3. A pair ($|\Psi+\rangle$ or $|\Psi-\rangle$) is successfully produced
   a. with fidelity $F \approx 1 - \alpha$
   b. with probability $p_{succ} \approx 2\alpha p_{det}$. Where $p_{det} \ll 1$ is the probability of emitting a photon followed by heralding success.



- **tattempt :** Time of an attempt (time preparing the communication qubit until receiving a reply from H, and completion of any post-processing such as moving to memory),
- **rattempt** : the maximum attempt rate (maximum number of attempts that can be performed per second not including waiting for a reply from H or post-processing).

Physical Layer Protocol : Midpoint Heralding Protocol (MHP)

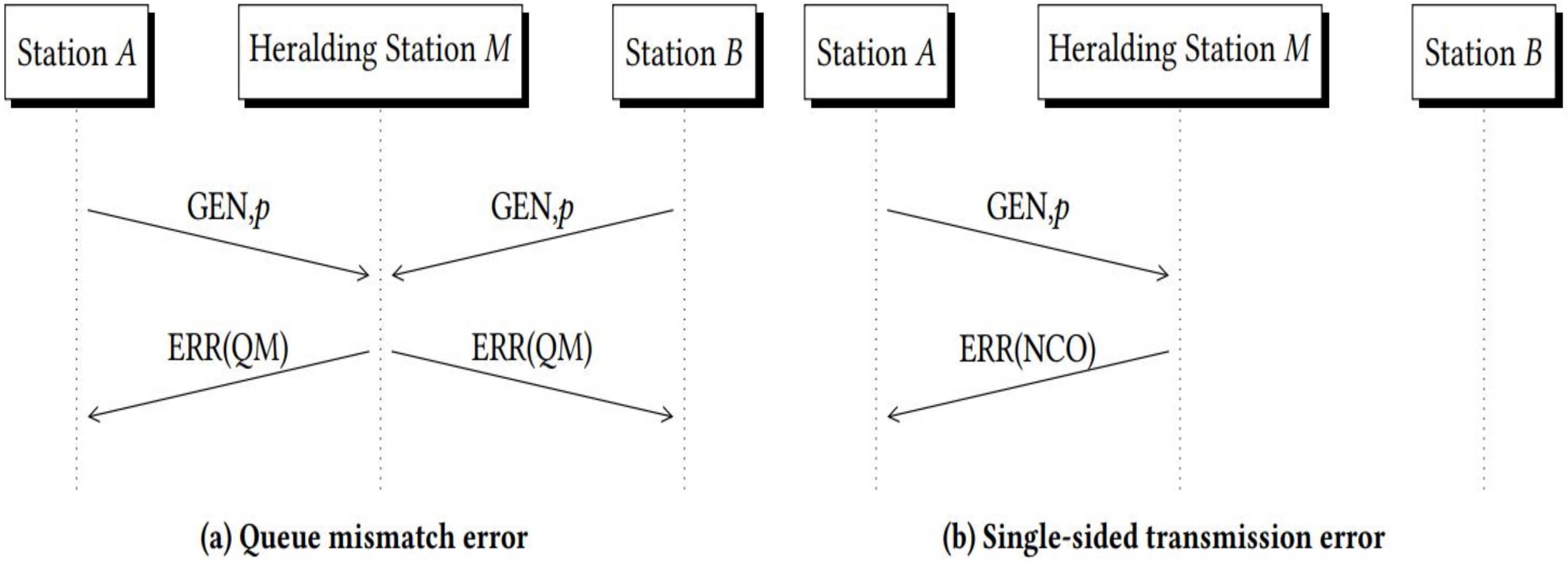
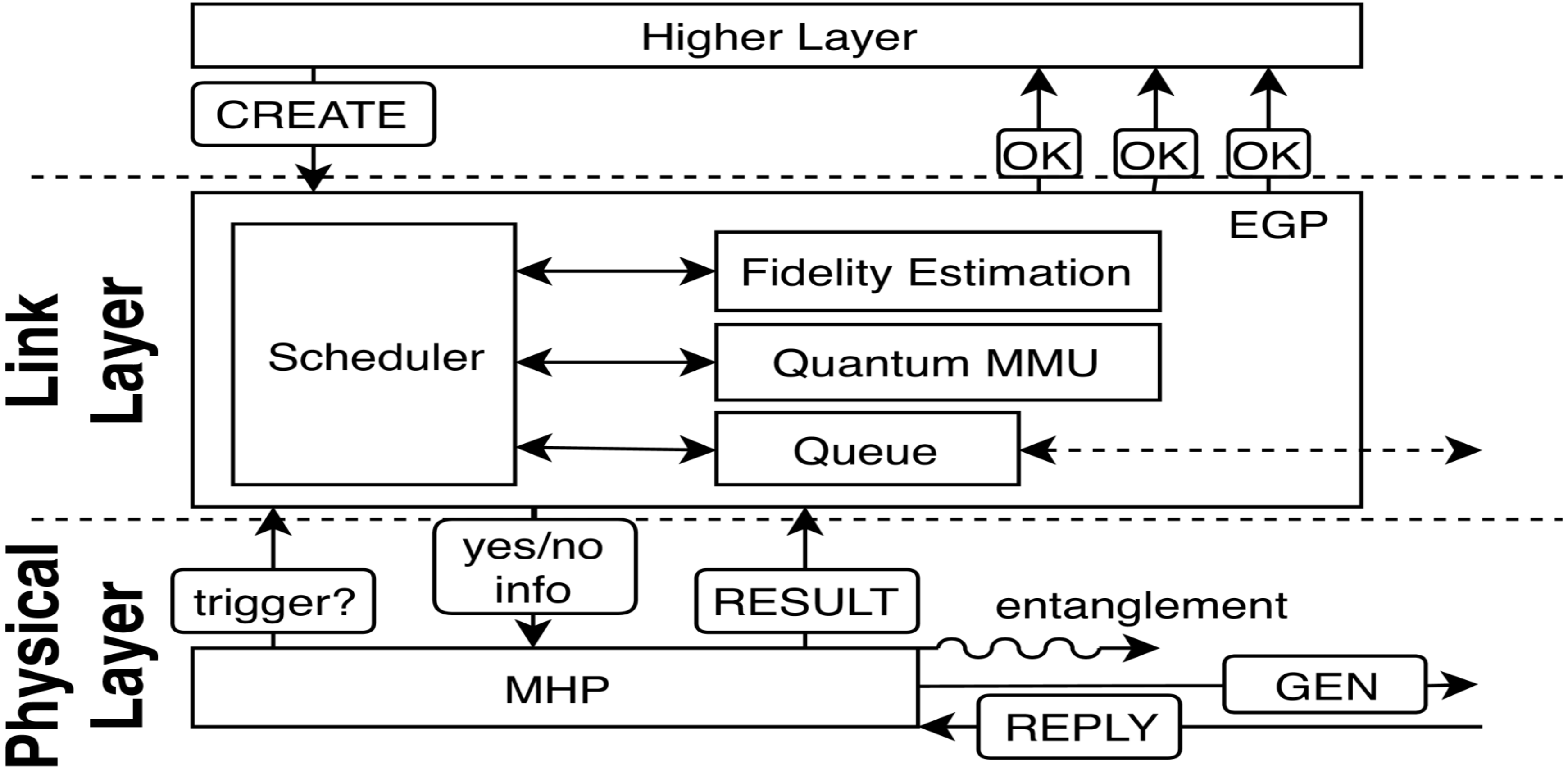


