DISTRIBUTED AND PRIVACY PRESERVING MACHINE LEARNING

A COMPARISON REVIEW OF APPROACHES

Advanced Topics in Distributed Systems

Sana Imtiaz

November 18, 2019

A PRIMER TO PRIVACY PRESERVATION

- Why is it important?

  - Machine learning is widely used in practice to produce predictive models for various applications
  - Models are more accurate when trained on huge amount of data
  - Massive data collection raises privacy concerns
  - ML model may inadvertently and implicitly store some of its training data
  - Careful analysis of the model may reveal sensitive information

Distributed and private ML to the rescue!

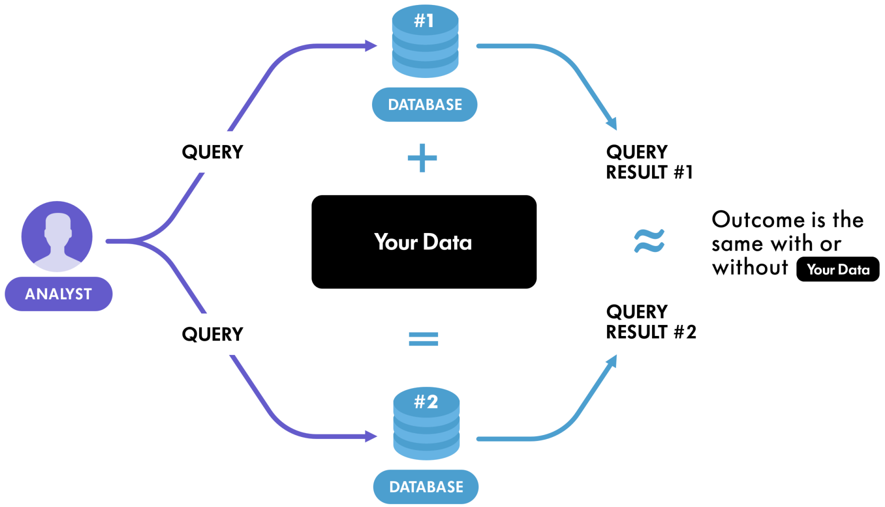
PAPERS TO BE REVIEWED

Problem: Distributed and privacy preserving machine learning (ML)

Approaches to address the problem:

1. Semi-supervised Knowledge Transfer for Deep Learning from Private Training Data (ICLR 2017)
2. SecureML: A System for Scalable Privacy-Preserving Machine Learning (IEEE S&P 2017)
3. Chiron: Privacy-preserving Machine Learning as a Service (arXiv; 2018)

DIFFERENTIAL PRIVACY

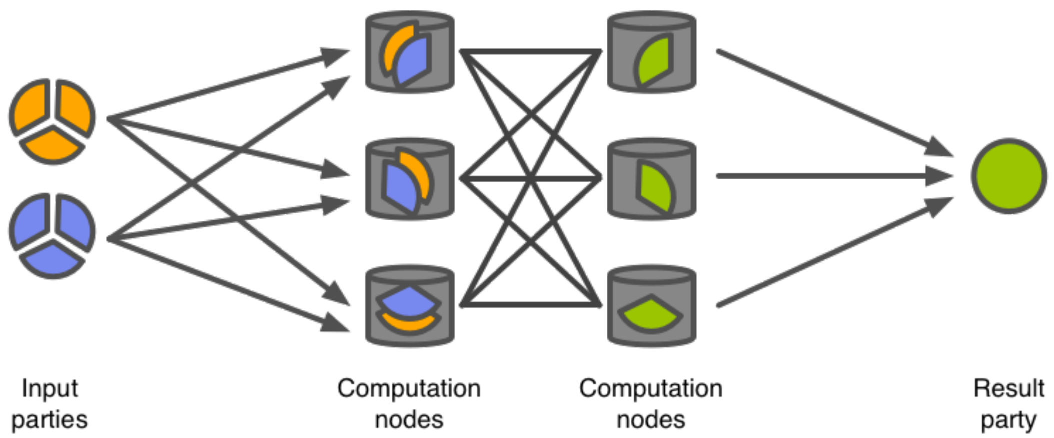


Definition 2.4 (Differential Privacy). A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ε, δ)-differentially private if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq \exp(\varepsilon) \Pr[\mathcal{M}(y) \in \mathcal{S}] + \delta,$$

