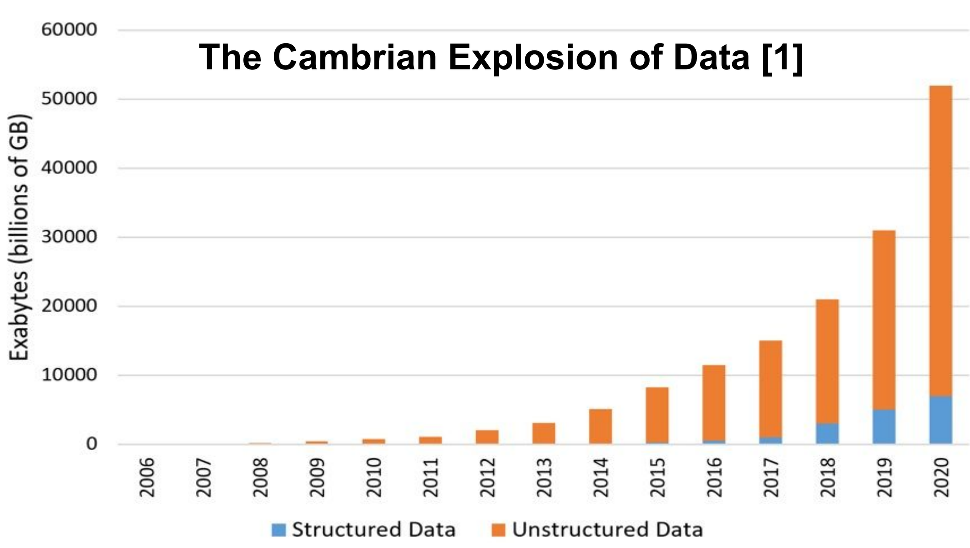
UBC Department of Computer Science

Biscotti: Private and Secure Decentralized Machine Learning

Muhammad Shayan, Clement Fung, Chris J. M. Yoon, Ivan Beschastnikh

"By 2020, the amount of data is predicted to sit at 53 zettabytes - increasing 50 times since 2003."

-- Hal Varian, Chief Economist at Google



Why not centralized ML?

Modern ML frameworks (TensorFlow, PyTorch) assume data is centralized which raises concerns:

- **Privacy:** Some data is sensitive and users may be uncomfortable with sharing or housing their data with other users' data
- **Scalability:** We are generating data at an unprecedented scale. Storing and processing this data centrally is increasingly expensive

Decentralization challenges

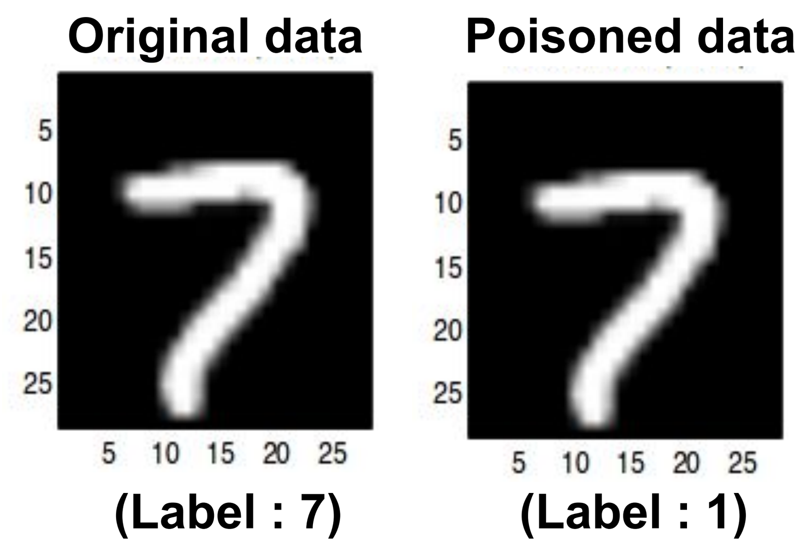
To minimize data transfer, decentralized solutions like Federated Learning have been proposed. These solutions have two issues:

- **Centralized coordination:** Requires a trusted centralized service to coordinate the distributed training at clients
- **Security:** Opens up the learning process to various types of attacks by malicious clients

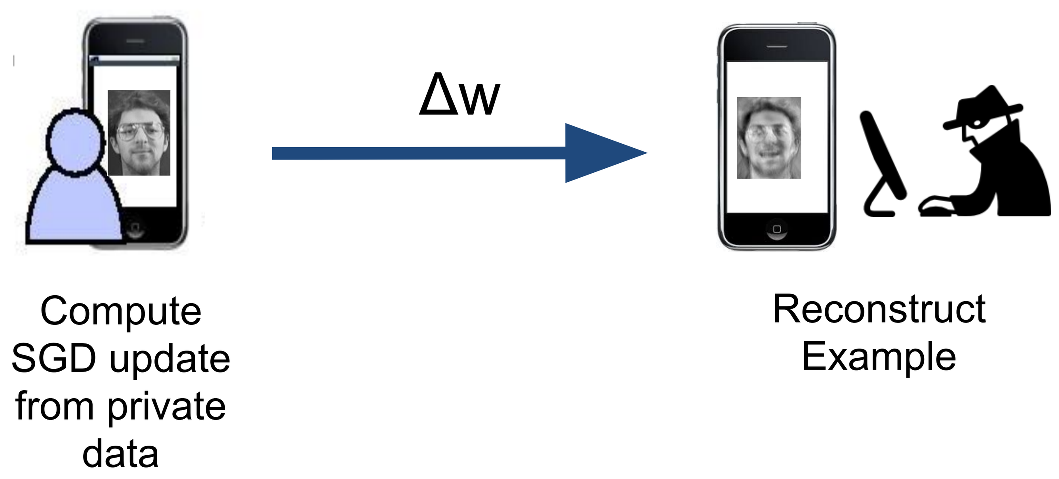
Biscotti: Peer-to-Peer secure and private ML system

Problem 1: Sybil attacks [2]

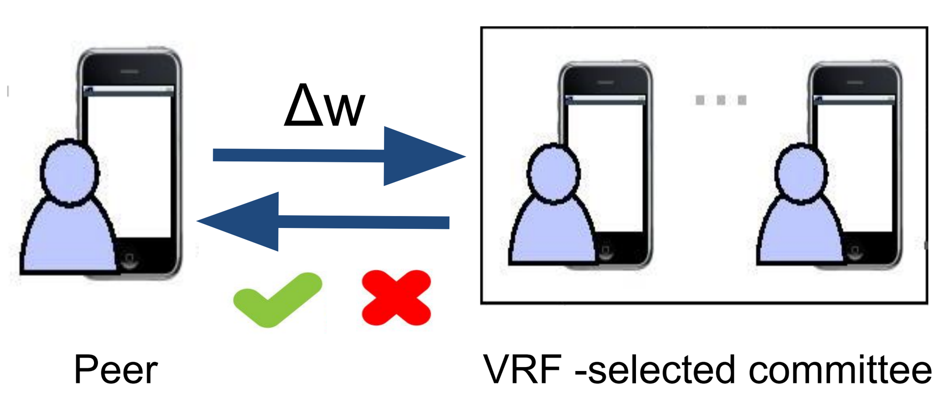
Problem 2: Poisoning attacks [3]



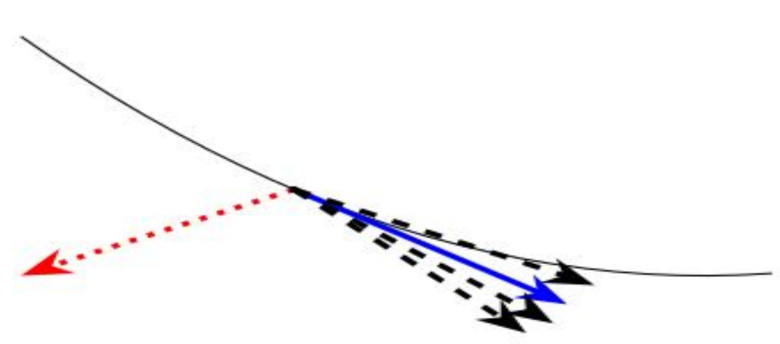
Problem 3: Privacy leakage from SGD updates [4]