Figure 26: Timeline of two types of errors within MHP. For a definition of GEN and REPLY message refer to Figure 27 and Figure 28, respectively. QM and NCO refer to specific fields of the REPLY message (i.e. OT field), i.e. QUEUE_MISMATCH and NO_MESSAGE_OTHER, respectively; both error types are explained in Protocol E.2.

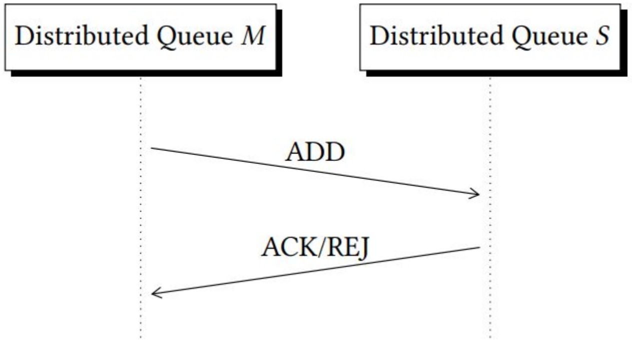
Dahlberg, Axel, et al. "A link layer protocol for quantum networks." *Proceedings of the ACM special interest group on data communication*. 2019. 159-173.

Link Layer Protocol : Entanglement Generation Protocol (EGP)



Dahlberg, Axel, et al. "A link layer protocol for quantum networks." *Proceedings of the ACM special interest group on data communication*. 2019. 159-173.

Link Layer Protocol : Entanglement Generation Protocol (EGP)