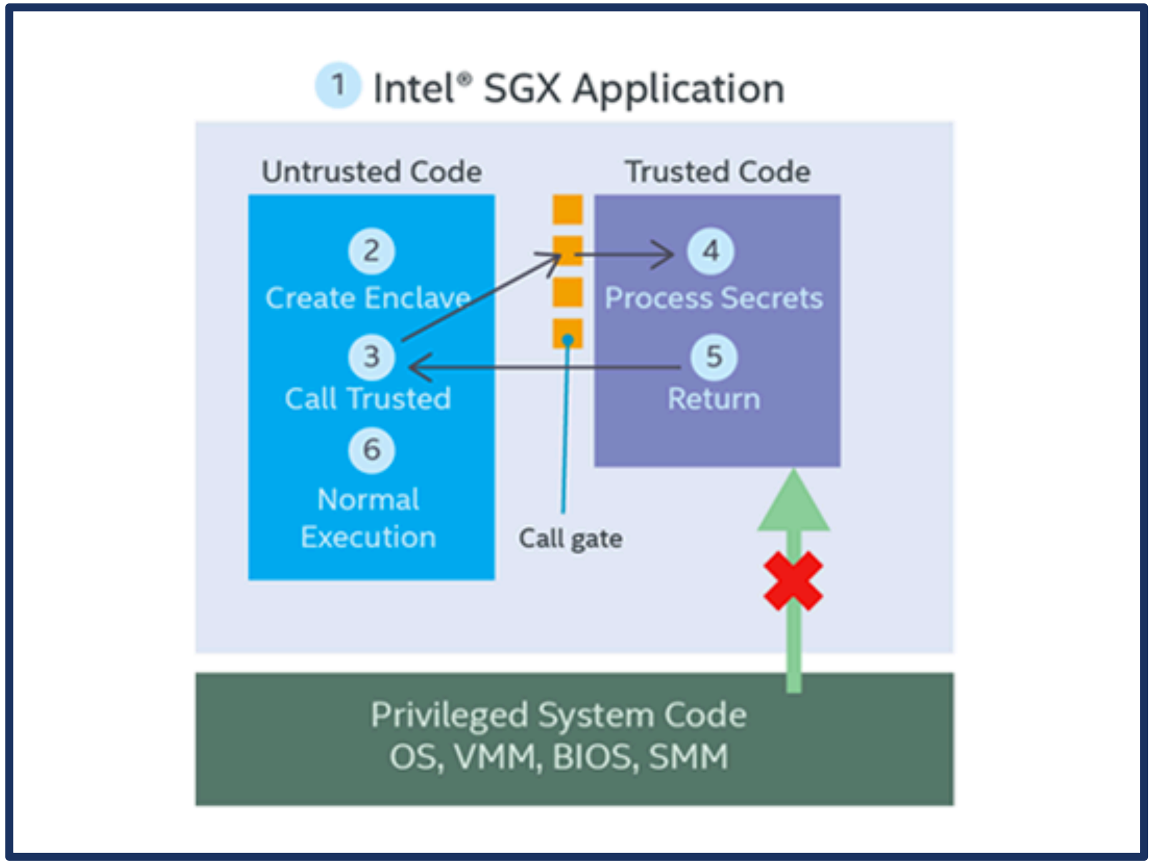
- Server is trusted party in this model
  - Data is fully disclosed to the provider/server
- Designed primarily for databases-based queries
- Laplacian noise added to the query result to ensure privacy preservation

SECURE MULTI-PARTY COMPUTATION



- Parties jointly compute a function using their inputs, while keeping these inputs private
- Uses secret sharing; divide and distribute one secret value over several nodes or users, so that no one knows anything about the secret value
- To retrieve the secret value, a minimum quorum of users must pool their data together

PRIVATE COMPUTE UNITS



1. App is built with trusted and untrusted parts
2. App runs and creates the enclave, which is placed in trusted memory
3. Trusted function is called, and execution is transitioned to the enclave
4. Enclave sees all processed data in clear; external access to the enclave data is denied
5. Function returns; enclave data remains in trusted memory
6. Normal execution resumes

