Post-Quantum Privacy-Preserving Data Analysis Pipelines

CERN openlab Technical Workshop 2022

José Cabrero-Holgueras (CERN, Universidad Carlos III de Madrid),

Gabriele Morello (CERN, Politecnico di Torino),

Alberto Di Meglio (CERN)

Who are we?

José Cabrero-Holgueras

Ph.D. Student at CERN

- Privacy-Preserving Computation Techniques for Secure Deep Learning in Healthcare
- Homomorphic Encryption and Secure Multiparty Computation for Deep Learning

Gabriele Morello

BSc Student: Computer Engineering at Politecnico di Torino

- Simulation of Quantum Key Distribution Protocols
- Federated Learning

Outline

- Introduction to the Quantumacy project.
- Quantumacy use cases:
  - Quantum Key Distribution and QKDSimkit.
  - Post-Quantum Secure Data Analysis Pipeline.
  - QKD Secured Federated Learning.
- Quantumacy Wrap-Up.
- Conclusions.

Processing Personal and Sensitive Data

From Classic Processing to Secure Processing

Large-scale healthcare research requires sharing and analysing increasing amounts of personal medical data on private and public computing resources, while at the same time preserving the privacy and the rights of the patients.

The Context

The Problem

Many different approaches to anonymisation are investigated today, but are subject to classical attacks. Providing secure, legally-compliant, user-centred data analysis is a major challenge and a fundamental step towards building trustable and trusted distributed AI systems.

QUANTUMACY will set up a proof-of-concept, end-to-end medical data analysis platform using state-of-the-art privacy-preserving deep-learning techniques and a block-chain-based user self-determination systems, secured with QKD-based encryption.

The Solution

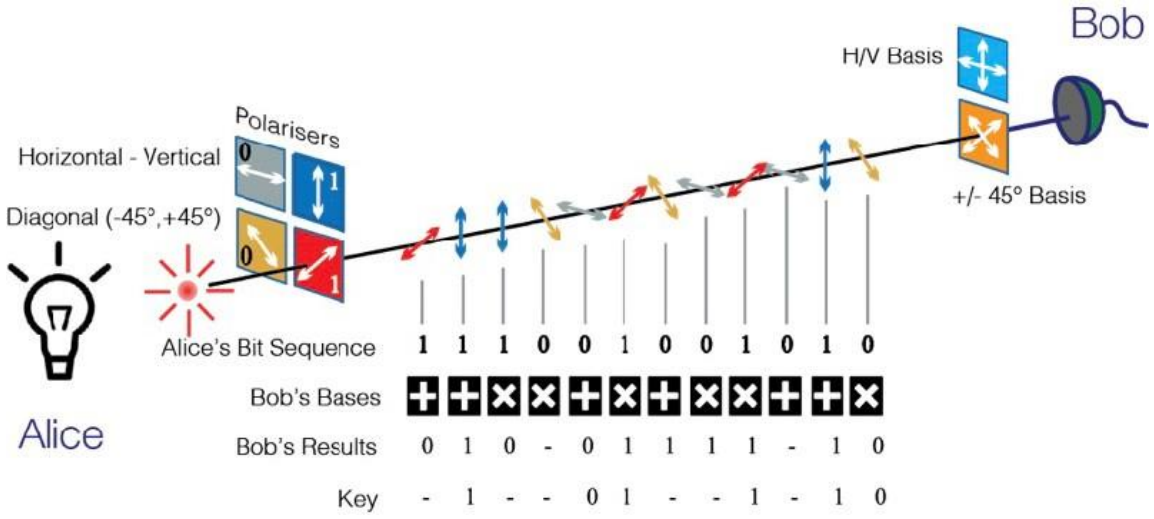
The Benefits

The **QUANTUMACY** approach will promote a move towards widespread use of distributed confidential computing and informational self-determination, enabling the possibility of sharing and processing data under solid principles of GDPR-compliant polymorphic privacy.

Quantum Key Distribution

- Quantum Key Distribution (QKD) enables the creation of a symmetric key profiting from quantum properties.

- Quantum security from the polarization of photons
  - Once a photon is observed, the state changes and a man in the middle can be detected.
- Implementations based on:
  - Fiber channel.
  - Free space - wireless.