Main Components

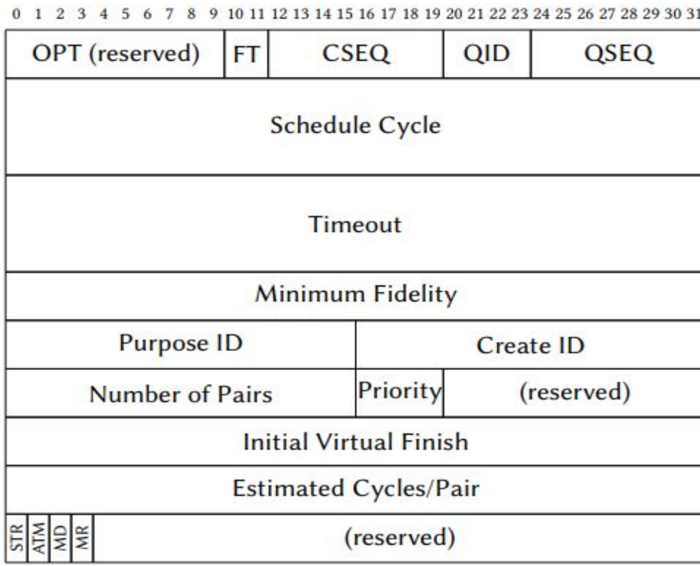
- Distributed Queue
  - queue comprised of synchronized local queues at the controllable nodes
  - separate requests based on priority
  - simple two-way handshake for enqueuing items
- Quantum Memory Manager (MMU):
  - Which qubits to use.
- Fidelity Estimation
  - Base on: known hardware capabilities, quality of the memory, quality of operations; and intersperses test rounds

Dahlberg, Axel, et al. "A link layer protocol for quantum networks." *Proceedings of the ACM special interest group on data communication*. 2019. 159-173.

Link Layer Protocol : Entanglement Generation Protocol (EGP)

Scheduler

- FCFS: First-come-first-serve with a single queue.
- LowerWFQ: NL are always service first (strict priority) and weighted fair queue (WFQ) is used between **CK (weight 2)** and MD (weight 1).
- HigherWFQ: NL are always service first (strict priority) anc a weighted fair queue (WFQ) is used between **CK (weight 10)** and MD (weight 1).



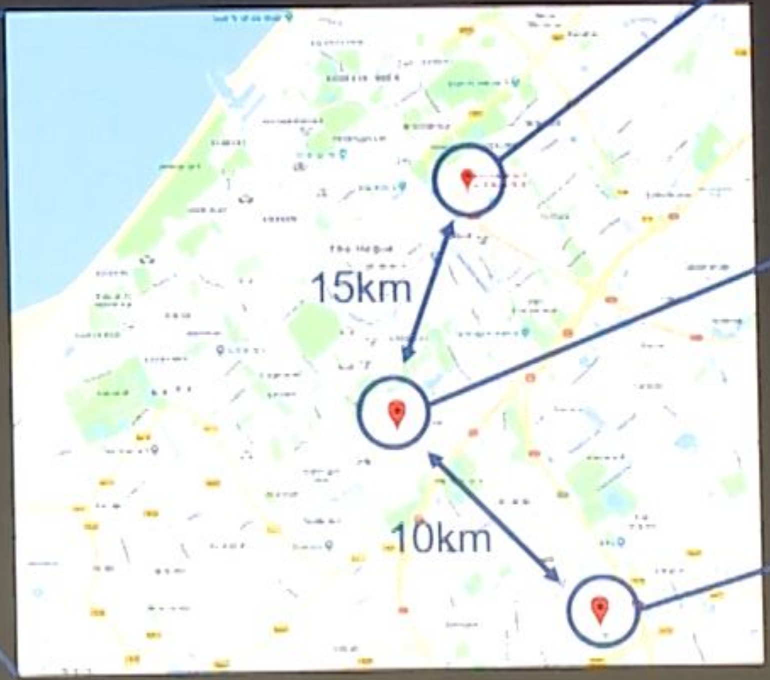
Dahlberg, Axel, et al. "A link layer protocol for quantum networks." *Proceedings of the ACM special interest group on data communication*. 2019. 159-173.

Experiments and Results

Experimental Scenarios

1. Two Scenarios
   a. LAB Scenario:
      i. distance to station 1m, , psucc ≈ α · 10−3
   b. The QL2020 scenario : long networks
2. Simulation: Implemented in purpose built discrete event simulator for quantum networks (NetSquid [1], Python/C++) based on DynAA [41]
   a. All simulations were performed on the supercomputer Cartesius at SURFsara [2], in a total of 2578 separate runs, using a total of 94244 core hours, and 707 hours time in the simulation (~250 billion MHP cycles).
   b. Long runs : 120 wall time hours
   c. Short runs: 24 wall time hours

Simulation Example: QL2020



KPN PB400
node location

KPN PBX
detector location

TU Delft
node location

Assumed loss
0.1 dB/splice
0.3 dB/km

How the simulation works

In each MHP cycle:

-New requests for k pairs (max kmax)

-Random kind of service : (NL, CK, MD)

- Probability is f · psucc/(E · k)
  - psuc: probability of an attempt being successful
  - f : a fraction determining load of the system
  - E: is the expected number of MHP cycles to make one attempt.

