ECAI 2023 Tutorial – Kraków, Poland

Decentralized Federated Learning: Enabling Collaborative AI With Enhanced Trust and Efficiency

Alberto Huertas Celdrán[1], Enrique Tomás Martínez Beltrán[2], Pedro Miguel Sánchez Sánchez[2], Gérôme Bovet[3], Gregorio Martínez Pérez[2], and Burkhard Stiller[1]

[1]Communication Systems Group CSG, Department of Informatics IfI, University of Zurich UZH, Switzerland

[2]Department of Information and Communications Engineering, University of Murcia, Spain

[3]Cyber-Defence Campus within armasuisse Science & Technology, Thun, Switzerland

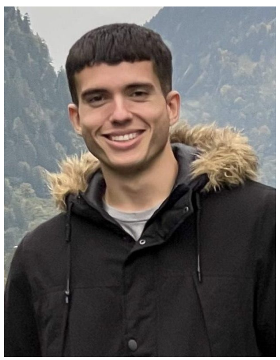
Presenters’ Information

Alberto Huertas Celdrán

Senior Researcher

Communication Systems Group (CSG) at the Department of Informatics (IfI), University of Zurich UZH, 8050 Zürich, Switzerland

huertas@ifi.uzh.ch

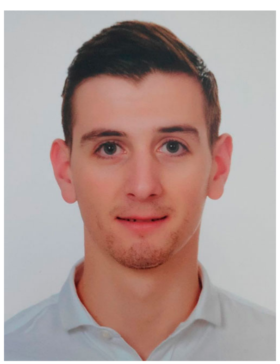


Enrique Tomás Martínez Beltrán

Junior Researcher

Department of Information and Communications Engineering, University of Murcia, 30100 Murcia, Spain

enriquetomas@um.es



Pedro Miguel Sánchez Sánchez

Junior Researcher

Department of Information and Communications Engineering, University of Murcia, 30100 Murcia, Spain

pedromiguel.sanchez@um.es

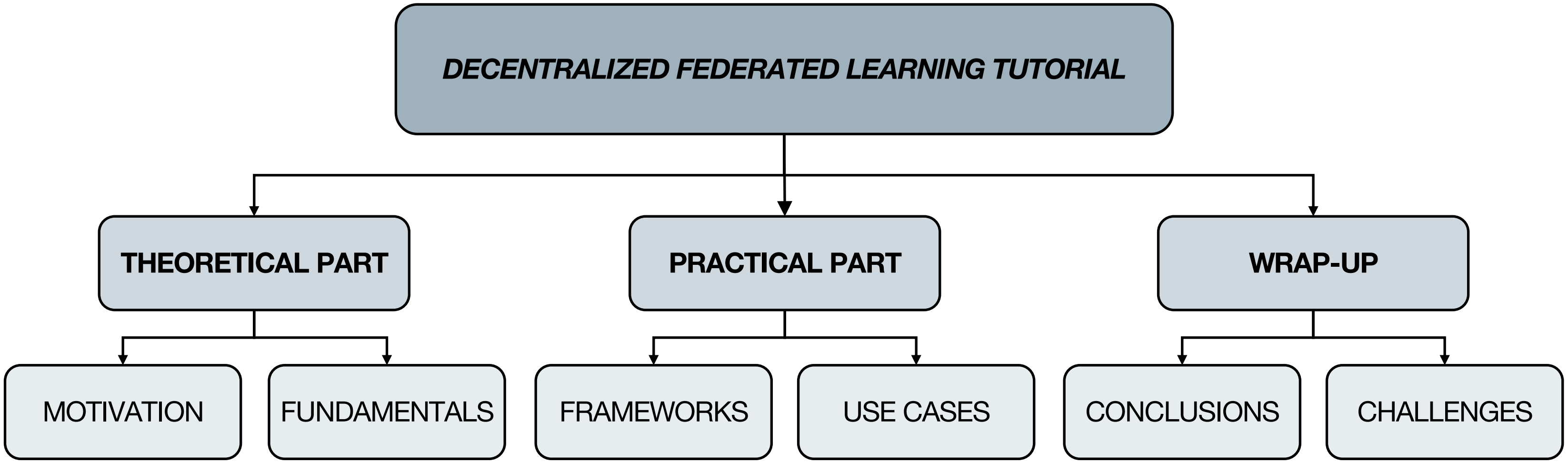
More information at https://cyberdatalab.um.es/dfl-ecai2023/

SCAN ME

Outline

- Tutorial structure → 3 main parts

1. Theoretical Part – *Decentralized Federated Learning Basics (40 min)*
2. Practical Part – Frameworks, Applications, and *Use Cases (30 min)*
3. Wrap-up – *Lessons Learned, Trends, and Conclusions (20 min)*



Background Details

- Analysis of 224 work items related to DFL
  - Fundamentals
  - Frameworks
  - Application Scenarios
  - Lessons Learned, Trends, and Open Challenges

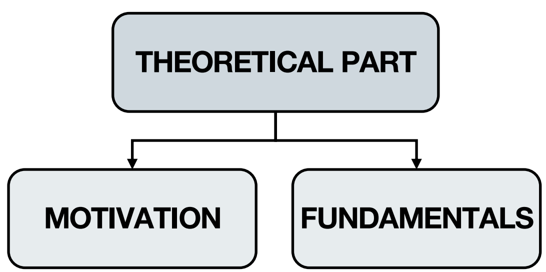
SCAN ME

- Published (September 2023)
  - IEEE Communications Surveys & Tutorials (35.6 IF, D1)