Solution 1: Verifiable Random Function (VRF) [5] Committees using Proof of Stake

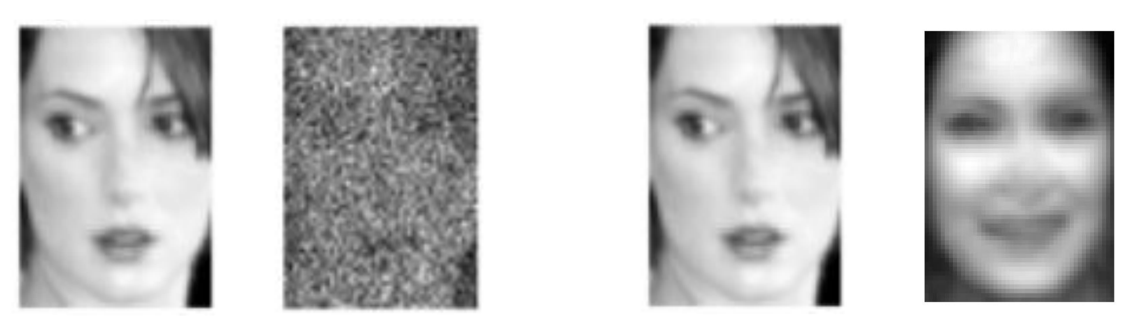


Solution 2: Filtering updates using Multi-KRUM [6]



Using pairwise euclidean distances to filter poisoned updates

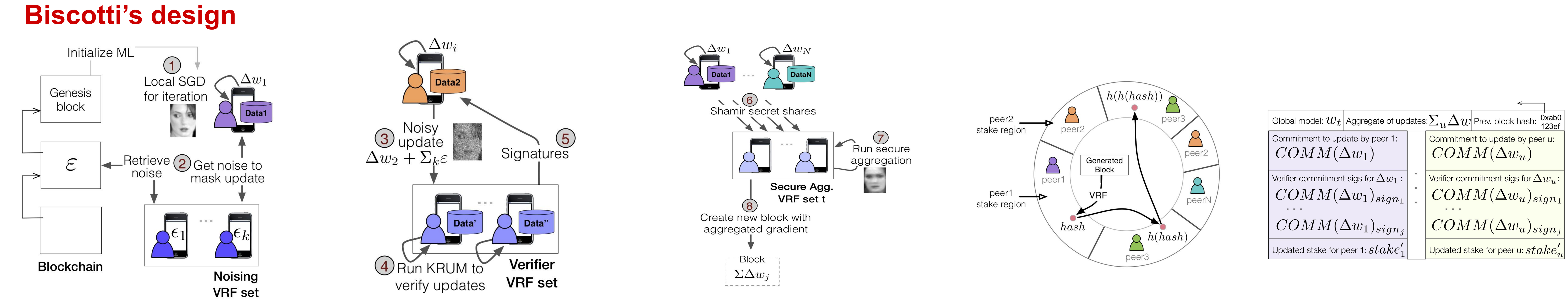
Solution 3: Differential privacy [7] and secure aggregation [8]



Adding noise to updates to protect contents

Aggregating multiple updates to protect privacy of individual update

Biscotti's design



Noising: Peers obtain pre-committed noise from a noising VRF committee to mask their update