-In the Lab: E = 1 for MD,1.1 for NL/CK
-In the QL2020: **16 cycles** for NL/CK (due to classical communication delays with H (145μs)

Usage pattern	*NL*	*CK*	*MD*
UNIFORM	$f = 0.99 \cdot 1/3, k_{max} = 1$	$f = 0.99 \cdot 1/3, k_{max} = 1$	$f = 0.99 \cdot 1/3, k_{max} = 1$
MORENL	$f = 0.99 \cdot 4/6, k_{max} = 3$	$f = 0.99 \cdot 1/6, k_{max} = 3$	$f = 0.99 \cdot 1/6, k_{max} = 256$
MORECK	$f = 0.99 \cdot 1/6, k_{max} = 3$	$f = 0.99 \cdot 4/6, k_{max} = 3$	$f = 0.99 \cdot 1/6, k_{max} = 256$
MOREMD	$f = 0.99 \cdot 1/6, k_{max} = 3$	$f = 0.99 \cdot 1/6, k_{max} = 3$	$f = 0.99 \cdot 4/6, k_{max} = 256$
NONLMORECK	$f = 0, k_{max} = 3$	$f = 0.99 \cdot 4/5, k_{max} = 3$	$f = 0.99 \cdot 1/5, k_{max} = 256$
NONLMOREMD	$f = 0, k_{max} = 3$	$f = 0.99 \cdot 1/5, k_{max} = 3$	$f = 0.99 \cdot 4/5, k_{max} = 256$

Experiments and Results

Simulation data

Experiments and Results

Experimental Parameters

M Request Type

1. We may choose to measure immediately before receiving a reply (here readout 3.7µs)
2. For M the communication qubit is measured before receiving the reply from H and thus allows for multiple attempts to overlap

Act the same for the Lab and QL200: Always measure immediately before parsing the response for H

- tattempt = 1/rattempt = 10.12 µs
  - Includes electron readout 3.7 µs, photon emission 5.5 µs and a 10 % extra delay to avoid race conditions.

Experiments and Results

Experimental Parameters

K Request Type

1. We may store the pair in the communication qubit, or move to a memory qubit (duration of **1040μs** for the qubit considered). The quality of this qubit degrades as we wait for H to reply.
2. for K, if the reply from H is failure, then no move to memory is done.

Lab Parameters

- tattempt = 1045 μs
- 1/rattempt ≈ 11 μs
- as memory qubits need to be periodically initialized (330 μs every 3500 μs)

QL2020 Parameters

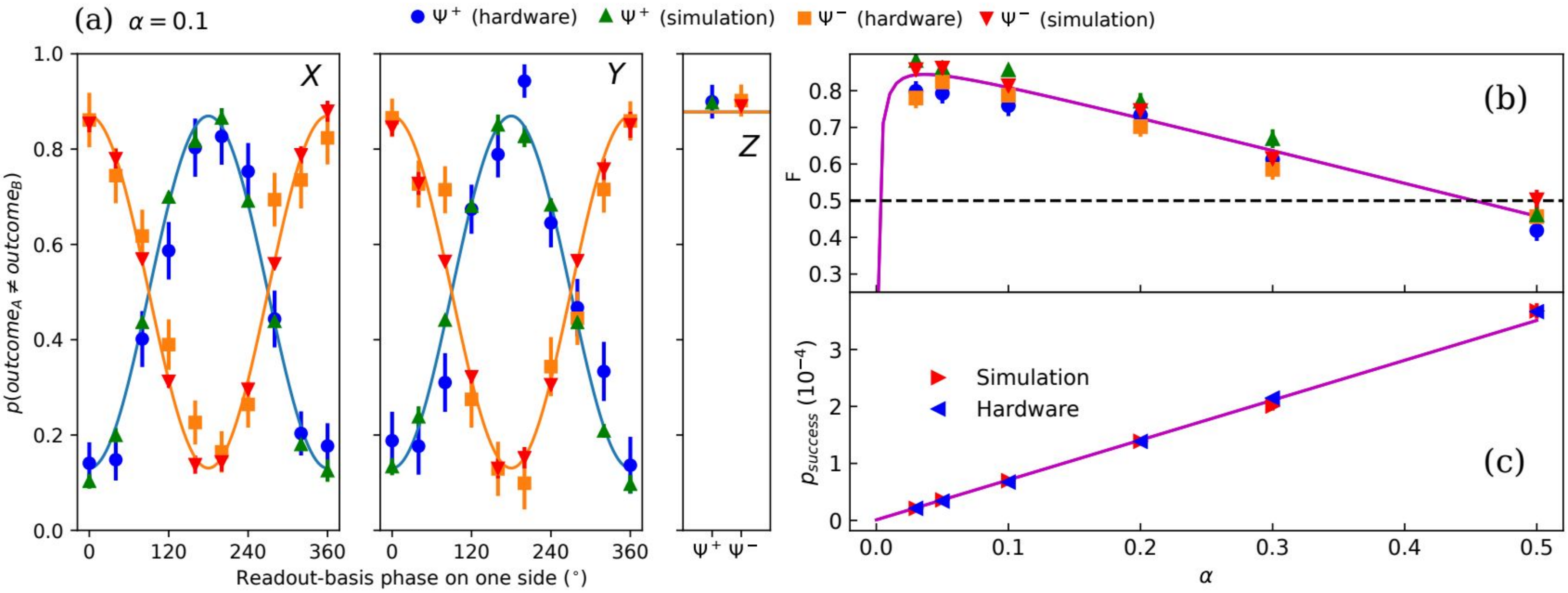
- A to H (10km), B to H (15km), delay of 72.6μs,fiber losses at 1588nm 0.5 dB/km
- tattempt = 1185 μs
- 1/rattempt ≈ 165 μs

