From Trust to Zero-Trust: Rethinking Privacy and Security in the Cloud Era

Andrei Tchernykh

CICESE Research Center, Ensenada, Baja California, México

chernykh@cicese.mx

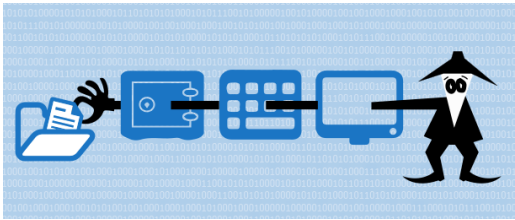
http://usuario.cicese.mx/~chernykh/

- Head of “Parallel and Cloud Computing Laboratory” at CICESE, Mexico
- Adjunct Chief Scientist (habilitation) of the Institute for System Programming of the Russian Academy of Sciences, Moscow, Russia
- Head of “Laboratory of Problem-Oriented Cloud Computing” at South Ural State University. Chelyabinsk, Russia.

CIBERTIC 2025

Congreso Internacional de Ciberseguridad, Tecnologías, Innovación y Ciencia

Guadalajara, Jalisco, México, May 20-22, 2025



Data Security Requirements (CIA triad)

1. **Confidentiality**
   Protecting from disclosure
2. **Integrity**
   Modification or deletion
3. **Availability**
   Users can access information
4. **Privacy**
   free from public attention, not observed or disturbed
5. **Scalability**
   ability to be used in a range of capabilities

"I'm no expert, but I think it's some kind of cyber attack!"

Data Security Requirements (CIA triad)

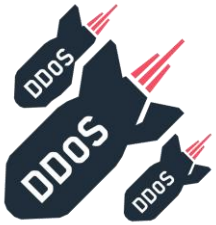
1. Confidentiality
   Protecting from disclosure
2. Integrity
   Modification or deletion
3. Availability
   Users can access information
4. Privacy
   free from public attention, not observed or disturbed
5. Scalability
   ability to be used in a range of capabilities

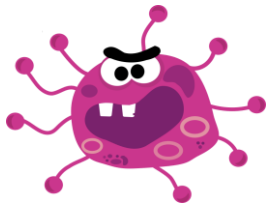
Solutions

1. Symmetric encryption: **AES, DES, 3DES**
2. Asymmetric encryption **RSA, ECDSA, DSA**
3. Digital signature
4. Access control
5. Data masking

Resiliency threads

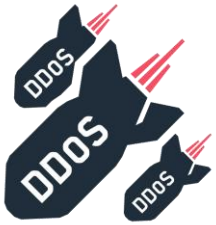
1. Environmental threats
   - Earthquakes, floods, fire, etc.
2. Deliberate threats
   - Interception, hacker attacks, etc.
3. Accidental threats
   - PC errors, Virus, Spam, etc.
4. Unfairness
   - User errors, carelessness, curiosity, falsification, etc.

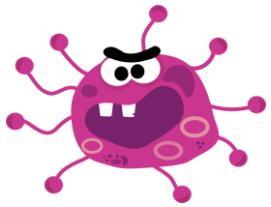




Resiliency threads

1. Environmental threats
   - Earthquakes, floods, fire, etc
2. Deliberate threats
   - Interception, hacker attacks, etc
3. Accidental threats
   - PC errors, Virus, Spam, etc.
4. Unfairness
   - User errors, carelessness, curiosity, falsification, etc.





Solutions

1. Replication
2. Erasure codes
3. Error detection-correction
4. Checkpointing

Security and Resiliency

Data security

Protecting data from

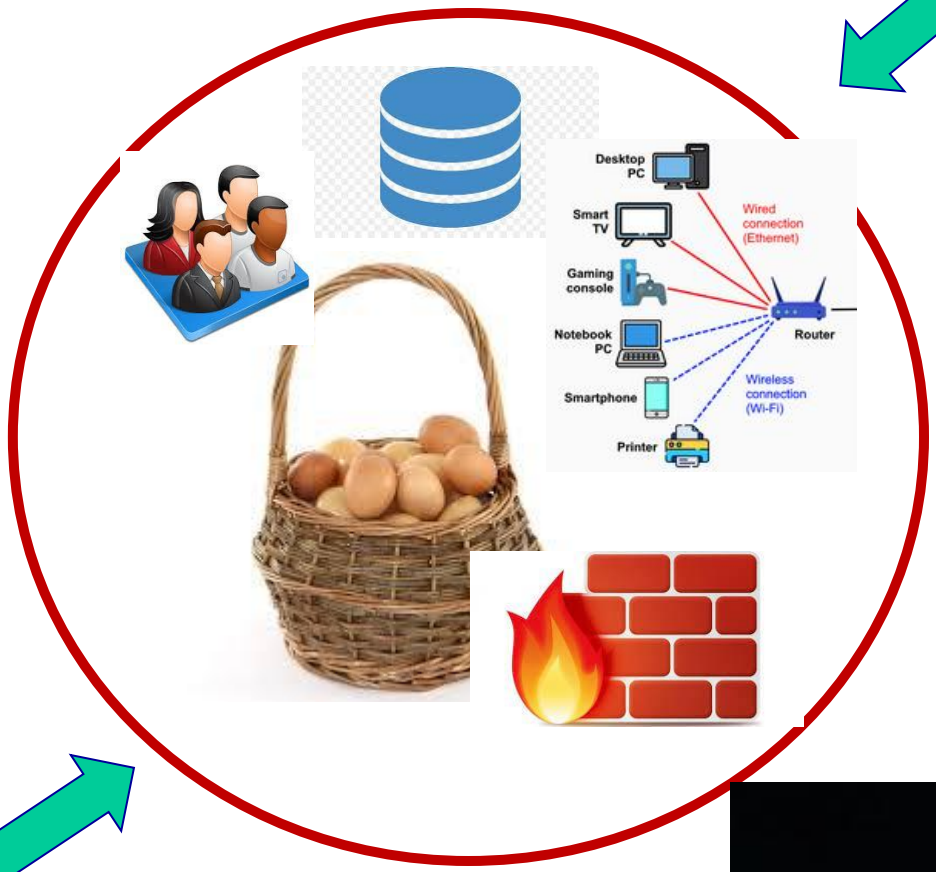
- Unauthorized access
- Data modification
- Corruption
- Loss
- Cyberattack or data breach
- Destructive force

Data Resilience

Ability

- Protect
- Maintain
- Recover after
  - equipment failure
  - power outage
  - disruption in
    - servers,
    - networks,
    - storages,
    - data centers

Data Storage



Trust Model (Perimeter-Based Security)

Trust-based models assume

- Everything inside the perimeter is safe
- Threats only come from outside the firewall
- Users and devices inside the perimeter are implicitly trusted
- Security is enforced at the entry point (e.g., firewalls, VPNs)
- “Castle and moat" approach

Limitations:

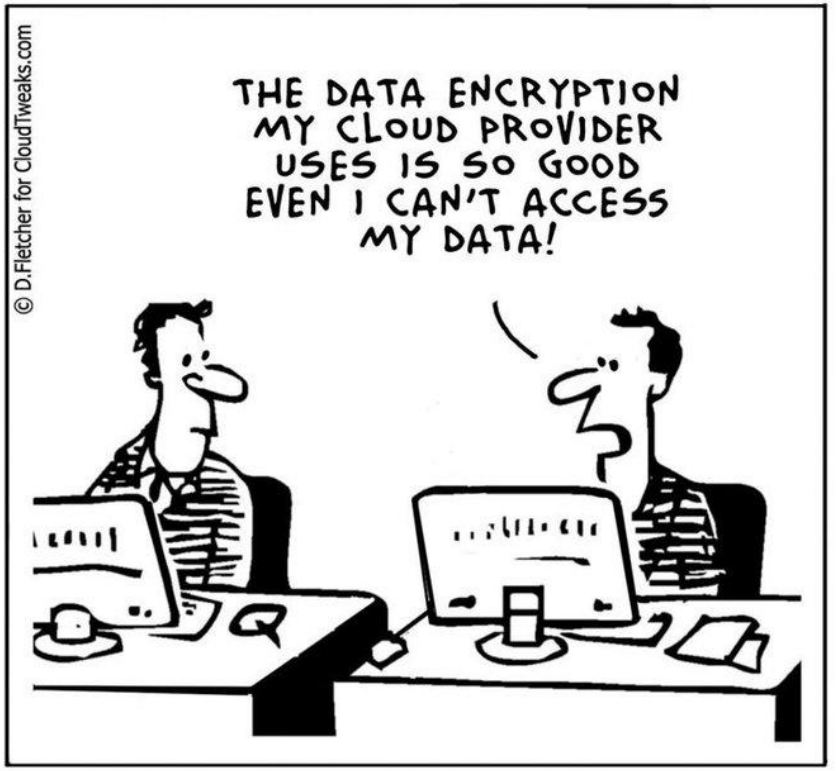
- Cannot protect from **insider threats** or compromised devices.
- **Lateral movement** within the perimeter is often unmonitored.
- Does not protect against
  - remote working
  - mobile devices
  - cloud services

In “Trust” we have to trust

- Computers under firewall
- Users
- Administrators
- Programmers
- Technicians

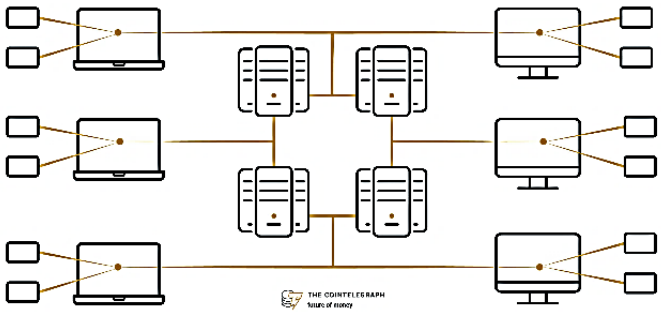
- Encryption methods
- Storages
- Backups mechanisms
- Disk-based hardware
- Transmission security and reliability

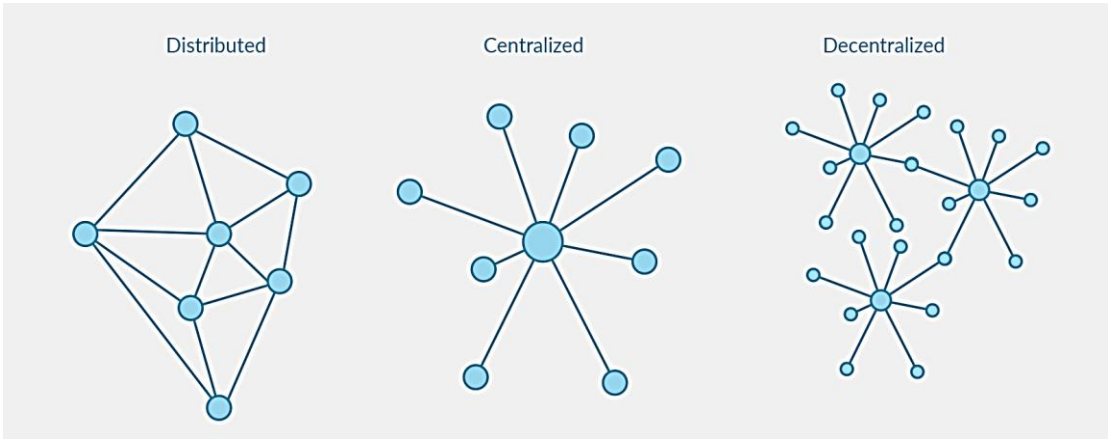
  etc.



