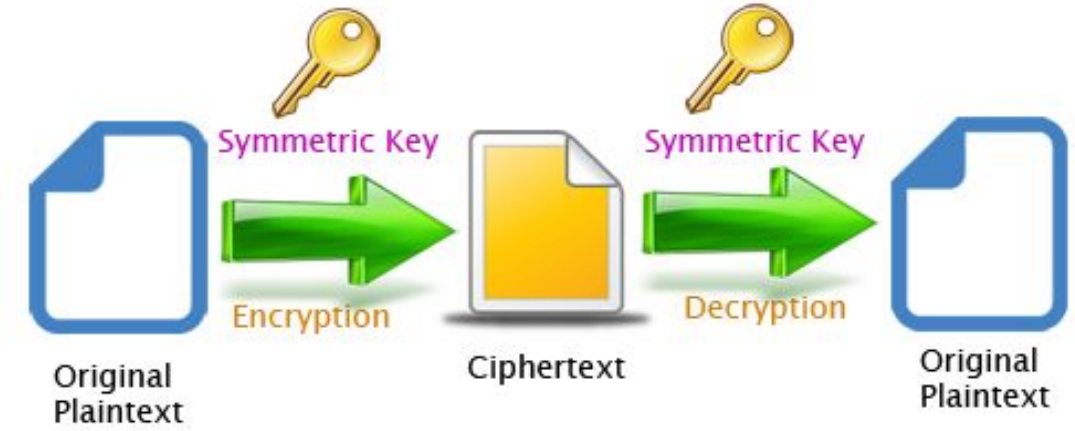
Homomorphic Encryption

By Raj Thimmiah

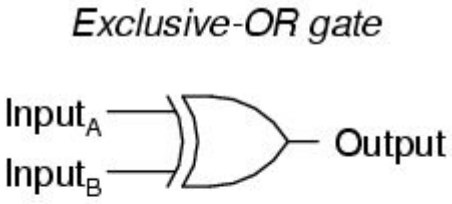
Symmetric Key Encryption

Symmetric Key Encryption



Symmetric Key Encryption: XOR Gates

- XOR gates are the simplest way to implement symmetric key encryption
- XOR gates have two inputs (A and B) and output a result from that



