Encrypted Speech Recognition using deep polynomial networks

Austin Zhang, **Dong Yu**

Tencent AI Lab

Yifan Gong

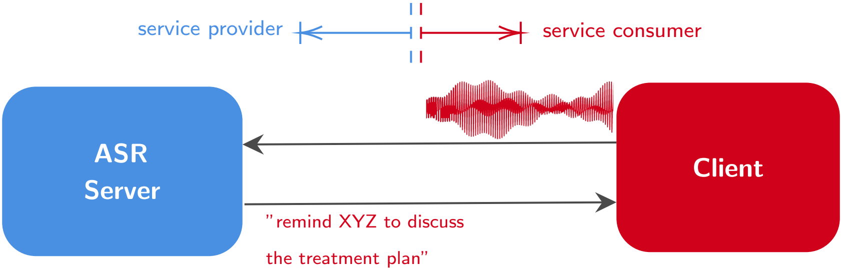
Microsoft

Table of contents

Motivation

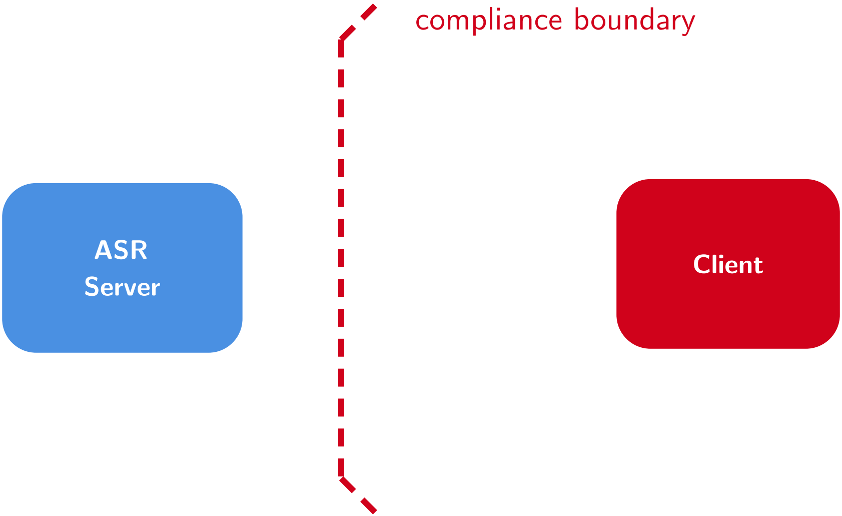
Why doing this?

- State-of-the-art speech recognition services are running on cloud
- However, this will leak the client's private information to the server.
  e.g., medical/financial/enterprise/sensitive data



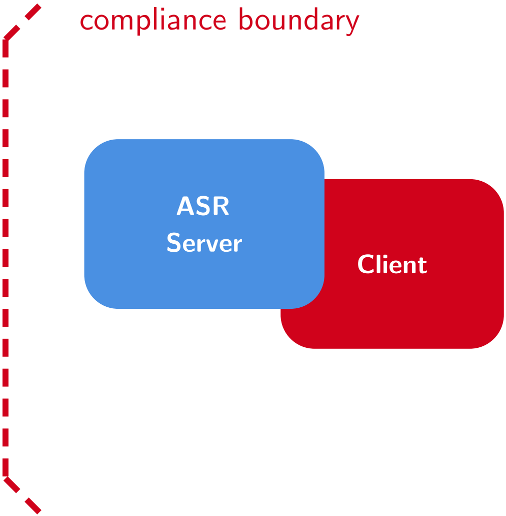
How to protect privacy? — A private-cloud solution

- Step 1: build a compliance boundary → prevent data to leak out



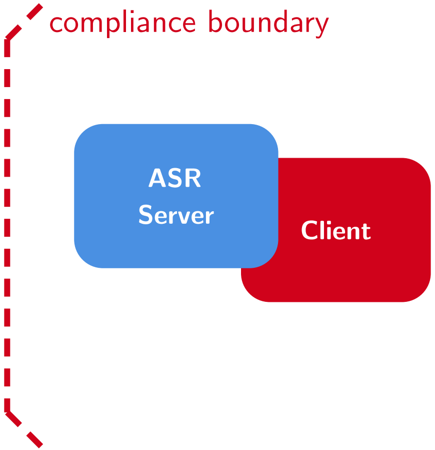
How to protect privacy? — A private-cloud solution