PAPER 1

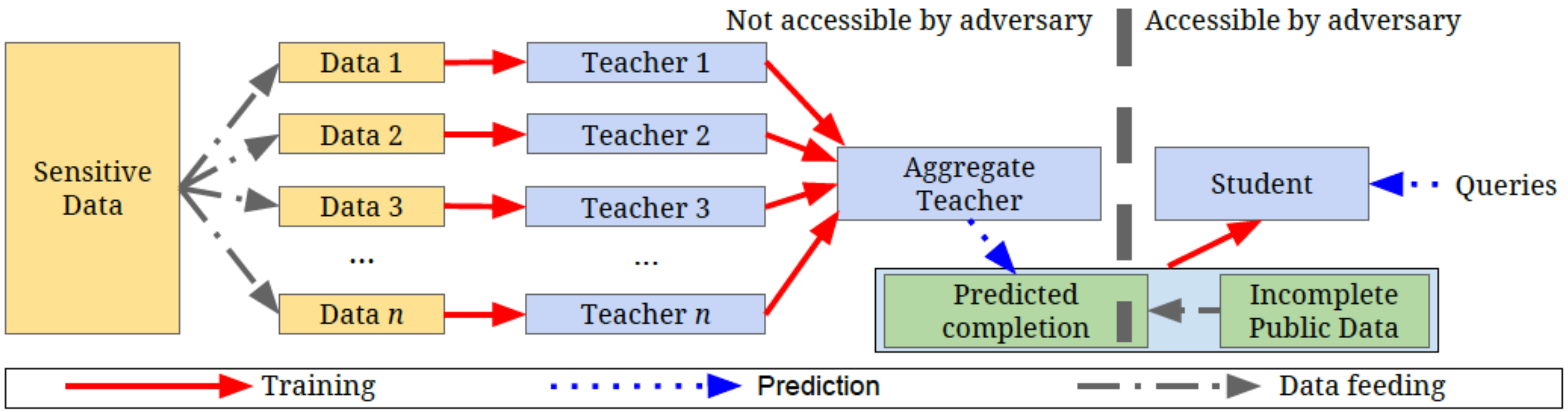
SEMI-SUPERVISED KNOWLEDGE TRANSFER FOR DEEP LEARNING FROM PRIVATE TRAINING DATA

SEMI-SUPERVISED KNOWLEDGE TRANSFER

Generally applicable approach to providing privacy guarantees for training data: Private Aggregation of Teacher Ensembles (PATE)

- Combines multiple models trained with disjoint datasets in a black-box fashion
- Models trained on sensitive data are not published
  - Used as "teachers" for a "student" model
- Student model learns to predict an output chosen by noisy voting among all of the teachers
- Student cannot directly access an individual teacher or the underlying data or parameters

SEMI-SUPERVISED KNOWLEDGE TRANSFER



- Uses differential privacy to limit the effect of any single sensitive data item on the student's learning
- Assumes the student has access to additional unlabeled public or non-sensitive data
- Variant using GANs for student learning (PATE-G) performs the best for different tested learning methods

SEMI-SUPERVISED KNOWLEDGE TRANSFER

Aggregation mechanism

- Counts majority vote for teachers in label classification
- Adds random Laplacian noise to the vote counts n_j to introduce ambiguity

$$f(x) = \arg\max_j \left\{ n_j(\vec{x}) + Lap\left(\frac{1}{\gamma}\right) \right\}$$

- Here, *j* is the assigned class label for input $\vec{x}$, γ is a privacy parameter, and *Lap(b)* is the Laplacian distribution with location 0 and scale b

PAPER 11

SECUREML: A SYSTEM FOR SCALABLE PRIVACY-PRESERVING MACHINE LEARNING

SECUREML: SCALABLE PRIVACY-PRESERVING MACHINE LEARNING

- Uses cryptography to ensure privacy preservation guarantees
- Server-aided setting where the clients outsource the computation to two untrusted but non-colluding servers S0 and S1
  - evaluator and a cloud service provider
- Clients distribute (secret-share) their inputs among the two servers in a setup phase and need not be involved in future computation
- Uses a combination of efficient techniques for boolean computation such as garbled circuits and Oblivious Transfer-extension, and arithmetic computation such as offline/online multiplication triplet shares