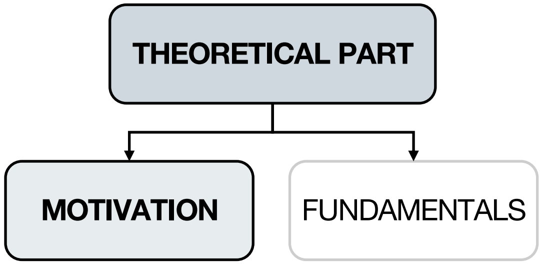
Martínez Beltrán, E. T., Quiles Pérez, M., Sánchez Sánchez, P. M., López Bernal, S., Bovet, G., Gil Pérez, M., Martínez Pérez, G., & Huertas Celdrán, A. (2023). Decentralized Federated Learning: Fundamentals, State of the Art, Frameworks, Trends, and Challenges. IEEE Communications Surveys & Tutorials. doi: 10.1109/COMST.2023.3315746

TUTORIAL – PART I

Decentralized Federated Learning
Motivation and Fundamentals



The Growing Importance of Decentralized Federated Learning

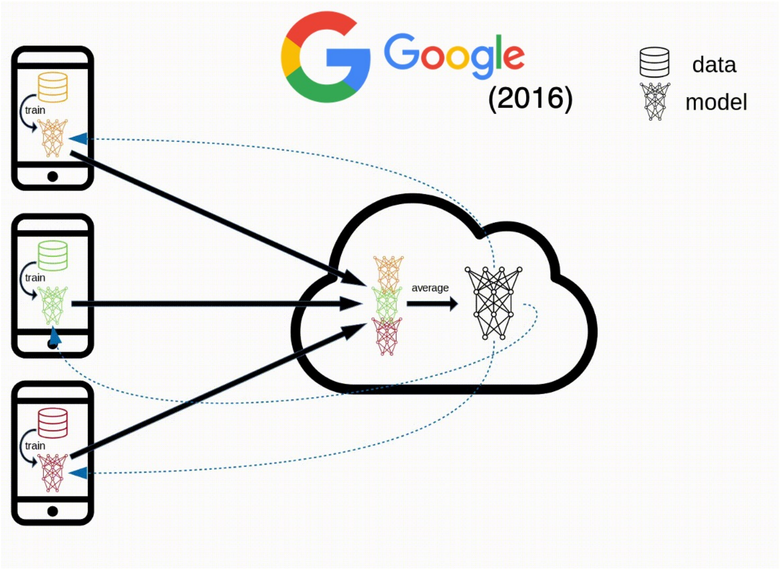


Federated Learning enables AI models to be trained directly on user devices, keeping data localized and private

- Typical lifecycle of FL process

1. ***Client Selection***: The **central server orchestrates** the training process samples from a set of clients
2. ***Broadcast***: The selected clients **download the current model weights** and training program
3. ***Client computation***: Each selected device **locally computes** an update to the model parameters
4. ***Aggregation***: The central server **collects all the model updates** from the devices and **aggregates them**
5. ***Model update***: The **server locally updates** the shared model based on the aggregated update

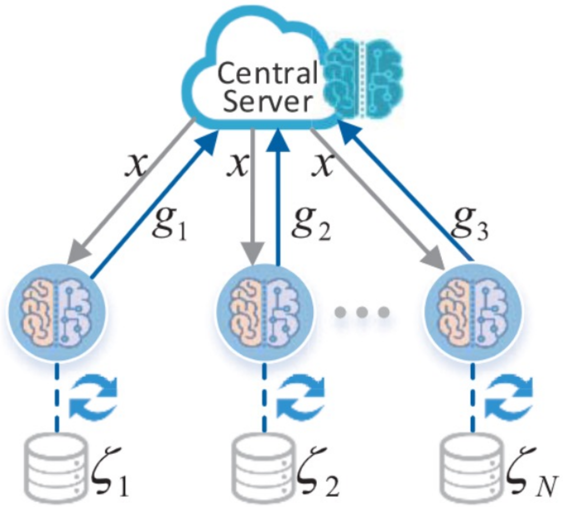
Steps 2-5 are repeated until our model has converged



MOTIVATION

Motivation

- Drawbacks of CFL
  - CFL has a **single point of failure**.
    - Any unavailability of the central server will cause an immediate and complete disruption of the training process
  - The server needs to have **reliable communication** with the devices
    - To support the transfer of potentially voluminous data with all of them
  - The server needs to be **trusted by all devices**
    - Also, it manages and guarantees the quality of service to orchestrate the process