	(Unsquared) fidelity	Duration/time	Experimentally realized
Electron T_1	-	2.86 ms	> 1h[3]
Electron T_2^*	-	1.00 ms	1.46 s[3]
Carbon T_1	-	∞	> 6m [21]
Carbon T_2^*	-	3.5 ms	$\approx$ 10ms [21]
Electron single-qubit gate	1.0	5 ns	> 0.995 (100 ns) [60]
E-C controlled-$\sqrt{X}$-gate (E=control)	0.992	500 μs	0.992 (500-1000 μs) fig 2 in [60]
Carbon Rot-Z-gate	0.999	20 μs	1.0 (20 μs) [93]
Electron initialization in $\|0\rangle$	0.95	2 μs	0.99 (2 μs) [82]
Carbon initialization in $\|0\rangle$	0.95	310 μs	0.95 (300 μs) [32]
Electron readout	0.95 ($\|0\rangle$), 0.995 ($\|1\rangle$)	3.7 μs	0.95 ($\|0\rangle$), 0.995 ($\|1\rangle$) (3-10 μs) [53]

Table 6: Gates and coherence times used in simulation. Values used in the simulation corresponding to LAB. We remark that since these are custom chips, no two are exactly identical. Individual values have since seen significant improvements (Experimentally realized), but not been realized simultaneously for producing entanglement that would allow a direct comparison to simulation. We have thus focused in simulation only what enables a comparison to data gathered from entanglement generation on hardware.

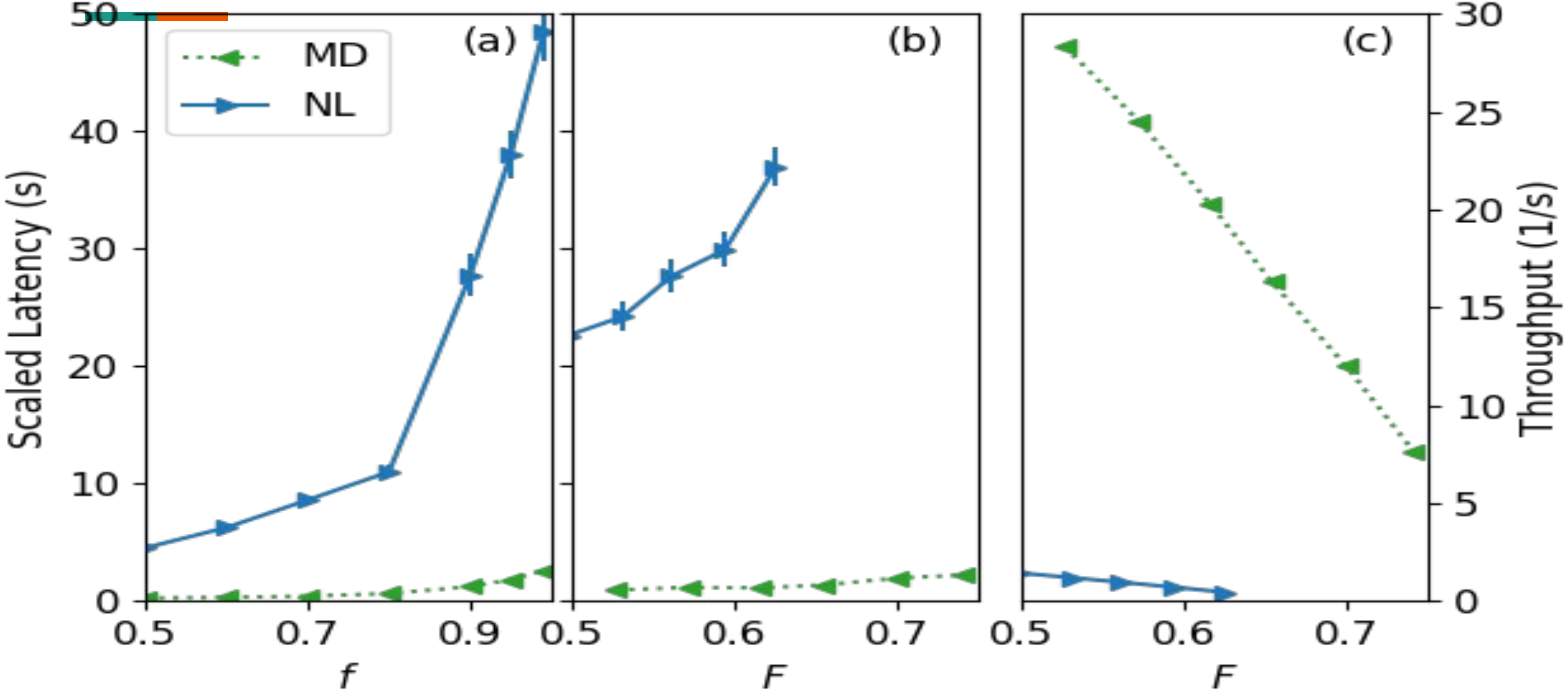
Experiments and Results

Validation of simulation: Comparison of simulation results with data from NV hardware from [53] (Lab scenario), showing good agreement



Experiments and Results

Performance trade-offs (with only a single kind of request (MD/CK/NL)



Dahlberg, Axel, et al. "A link layer protocol for quantum networks." *Proceeding*
ACM special interest group on data communication. 2019. 159-173.