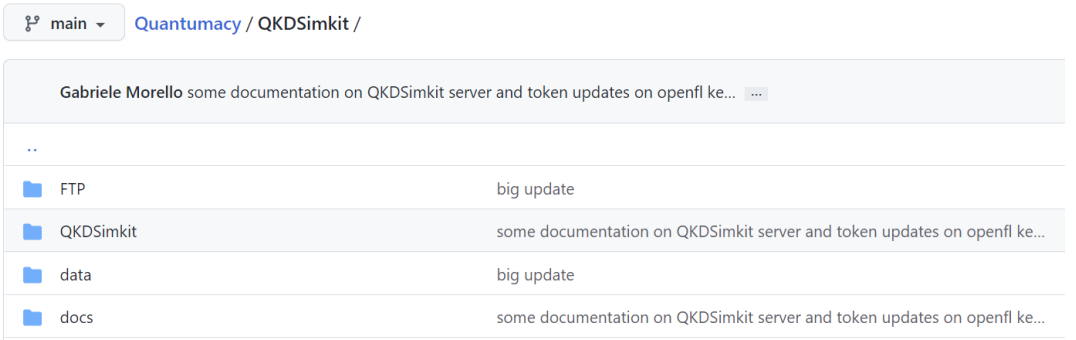


Source: (Quantum Flagship)

QKDSimkit

A working bleeding-edge QKD simulator

- Quantum Key Distribution (QKD) hardware is very complex and requires specific setup protocols.
- Many state-of-the-art quantum simulators support QKD but present problems at runtime.
- QKDSimkit is a QKD simulation library that permits deploying simulated QKD infrastructure.
  - It enables an easy migration to real QKD infrastructure.



Project soon to be released Open-Source on Github

Pre-release version available on PyPi

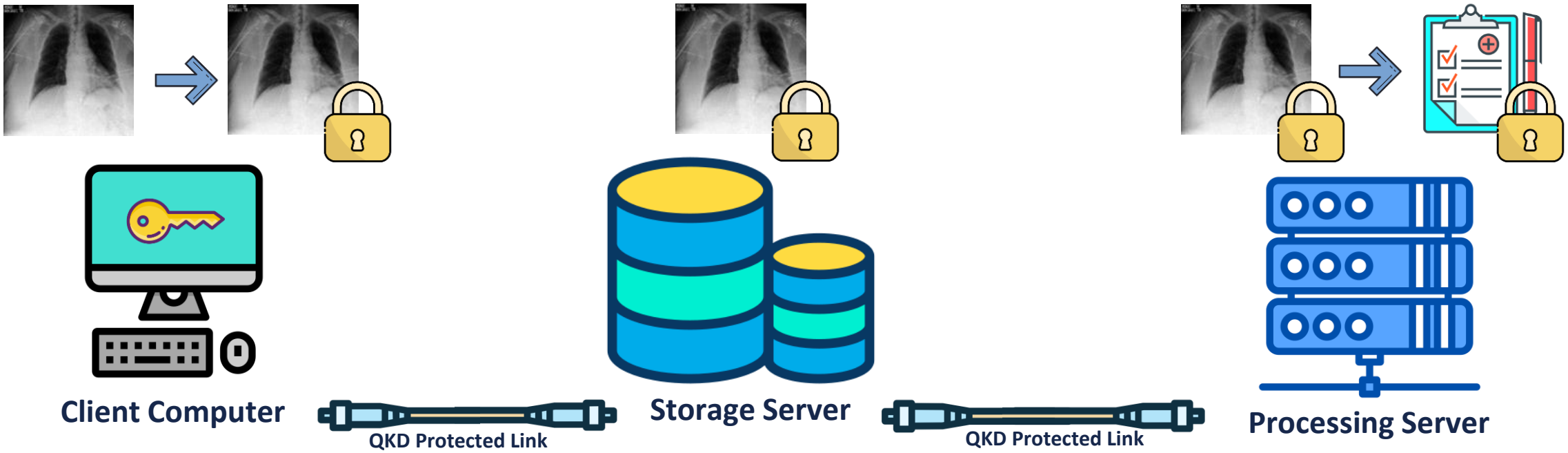
Post-Quantum Secure Data Analysis Pipeline

- Machine Learning Analysis of Chest Radiography Scans.
  - Logistic Regression: simple model, relevant accuracy.

- Homomorphic Vectorization of the Model:
  - Packed CKKS Homomorphic Evaluation of Logistic Regression.
    - Evaluation time of Packed HE Logistic Regression: 0.6 s.
    - Potential for Real Time Diagnosis.
  - Quantum Compliant Parametrization.

- Quantum Key Distribution for Pipeline:
  - Connections between services are authenticated by QKD.
  - Three-way Quantum Encryption: Rest, Transport and Processing.