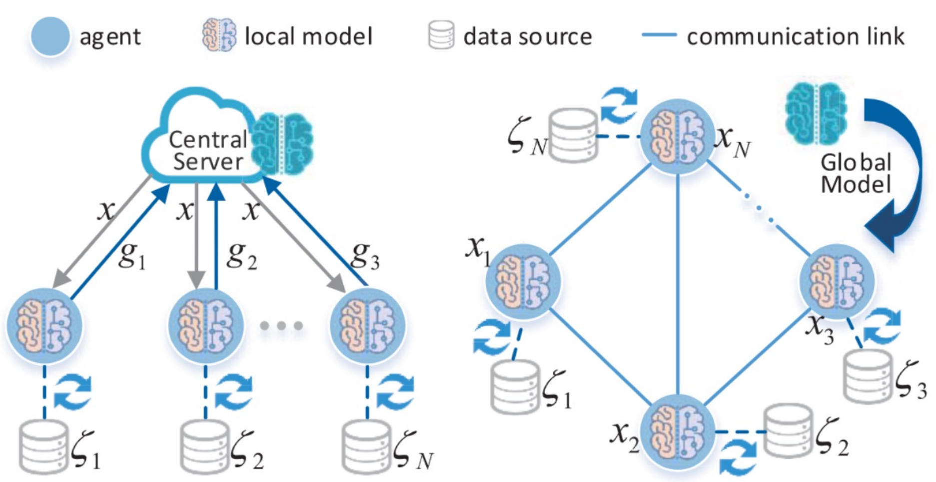


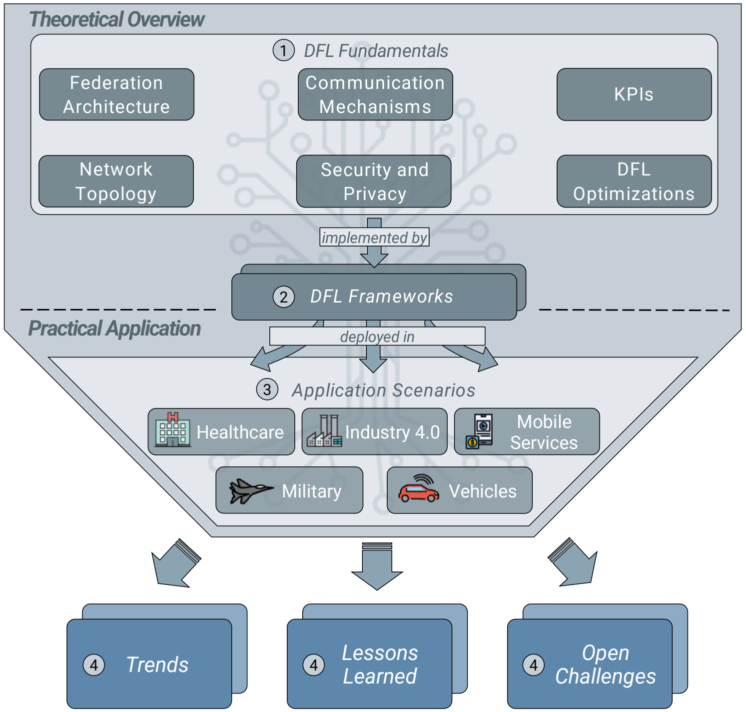
MOTIVATION

Decentralized Federated Learning

Decentralized Federated Learning removes the central server, allowing devices to collaborate directly in model training

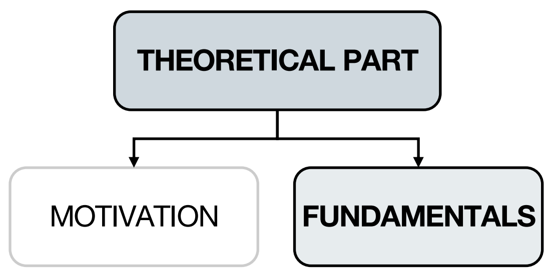
- Also known in the literature as "**Fully decentralized Federated Learning**" or "**Serverless Federated Learning**"