A	B	Output
0	0	0
0	1	1
1	0	1
1	1	0

Symmetric Key Encryption with One Time Pads

Message	0	1	1	0	0	0
	⊕					
Key	1	1	0	1	1	0
	=					
Ciphertext	1	0	1	1	1	0

Exclusive-OR gate

A	B	Output
0	0	0
0	1	1
1	0	1
1	1	0

Symmetric Key Encryption with One Time Pads

Message	1	0	1	1	1	0
	⊕					
Key	1	1	0	1	1	0
	=					
Ciphertext	0	1	1	0	0	0

Exclusive-OR gate

A	B	Output
0	0	0
0	1	1
1	0	1
1	1	0

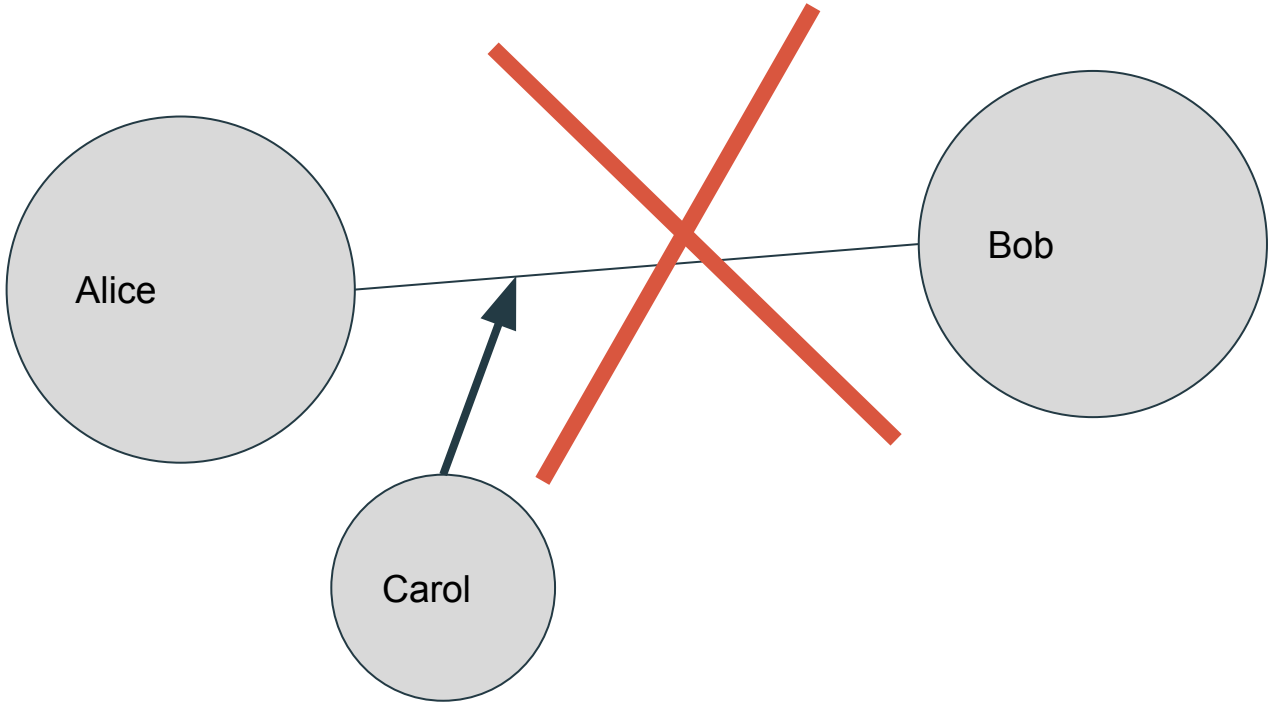
Modern Symmetric Key Encryption

- Since one time pads must scale with the size of the message, they are impractical for large pieces of data
- There are many modern encryption algorithms that can use keys commonly with a size of 128, 268, and 256 bits
- AES (advanced encryption standard) is one of the most common of such algorithms and the only real (current) attack vector is brute forcing which is still impractical as it would take a very long time to break even a single message