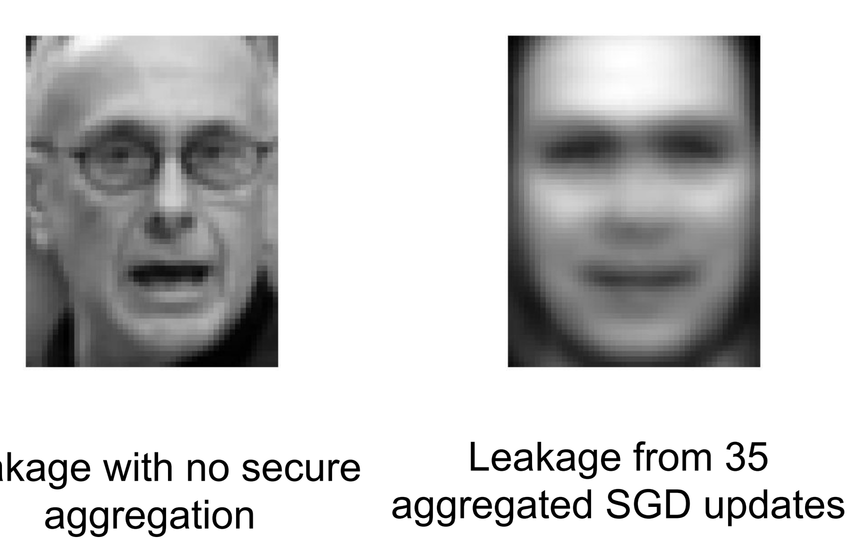
Verification: Peers send their noisy updates to a verifier set that filters out poisoned updates using KRUM

Aggregation: A committee creates the next block with the aggregate of accepted updates via secret sharing

VRF committee [9]: Selected by consistent hashing of VRF output from last block hash

Block: Contains commitments to accepted updates that can be used to verify the aggregate