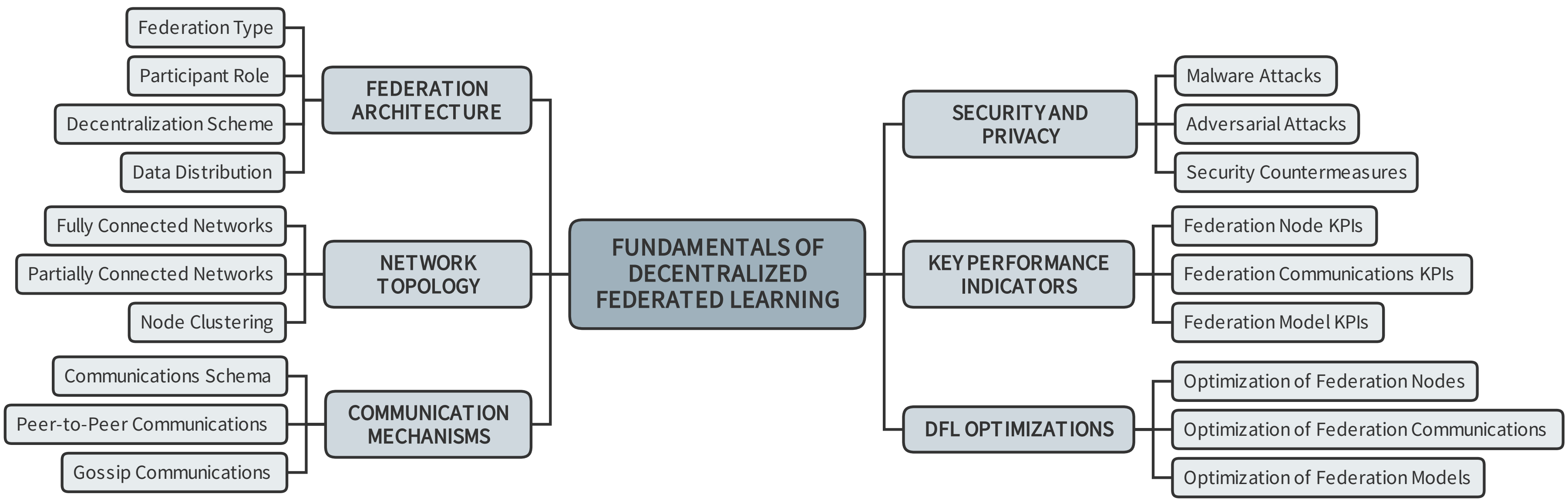
MOTIVATION

Fundamental aspects of Decentralized Federated Learning



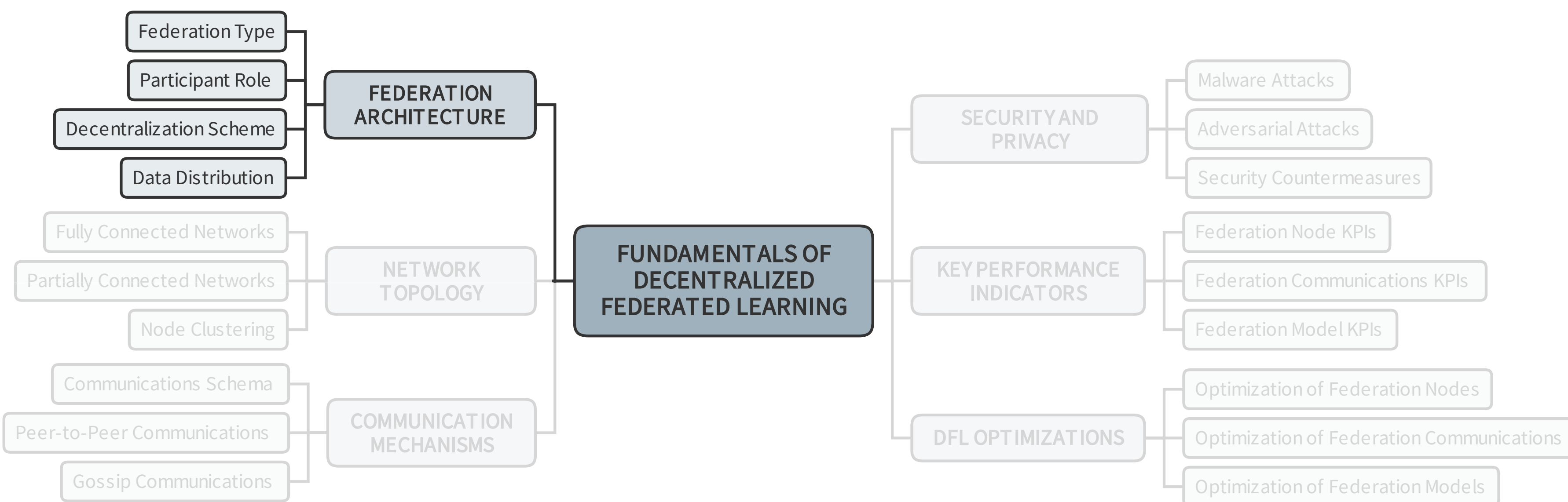
Fundamentals Overview

- Analyze the **key aspects** presented in the literature
  - **Federation Architecture**: Scheme dictating how DFL devices interact collaboratively
  - **Network Topology**: Defines layout and efficiency of node connections
  - **Comm. Mechanisms**: Protocols guiding communication among DFL nodes
  - **Security and Privacy**: Ensuring data protection in a decentralized setting
  - **KPIs**: Metrics to measure DFL efficiency and effectiveness
  - **Optimizations**: Refining DFL algorithms boosts performance outcomes



FUNDAMENTALS

Federation Architectures



FUNDAMENTALS

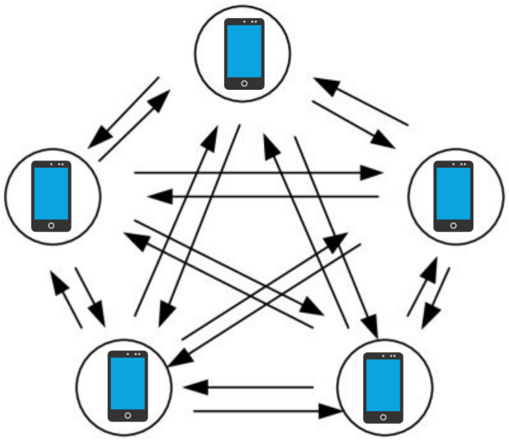
Fundamentals | Federation Architecture

UNIVERSIDAD DE MURCIA

- Analyze the key aspects presented in **Federation Types**

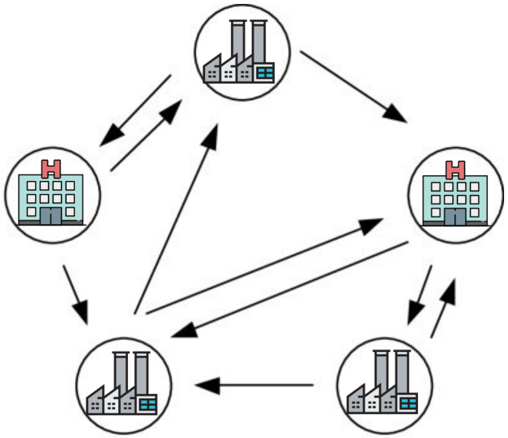
Networks of Remote Devices
Cross-device DFL

- Nodes number >100, each with thousands of samples
- Limited computational power per node
- Power consumption and complex training
- Nodes may periodically disconnect



Networks of Isolated Organizations
Cross-silo DFL

- Nodes are organizations or data centers
- Usually, <100 nodes with millions of samples
- Distributed from diverse business consumers
- Robust and scalable computing over time
- High network performance, minimizing failure points



FUNDAMENTALS

UNIVERSIDAD DE MURCIA

- Participants can be in different **roles** during the federation

Trainer

- Aims to train a local model with its local dataset
- Transmits parameters to neighbors and expects updated federated model parameters

Aggregator

- Responsible for obtaining and aggregating parameters in the global model
- Transmits them to neighboring nodes

Proxy

- Relays received model parameters to neighboring nodes
- Allows interconnection between different nodes or network topologies (e.g., clusters)

