# DECML: Distributed Edge Consensus Machine Learning

Cyrile Verdeyen - Supervised by: Raja Naeem Akram, and Konstantinos Markantonakis

Information Security Group, Smart Card and IoT Security Center
Royal Holloway, University of London

The Smart Card and Internet of Things Security Centre

## Objectives

More and more it is feasible to push the machine learning to the edge devices in our systems. But then there is still a requirement for a privacy preserving form of communication for sharing the knowledge gained by machine learning.

- Propose a new architecture that would support many different nodes to be able to contribute their machine learning knowledge without having to share their data or models with the other nodes.
- Measure time it takes to learn the model and accuracy of the models in comparison to centralised machine learning.

## Introduction

As more processing power and hardware is being shifted to the edges of our technologies there has been an increase in focus towards creating privacy preserving models to be able to share information safely. Whether this be Federated, Decentralised, or a Centralised Machine Learning approach, large amounts of research is being put forward to tack on algorithms to these models to allow them to safely share the information learnt in a privacy preserving manner.

To avoid the need of adding more layers to already complex systems, a new architecture all together was come up with that innately preserves each nodes privacy.

The core concept behind DECML is that there are clusters of nodes, where each node has its own separate data that it runs a machine learning model on. For each cluster there is a Cluster Orchestrater (CO) which can receive questions to ask to the cluster. Each node then answers the question using its own model, and sends it back to the CO, where a consensus of the answers is taken, and the response is sent back.

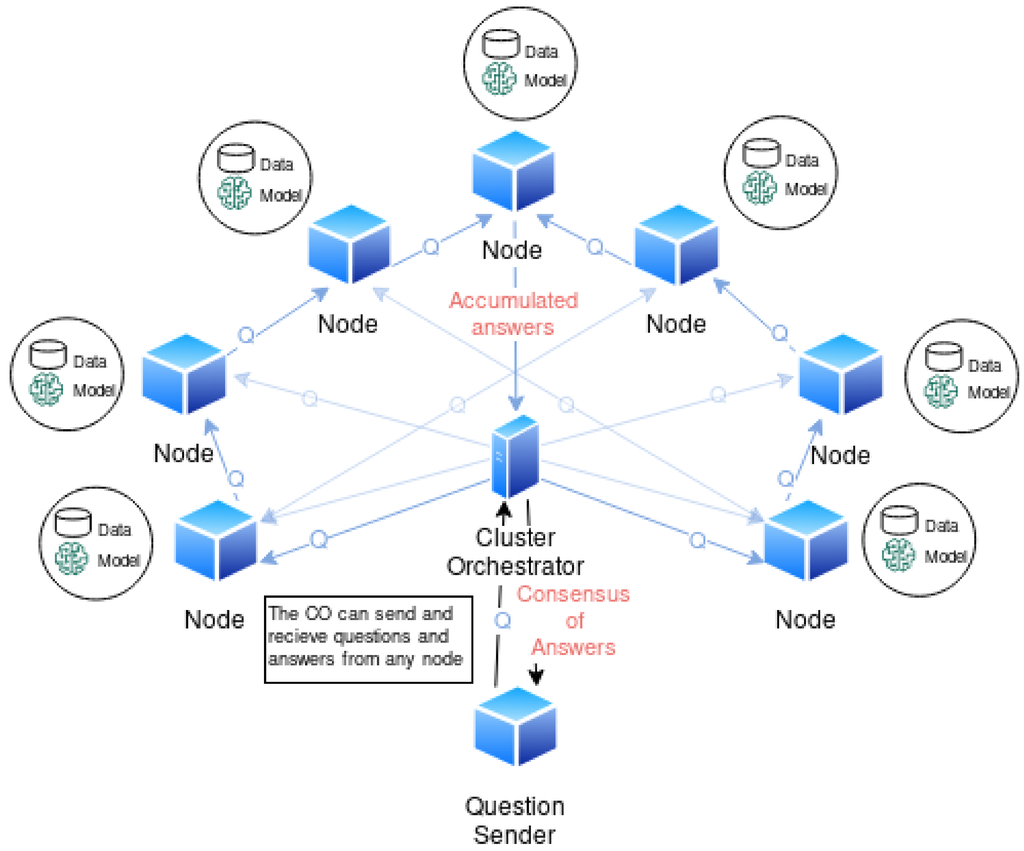


Figure 1: DECML Diagram

## Implementation

**Setup** Each node $n_i$ runs a machine learning algorithm on its own data set until it has a model that can make predictions with certainties.

**Round 1** Node $n_i$ connects to the CO *co* and learns about and connects to its peers, creating a peer to peer network.

**Round 2** *co* receives a question in the form of a data point for it to present to the cluster to answer.

**Round 3** *co* sends the question to $x$ amount of nodes, where $x$ scales upwards with amount of nodes in the cluster.

**Round 4** The $x$ nodes answer the question, and propagate the question to the rest of the nodes. These add their answers to the message and continue propagating the question.

**Round 5** When there are no more nodes that have not answered, a certain time limit has been reached, or enough nodes have responded, the answers all get sent back to the *co*.

**Round 6** The *co* waits a set amount of time $t_r$ after receiving the first set of answers to allow other answer sets to arrive, and then does a consensus of all the answers it received, and chooses the answer with the highest certainty.

**Round 7** The *co* sends the answer back to the question sender.

## Benefits & Uses

The proposed protocol enables privacy preserving cooperation between non trusted parties who wish to share their ML models and data. This allows for cooperation in many areas where the data itself needs to remain highly confidential, but where collaboration is highly beneficial, such as:

- Hospital patients disease classification;
- Network traffic analysis of unknown packages;
- Topic classification of texts and messages

## Results

SVM Model

| Set Up | Training Time (Hr) | Accuracy % |
|---|---|---|
| Singular Machine | 10.88 | 48.5 |
| DECML | 12.32 | 52.7 |

SGD Model

| Set Up | Training Time (Min) | Accuracy % |
|---|---|---|
| Singular Machine | 11.02 | 13.2 |
| DECML | 59.53 | 30.3 |

## Conclusion

Having done the tests using the Cifar-10 data set with two machine learning models,two main conclusions can be drawn, those being, training times are longer using DECML architecture, and accuracy is greater when using the DECML architecture. And of course, the data and model of each node remains private.

## Contact Information

- Web: https://scc.rhul.ac.uk/
- Email: Cyrile.Verdeyen.2016@live.rhul.ac.uk