SECUREML: SCALABLE PRIVACY-PRESERVING MACHINE LEARNING

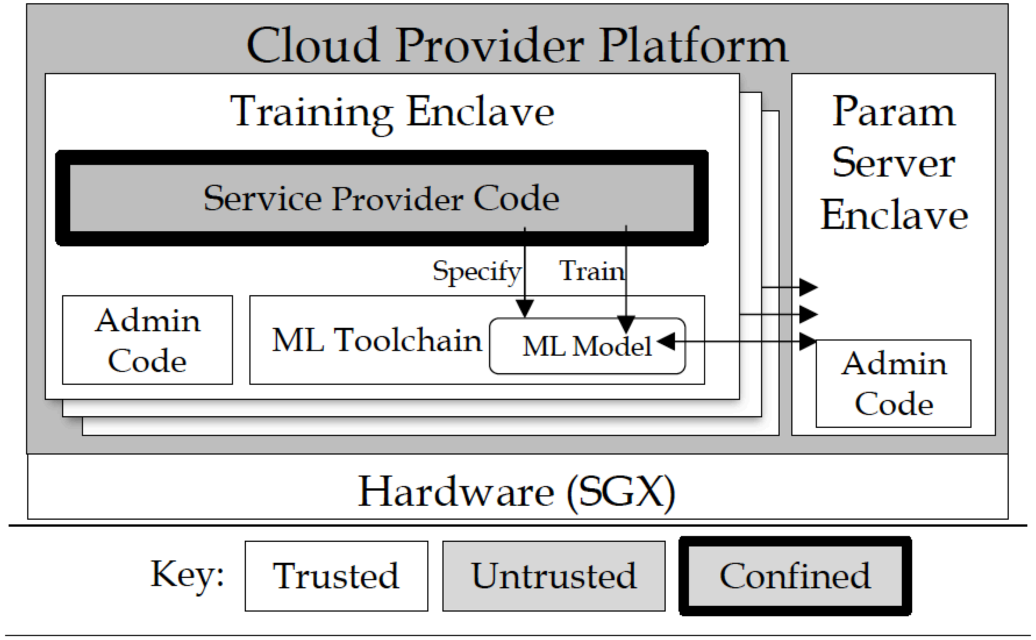
Note:

- By scalability, the authors imply that the system can handle large-scale datasets
- The system itself works in a 2-party scenario, where we may be able to scale the individual components of the system
- Not scalable in terms of traditional distributed systems

PAPER III

CHIRON: PRIVACY-PRESERVING MACHINE LEARNING AS A SERVICE

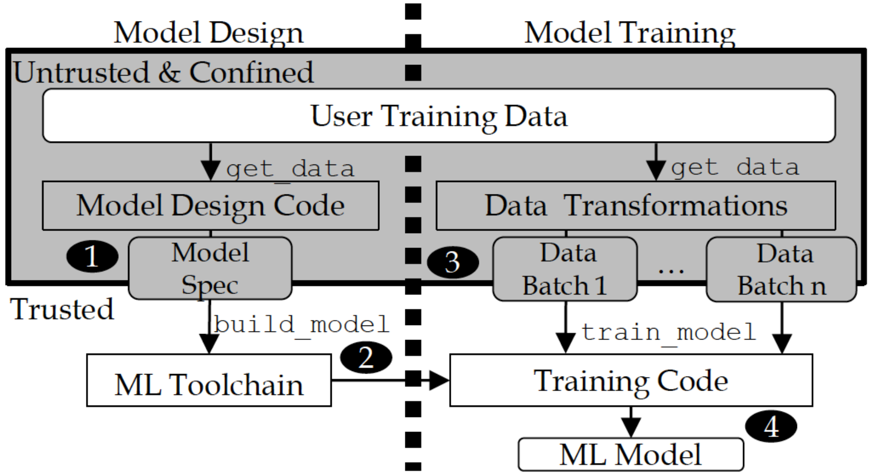
CHIRON: PRIVACY-PRESERVING ML AS A SERVICE



- Uses Intel SGX architecture and sandbox technology to provide privacy guarantees
- Applicable to ML as a service mechanisms provided by cloud providers
  - Users can not see inherent details of trained model and parameters
  - Providers do not see the training data
  - Clients can verify the validity of ML models offered
  - Third-party provider used as a trusted base to verify the validity of the model and the client