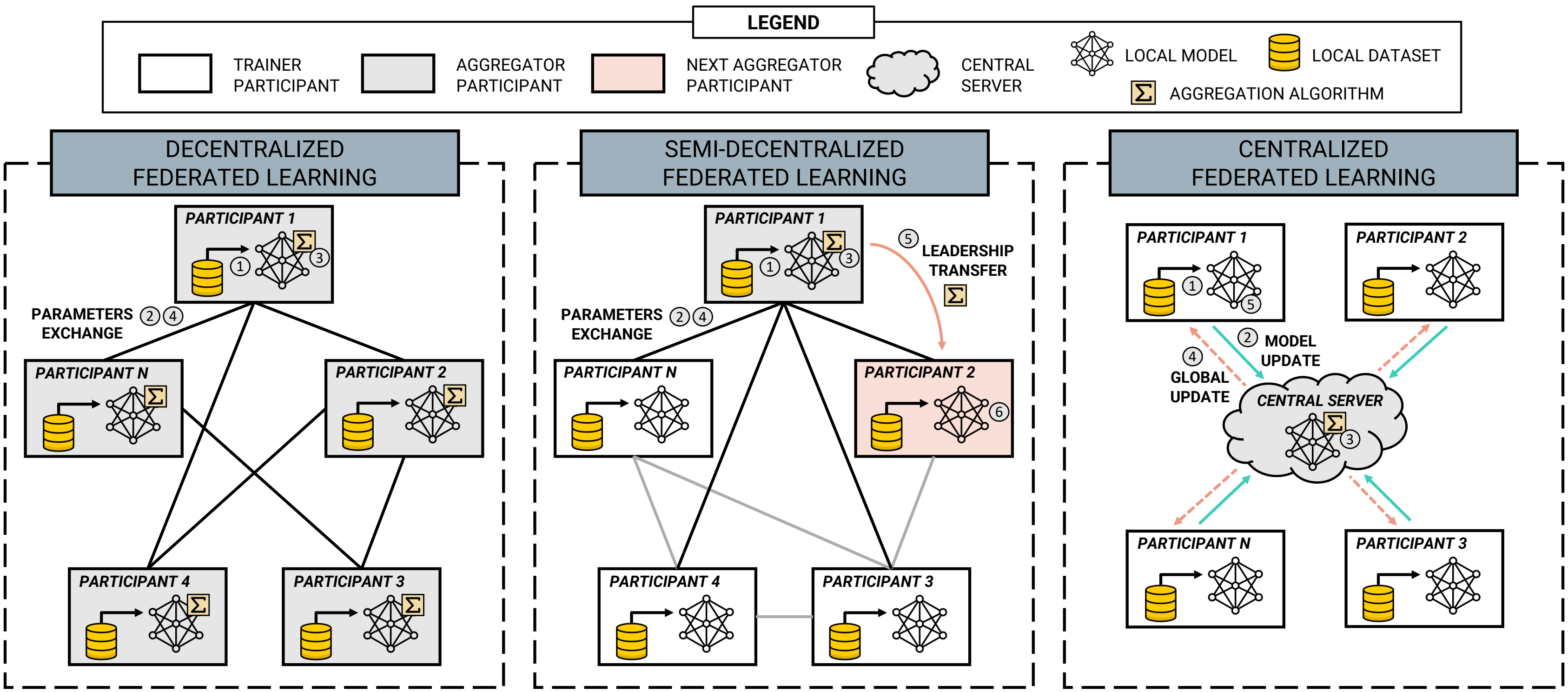
Idle

- May not have any of the roles, not participating in the federation

Fundamentals | Federation Architecture

UNIVERSIDAD DE MURCIA

- Analyze the key aspects presented in **Decentralization Scheme**



▪ Device-to-device communication	▪ Device-to-device communication	▪ Server-client communication
▪ No global coordination, **local aggregation in all devices**	▪ No global coordination, local aggregation, and **rotating role**	▪ Global coordination, **global aggregation**
▪ Asynchronous exchange	▪ Asynchronous exchange	▪ Single point of failure and bottleneck
		▪ Need to trust a central device

FUNDAMENTALS

- Explore various configurations in DFL depending on **data distribution**

Independent and Identically Distributed (IID) vs. Non-IID Data

Data vary in quality, diversity, and quantity in the network, increasing the complexity of training, analysis, and evaluation

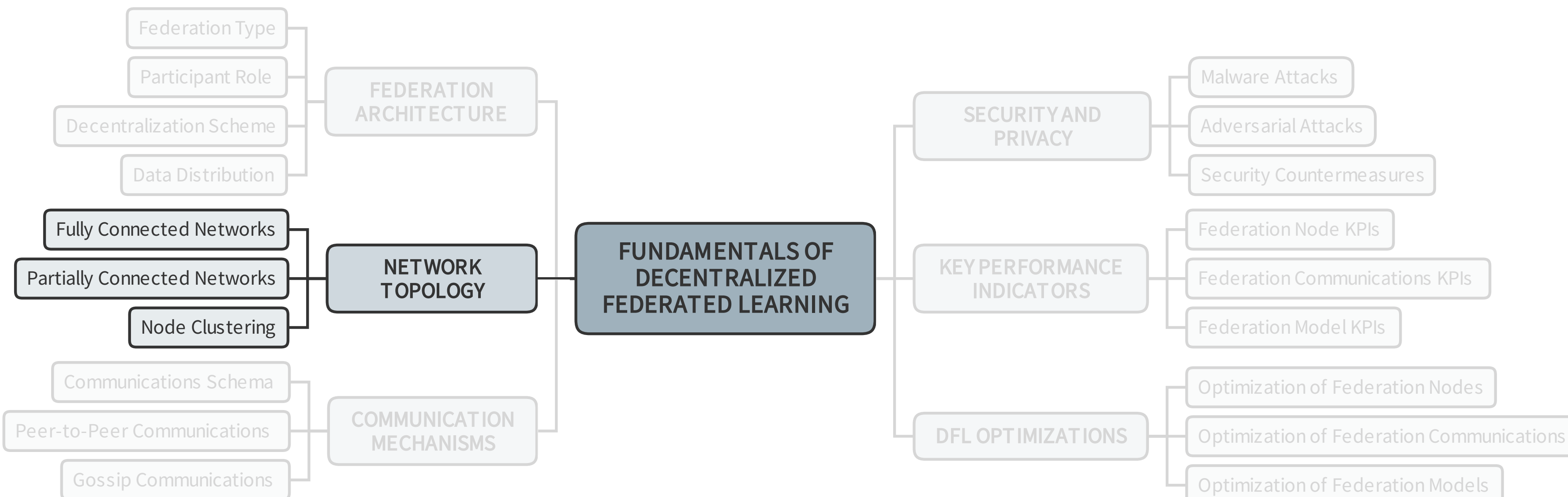
- Accepting the presence of different federated models
  - Adopt a flexible architecture to accommodate variations in data and model structures