Data Storage

Distributed data storage





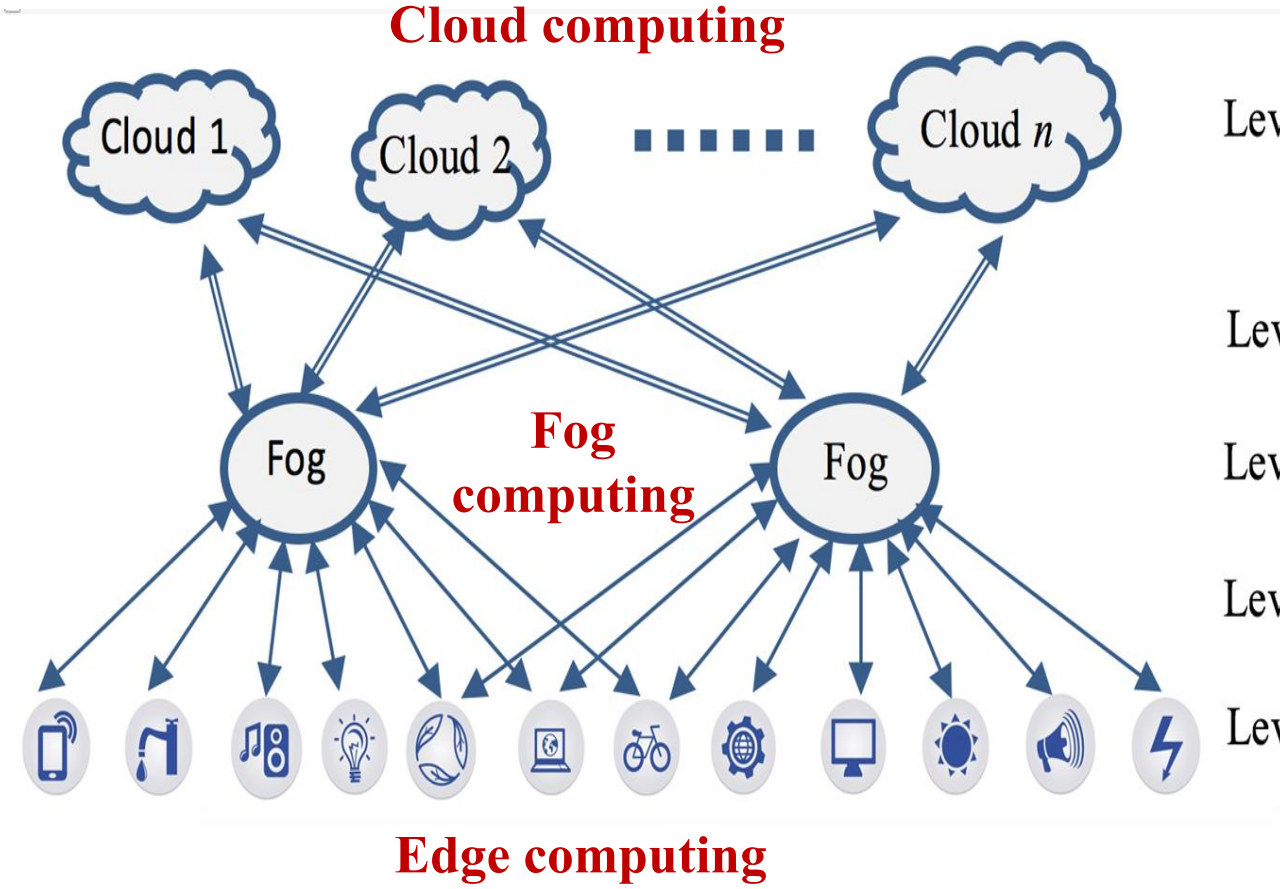
Multi-cloud distributed data storage

- Multiple cloud computing and storage services

=

Single heterogeneous architecture

Internet of Things with Multi-cloud



Level 5 **Storage-processing**

Level 4 **Communication**

Level 3 **Processing**

Level 2 **Communication**

Level 1 **Intelligent sensors**

Security collusion

Improper secret agreement between two or more malicious entities**, to obtain unauthorized access** to confidential data

Collaborate to

- decrypt sensitive data
- compromise security mechanisms

Result

- Account hijacking
- Data loss
- Abuse and illegal use of cloud services
- Denial Of Service

Traditional solutions:

- Secret sharing scheme
- Asymmetric and Symmetric cryptosystems
- Access structure

Zero-Trust Model: “Never trust, always verify”:

- No trust any user, device, or system—inside or outside the network.
- Every access must be continuously verified.
- Security is enforced at every layer: user, device, app, and data.

Core Principles of Zero-Trust:

- **Least privilege access**: Users get only the minimum access necessary.
- **Micro-segmentation**: Network is divided into small zones.
- **Strong identity verification**: Multi-factor authentication (MFA), biometrics, etc.
- **Real-time monitoring** and analytics: Constant assessment of trustworthiness.

Benefits of Zero-Trust:

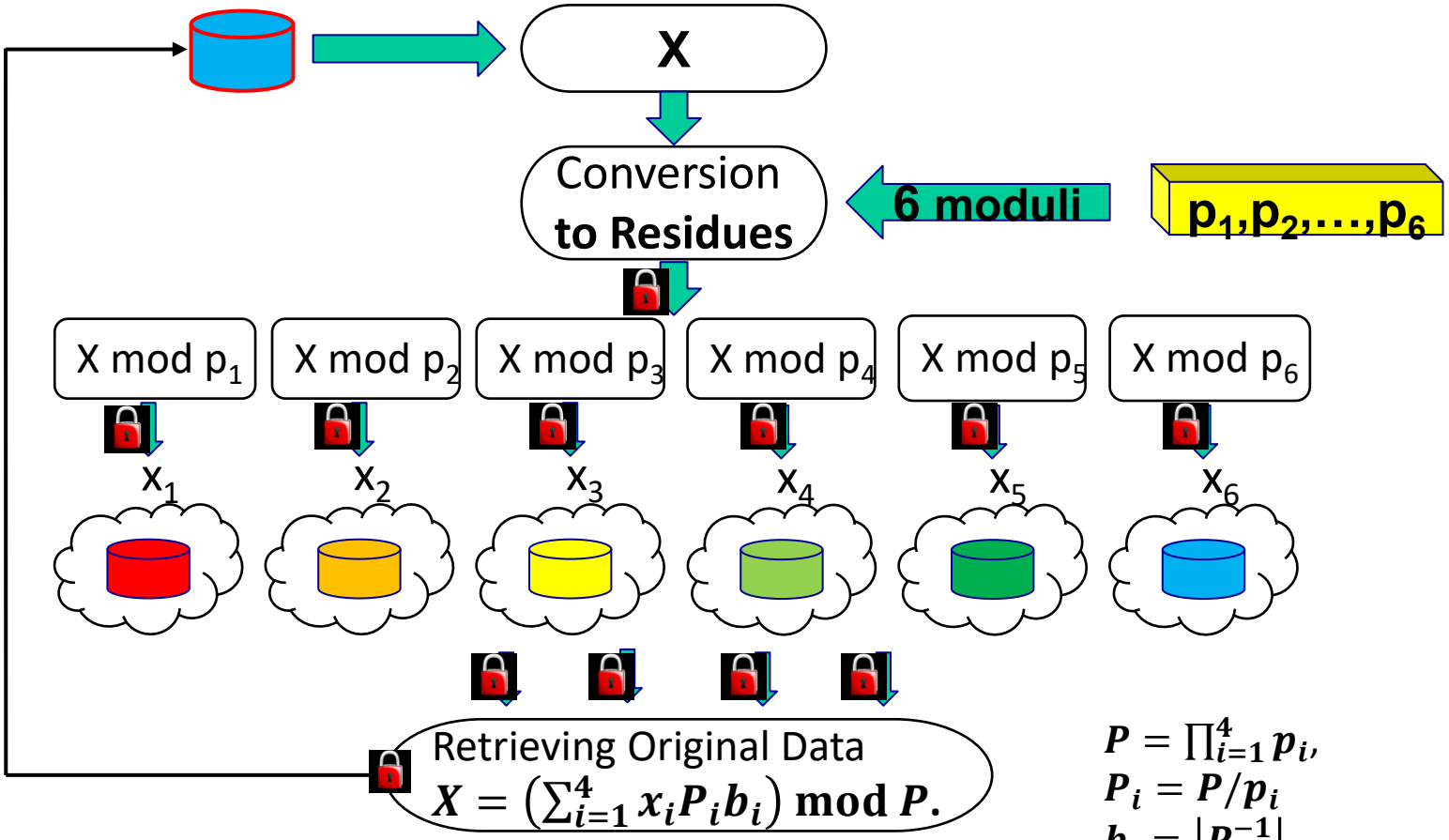
- **Reduces risk from insider threats** and compromised credentials.
- **Better suited for cloud, BYOD “Bring Your Own Device”, and hybrid environments.**
- Enhances visibility, control, and compliance.

Challenges in Implementation:

- Requires new policies.
- Needs investment in identity management, monitoring tools, and network segmentation.
- **Integration** with existing systems **can be complex**.

Privacy Preserving with Zero Trust

RRNS Multi-Cloud Storage

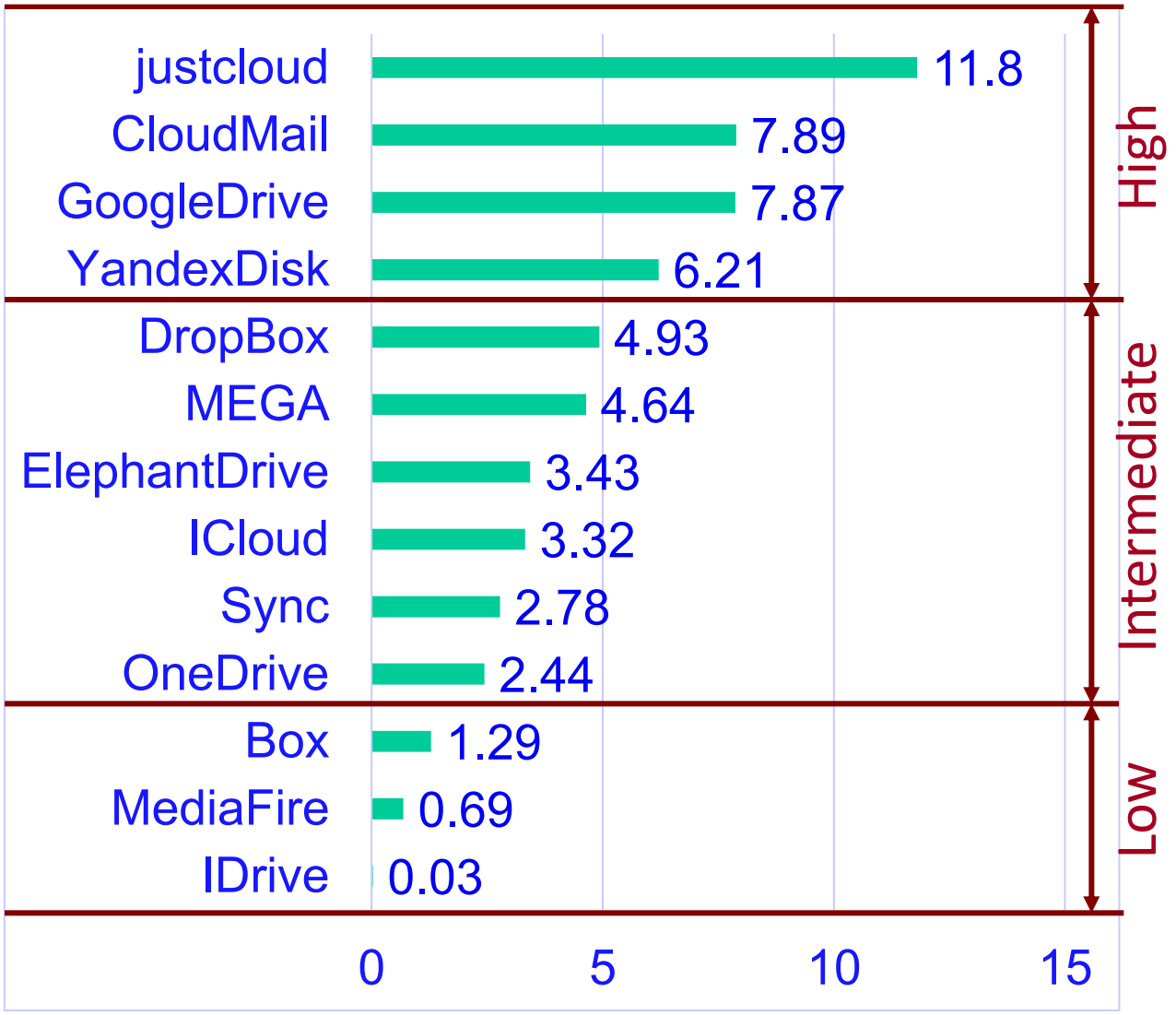


- own encryption
- full control over data storage

Cloud storage as a service (StaaS)

Name	URL	Name	URL
Alibaba Cloud	https://www.alibabacloud.com/	MediaFire	https://www.mediafire.com/
Amazon Drive	https://www.amazon.com/gp/drive/about	Mega	https://mega.nz/
Box	http://box.com/	one backup	https://mozy.com/
certain safe	https://certainsafe.com/	One Drive	https://onedrive.live.com/
Dropbox	http://dropbox.com/	pCloud	https://www.pcloud.com/
Egnyte	https://egnyte.com/	Rackspace	https://www.rackspace.com/cloud
Elephant drive	https://home.elephantdrive.com/	Salesforce	https://www.salesforce.com/
FlipDrive	https://flipdrive.com/	Sharefile	https://www.sharefile.com/
Google Drive	https://www.google.com/drive/	spideroak	https://spideroak.com/one/
Hubspot	https://www.hubspot.com/	storegate	https://www.storegate.com/gl/
iCloud	https://www.icloud.com/	SugarSync	https://www2.sugarsync.com/
IDrive	https://www.idrive.com/	sync	https://www.sync.com/
Jumpshare	https://jumpshare.com/	Windows Azure	https://azure.microsoft.com/en-us/services/storage/
JungleDisk	https://www.jungledisk.com/		
Justcloud	http://www.justcloud.com/	Yandex Disk	https://disk.yandex.com/

