# Scaling the Federated Edge

Jon Crowcroft

# Extreme Federation…

- Edge processing data in networked systems becoming mainstream:
  - It reduces load on the uplinks,
  - it saves energy &
  - potentially provides better privacy for personal data.
- Federation originally just meant
  - local model acquisition, but central aggregation
- More useful to think in terms of hierarchy, clustering, p2p

# A variety of edge techniques

- simple aggregation,
- compressive sensing, &
- edge-machine learning
  - where models are locally acquired, and
  - model parameters are distributed,
  - so nodes can further refine their models.

# Challenges #1

- Firstly to scale federated learning to billions of nodes needs some way to scale
- even just sharing model parameters e.g. https://arxiv.org/abs/1907.08059

- including sampling of model parameters
  - thinning,probabilistic update &
  - self organising hierarchies of aggregation (model parameter servers). e.g. https://arxiv.org/abs/1709.07772
  - For some Machine Learning algorithms, there may be updates from the federated model back to nodes to adjust their learning (e.g. regret) as well.
  - indeed, what even is initial placement system?
  - it sure isn't Kubernetes – meta-scaling tools!
  - Could be

https://www2.eecs.berkeley.edu/Pubs/TechRpts/2018/EECS-2018-119.pdf

# Challenges #2

- Some schemes may require synchronisation of learning steps.
- All these need to scale out, &
- techniques from data centers may,
  surprisingly be applicable, even though
  - we are often in a much less rich networking environment,
  - even without full connectivity or symmetric bandwidth or reachability.

# Challenges #3

- Federation alone is not a complete solution to privacy, &
- some further techniques may be needed to reduce the loss of confidentiality –
  - e.g. differential privacy is useful, but also
  - more fundamental approaches such as secure multi-party computation, in extreme cases, or FHE.

# Challenges #4

- Secondly, there is the problem of bad actors injecting false data
  - pollution, as well as non-IID data
  - Clustering helps with detecting pollution
  - And heterogenous node/data distribution
- Then there is the omnipresent presence of possible DDoS attacks.
- Scale federated trust?
- Decentralised trust  (transparency) – what tools?
- Proof-of-X is not currently a scalable solution for most X

# Challenges #5

- Thirdly, a federated model may present some challenges to model
  - explain-ability or interpret-ability.
- Interesting trade-offs between these requirements & privacy.

# Promise

- Mastodon, Matrix etc promise of federated instances for
  - social media and secure messaging
- Interop of key management seems to be one good area.
  - Scaling consensus does not seem too hard either…

# Conclusions

10 thousand data centres with a million cores

10 billion edges – add some structure?