Organization of Data

The strategic alignment of data forms the foundation for effective DFL, balancing the diversity of sources and the unified learning objectives

- Horizontal Federated Learning (HFL)
  - Applicable when there are many overlapping features and few overlapping nodes
  - Common in cross-device scenarios
- Vertical Federated Learning (VFL)
  - Focuses on feature binding with many overlapping nodes and few overlapping features
- Transfer Federated Learning (TFL)
  - Used when there is a limited feature and sample intersection between nodes.
  - Aims to build efficient models in cases where data are sparse

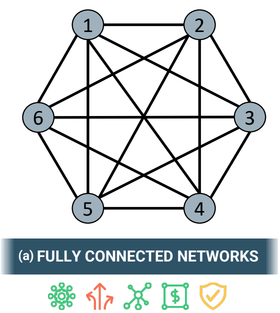
Network Topology



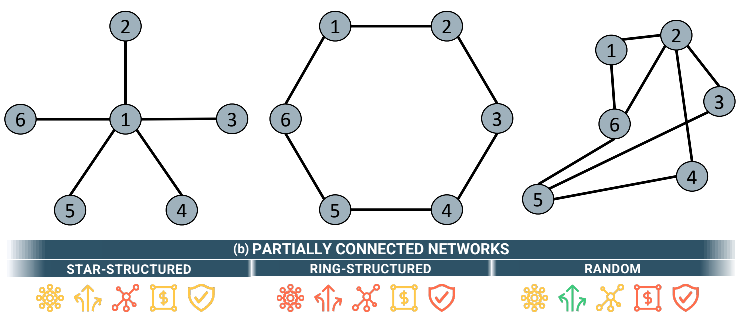
Fundamentals | Network Topology

CHARACTERISTICS: ROBUSTNESS · FLEXIBILITY · FAULT TOLERANCE · COMMUNICATIONS COST · SECURITY

IMPORTANCE: HIGH · MEDIUM · LOW



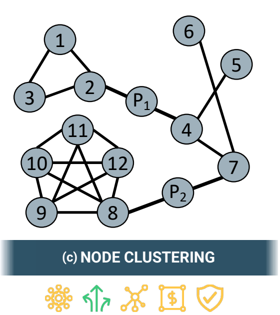
- High communication cost and complexity
- Low flexibility with new nodes
- High reliability and robustness despite failures



❑ **Star-structured**
- Communication cost grows linearly
- Potential bottleneck at the central node. Low fault tolerance.

❑ **Ring-structured** (unidirectional or bidirectional)
- Increased transmission delays as nodes grow

❑ **Random**
- Connections based on heuristics
- High flexibility and moderate fault tolerance

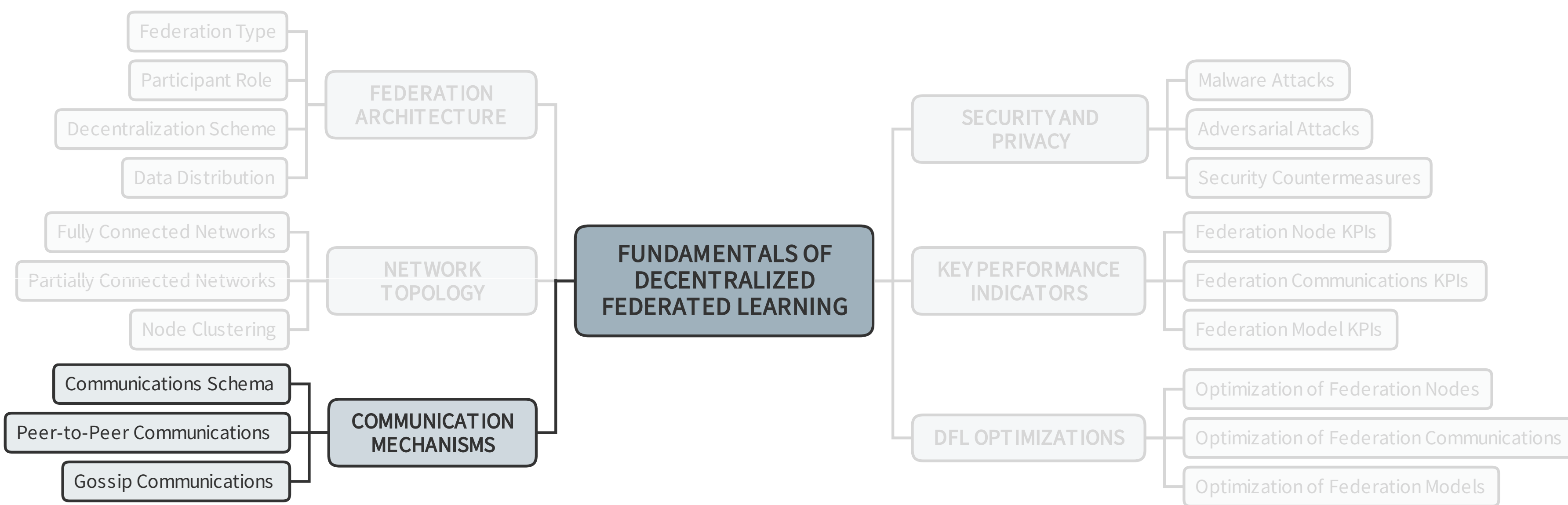


❑ **Similarity-based Clusters**
- Based on local model parameter similarity
- More individualized clusters