Flaws of Symmetric Key Encryption

- Cannot be used on an unsafe channel

Flaws of Symmetric Key Encryption

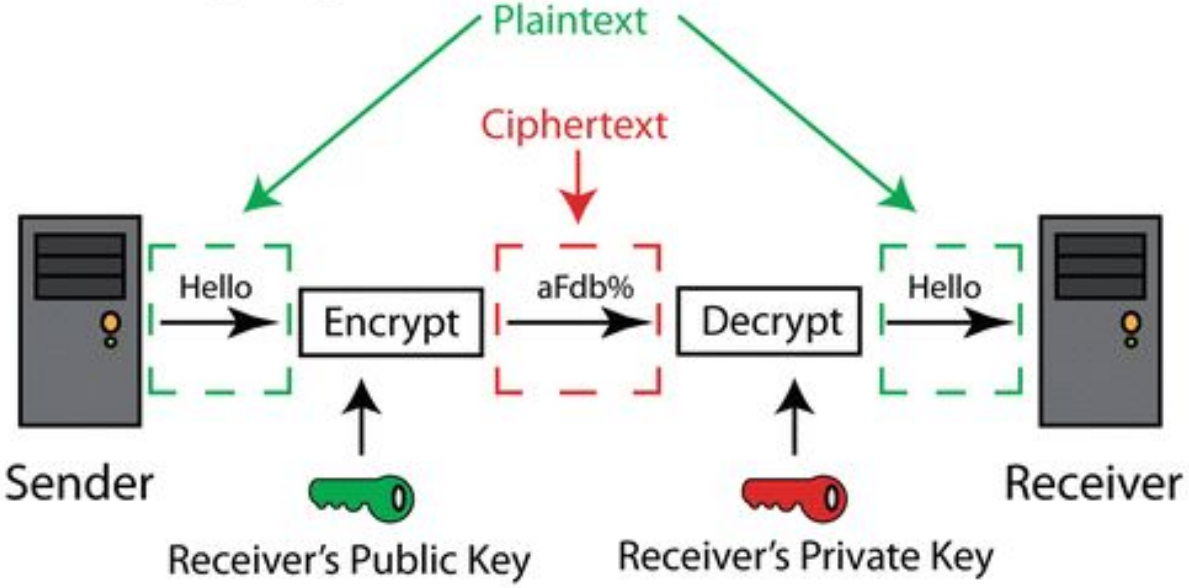


Asymmetric Key Encryption

Asymmetric Key Encryption

- Can communicate on an untrusted channel
- Each person now has 2 keys
  - Public Key
    - Shared publicly
  - Private key
    - Held securely by user

Asymmetric Key Encryption



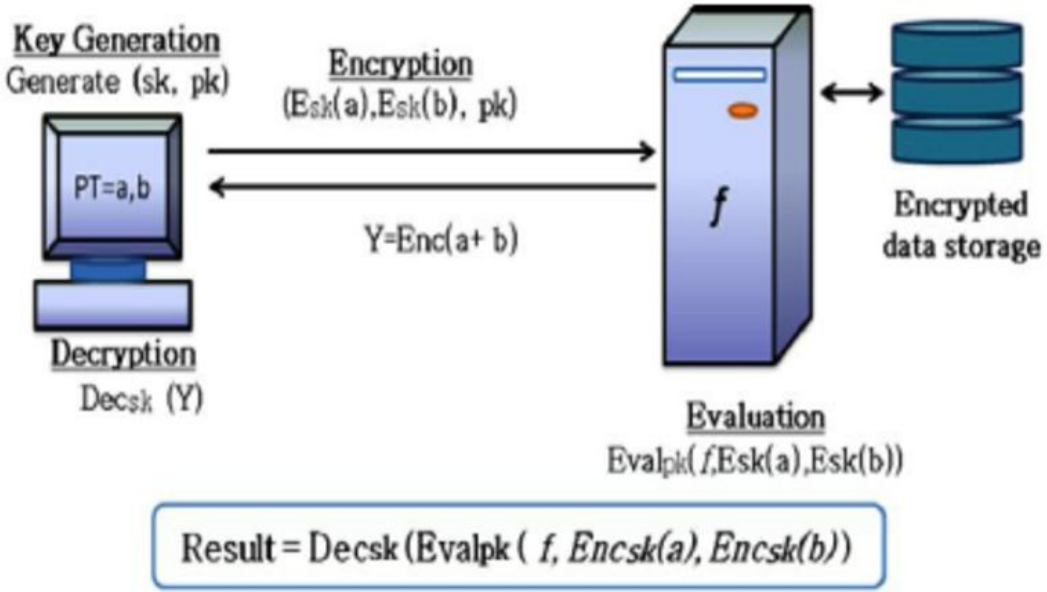
http://www.springer.com/kr/book/9783319122281

Weaknesses

- Must trust other party with the data you send
- Data must be decrypted to be processed
    - While decrypted, data is not secure

Homomorphic Encryption

Homomorphic Encryption



Alice’s Jewelry Shop

https://crypto.stanford.edu/craig/easy-fhe.pdf

Alice’s Jewelry Shop

- Alice wants her workers to process her materials without having access to them
- She creates an (impenetrable) box with gloves for access to the materials inside
- The end contents of the box can only be accessed by her as only she has the key
- The only option is for the worker to process the materials and then give Alice the result

https://crypto.stanford.edu/craig/easy-fhe.pdf

Parallels with Traditional Client-Server Architecture

- Data can be sent securely but there are no guarantees it is processed securely
- Must either do all processing yourself or trust blindly


GETINGE


GETINGE

Somewhat Homomorphic Encryption (SWHE)