Data access speed



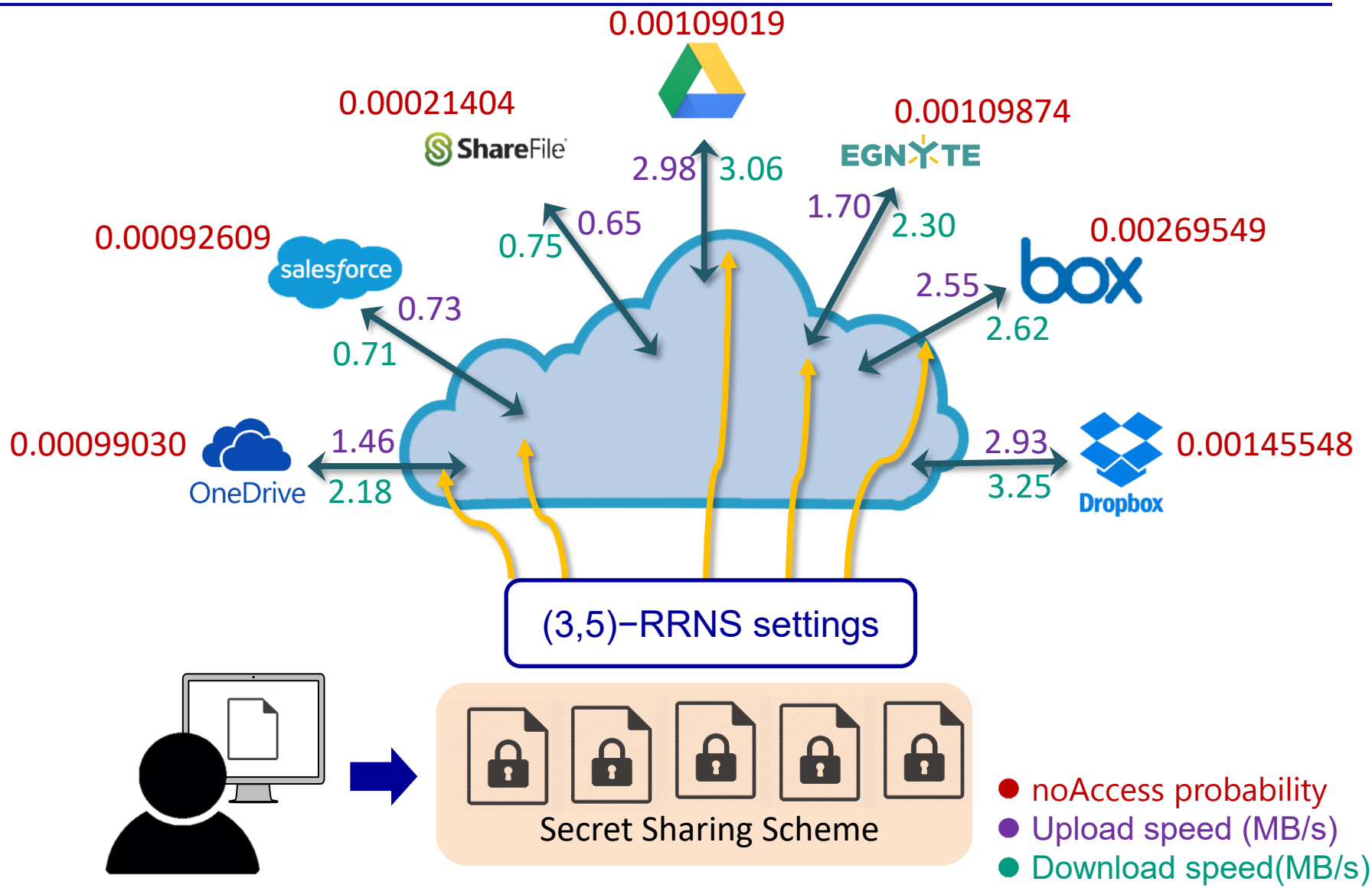
Reliability

comparison of cloud services by providing reliable and objective performance analysis

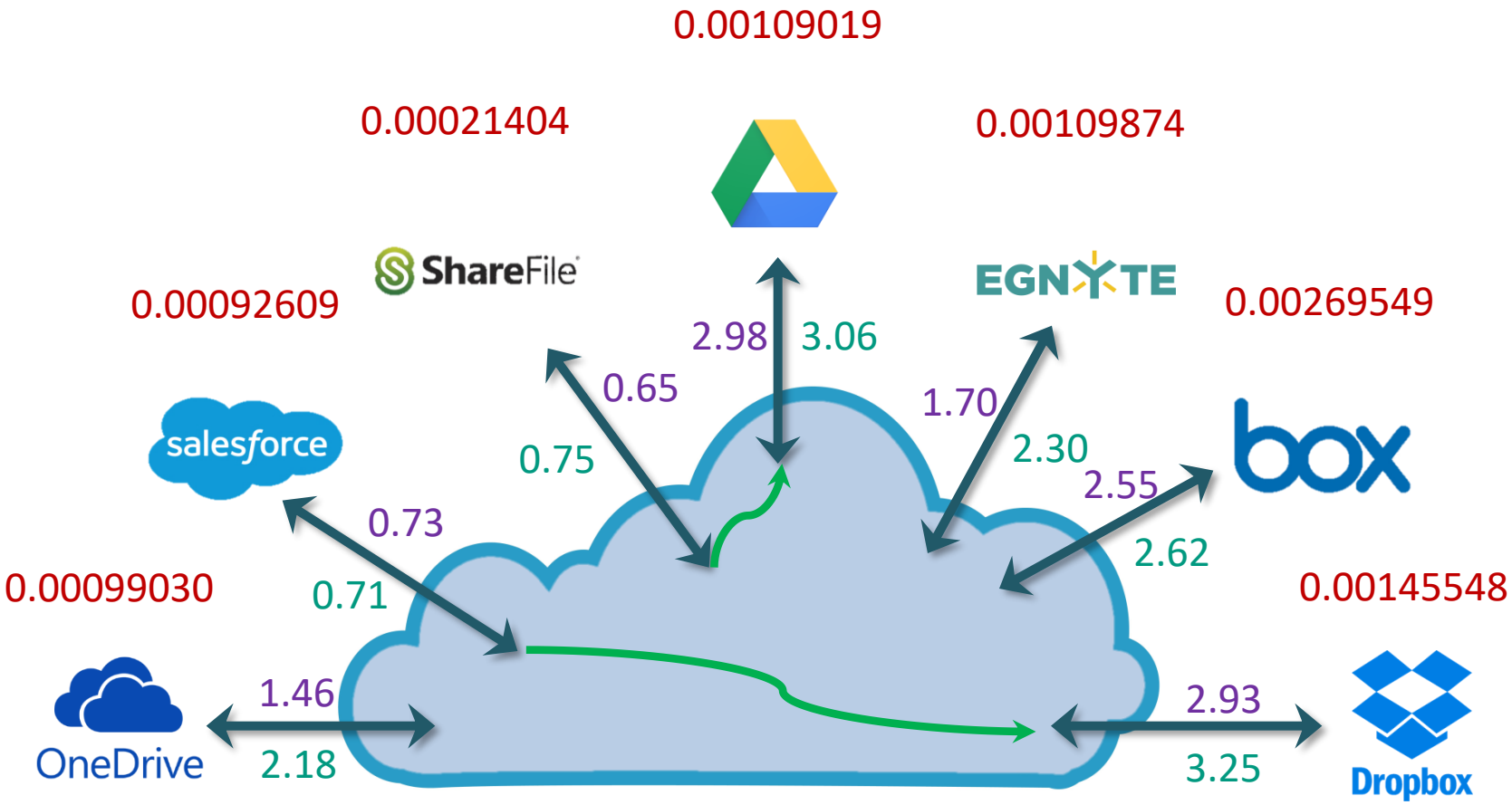
Service Name	Region	30 Day Availability *1 block = 1 mins*	Outages	Downtime
Faction Cloud	seattle	83.1928%	18	121.01 hours
eApps Cloud	richmond	99.8051%	1	1.4 hours
UpCloud	london	99.8123%	1	1.35 hours
Linode Cloud Hosting	singapore	99.9523%	1	20.6 mins
Vultr	silicon-valley	99.9751%	2	10.77 mins
UpCloud	singapore	99.976%	5	10.37 mins
Vultr	miami	99.9802%	11	55.13 mins
Linode Cloud Hosting	london	99.9856%	1	6.23 mins
Exoscale Compute	DE-FRA-1	99.9883%	1	5.05 mins
Cloud Central	canberra	99.9914%	2	3.73 mins
Exoscale Compute	BG-SOF-1	99.9918%	1	3.3 mins
DigitalOcean	ny1	99.9938%	2	2.68 mins
Vultr	dallas	99.9951%	1	2.12 mins
StratoGen VMware Cloud	denver	99.9976%	1	1.02 mins
IBM Cloud Compute	MEL	99.9984%	1	41 secs
StratoGen VMware Cloud	docklands	99.9985%	1	39 secs
Alibaba Elastic Compute Service	ap-northeast-1	100%	0	None

https://cloudharmony.com/status

Adaptive Multi-Cloud Storage

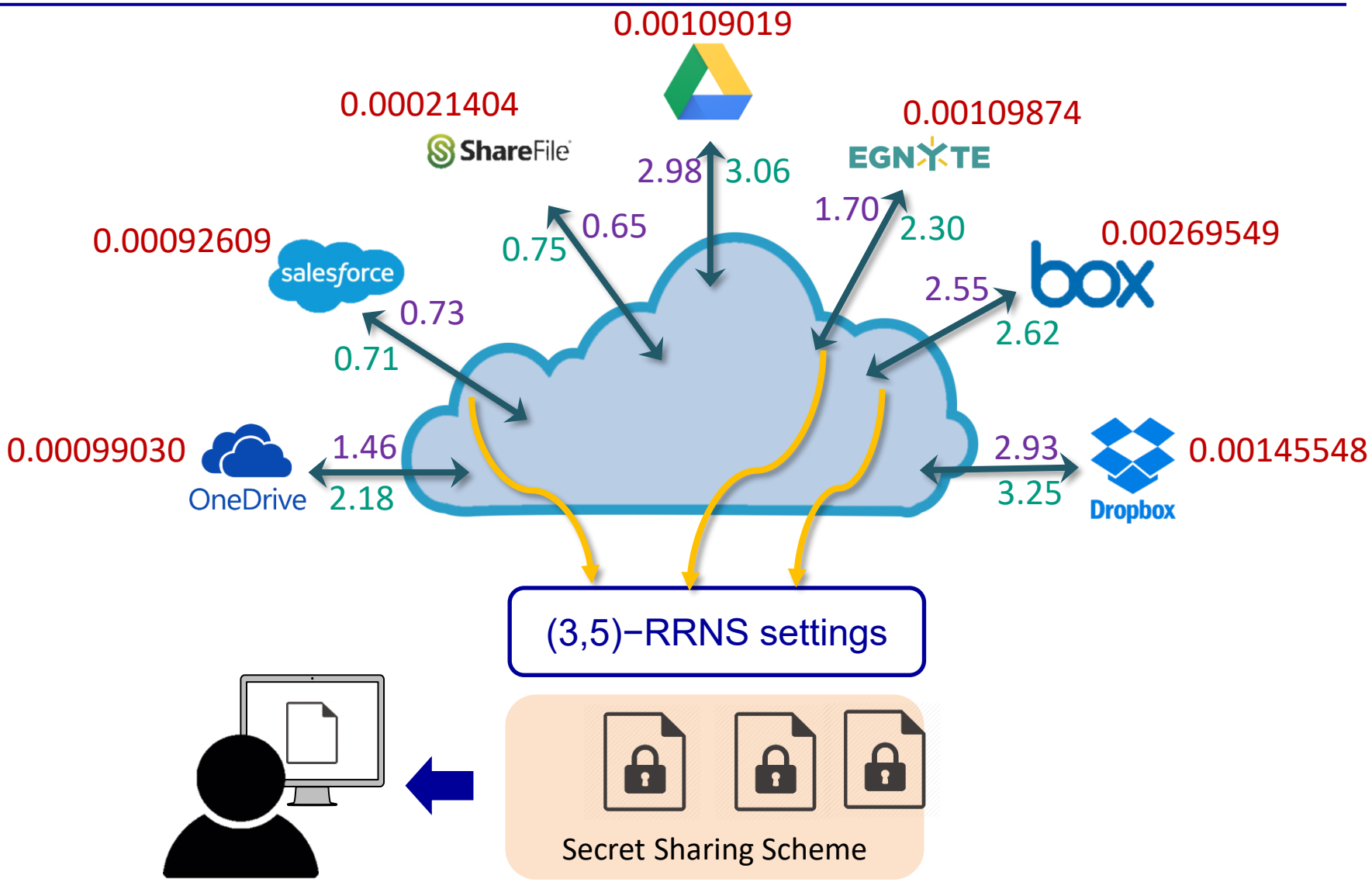


Adaptive Multi-Cloud Storage



- noAccess probability
- Upload speed (MB/s)
- Download speed(MB/s)

Adaptive Multi-Cloud Storage



Zero Trust uncertainty in clouds

- Shared resources
- Hybrid infrastructure
- Illusion of infinite computing resources on demand
- Scalability and flexibility (dynamic elasticity)
- Massive, diverse, incomplete, heterogeneous data
- Virtualization, loosely coupling applications to the infrastructure
- Resource provisioning time variation
- Variation in data transmission
- Workload uncertainty

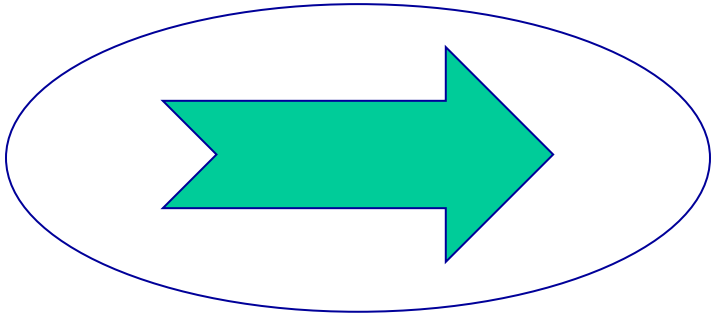
Optimization problem

- How to select parameters of the storage?
- How many storages should be used?
- How to select clouds from available?

