LAIDLAW SCHOLARS RESEARCH AND LEADERSHIP PROGRAMME 2024 – 2025

SUMMER 1 RESEARCH REPORT

RESEARCH PROJECT TITLE:
UTILISING HOMOMOPRHIC ENCRYPTION, ZK-SNARKS AND SMART CONTRACTS TO DEVELOP A PRIVACY PRESERVING QUANTUM SAFE FEDERATED LEARNING MODEL FOR HEALTHCARE APPLICATIONS

REPORT BY: DHRUV RANAJIT CHOUDHURY
SUPERVISOR: DR. HITESH TEWARI

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

The inspiration for my research project emerged while going through a reading from my "Information Management" class at college. It was an article by the Harvard Business Review titled "Why AI Failed to Live Up to Its Potential During the Pandemic". It begins by highlighting how Artificial Intelligence (AI) was a promising candidate to help combat the pandemic's adverse impact on lives around the globe. AI enjoyed some initial optimism and sparse success in some industries for example automation in airports and warehouses, enhancing efficiencies and freeing up human resources. AI also aided vaccine trial site selections and other front-line efforts. However, when it came to the crucial task of diagnosing and managing COVID-19 in core healthcare settings, AI failed to deliver. Predictive AI models for diagnosis and prognosis struggled due to poor-quality datasets, biases in decision-making algorithms, and the complexity of the pandemic's global context. The dichotomy between the early scattered successes and the broader failure highlighted the limitations of AI during a global crisis like the pandemic.

The idea of distributed computing and distributed Machine Learning (ML) has existed since the late 1990s, most of the research in this area was focused on how multiple computers could process data, or how multiple computers could collaborate on a shared model. This research accelerated further with the emergence of grid computing. However, this often involved sharing data with central servers and exposing it to cyber threats, and developments to protect data privacy were of minimal concern at the time.

In the early 21ST Century, this started to change with increasing cyber threats and growing importance of policymakers placed on protecting data privacy around the world. A great example of this is the European Union's General Data Protection Regulation (GDPR). With this, resources were being increasingly employed to make systems that complied with data protection laws and employed data to commercial use, while protecting individual user and institutional data. This led to the emergence of edge computing, where data computations on data were increasingly pushed closer to the source of data (edge devices), such as user devices. In 2016, researchers at Google formalised the concept of multi-party ML and coined the term "Federated Learning" (FL). It was introduced as a way for mobile devices to collaboratively train ML models without uploading their data to a central server. Over the years FL has been thoroughly researched, it has proven to be an exemplary candidate as a core framework for shared models, but research has also shed light onto FL's predisposition to two core

vulnerabilities: inference attacks, model poisoning attacks, and the central servers that act as aggregators can become single points of failure.

These vulnerabilities have hindered FL's adoption in sensitive fields like healthcare, where inference attacks can compromise patient confidentiality; and with the increasing threat of quantum computing potentially undermining classical cryptosystems in the future, I decided to work on developing a schema that fosters collaboration in domains where it could have a transformational impact.

The core of my research started as follows: taking state-of-the-art cryptographic technologies, namely, Homomorphic Encryption (HE) and Blockchain (primarily smart contracts) and employing them at different stages of the FL process to make it privacy preserving and quantum safe. HE safeguards the models from each participant or device in the training, i.e. the model is encrypted before being sent for aggregation. Unlike classical encryption techniques, HE allows for computation on ciphertext, i.e. computation on data in its encrypted state. Concretely it allows for multiplication or addition, or both (depending on the scheme) in the encrypted space. Furthermore, HE has been proven to be Quantum-safe, making it future-proof. Blockchains are networks of various participants, they act as decentralised ledgers and provide the ability to move computations away from the central server to this decentralised network. This combats the single point of failure issue in standard FL. Blockchains can be used as a facilitator for aggregation instead of a single device (server) acting as the aggregator. This is also beneficial in the fact that no single actor has control of all the shared models and the aggregated global model. Smart Contracts (SCs) are programs (written in code), that execute automatically on a blockchain when certain conditions (terms of the contract) are met. Instead of needing a trusted party for aggregation, we propose that a SC (accessible to all participants) is used for aggregation.