- Can do basic operations
- Each operation adds noise to the result
- Beyond a certain point, output cannot be decrypted
- Thus only computations up to a certain complexity can be done

https://crypto.stanford.edu/craig/easy-fhe.pdf


GETINGE

Alice’s Jewelry Shop

- The impenetrable boxes that Alice has been giving to her workers have gloves that stiffen after 1 hour meaning that they can only make basic jewelry in that time unless she takes out the jewelry and puts it in a new box which wastes time
- To get around this issue, she uses multiple boxes (eg. 1, 2, 3, 4 with 4 being the outermost box)
- The worker first works within box #1 and when box #1’s gloves stiffen, they can then unlock the first box with a key that is present in box #2 and be worked on further.
- Regardless of creation time, any piece of jewelry can now be worked on in a theft-free environment

https://crypto.stanford.edu/craig/easy-fhe.pdf

Fully Homomorphic Encryption (FHE)

- To transform the previous system into a system that allows for unlimited operations, Gentry proposed “bootstrapping” in 2009

Use Cases

Computing On Private Information

SECURE GENOME ANALYSIS COMPETITION

Three tracks of competition tasks

Track 1: Practical Protection of Genomic Data Sharing through Beacon Services (privacy-preserving output release)

Given a sample Beacon database, we will ask participating team to develop solutions to mitigate the Bustamante attack. We will evaluate each algorithm based on the maximum number of correct queries that it can respond before any individual can be re-identified by the Bustamante attack. (data link)

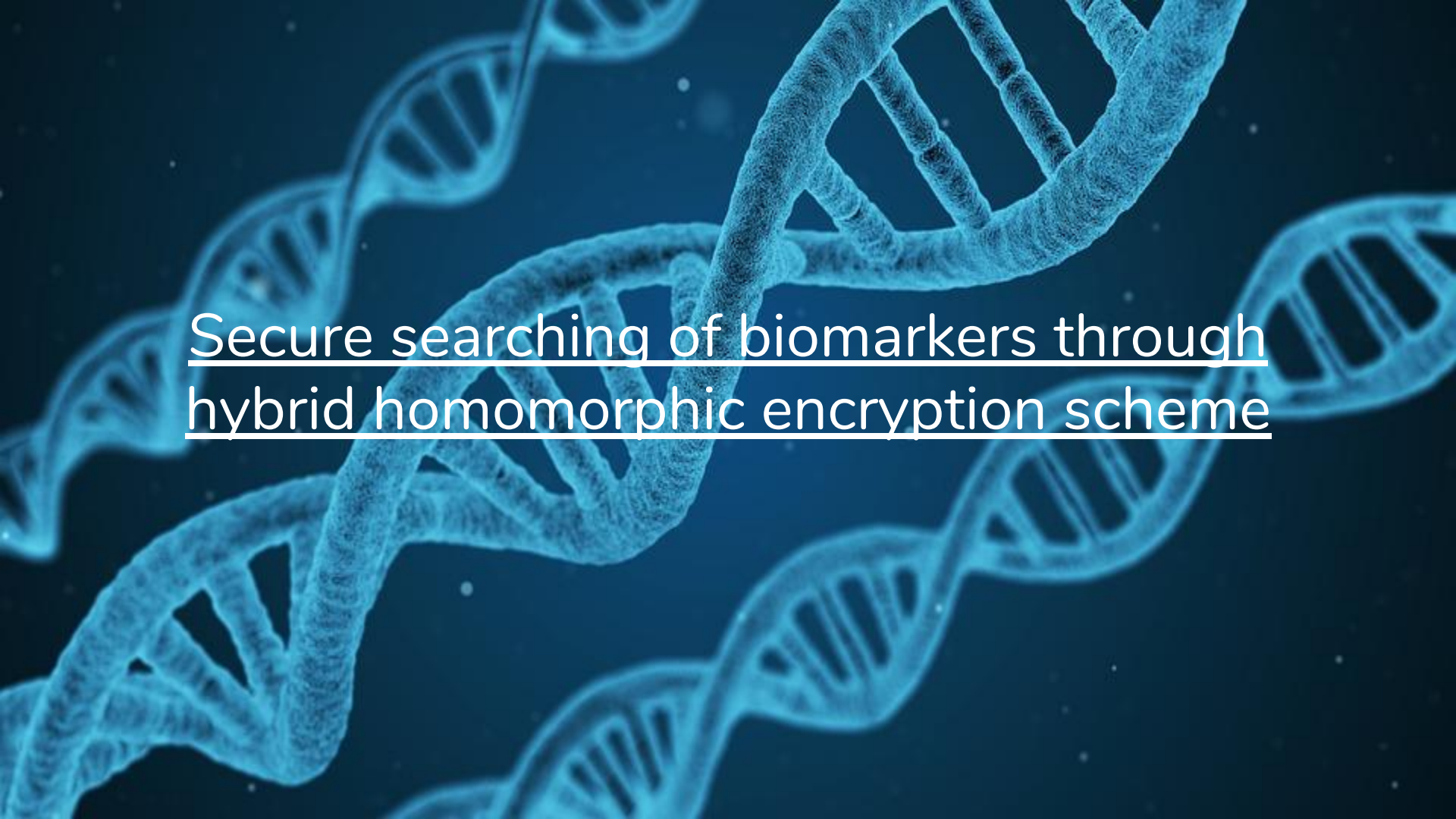
Track 2: Privacy-Preserving Search of Similar Cancer Patients across Organizations (secure multiparty computing)