Dynamic Cloud Selection

$C = \{c_1, c_2, \ldots, c_N\}.$

$c_j = \{err_j\}$



RRNS (k, n)

$P_r(k, n)$

$\{c_{i1}, c_{i2}, \ldots, c_{in}\}$

selected storages

- **Estimate and classify reliability levels**
- **Find adequate settings.**

Dynamically adapt settings for security concerns

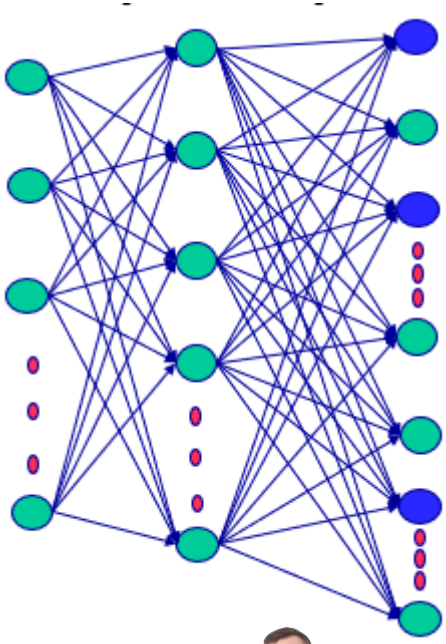
Dynamic Cloud Selection

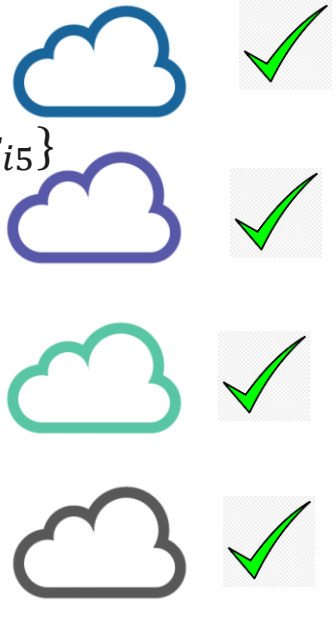
$C = \{c_1, c_2, \ldots, c_7\}.$ $c_j = \{err_j\}$ RRNS $(3, 5)$

NN for adaptive security

selected storages

$\{c_{i1}, c_{i3}, \ldots, c_{i5}\}$

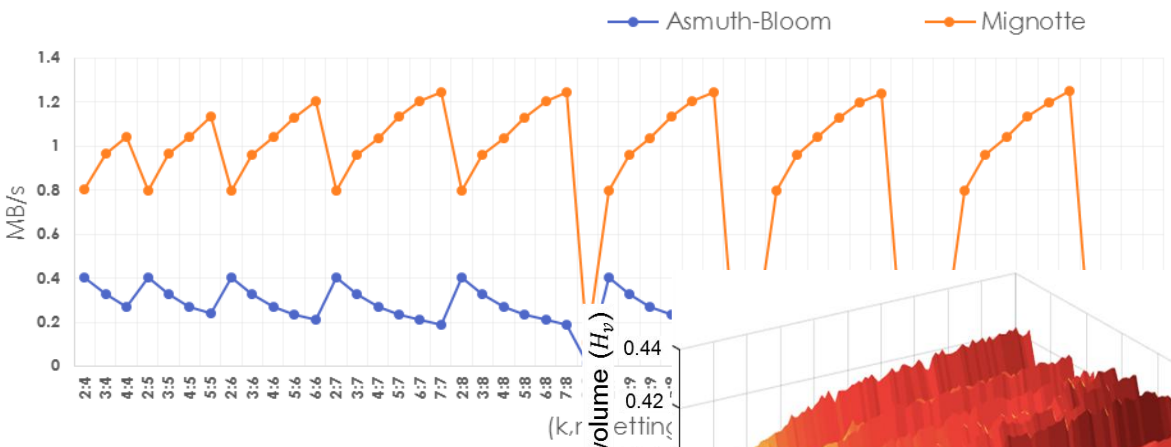


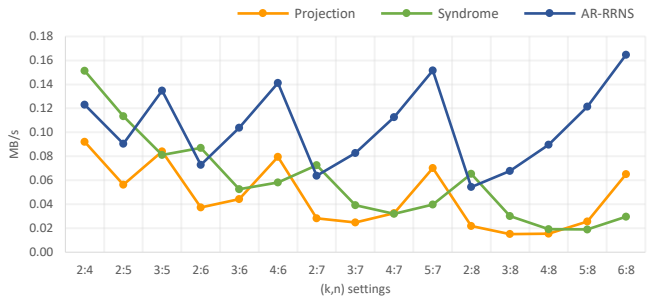


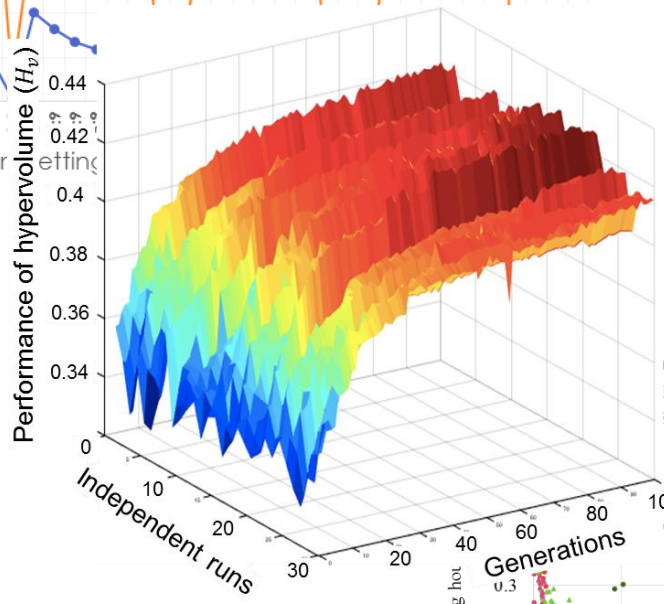
K = 3

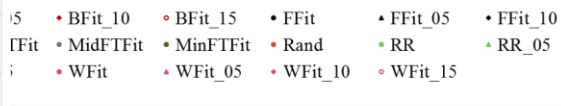
Google drive parameters

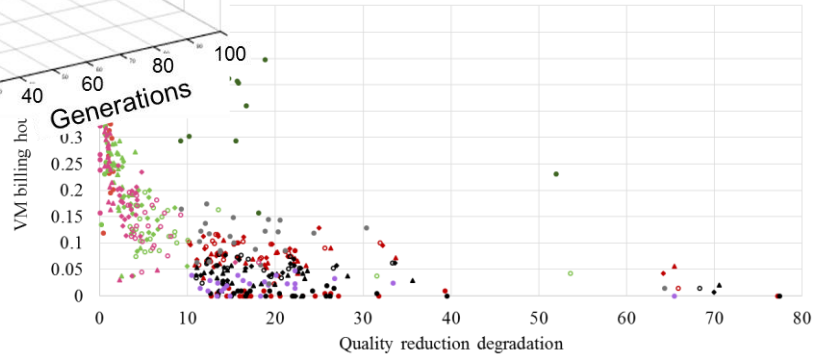
Experimental Analysis











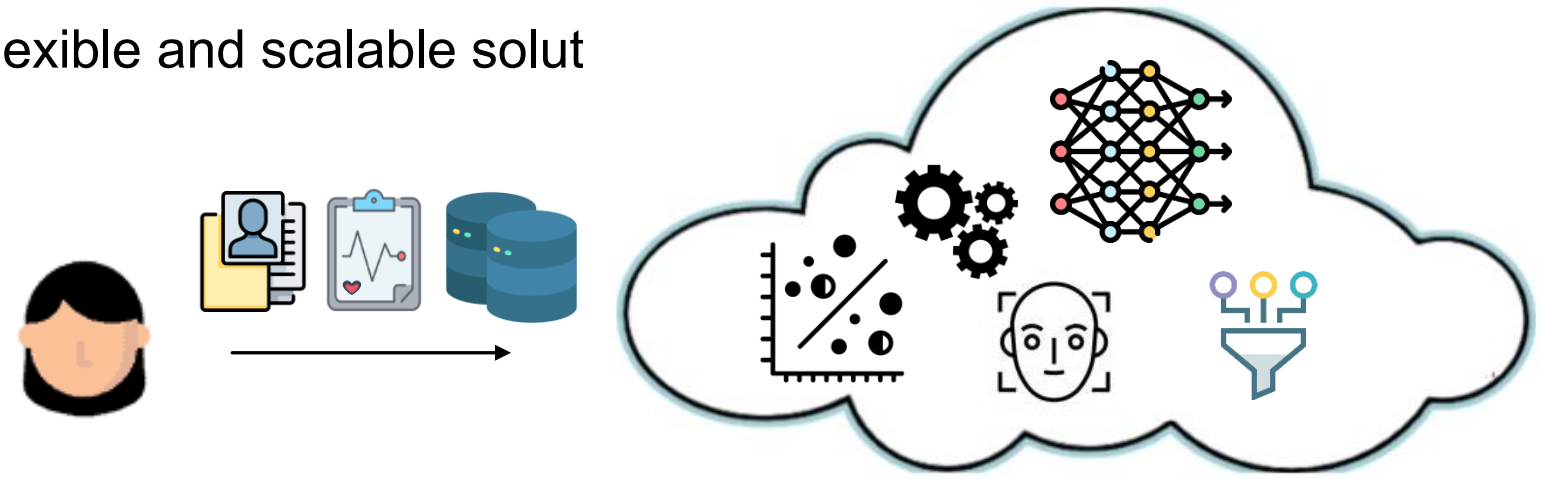
- Encoding/Decoding speeds
- Upload speed
- Download speed
- Redundancy
- Error detection-correction methods
- Probability of the data loss

Zero Trust in AI

Health Care case

Machine Learning as a Service (MLaaS)

- Machine Learning tools as a component of cloud computing.
- Flexible and scalable solut



A critical limitation of the adoption of MLaaS

- low protection of sensitive data in **an unsecured shared environment**

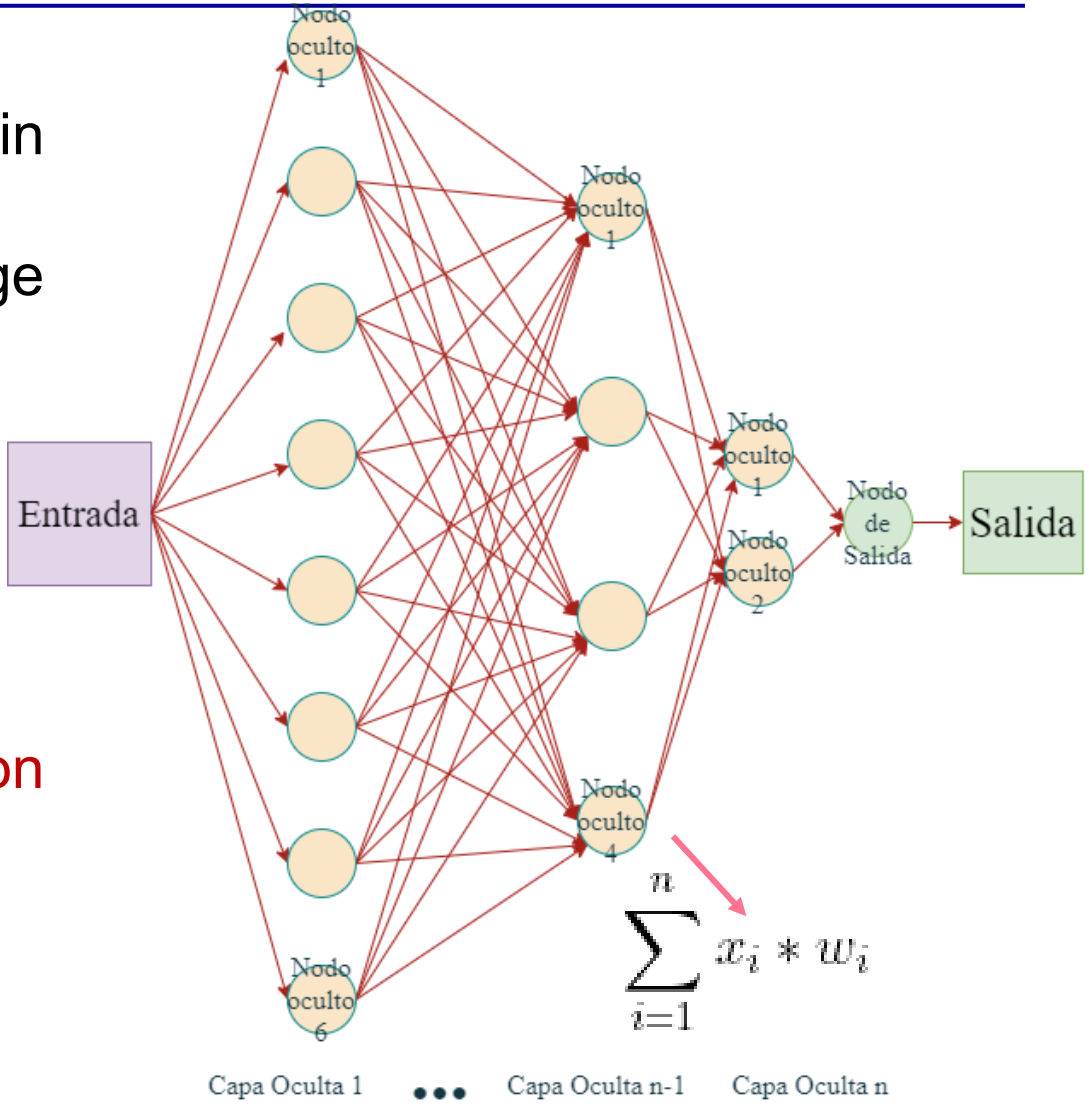
Open problem: Privacy-preserving ML

Deep Neural Networks (DNN)