I began my research summer by conducting an literature review, which meant navigating a plethora of academic papers and journal articles, to understand the current state of federated learning, homomorphic encryption and blockchain technology. In all honesty, this was probably one of the most daunting phases of my research project, this helped me identify the gaps and vulnerabilities in existing systems but at the same time, the first few papers I went through seemed incomprehensible, jumble filled with intricate polynomials and dense technical jargon. However, just within a couple of weeks of combing through the literature, I started to

coherently understand the years of research behind these brilliant technologies. I truly felt like I was able to do a deep dive into a subject I am passionate about. But I realised there are a myriad of studies already done on FL that incorporated HE and Blockchains, this led me to redefining the scope of my project, I want to make material contribution to research. I decided to investigate how participants in FL could be protected against model poisoning attacks. Model poisoning attacks are a threat that a malicious participant uploads manipulated models to degrade the model. Hence, on top of privacy preservation and quantum resistance, I added a third core aspect to the framework: protection against malicious intent in the absence of trust between the participants.

Eventually I came across Zero Knowledge Succinct Non-Interactive Arguments of Knowledge (zk-SNARKs), a form of Zero Knowledge Proofs (ZKPs). I provide a simple introduction to ZKPs in my summer 1 blog. Please find it here. zk-SNARKs form the core of various blockchains, but research investigating zk-SNARKs in the context of FL is scarce. After presenting my idea with my supervisor and some discussions, I decided to probe into how zk-SNARKs could be integrated with FL. This was a very exciting turning point in my research, but it also meant I had to revise the final goals of my research for the summer. With the added complexity it meant I had to add extra steps onto my original proposal and trim some aspects, moving to a more theoretical approach to my research for the time being with a narrower scope for simulations and experiments.

zk-SNARKs are the most widely used form of ZKPs, known for their non-interactivity and succinctness, i.e. the prover needs to only one transaction with the verifier to convince them of some statement, i.e. no further interaction between the parties is necessary after the initial proof submission. I realised that a zk-SNARK could be used (along with HE and blockchain) to create a verifiable computation protocol that could be used to verify the computations done on the models submitted by the participants during FL.

Subsequent to the literature review phase and outlining an overview of my proposed system I moved on to implement this framework in a simplified healthcare environment. A brief overview of how this is effectuated is as follows:

Start by constructing ECNN by combining the architecture three CNNs (DenseNet201, ResNet50, and EfficientNetB5) to enhance feature extraction from chest X-ray images. CNNs

are proficient in handling visual data due to their ability to capture spatial hierarchies. This ECNN serves as the architecture that all participants train their model on.

Once we have this architecture defined, we have to perform a setup for zk-SNARKs. zk-SNARKs cannot be simply applied to any type of information, computation or problem, it needs to expressed as a Quadratic Arithmetic Program (QAP). There are some intermediate steps that the computation has to go through before this (as seen in the figure below), but a QAP is at the core of verifiable computation protocol based on zk-SNARKs. A QAP is essentially the computation (in our case the main model training function/program) broken down into multiplications, additions and then represented as a polynomial. After this setup, a smart contract can be written to verify proofs: to ensure adherence to the agreed-upon model architecture and prevent poisoning attacks.

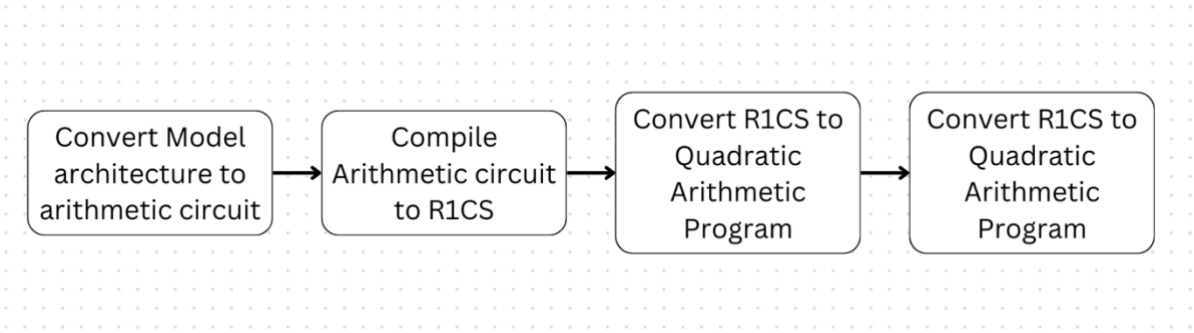

Figure: Representing a model as QAP

Next, I collected Publicly available chest X-ray datasets to simulate a federated learning environment akin to multiple healthcare institutions, the dataset was partitioned into five distinct groups. Each group representing a separate institution, maintaining local data storage without centralised sharing. Each dataset would be used to train, validate and test models separately (or "locally"). During training, each participant generates a zk-SNARK proof demonstrating compliance with the established architecture without revealing model parameters.