The scenario of this track is to find top-k most similar patients in a database on a panel of genes. The similarity is measured by the edit distance between a query sequence and sequences in the database. We expect participating teams come up with different algorithms that can provide good approximation to the actual edit distance and also be efficient. (data link)

Track 3: Testing for Genetic Diseases on Encrypted Genomes (secure outsourcing)

This is to calculate the probability of genetic diseases through matching a set of biomarkers to encrypted genomes that stored in a commercial cloud service. The requirement is that the entire matching process (only consider the exact match for each variation) needs to be carried out using homomorphic encryption so that no trace is left behind during the computation. (data link (example code updated on 5/22/2016))

- Genome sequencing widely available but there are a number of personal privacy issues
- Can find a matching set of biomarkers in DNA sequences homomorphically
- Takes around 4 seconds to search a database of size 4 million


Secure searching of biomarkers through hybrid homomorphic encryption scheme

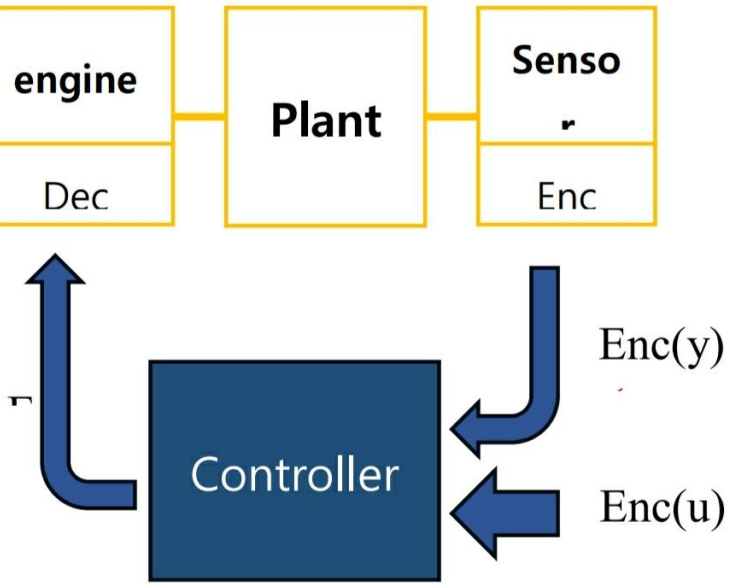
Outsourcing IoT/Mobile Processing and Storage

Outsourcing IoT/Mobile Processing and Storage

https://github.com/Talos-crypto/Pilatus http://www.vs.inf.ethz.ch/publ/papers/mshafagh_SenSys17_Pilatus.pdf