DNN has been used in several disciplines, including

- medical image classification
- segmentation tasks:

- X-ray
- MRI
- Histopathology
- Positron Emission Tomography (PET)



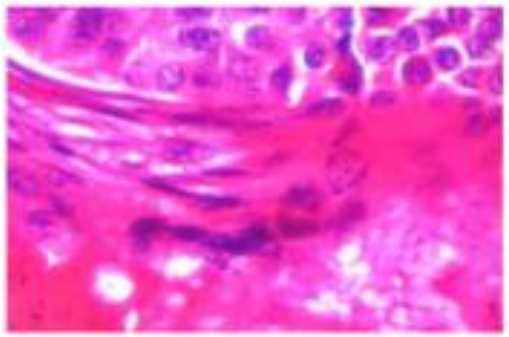
Open challenges in the DNN TRUST:

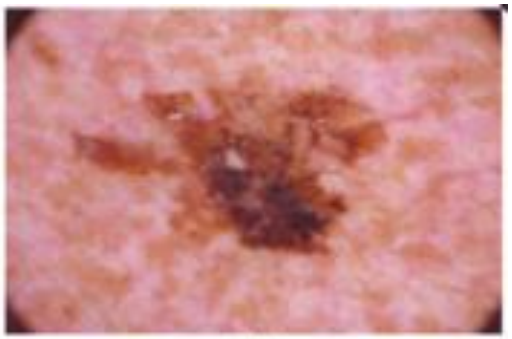
- Privacy and security issues
- Large volume of images
- Variability in image quality

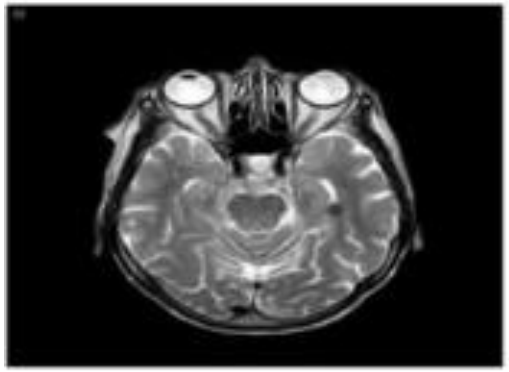
AI for Health Care

a. Breast cancer
b. Skin cancer
c. Brain tumor
d. COVID-19 screening
e. Thyroid ultrasound
f. Alzheimer's disease

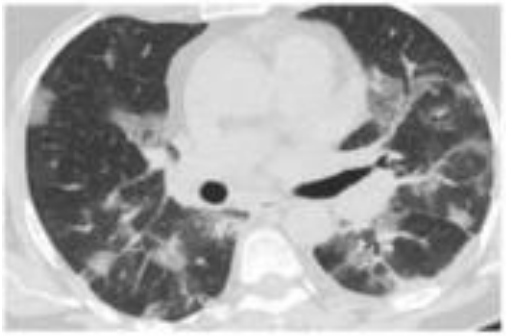
etc,

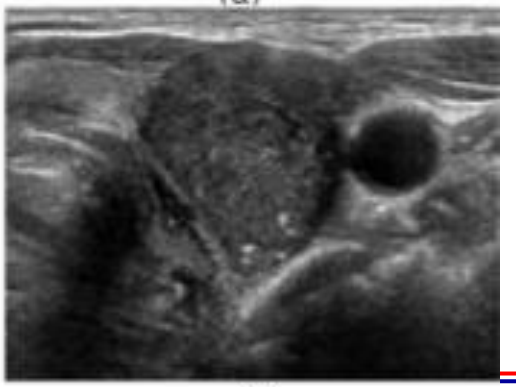

(a)


(b)


(c)

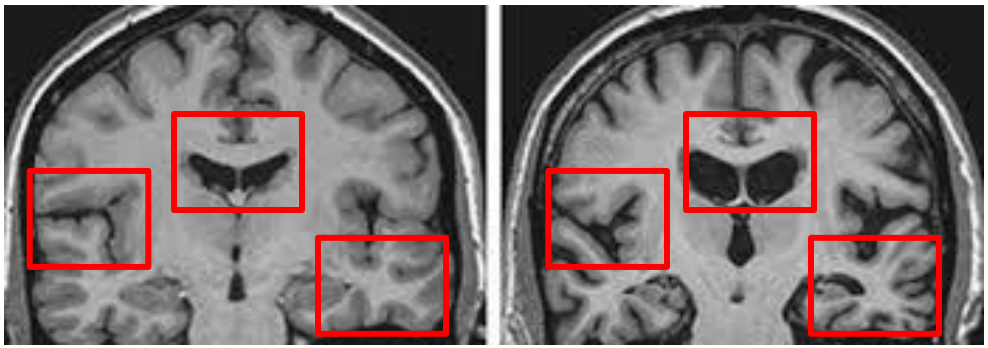

(d)


(e)


(f)

Alzheimer's Disease (AD)

- progressive neurodegenerative disease that affects
  - memory, thinking, orientation, and behavior
- most common form of dementia worldwide
  - 10% of older adults (Seniors) in México suffer from it



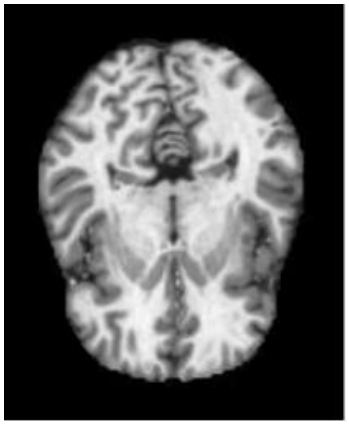
Sano **Alzheimer**

AD causes gradual **morphological changes** in a brain

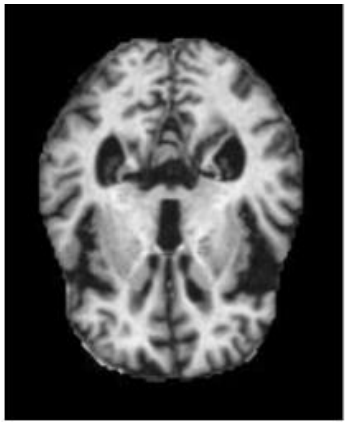
Magnetic Resonance Imaging (MRI)

- standard method of AD detection

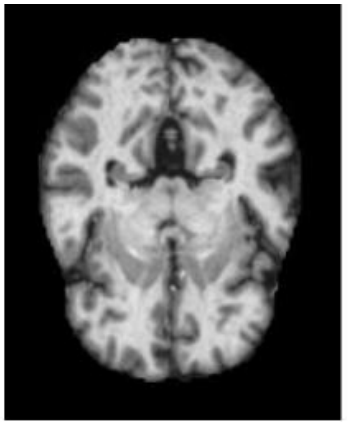
Alzheimer 4 classes



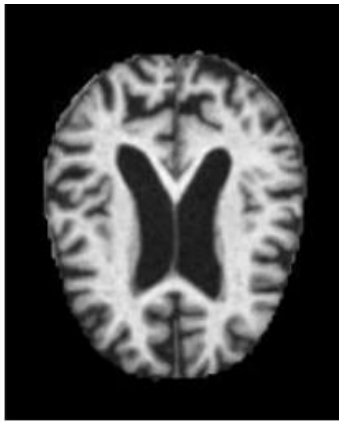
Normal



Mild



Moderate

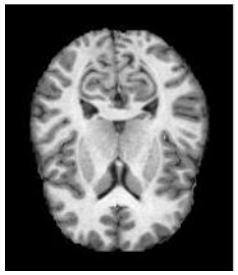


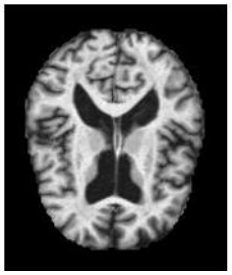
Severe

AD classification

Para la clasificación de la EA por IRM existen dos vías

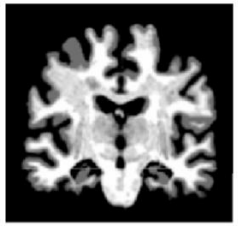
- Clasificación Binaria

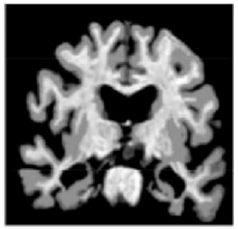

S

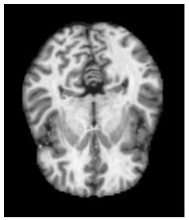

EA

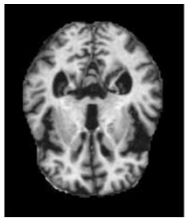
- Clasificación Multiclase

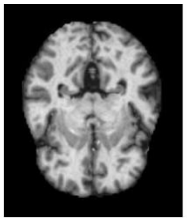

S

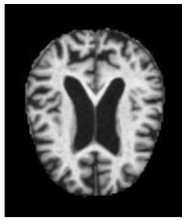

DCL


EA


ND


DML


DL

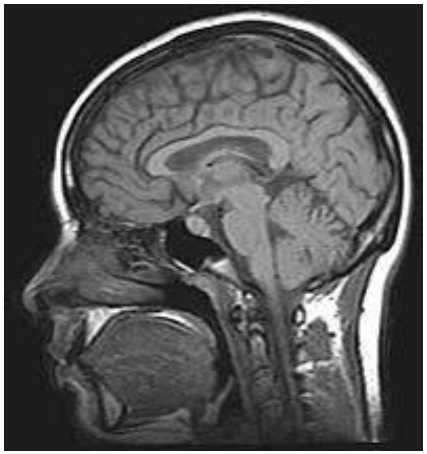

DM

Clases	Abreviación
Sano	S
Deterioro cognitivo leve	DCL
Alzheimer	EA
No demente	ND
Demente muy leve	DML
Demente leve	DL
Dementes moderados	DM

Regulations

Hospitals handle a large volume of data such as:

- Personal data
- Medical Records
- Studies
  - Laboratory
  - Electrocardiograms
  - **Image**



Magnetic Resonance Imaging MRI

There are regulations in place to manage this data.