- Step 2: move ASR service inside of compliance boundary



How to protect privacy? — A private-cloud solution

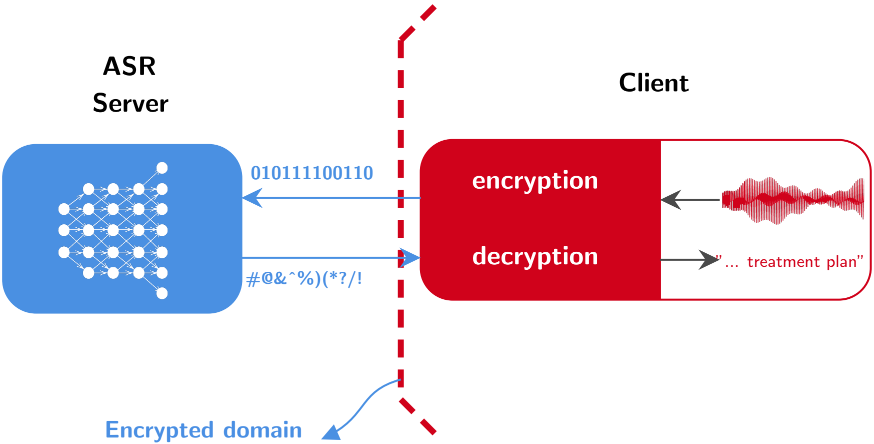
- issues 1: hard to deploy an update to the private cloud
- issues 2: costly for some small business/individual users
- issues 3: service provider may divulge the model and decoder to the service consumer who may resale to others



Encrypted Speech Recognition

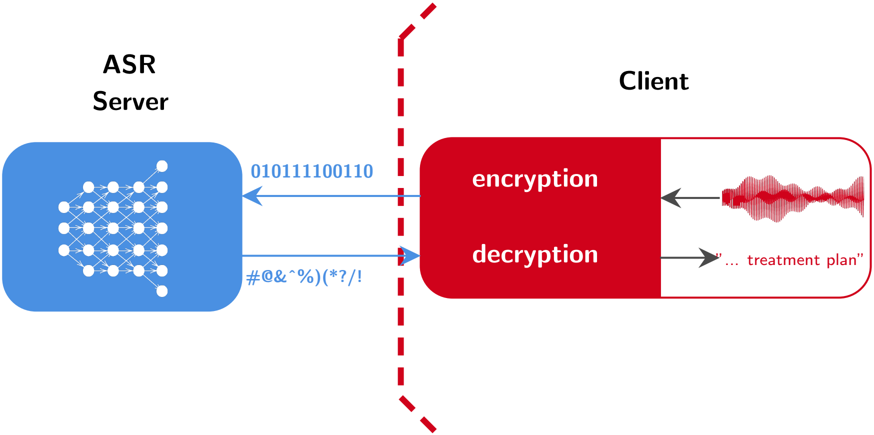
What if we can encrypt the data

- Ideally, we want ...



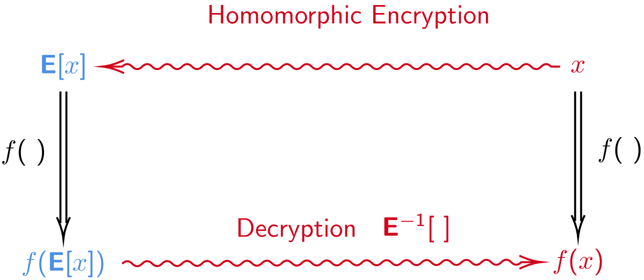
What if we can encrypt the data

- Does this encryption exist?