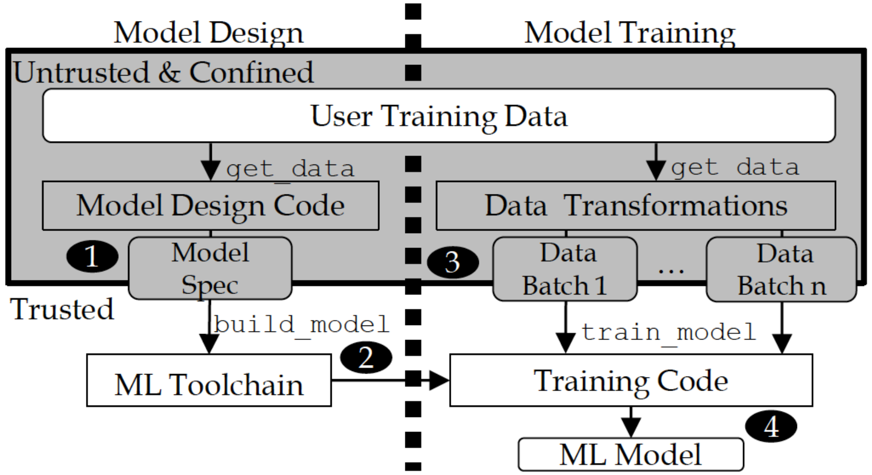
CHIRON: PRIVACY-PRESERVING ML AS A SERVICE

- Training code is public - its integrity can be remotely attested
- Users get a public-private key pair to communicate with the enclaves
- Service provider loads code in the sandbox and makes one or more training enclaves available to the user
- User connects to training enclaves and submits data



CHIRON: PRIVACY-PRESERVING ML AS A SERVICE

1. Service provider code examines data, then generates a model specification (model architecture, loss function, optimization function, and training hyperparameters) and passes it to the ML toolchain
2. ML toolchain uses the specs to generate model-training code
3. Service provider code transforms data and breaks it into batches for training
4. Model-training code is invoked for each batch, updating the model
5. After the model has been created, the user measures its test accuracy on a validation set and and proceeds to use the model



COMPARISON OF THE REVIEWED APPROACHES

COMPARISON OF THE REVIEWED APPROACHES

Trust Assumptions

1. PATE-G
   - Completely trusts the ML service provider
   - Does not trust the clients that may query the results of model
2. SecureML
   - No trust assumptions on clients
   - Adversary may collude with one of the serves but the servers do not collude among each other
3. Chiron
   - Trust assumption on clients
   - Protect users’ data from malicious providers of ML-as-a-service
   - Places trust on a verifiable and trusted third party

COMPARISON OF THE REVIEWED APPROACHES

Performance

- Paper I: For training data partitions n = 250
  - Average test accuracy of individual teachers is 83:86% for MNIST and 83:18% for SVHN
- Paper II:
  - MNIST dataset, the model trained by Tensorflow (with softmax) can reach 94.5% accuracy on all 10 classes, while we reach 93.4% using our proposed function

Dataset	ε	δ	Queries	Non-Private Baseline	Student Accuracy
MNIST	2.04	10^{-5}	100	99.18%	98.00%
MNIST	8.03	10^{-5}	1000	99.18%	98.10%
SVHN	5.04	10^{-6}	500	92.80%	82.72%
SVHN	8.19	10^{-6}	1000	92.80%	90.66%

COMPARISON OF THE REVIEWED APPROACHES

Performance

- Paper III: Using 16 training enclaves and ImageNetLite dataset
  - Chiron slows down ImageNetLite training by 16%, while preserving the accuracy of the trained model
  - Other results in Table 2

	Top1(%)	Top5(%)	Train(hr)	Query(sec)
Baseline	55.12	78.51	39.83	3825.30
Chiron	52.41	76.42	38.85	3843.53

Table 2. Model accuracy, training time, and query time for ImageNetLite. Top 1 is the accuracy for the most likely prediction. Top 5 is the accuracy of the five most likely predictions. Query shows time for querying 100,000 images in batches of 1,000.

COMPARISON OF THE REVIEWED APPROACHES (2)

Shortcomings and weaknesses

- Semi-supervised knowledge transfer:
  - large set of trials needed to figure out the right size for data partitions and privacy parameters
  - Public and private portions of datasets must be available

- SecureML:
  - No real scalability shown, though claimed in the paper

- Chiron: dedicated ML toolchain
  - Service provider code must use Chiron's ML toolchain to define the model
  - All other forms of output are disallowed by Chiron's confinement