- Sensitive Data
- Patient privacy must be ensured.
- Data anonymization is required

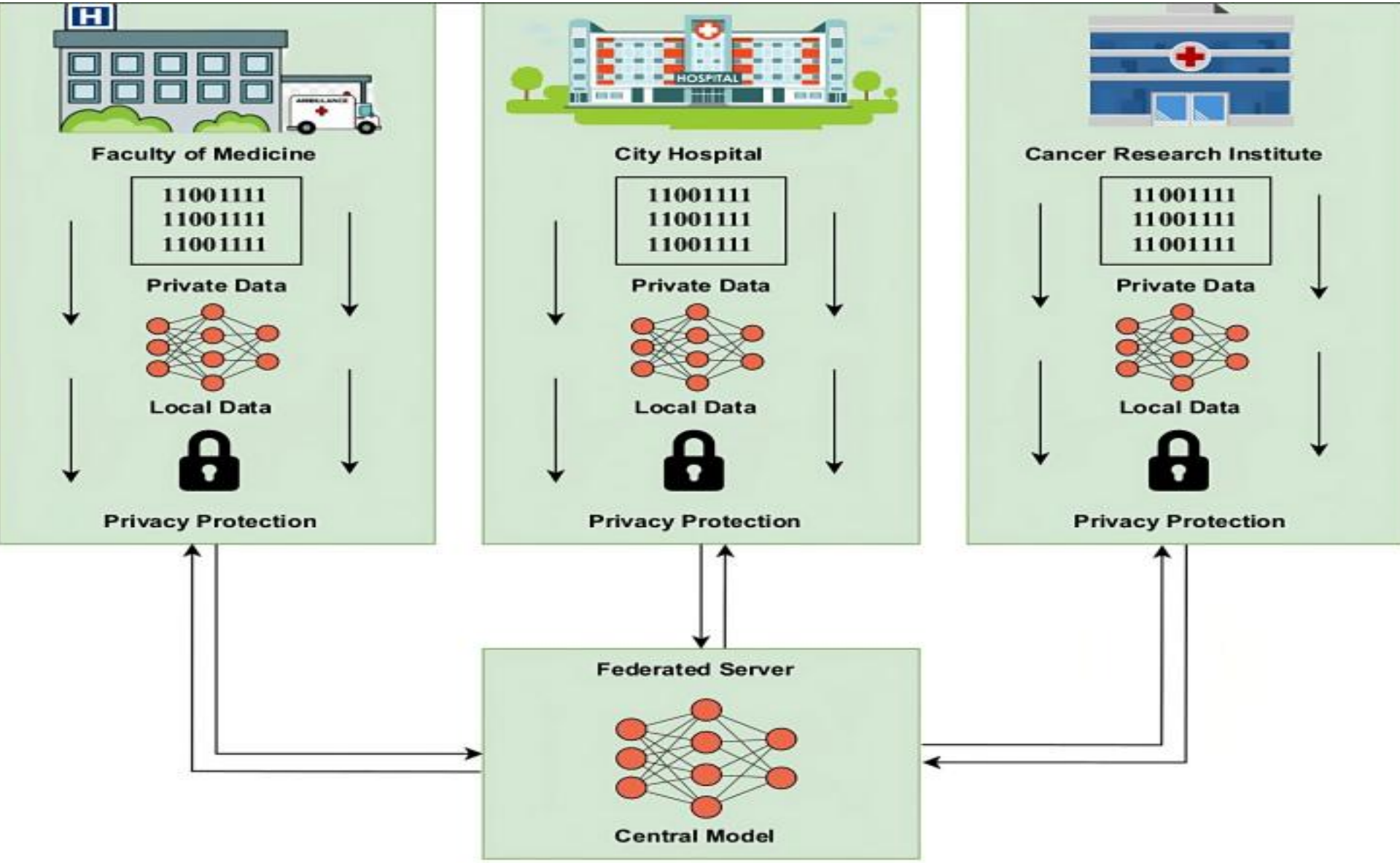
There are challenges in accessing this data:

- Data sharing
- Database creation
- Training computational models

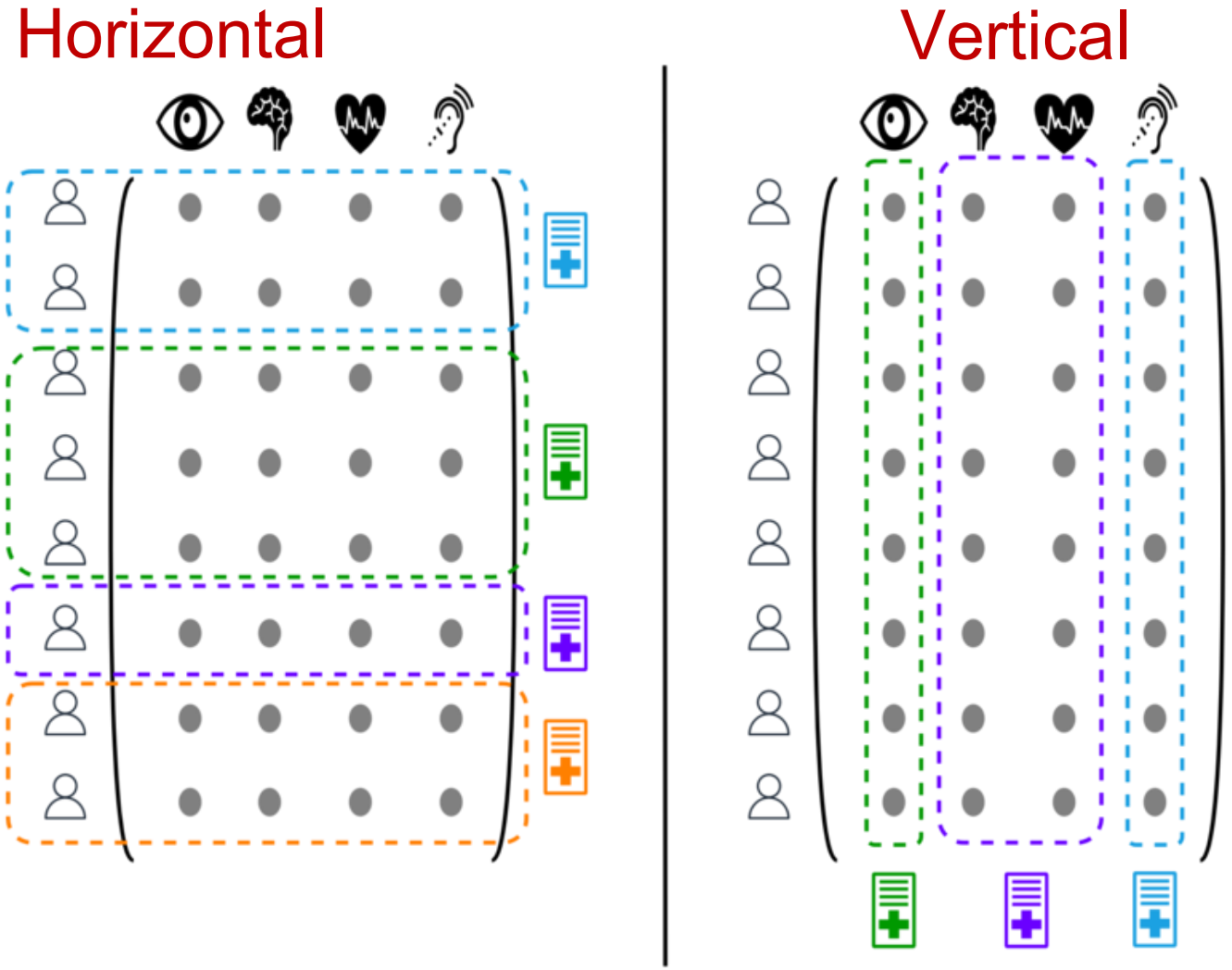
Federated Learning

A way to Zero Trust

Federated Learning in Health Care

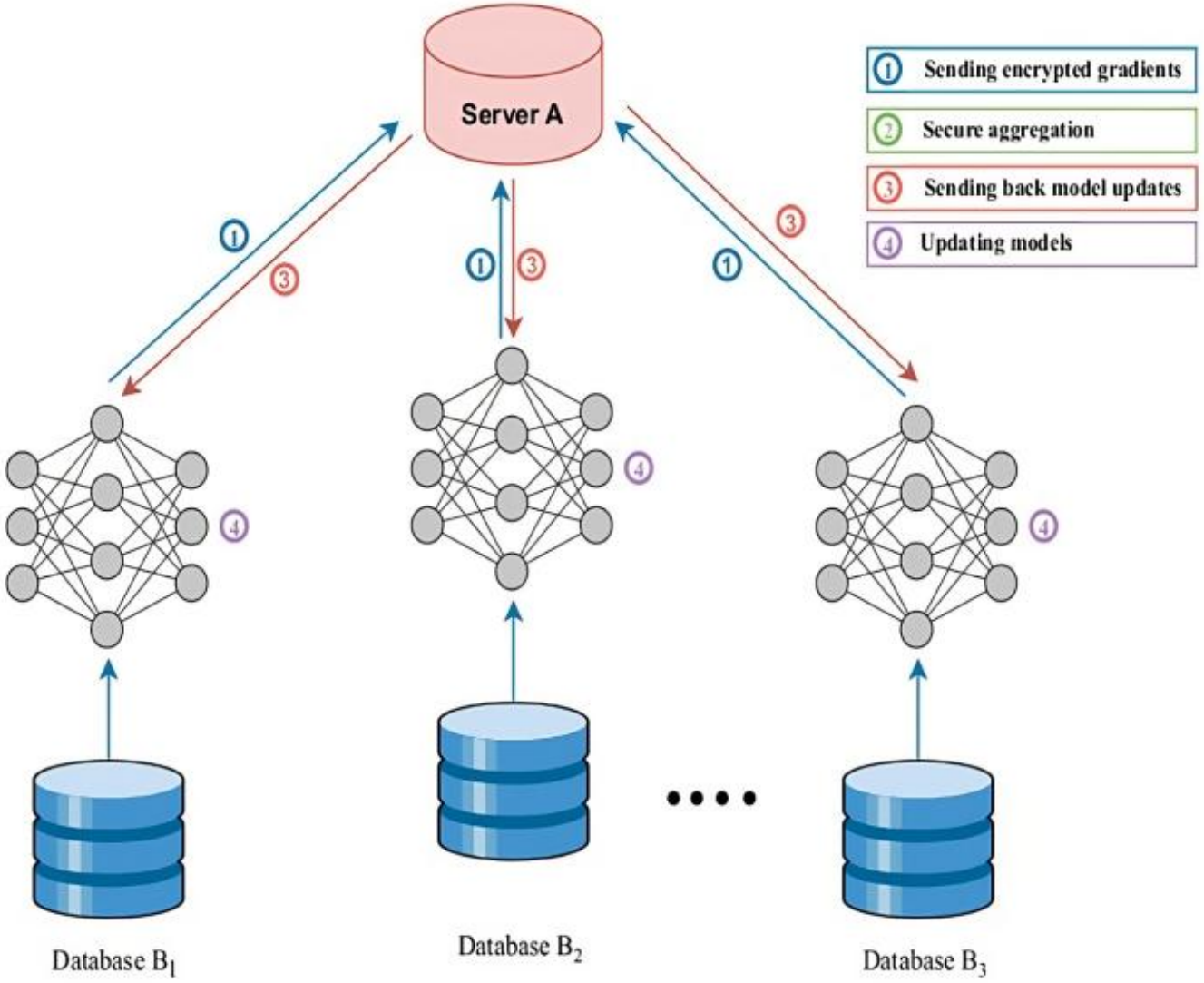


Types of Federated Learning



- Horizontal: Mismas características, pero diferentes instancias
- Vertical: Mismas instancias, pero con características diferentes

Horizontal FL



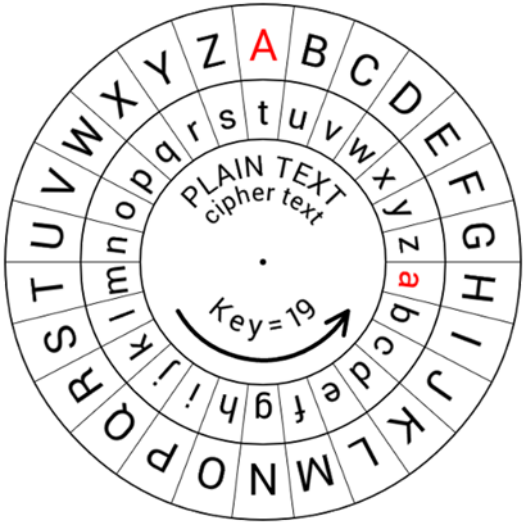
Federated Learning Models

Federated Learning Type	Data Distribution	Description
Horizontal (HFL)	Same columns, different rows.	Multiple nodes with the same features but different instances.
Vertical (VFL)	Different columns, same rows.	Multiple nodes with different features about the same instances.
Federated Transfer Learning	Different columns and different rows.	Combines FL with transfer learning for heterogeneous data.
Semi-Supervised	Combination of labeled and unlabeled data.	Utilizes labeled and unlabeled data across nodes.
Asynchronous	No synchronization required between nodes.	Allows asynchronous model updates between nodes.
Hybrid	Combination of different columns and rows.	Combines characteristics of HFL and VFL for heterogeneous data.