Homomorphic Encryption

$\mathbf{E}^{-1}\Big[f\big(\mathbf{E}[x]\big)\Big] \equiv f(x) \rightarrow$ An elegant solution for all above questions



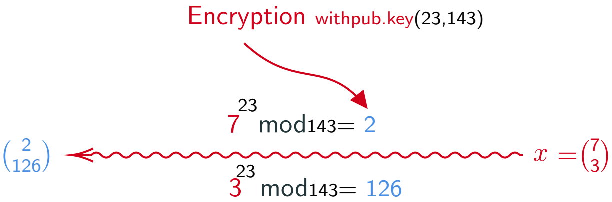
Homomorphic Encryption — example

$$x = \binom{7}{3}$$

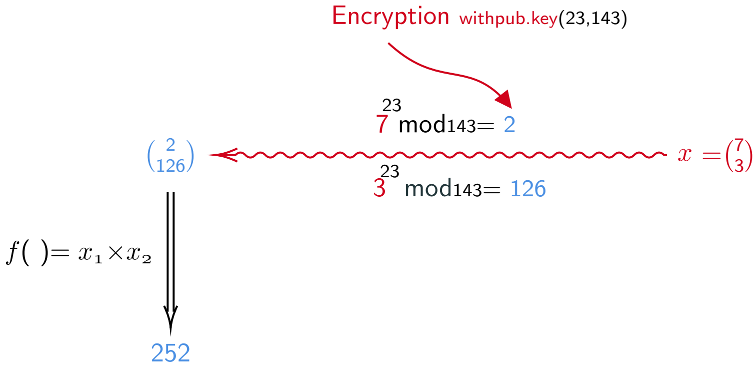
$$f(\ ) = x_1 \times x_2$$

$$f(x) = 21$$

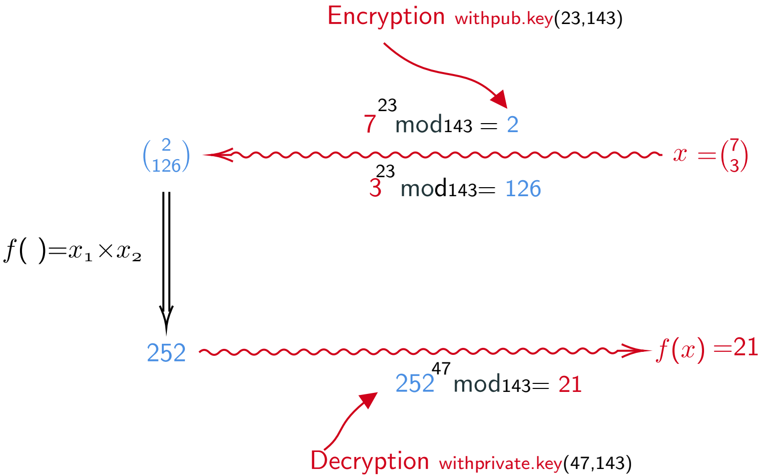
Homomorphic Encryption — example



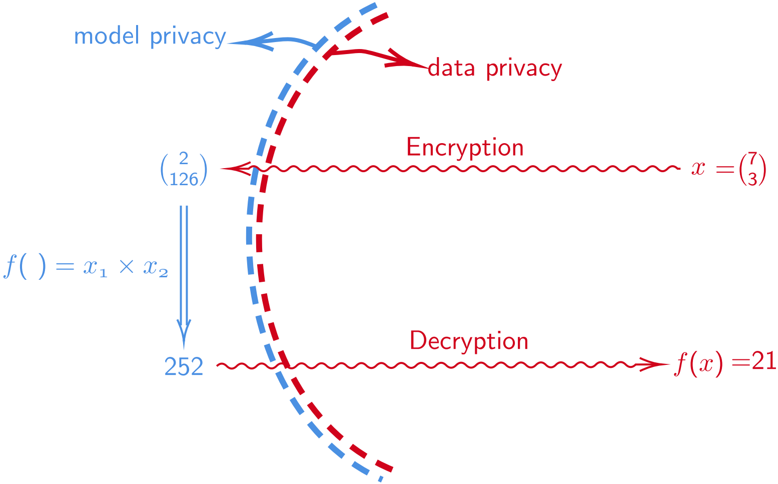
Homomorphic Encryption — example



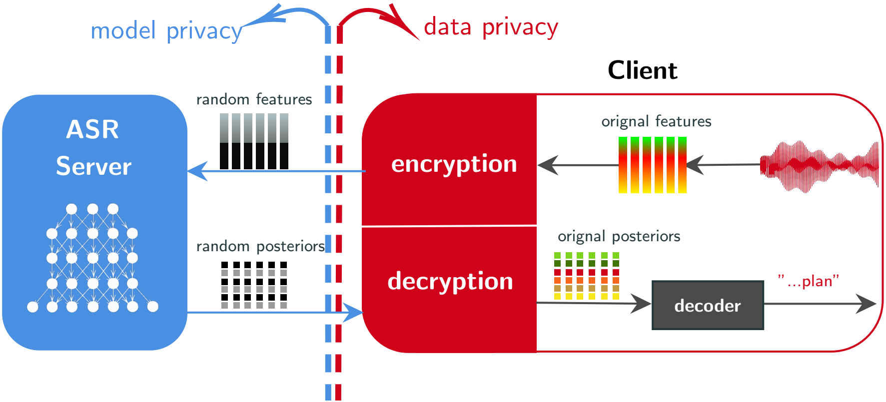
Homomorphic Encryption — example



Homomorphic Encryption — example



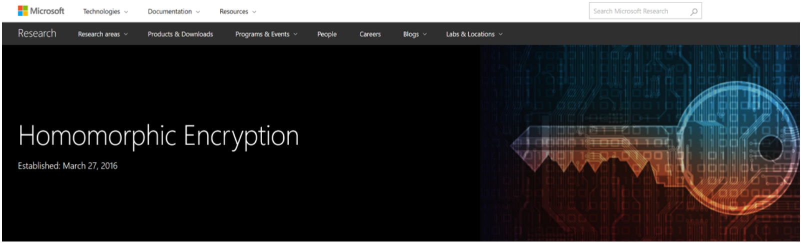
Proposed Framework