Smart contracts automatically generate public/private key pairs required for HE. The use of smart contracts ensures that the key management process is transparent, secure, and tamper-proof, leveraging the inherent properties of blockchain technology. Post-training, participants

encrypt their model updates using HE. These model updates are the result of training and primarily comprise of decision weights and biases.

The encrypted model updates are then shared along with their corresponding zk-SNARK proof. Verified encrypted model updates are homomorphically aggregated to form a global model (using a smart contract). The aggregated model is evaluated against predefined performance metrics using a validation set. If performance thresholds are met, training concludes. Otherwise, additional training rounds are initiated, with participants receiving updated global model parameters to continue local training. This iterative process can be continued until satisfactory performance is achieved.

Drawing from initial simulations and analyses, I am sharing the following findings and look forward to advancing this research through more comprehensive simulations, implement and test additional components of the framework:

Despite the volatility of local models, the shared global model in FL achieves higher accuracy over multiple iterations. This improvement is due to the incorporation of broader and more diverse data, which helps reduce both overfitting—where a model becomes too tailored to specific data and loses generalisation ability—and underfitting—where a model fails to capture underlying patterns in the data.

While incorporating Homomorphic Encryption seems like an ideal solution for protecting local models in a federated learning environment, the computational cost of HE increases exponentially with model complexity. For some models, encryption and decryption can take days or become infeasible. In light of these challenges, it is crucial and I hope to explore optimisation strategies for Homomorphic Encryption in machine learning contexts.

Similarly, in the generation of Zero-Knowledge Succinct Non-Interactive Argument of Knowledge (ZK-SNARK) proofs, increasing model complexity results in larger proofs that consume substantial amounts of memory, often reaching gigabyte scales. Note that this is computational complexity increases for the prover, but complexity for the verifier remains independent of the size of the computations because of the succinctness property of zk-SNARKs. Temporarily storing these proofs on decentralised networks can incur significant costs. zk-SNARKs have emerged as an exemplary solution for safeguarding models against

model poisoning and have demonstrated high effectiveness with simple models such as Linear Regression and Naive Bayes classifiers. However, their scalability to more complex models remains a challenge. Consequently, I would like to see future research focus on optimising zk-SNARKs for complex models like Convolutional Neural Networks (CNNs).

This research project has been a transformative experience that not only advanced my academic knowledge but also fostered personal growth. I navigated complex technical domains, and I really felt like I challenged myself. It reinforced the importance of adaptability, collaboration, and perseverance. While I was unable to implement all aspects of my initial proposal, I found that reevaluating the scope during the literature review period led me down a different path. Although this shift was made me feel somewhat inadequate initially, I came to realise that it is not necessary to address every issue within secure federated learning. Instead, I can still make meaningful contributions by establishing a comprehensive framework that paves the way for future research. I am very grateful that supervisor was a vital source of guidance and encouragement. Regular meetings provided clarity and kept the project on moving. Finally, I am proud that I was able to design a promising framework that aligns with my vision for consequential and transformational collaboration; conclude with preliminary yet realistic and insightful simulation results; demonstrating the potential of this work; something that I can and hope to build upon moving forward.

REFERENCES

- Chakravorti, B. (2022) Why AI failed to live up to its potential during the pandemic. https://hbr.org/2022/03/why-ai-failed-to-live-up-to-its-potential-during-the-pandemic.
- McMahan, B., Moore, E., Ramage, D., Hampson, S. and y Arcas, B.A., 2017, April. Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics (pp. 1273-1282). PMLR.
- Petkus, M., 2019. Why and how zk-snark works. arXiv preprint arXiv:1906.07221.
- G. Keshavarzkalhori, C. Pérez-Solà, G. Navarro-Arribas, J. Herrera-Joancomartí and H. Yajam, "Federify: A Verifiable Federated Learning Scheme Based on zkSNARKs and Blockchain," in IEEE Access, vol. 12, pp. 3240-3255, 2024, doi: 10.1109/ACCESS.2023.3347039
- https://medium.com/@alexppppp/how-to-train-an-ensemble-of-convolutional-neural-networks-for-image-classification-8fc69b087d3
- https://medium.com/@VitalikButerin/quadratic-arithmetic-programs-from-zero-to-hero-f6d558cea649