Experiments and Results

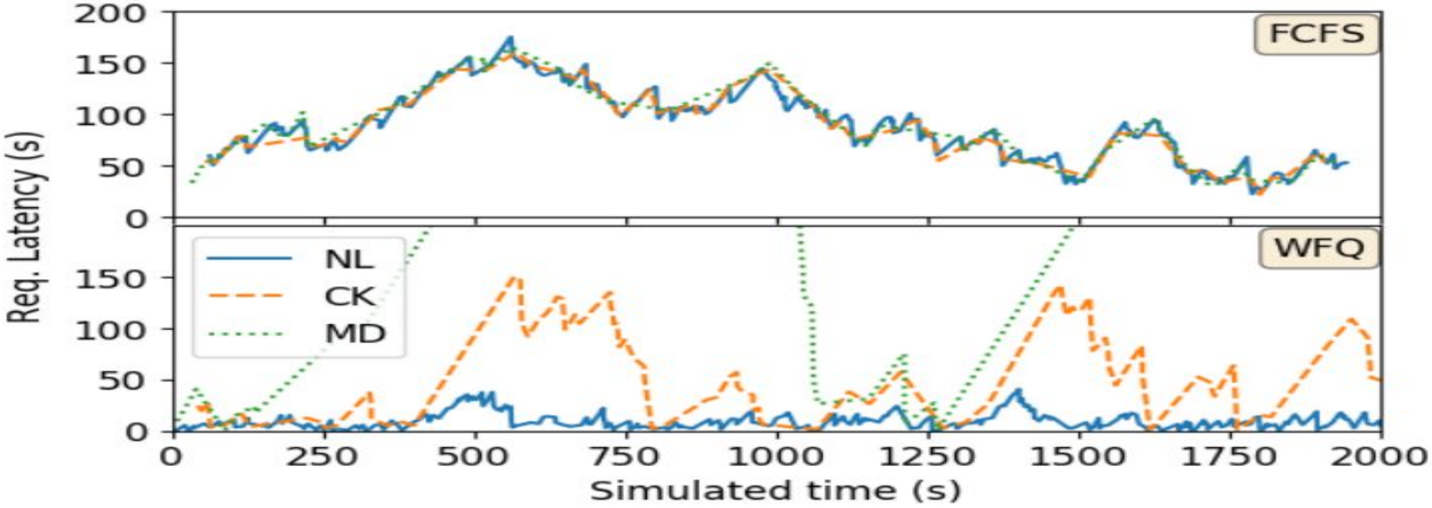
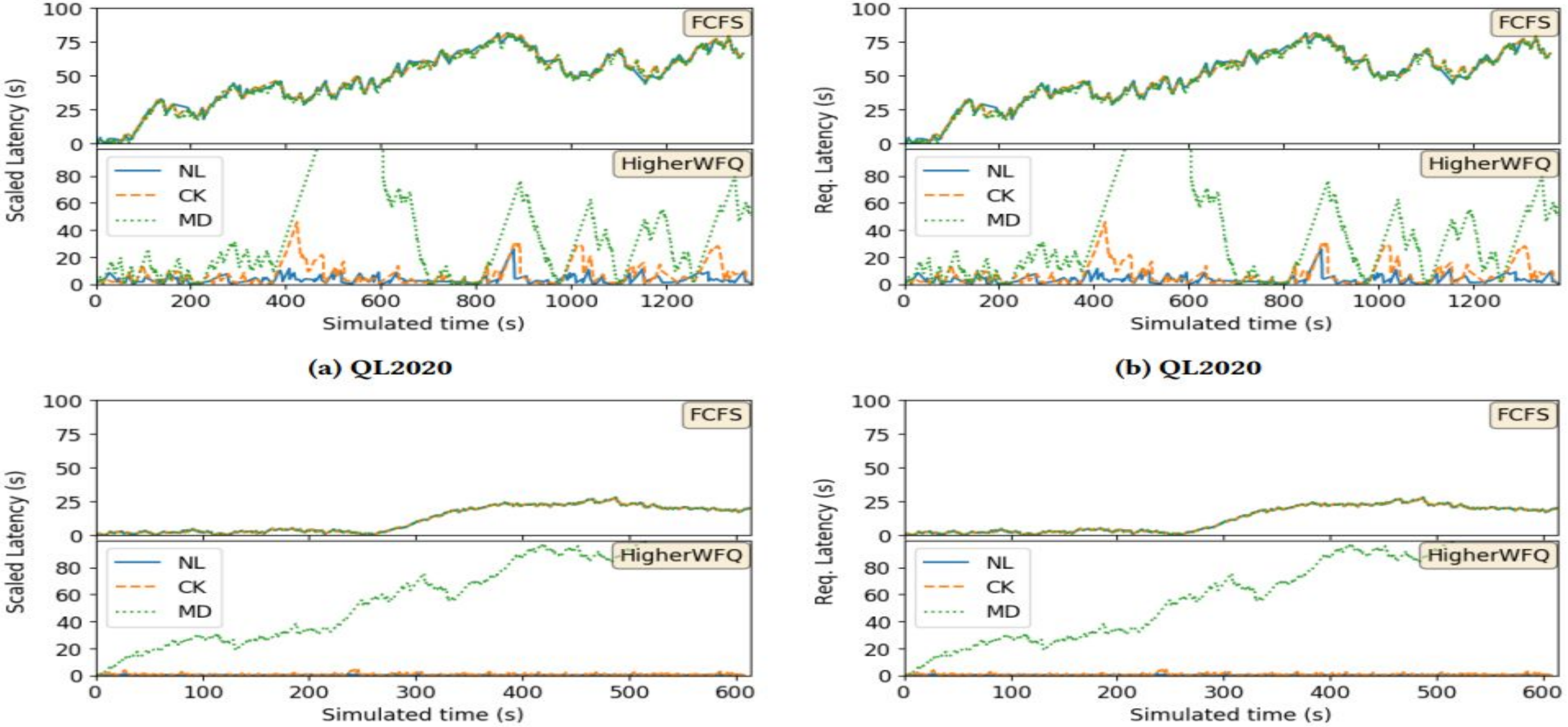