Post-Quantum Secure Data Analysis Pipeline

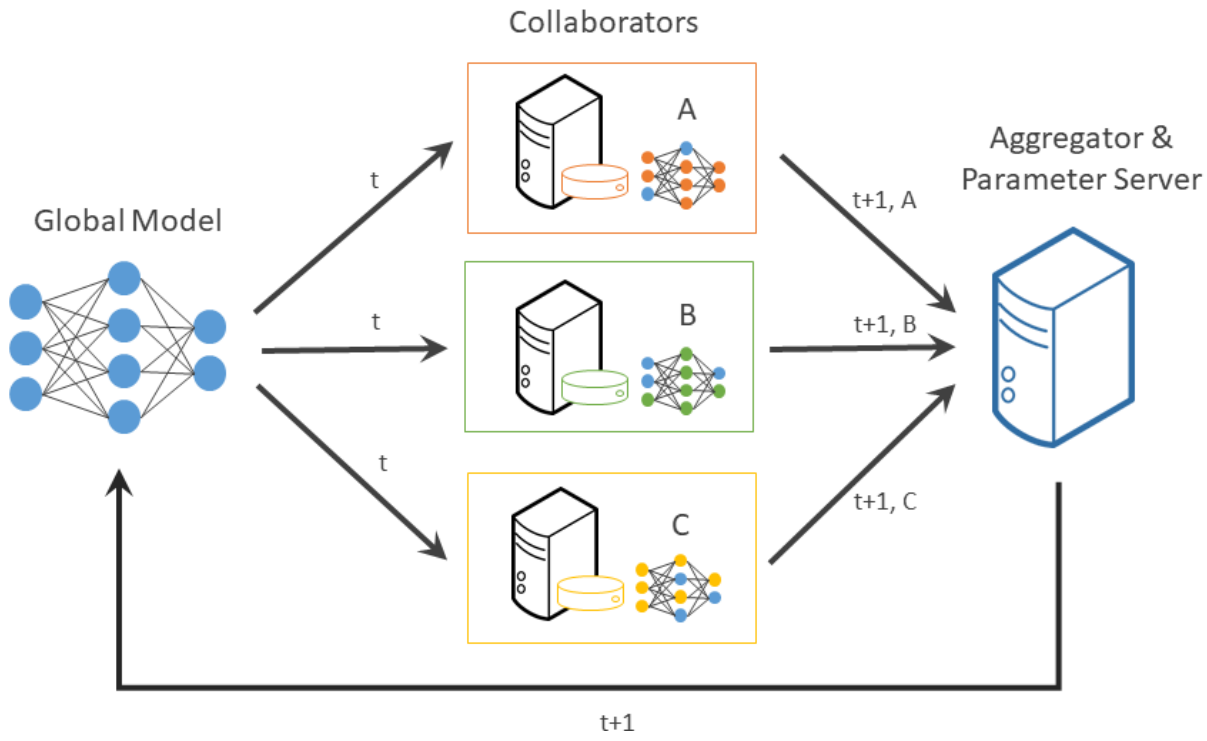


Federated Learning

Introduction to Federated Learning

- Federated Learning is a private distributed training procedure for Machine and Deep Learning models.

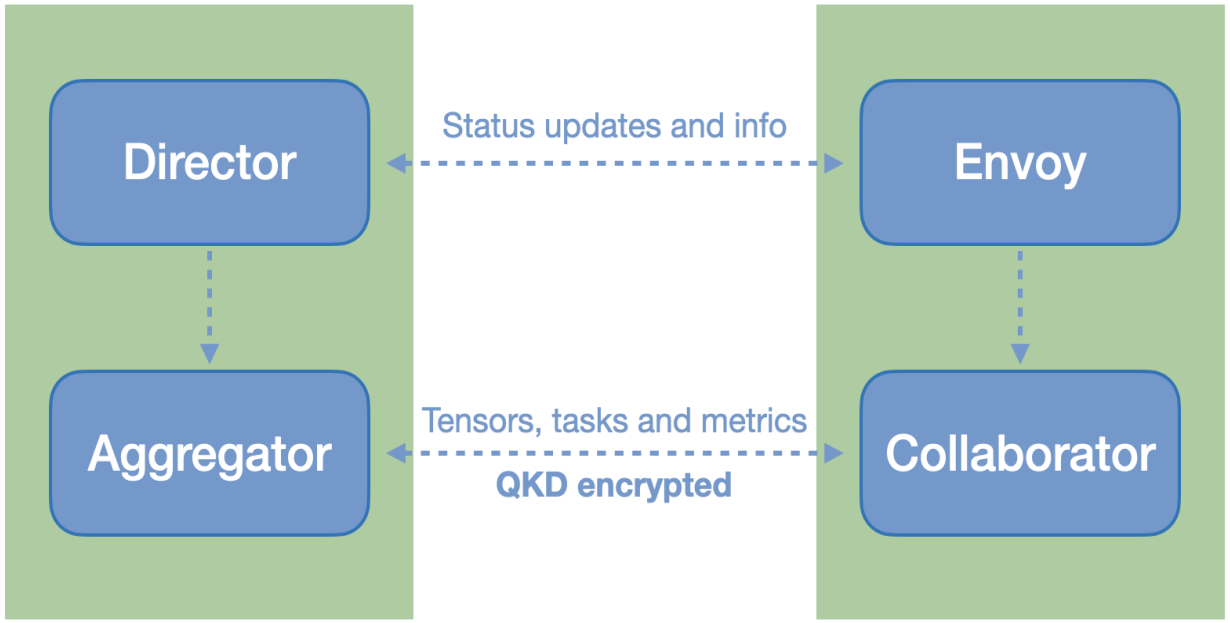
- Steps:
  - The **collaborators** hold data and local copies of the model that they use to locally train the model.
  - The **aggregator** combines the data from the different models in a central model comprising the different distributions.