❑ **Proxy-based Clusters**
- Nodes interconnect different topologies
- A potential bottleneck in the overall architecture

FUNDAMENTALS

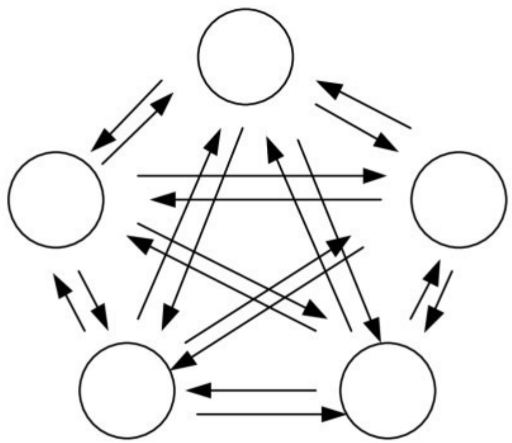
Communication Mechanisms



FUNDAMENTALS

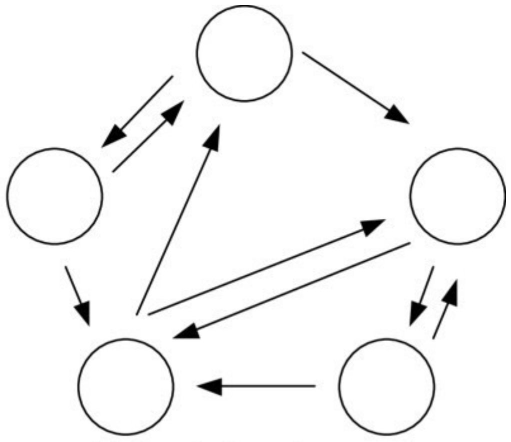
UNIVERSIDAD DE MURCIA

P2P approach



- Each peer is in direct contact with the rest

Gossip approach



- Peers operate in parallel, and with one or more randomly selected neighbors

Communications Scheme

Synchronous

- ❑ Nodes perform local multi-step training
- ❑ Parameters exchanged at synchronization points
- ❑ Slow convergence due to waiting time

Semi-synchronous

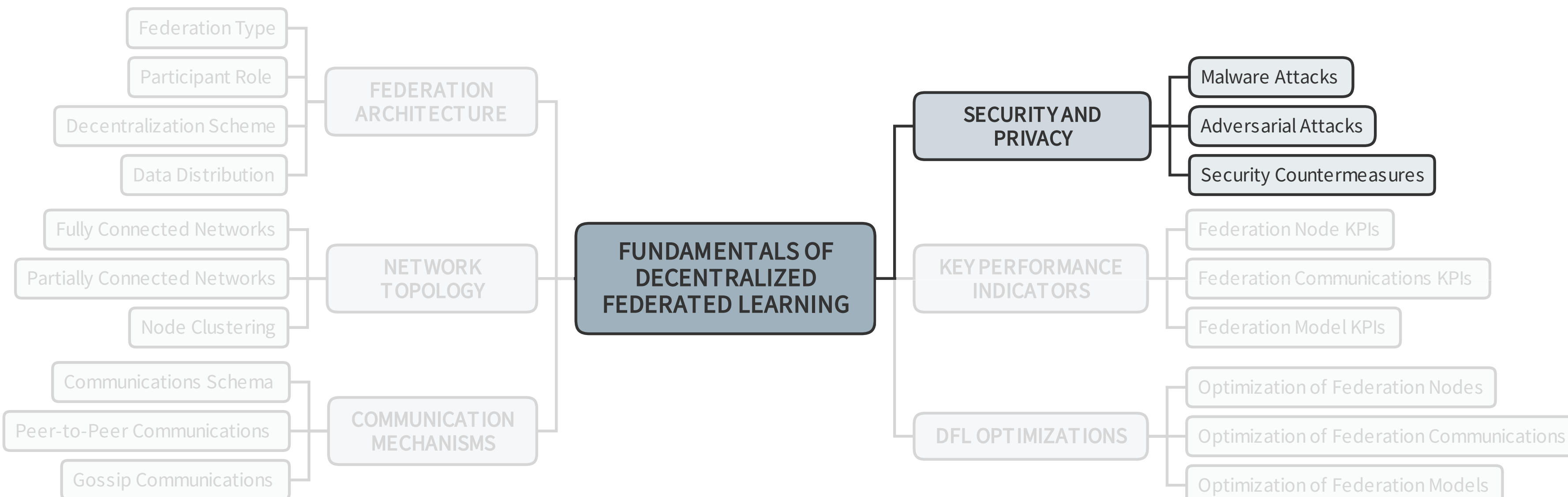
- ❑ Local training until a preset synchronization point
- ❑ Balances resource usage and communication costs
- ❑ Uses thresholds for beneficial aggregation

Asynchronous

- ❑ Independent parameter transmission and reception
- ❑ Offers faster convergence speed
- ❑ Faces higher costs and lower generalization due to staleness

FUNDAMENTALS

Security and Privacy



UNIVERSIDAD DE MURCIA

❑ Vulnerability in Decentralized Scenarios

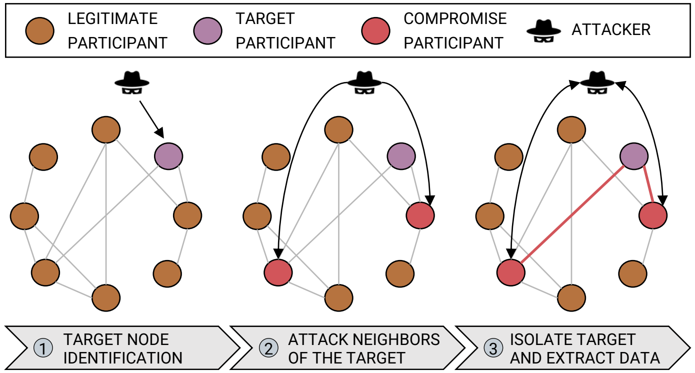
Unveiling the Critical Vulnerability Landscape in Decentralized Configurations