- Only AM scores are computed on server side.
- Original result guaranteed after decryption.
- No need to retrain the DNN on encrypted data.

What about Latency?

It is extremely slow and not feasible before, until ...

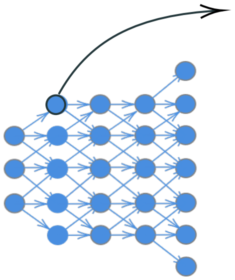


Microsoft researchers smash homomorphic encryption speed barrier!

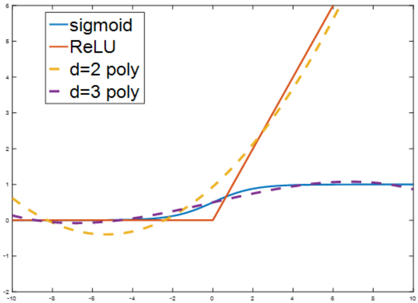
- But $f(\cdot)$ must be polynomial
- must be fixed point operation
- open source $\rightarrow$ http://sealcrypto.org/

Deep Polynomial Network

Deep polynomial network

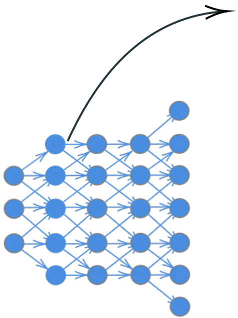


Replace ReLU and Sigmoid as polynomials



- unbounded polynomial approximation $\rightarrow$ batch norm is a must.