QKD Secured Federated Learning

QKDSimkit – Intel OpenFL integration

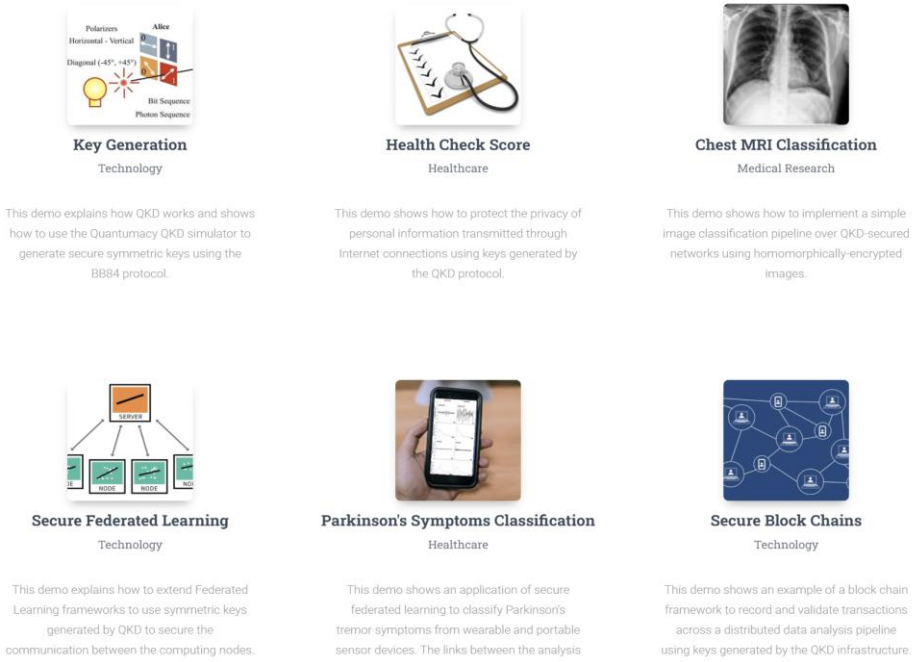
- Interaction between components in complex networks could be established.
- Federated Learning with QKD connectivity between aggregator and collaborators.
- The feasibility highly depends on the reliability of the channel.



Basic Components of Intel OpenFL

Project Wrap-up

- First version of Post-Quantum Secure Data Analysis Pipelines for Training and Inference based on Federated Learning and Homomorphic Encryption based Deep Learning.



Webpage soon to be released on QTI website.

Conclusions

- Data analytics in its various forms prove to be valuable in multiple domains.
- Domains we deal with sensitive data require special attention.
- Through Privacy-Preserving Technologies, we can reach the security standard needed to have legal compliance.
- We need to ensure security against future threats to current techniques such as Quantum Computers.
- Through collaborations with healthcare practitioners, we can develop tools for their benefit.

References

- Quantumacy Github
- QKDSimkit Pip Package
- CERN Quantum Technology Initiative Webpage
- Images from Intel OpenFL
- Icons from Icons8

QUESTIONS?

jose.cabrero.holgueras@cern.ch

gabriele.morello@cern.ch

alberto.di.meglio@cern.ch