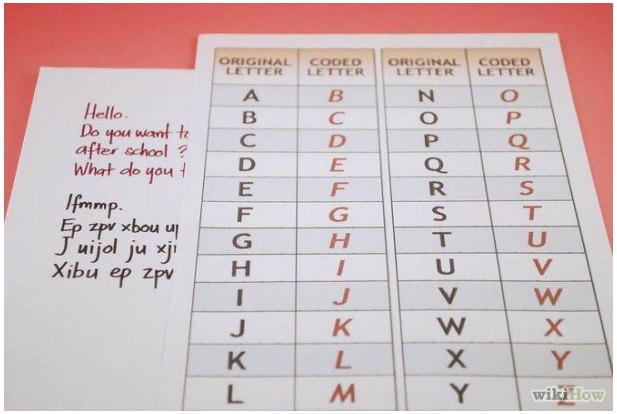
Homomorphic Encryption

Privacy Preserving Processing

Privacy Preserving Processing: Cryptography does not help



Caesar cipher



Enigma

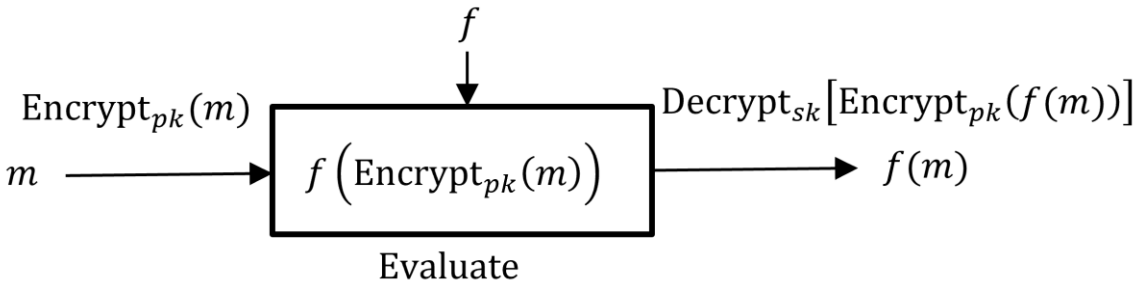
A●-
B-●●●
C-●-●
D-●●
E●
F●●-●
G--●
H●●●●
I●●

J●---
K-●-
L●-●●
M--
N-●
O---
P●--●
Q--●-
R●-●

S●●●
T-
U●●-
V●●●-
W●--
X-●●-
Y-●--
Z--●●

Homomorphic Encryption foundation

- A function applied to ciphertexts provides the same (after decryption) result as applying the function to the original unencrypted data.



- Lattice-based schemes whose security is based on the hardness of the Learning with Errors (LWE) or Ring LWE (RLWE) problems.
- Homomorphic Encryption (HE) exploits the hardness of identifying a secret $\boldsymbol{sk}$ from noisy pairs of the form $\boldsymbol{pk} = (b, a) = ([-(a \cdot sk + e)]_q, a)$

  where $sk, a, e \in R_q = \mathbb{Z}_q[X]/(X^n + 1)$,

  - $a \leftarrow U(R_q)$ and $e \leftarrow \chi_{\text{err}}$.
- **Encrypt**: $(m, 0) + pk = (m - a \cdot sk + e, 0 + a) = (c_0, c_1) = \boldsymbol{c}$ (two polynomials)
- **Decrypt**: $m' = c_0 + c_1 \cdot \boldsymbol{sk} = m - a \cdot sk + e + a \cdot sk = m + e \approx m$

Homomorphic Encryption foundation

Polynomial ring R= $\mathbb{Z}[X]$/f(X), where

- $\mathbb{Z}[X]$ - polynomial ring with coefficients in $\mathbb{Z}$
- f(x) is a cyclotomic polynomial of degree d that is the unique irreducible polynomial with integer coefficients that is a divisor of x^n-1

In practice, $f(x)=X^d+1$ and $d=2^n$.

Elements of R are polynomials of degree less than d and coefficients in $\mathbb{Z}$.

- q - coefficients modulus
- $R_q[x]= \mathbb{Z}_q[x]/f(x)$
- $\mathbb{Z}_q[x]$ - polynomial ring with coefficients modulo q.
- $R_q[x]$ are polynomials of degree less than d and coefficients modulo q.
- [x] rounding to the nearest integer.
- χerr and χkey are distributions

Secret key is a polynomial sk <-χ_{key} with binary coefficients randomly sampled from χ_{key}.

Public key pk is a couple of two polynomials $\boldsymbol{pk} = (b, a) = ([-(a \cdot sk + e)]_q, a)$

Public key is sampled ***a*** from the ring R_q , a <- R_q, and a random error e from χerr e <- Xerr.

HE taxonomy

By supported arithmetic operations

Partially homomorphic Encryption (PHE):
- Supports only addition **or** multiplication.

Somewhat Homomorphic Encryption (SWHE):
- Bounded additions and multiplications
- Computationally cheap.
- No bootstrapping
- Pre-2009 schemes were *somewhat homomorphic*.

Fully Homomorphic Encryption (FHE)
- Unbounded additions and multiplications
- Computationally expensive
- Bootstrapping notion

By type of data

Logical (Boolean)
- FHEW
- TFHE

Integer based
- Brakerski-Gentry-Vaikuntanathan (BGV)
- Brakerski/Fan-Vercauteren (BFV)
- Lopez-Tromer-Vaikuntanathan (LTV)
- Doroz-Hu-Sunar (DHS)

Fixed-precision numbers
- Cheon-Kim-Kim-Song (CKKS)

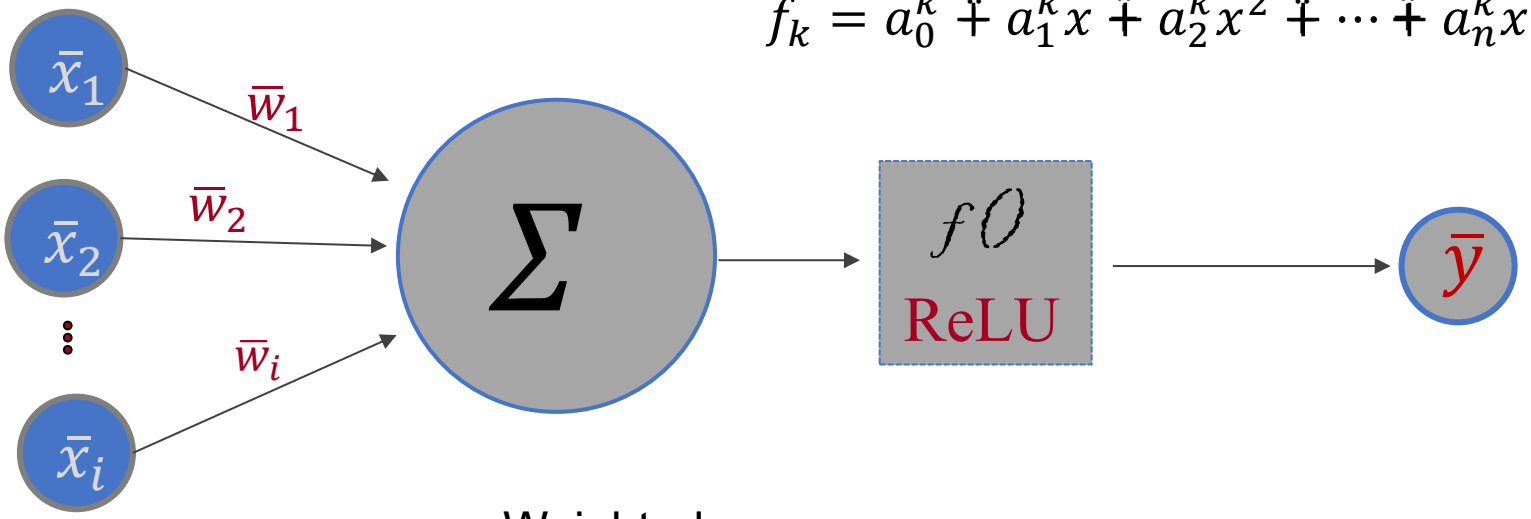
General-purpose HE libraries

Tool	Support	Pros	Cons
SEAL	Microsoft	Well-documented Easy security parameters setting	Poor flexibility Limited number of supported schemes
HElib	IBM	Efficient homomorphic operations	Low bootstrapping performance Complicated security parameter setting
TFHE		Fast bootstrapping	Poor performance for simple tasks
PALISADE **OpenFHE**	DARPA, MIT, UCSD, etc.	Multiple HE schemes Cross-platform	
cuHE		Mass parallelism and high memory bandwidth of GPUs	Poor documentation and support
HEAAN	Seoul National University	Operations between rational numbers	
HE-transformer	Intel	Integration with deep learning libraries	Extension of SEAL

Homomorphic Neural Network

Approximating an activation function

$$\ddot{f}_k = a_0^k \ddot{+} a_1^k x \ddot{+} a_2^k x^2 \ddot{+} \cdots \ddot{+} a_n^k x^n$$



Weighted sum.

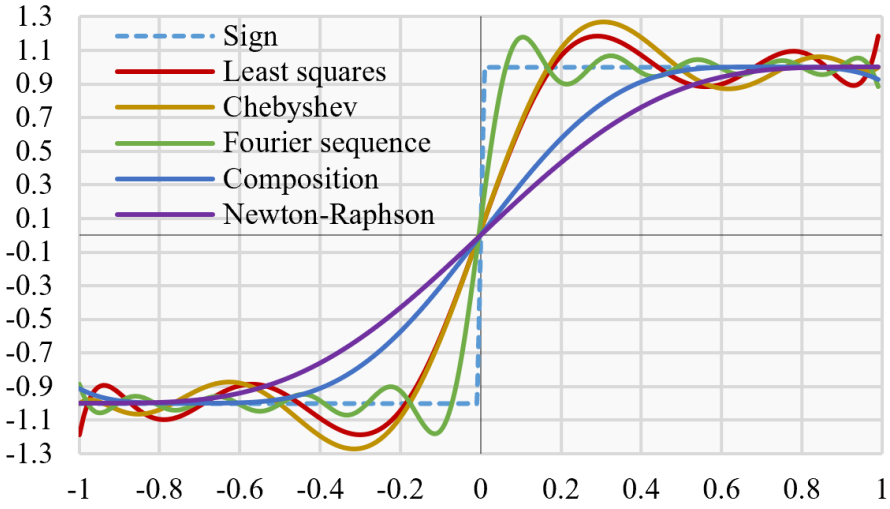
$$\bar{y} = \bar{w}_1 \, \bar{x}_1 \ddot{+} \bar{w}_2 \, \bar{x}_2 \ddot{+} \cdots \ddot{+} \bar{w}_i \, \bar{x}_i$$

Aproximating $sign$ (time)

Homomorphic operations:

Timings (ms) for the four processes in a HE scheme

	ε	$KeyGen_\varepsilon$	$Encrypt_\varepsilon$	$Decrypt_\varepsilon$	$Evaluate_\varepsilon$	
					∔	※
P_1	BFV	60.076	5.0547	0.8883	1.3172	4.0361
	CKKS	84.378	7.1897	0.7021	1.7650	2.9165
P_2	BFV	404.897	12.5529	2.9475	3.5390	33.7440
	CKKS	1,351.270	26.0943	6.2270	11.7310	30.2540



Operations:

- Addition: **0.087 ms, HE 11.7310 ms**
- Multiplication: **0.099 ms, HE 30.2540**
- Comparison: **0.0464 milliseconds, HE 143.28 ms**

HE comparison needs to be optimized; otherwise, it is inapplicable.

Performance (ms) of state-of-the-art homomorphic comparison approaches

Approach	*Generation time*	*Comparison time*
Least-squares	0.61	143.28
Chebyshev	0.56	121.65
Iterative	-	1,671.14
Fourier sequence	1.26	132.91
Newton-Raphson	0.38	98.91
Composition	0.42	99.02

Slight improvements of homomorphic comparison are high steps toward a privacy-preserving model

Self-Learning Activation Functions

Self-Learning Activation Functions (SLAF)

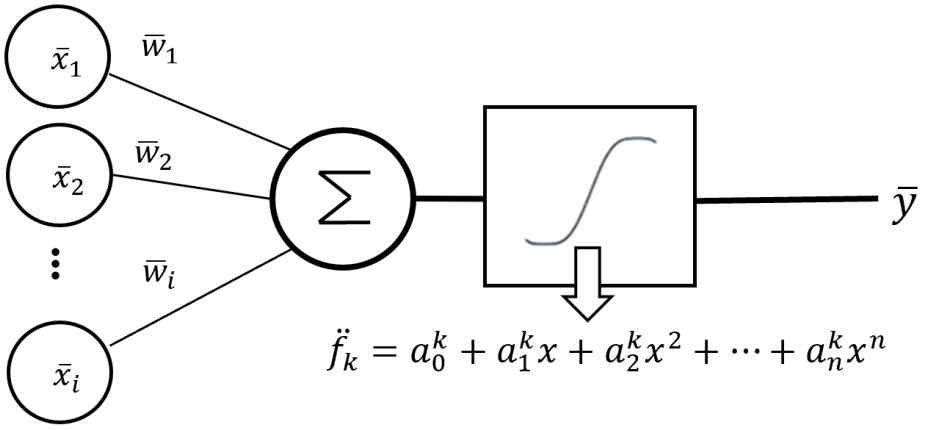
- We approximate the activation function by a polynomial at each neuron independently with trainable coefficients as

$$\bar{y}_k \leftarrow \ddot{f}_k\left(\sum\left(\bar{w}_i^k \ddot{\times} c_i^k\right) \ddot{+} \bar{\beta}_k\right)$$