Biscotti's aggregation privacy

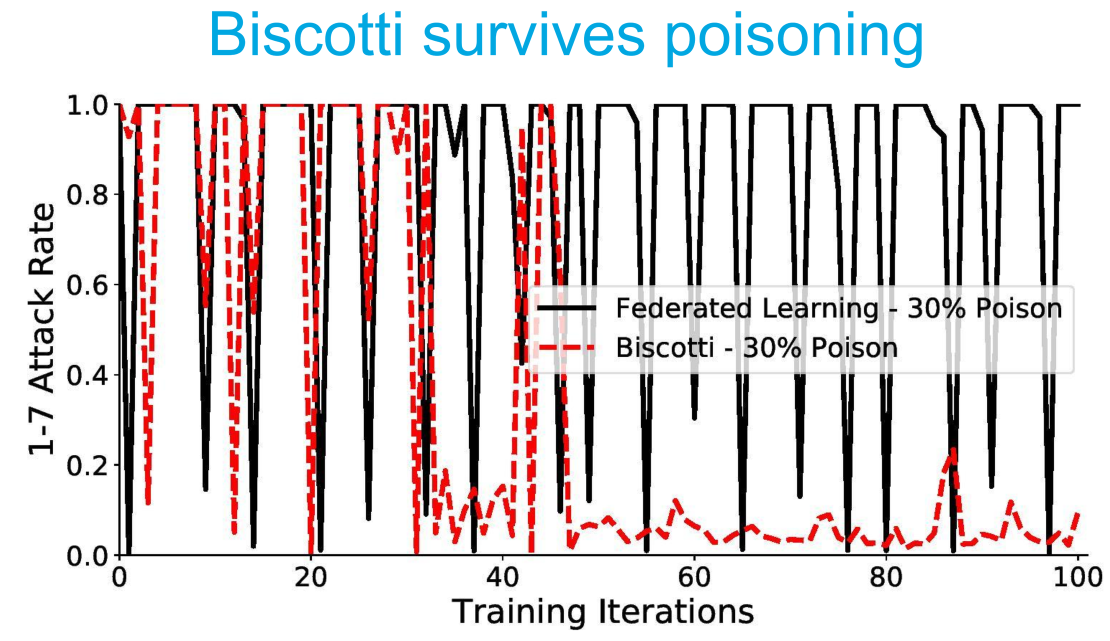


Leakage with no secure aggregation

Leakage from 35 aggregated SGD updates

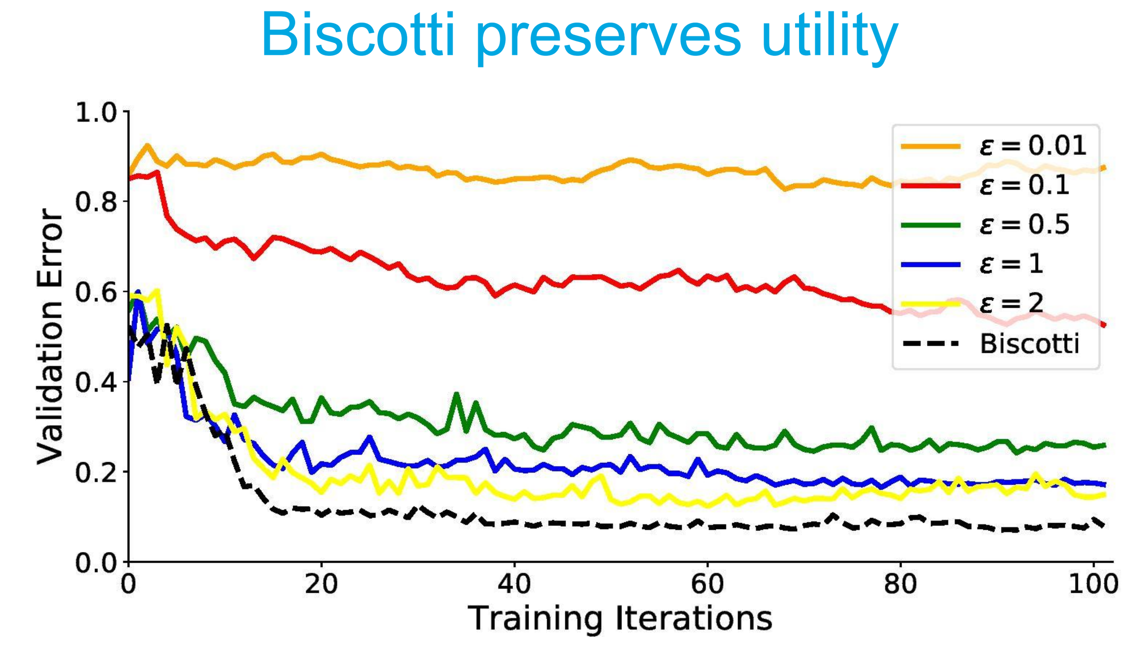
Biscotti protects privacy of individual training examples via secure aggregation

Biscotti survives poisoning



Biscotti protects against an MNIST 1-7 poisoning attack from 30% poisoners while Federated Learning struggles

Biscotti preserves utility



Biscotti achieves optimal performance compared to loss of utility when training with differential privacy

1. Rizzatti, L. (24 Sept 2016) "Digital Data Storage is Undergoing Mind-Boggling Growth." EE Times.
2. Wang et al. "Defending against Sybil Devices in CrowdSource Mapping Services", MobiSys 2016
3. Huang et al. "Adversarial Machine Learning" ICLR 2015
4. Melis et al. "Exploiting unintended feature leakage in collaborative learning" IEEE S&P 2019
5. Micali et al. "Verifiable Random Functions"FOCS 1999
6. Blanchard et al. "Byzantine Tolerant Gradient Descent" NIPS 2017
7. Dwork et al. "The foundations of algorithmic differential privacy" TCS 2014
8. Bonawitz et al. "Practical Secure Aggregation for Federated Learning on User-Held Data" CCS 2017
9. Gilad et al. "Algorand: Scaling Byzantine Agreements for Crypto Currencies" SOSP 2017

UBC ICICS NSERC CRSNG HUAWEI