Deep polynomial network



Dense layer (polynomial)

$$\mathbf{E}[\mathbf{W}]^{\mathsf{T}}\mathbf{E}[\mathbf{x}] \xrightarrow{\mathbf{E}^{-1}} \mathbf{W}^{\mathsf{T}}\mathbf{x}$$

Convolution layer (polynomial)

Batch norm (merged to dense layer)

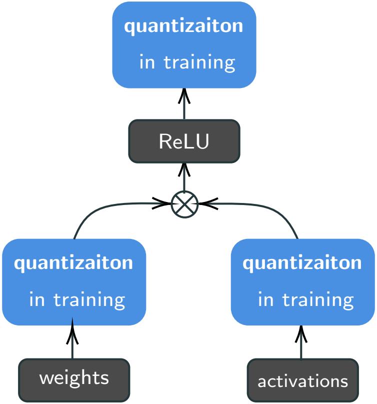
$$\mathbf{W}^{\mathsf{T}}\left(\gamma\frac{\mathbf{z}-\mu}{\sigma}+\beta\right)+\mathbf{b} \;=\mathbf{W}'^{\mathsf{T}}\mathbf{z}+\mathbf{b}'$$

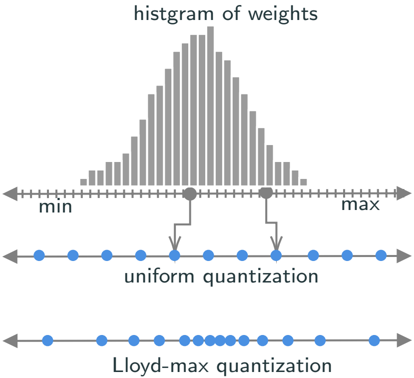
Max pooling layer(approximate)

$$\max\left(x_1,\ldots x_n\right)=\lim_{d\to\infty}\left(\sum_{i=1}^{n}x_i^d\right)^{\frac{1}{d}}$$

Low-bit quantization

low-bit model is critical for encryption speed





Experimental Results

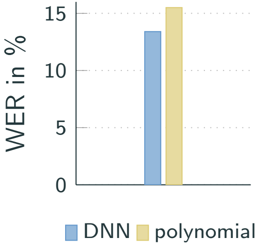
WER on Switchboard

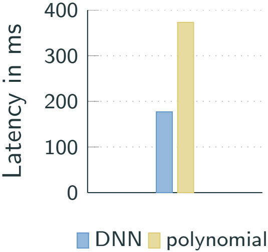
WER in %		16-bit	8-bit	4-bit	2-bit
DNN	quantized train	14.7%	14.7%	14.9%	30.3%
	→ polynomial	15.8%	15.8%	16.1%	30.8%
CNN	quantized train	12.2%	12.3%	12.7%	–
	→ polynomial	13.5%	13.6%	14.0%	–

- with proper quantized training, 4-bit is sufficient.
- the polynomial networks increase WERs by a little as a cost.

WER and Latency on Cortana Task

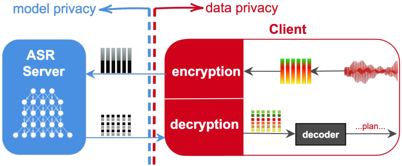
	16-bit	4-bit	avg. latency per utterance: encryption	decryption	overall
DNN	12.9%	13.4%	–	–	177ms
polynomial	14.8%	15.5%	202ms	16ms	373ms



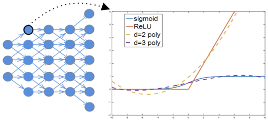


Summary

- a framework that enables privacy-preserving speech recognition



- a polynomial network that can make predictions over the encrypted speech in real time.



- with quantized training, 4-bit is sufficient for DNN/CNN.

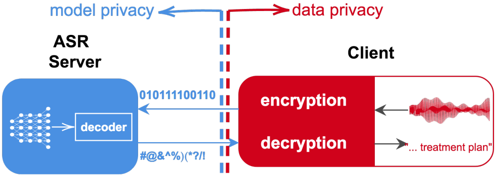
Thanks.

Questions?

Tencent AI Lab

Future work

- make the decoder also work on encrypted domain, so that we could run everything on the cloud.



- investigate training on encrypted data so that multiple parties (e.g.Microsoft, Google and Amazon) can encrypt and combine their data together to train models without sacrificing users privacy.