CONCLUSIONS

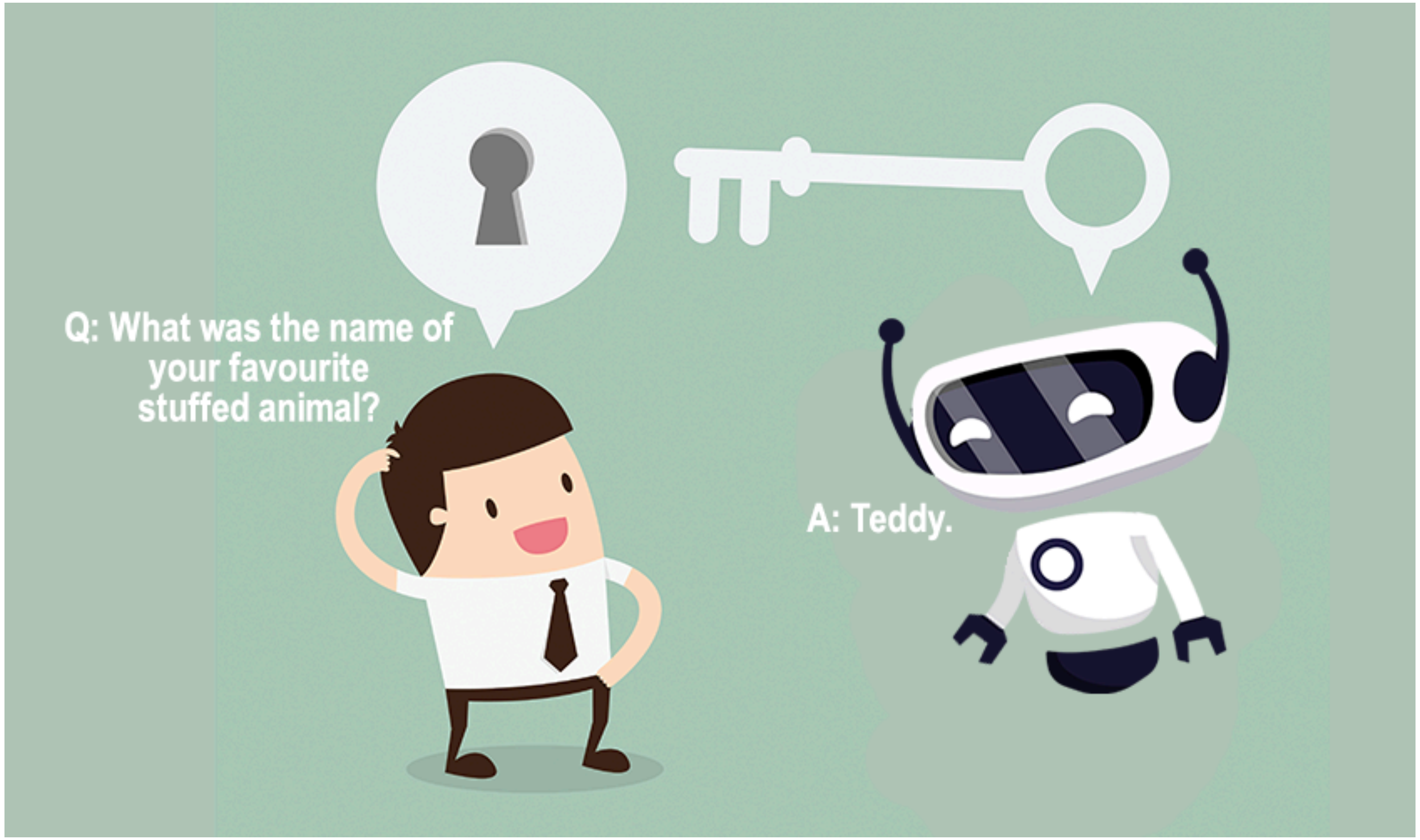
TAKEAWAYS AND FUTURE WORK

CONCLUSIONS

- The field of distributed privacy-preserving machine learning is nascent and has huge room for research
- The tuning of privacy-preservation parameters often needs to be done on a hit-and-trial basis, which might not be ideal for big data scenarios
  - For example, Paper 1 used 250 partitions of data (determined by experiment) and fine turned different privacy settings for the best results quoted in the paper
- Cryptography based solutions have huge impact on efficiency of the system though they provide very strong privacy guarantees

CONCLUSIONS

- Privacy preserving ML as a service is a good solution but places trust on third-party and efficiency requirements need further investigation
- Appropriate choice for privacy preserving method can be applied under these considerations:
  - Availability of public and private versions of dataset
  - Efficiency and accuracy constraints
  - Adversary model and trust assumptions



QUESTIONS