Figure 7: Request latency vs. time for two scheduling scenarios (long runs simulated 120 h wall time). As expected the max. latency for *NL* is decreased due to strict priority. In this scenario, there are more incoming *NL* requests ($f_{NL} = 0.99 \cdot 4/5$, $f_{CK} = 0.99 \cdot 1/5$ and $f_{MD} = 0.99 \cdot 1/5$).

Dahlberg, Axel, et al. "A link layer protocol for quantum networks." *Proceedings of the ACM special interest group on data communication*. 2019. 159-173.

Experiments and Results

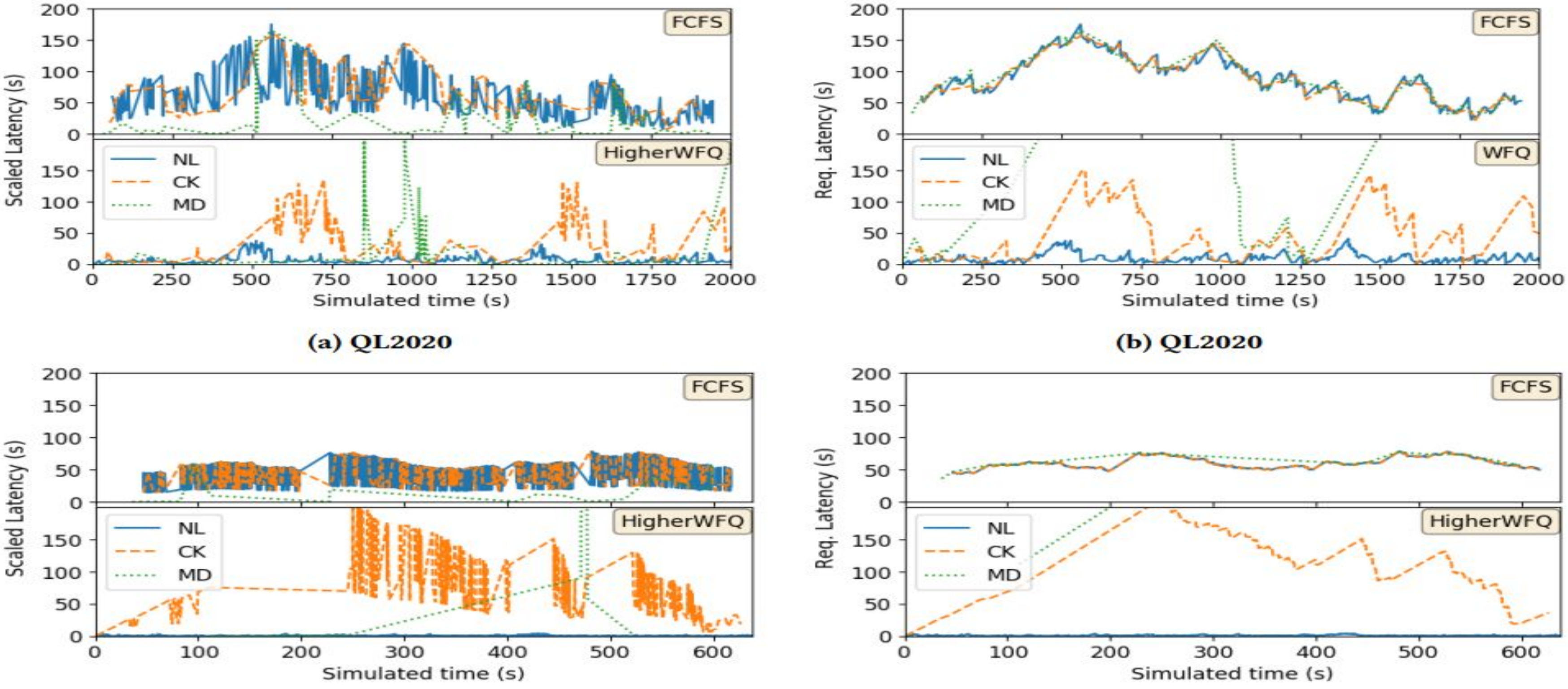


(a) QL2020 (b) QL2020

(c) LAB (d) LAB

Figure 11: Latencies for UNIFORM

Experiments and Results



(a) QL2020

(b) QL2020

(c) Lab

(d) Lab

Figure 12: Latencies for MoreNL

Experiments and Results

Table 1: Throughput (T) and scaled latency (SL) using scheduling strategies FCFS and WFQ for two request patterns: (i) with $f_{NL} = f_{CK} = f_{MD} = 0.99 \cdot 1/3$, i.e. a uniform load of the different priorities and (ii) with $f_{NL} = 0$, $f_{CK} = 0.99 \cdot 1/5$ and $f_{MD} = 0.99 \cdot 4/5$, i.e. no *NL* and more *MD*. The physical setup: QL2020 and number of pairs per request: 2 (*NL*), 2 (*CK*), and 10 (*MD*). Each value average over 102 short runs each 24 h, with standard error in parentheses.

T (1/s)	*NL*	*CK*	*MD*
(i) FCFS	0.146 (0.003)	0.144 (0.003)	2.464 (0.056)
(i) WFQ	0.154 (0.003)	0.156 (0.003)	2.130 (0.063)
(ii) FCFS	-	0.086 (0.003)	5.912 (0.033)
(ii) WFQ	-	0.096 (0.003)	5.829 (0.049)

SL (s)	*NL*	*CK*	*MD*
(i) FCFS	10.272 (0.654)	10.063 (0.631)	1.740 (0.120)
(i) WFQ	3.520 (0.085)	6.548 (0.361)	4.331 (0.336)
(ii) FCFS	-	5.659 (0.313)	0.935 (0.062)
(ii) WFQ	-	2.503 (0.100)	1.194 (0.093)

- A Layered Stack helps focus and develop protocols for Quantum Internet
- The link layer protocol works well in different experimental setups and in the simulation as well
- **Future Works** Could address:
    - The purification/ entanglement swap process
    - A SDN control plane

A Link Layer Protocol for Quantum Networks

Axel Dahlberg , Matthew Skrzypczyk , Tim Coopmans , Leon Wubben Filip Rozpędek, Matteo Pompili , Arian Stolk , Przemysław Pawełczak , Robert Knegjens , Julio de Oliveira Filho , Ronald Hanson, Stephanie Wehner,

QuTech, Delft University of Technology and TNO
Kavli Institute of Nanoscience, Delft University of Technology

Diego Abreu (Federal University of Pará) - Brazil
Advisor: Prof. Antônio Abelém

Desired Service

4.1.1 Requesting entanglement

-Rquest purpose ID

-Remote node

-Type of request : create and keep (K) , create and measure (M) , ou Network Layer (NL)

-Number of entangled pairs to be created

-Waiting time (Max)

-Flags: Atomic (all pairs be made available at the same time, for CK), Consecutive (OK, l for NL use case)

-priority: to be used by a scheduler.

-Desired Minium Fidelity.

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Remote Node ID					
Minimum Fidelity				Max Time	
Purpose ID				Number	
Priority	T	A	C	reserved	

Desired Service

4.1.2 **Response to entanglement requests**

-IF success, OK. ELSE: TIMEOUT, UNSUP (fidelity no achivable), MEMEXCEEDED/OUTOFMEM (not enough memory), DENIED, EXPIRE (EPR not available).

-Entanglement ID:

-Qubit ID

-Goodness: Fidelity estimation, where G >= Fmin

-Measure outcome

-Time of entanglement creation

-Time of the goodness: when the fidelity estimation was made

Desired Service

Fixed hardware parameters

- The number of available qubits.
- The qubit memory lifetimes.
- Possible quantum operations.
- Attainable fidelities and generation time
- The class of states that are produced

Desired Service

Performance Metrics

- Throughput (entangled pairs/s)
- **Latency**
  - Latency per request
  - Latency per pair
  - **scaled latency (**Latency per request per number of requested pairs