Protecting Control Systems

Protecting Control Systems

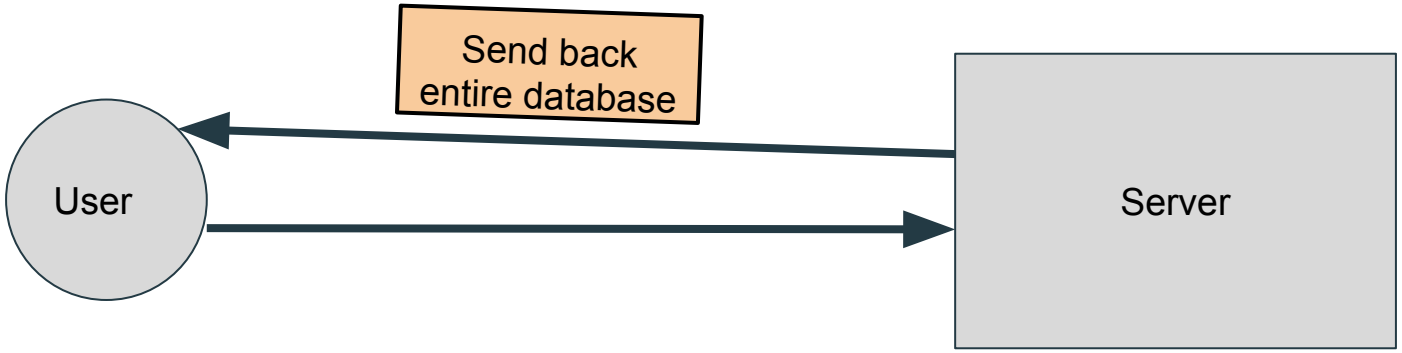


Private Information Retrieval (PIR)

Traditional Information Retrieval (non-private)

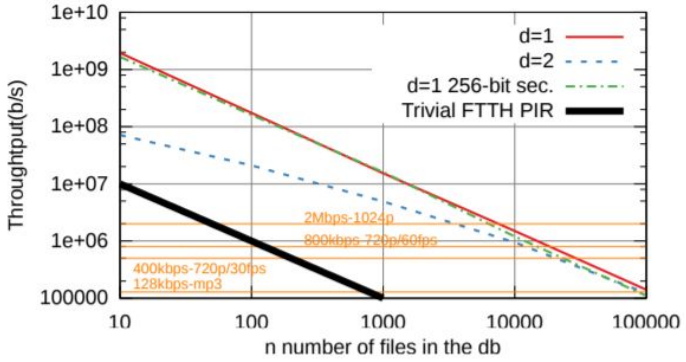
- Simply send a request to a server for a specific file and receive a result quickly
- Processing speed is minimal

Trivial Private Information Retrieval



PIR Using Homomorphic Encryption

- While there are many *many* papers on this topic one of the best implementations I've seen (at least based on the performance they claim) is XPIR
- Added latency and lower bandwidth compared to traditional solutions in exchange for privacy
- Github Repo
- Paper