$$\ddot{f}_k = a_0^k + a_1^k x + a_2^k x^2 + \cdots + a_n^k x^n$$

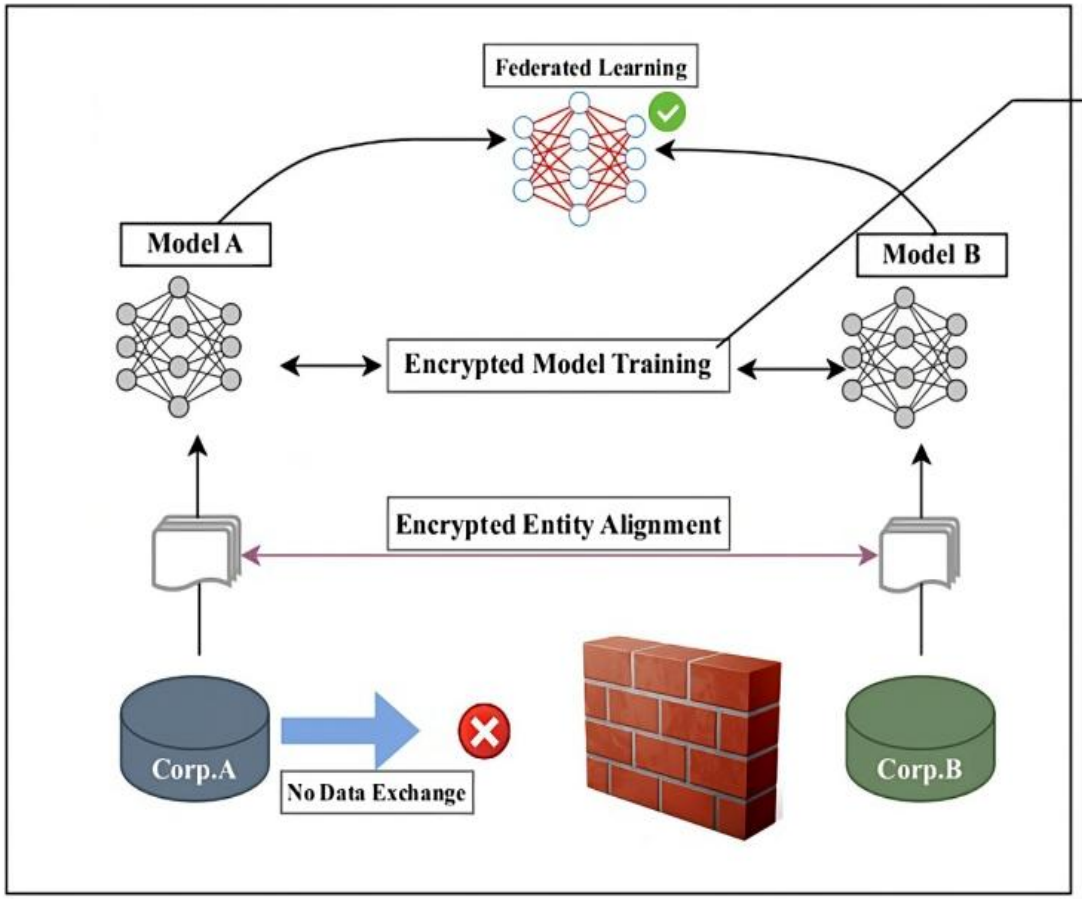
where $a_0^k, a_1^k, \ldots, a_n^k$ denote the trainable coefficients of the polynomial $\ddot{f}_k$ at neuron k.

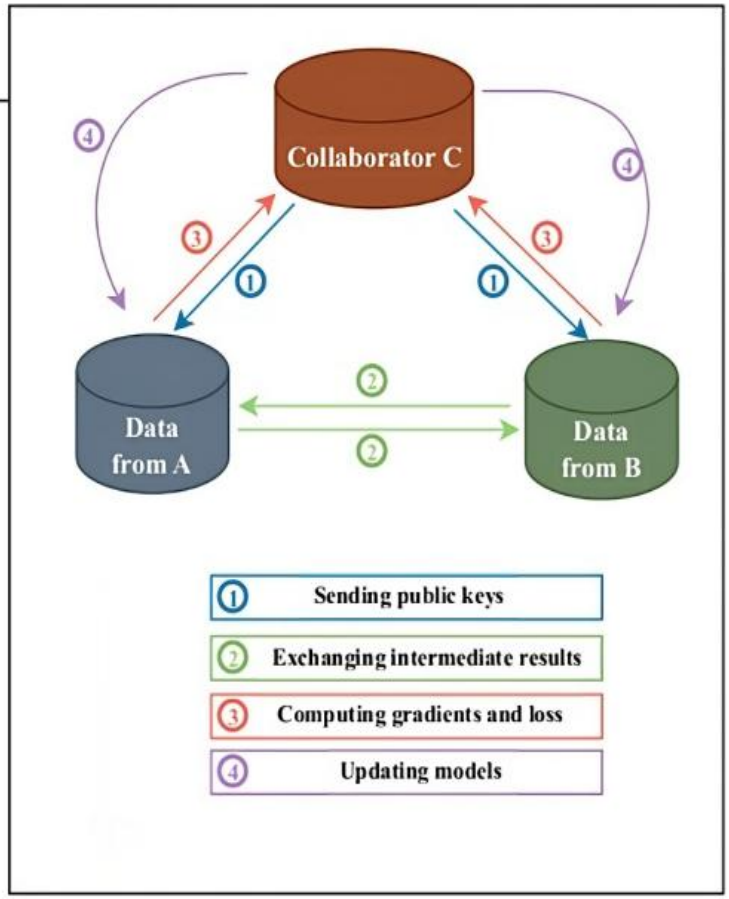
- Activation function approximation is optimized for a given problem and dataset.
- Training process aims to find:
  - ✓ **Weights** w
  - ✓ **Coefficients** $a_0^k, a_1^k, \ldots, a_n^k$



Homomorphic neuron k with a SLAF $\ddot{f}_k$

Federated Learning privacy preserving



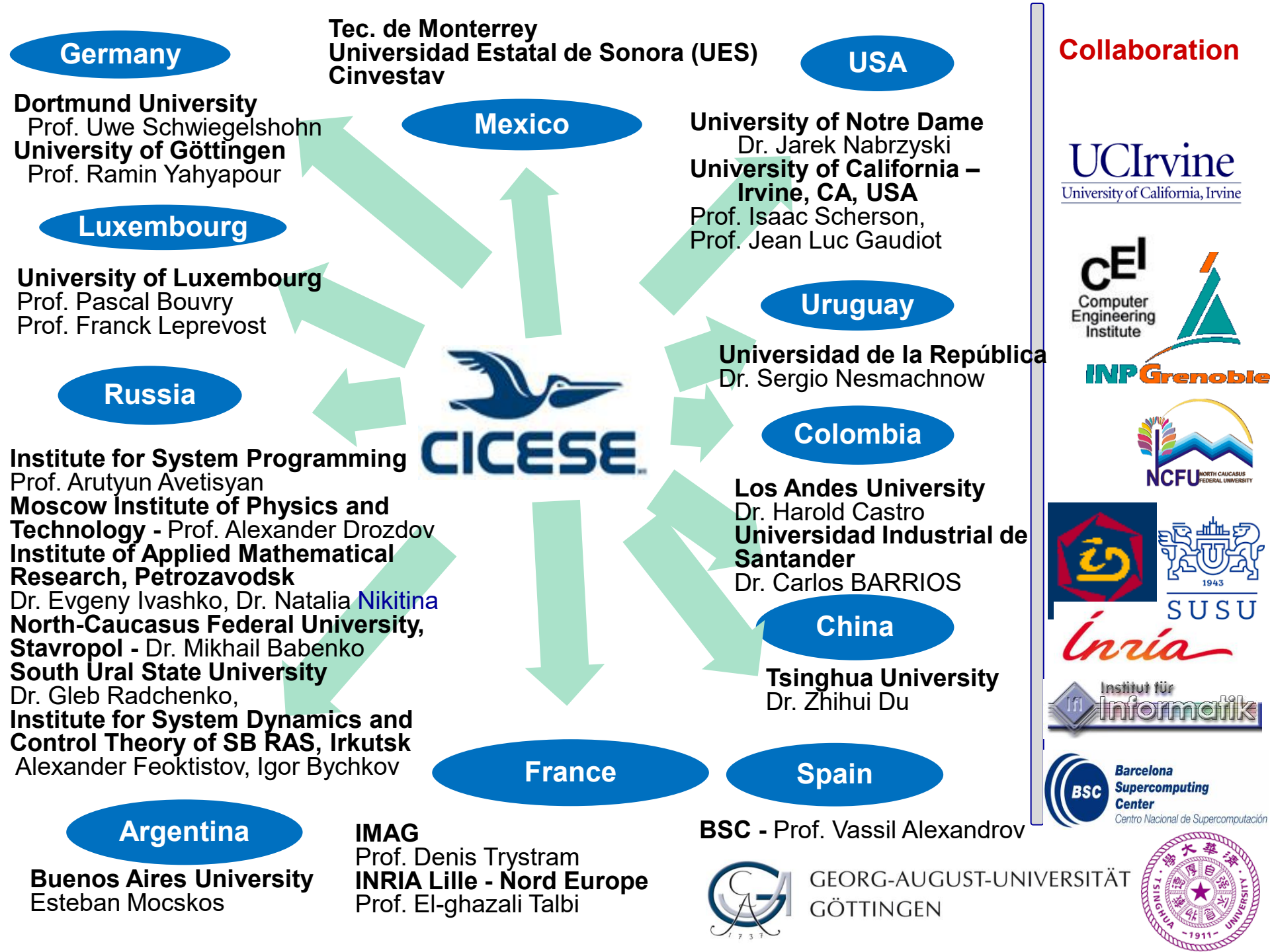


Thanks for the team

Team

and collaborators

Collaboration



Germany

Dortmund University
Prof. Uwe Schwiegelshohn
University of Göttingen
Prof. Ramin Yahyapour

Luxembourg

University of Luxembourg
Prof. Pascal Bouvry
Prof. Franck Leprevost

Russia

Institute for System Programming
Prof. Arutyun Avetisyan
Moscow Institute of Physics and Technology - Prof. Alexander Drozdov
Institute of Applied Mathematical Research, Petrozavodsk
Dr. Evgeny Ivashko, Dr. Natalia Nikitina
North-Caucasus Federal University, Stavropol - Dr. Mikhail Babenko
South Ural State University
Dr. Gleb Radchenko,
Institute for System Dynamics and Control Theory of SB RAS, Irkutsk
Alexander Feoktistov, Igor Bychkov

Argentina

Buenos Aires University
Esteban Mocskos

Tec. de Monterrey
Universidad Estatal de Sonora (UES)
Cinvestav

Mexico

USA

University of Notre Dame
Dr. Jarek Nabrzyski
University of California – Irvine, CA, USA
Prof. Isaac Scherson,
Prof. Jean Luc Gaudiot

Uruguay

Universidad de la República
Dr. Sergio Nesmachnow

Colombia

Los Andes University
Dr. Harold Castro
Universidad Industrial de Santander
Dr. Carlos BARRIOS

China

Tsinghua University
Dr. Zhihui Du

France

IMAG
Prof. Denis Trystram
INRIA Lille - Nord Europe
Prof. El-ghazali Talbi

Spain

BSC - Prof. Vassil Alexandrov

UCIrvine
University of California, Irvine

CEI Computer Engineering Institute

INP Grenoble

NCFU North Caucasus Federal University

1943 SUSU

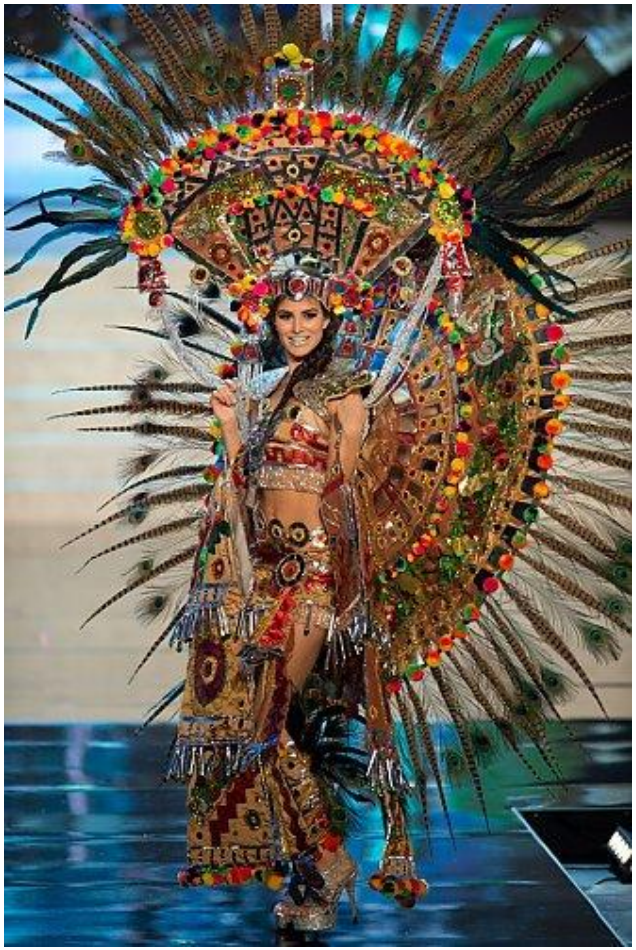
Inria

Institut für Informatik

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

TSINGHUA UNIVERSITY 1911



Thanks for your attention!

Thanks for your attention!