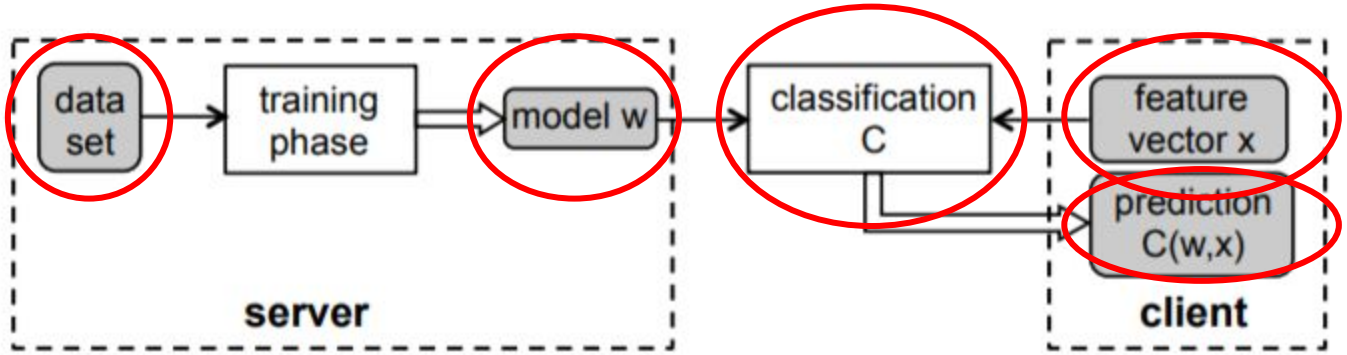


Predictions on Data from Multiple Sources

Data Aggregation and Querying

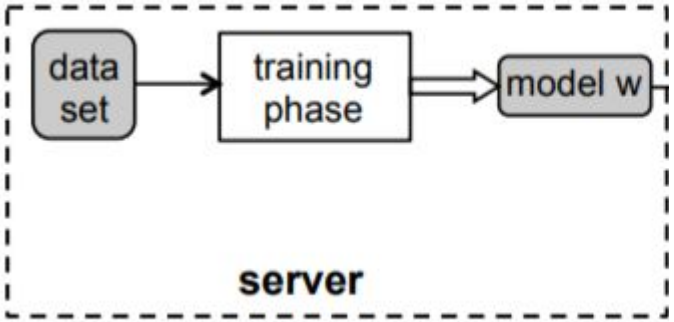
- A teacher might want to know a student's likelihood of dropping out
- This likelihood might be based on private data from multiple sources such as the police department, hospital, welfare information
- This data is all encrypted with different keys and is from many different sources
- Using multi party computation (MPC) a teacher can aggregate this data and with homomorphic encryption, process and retrieve some result

Private Machine Learning


data set
training phase
model w
classification C
feature vector x
prediction C(w,x)
server
client

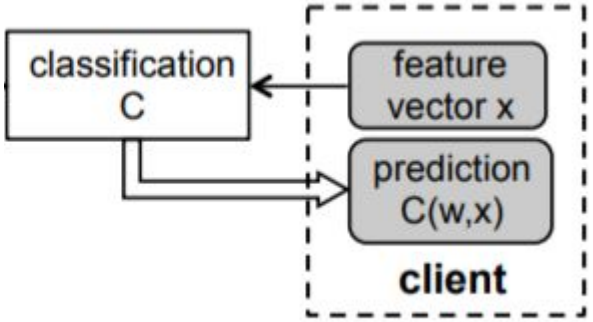
Hiding Training Data and Model from Server

- Outsourced Training
  - Computationally bounded training
- Training on private data
  - Train a model on private individual data
  - Data aggregation by a cloud service provider and selective training/analysis



Hiding Classifier from User and Input from Classifier

- Sensitive private data
  - Prediction for medical data
- Outsourced classification
  - Computation bound predictions
- Expensive and proprietary classifier
  - Hide algorithm from hospitals/competitors while being HIPAA compliant
  - Location Privacy
- Secure tamper proof control systems



Successful Projects

- CryptoDL: Deep Neural Networks over Encrypted Data
  - 99.52% accuracy in optical character recognition on encrypted inputs (trained on MINST dataset)
    - 16,400 predictions per hour
- Privacy-Preserving Visual Learning Using Doubly Permuted Homomorphic Encryption
  - Update model weights on private data locally and average locally updated classifiers in the cloud
- Privacy Preserving Multi-party Machine Learning with Homomorphic Encryption
  - Enables multiple parties to train a machine learning model from aggregated data
  - Ensures privacy of each party's datasets

Libraries

Libraries/Projects

- HELIB (Paper)
- R Stats thing (Paper)
- LoL (Paper)
- Pilatus (Paper)
- XPIR (Paper)
- PySyft Private Deep Learning Client
- Suggested Use Cases

https://arxiv.org/pdf/1508.06574.pdf