- Topology-based Impact
  - Fluctuating vulnerability levels are influenced by specific network structures
- Amplified Risks
  - Escalation in hazards due to profuse, sporadic, and fragile connections

❑ Diverse Attack Types

Exploring the Diverse and Complex Spectrum of Potential Attacks in Decentralized Systems

- Adversarial Attacks
  - Potential for unintended model and data manipulation and operational damage
  - Encompasses model inversion, membership inference, etc.
- Communications Attacks
  - Impact on behavioral and communicational aspects
  - Encompasses **eclipse attacks**, free-rider attacks, etc.



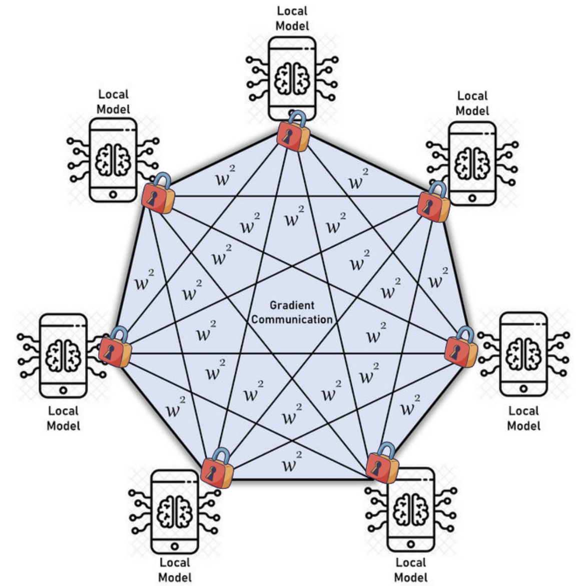
FUNDAMENTALS

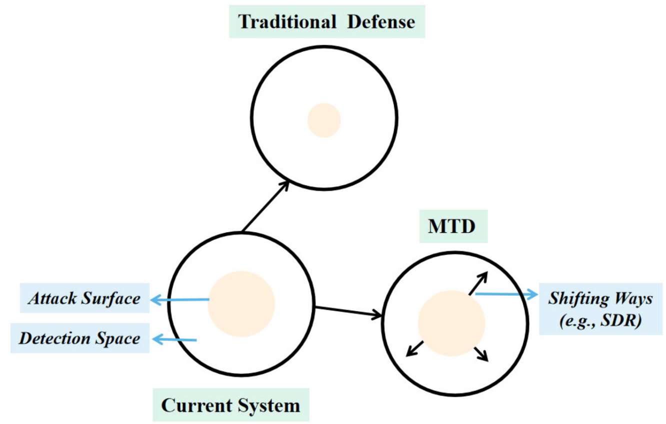
Fundamentals | Security and Privacy

❑ Strategic Security Countermeasures

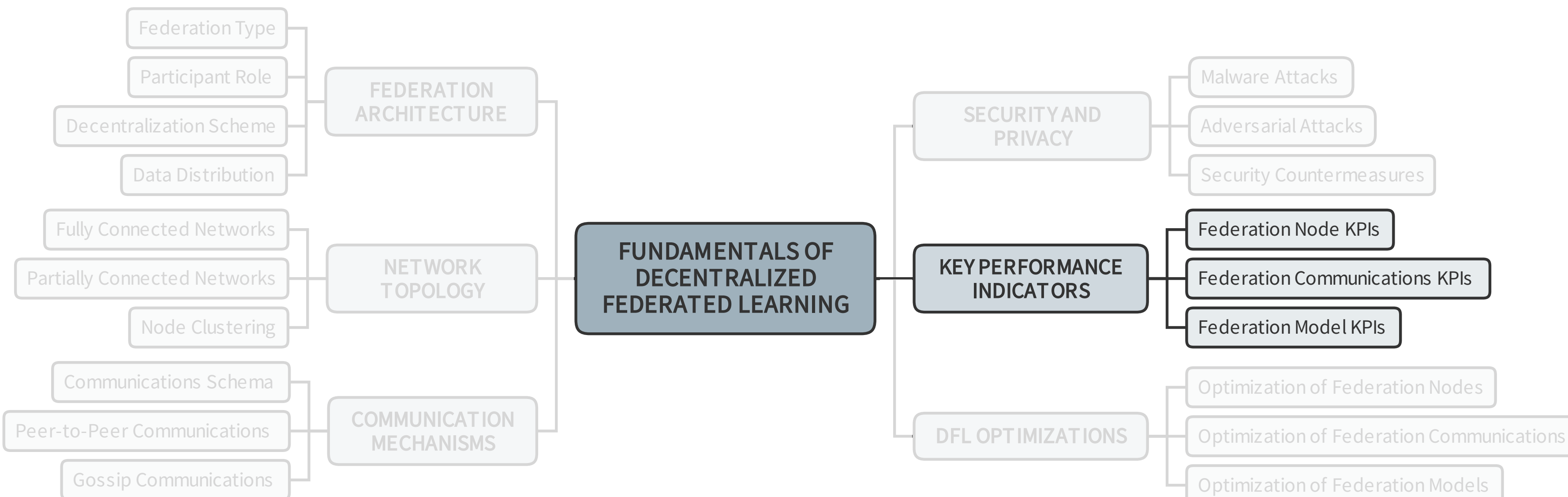
Forging Ahead with Strategic and Comprehensive Security Countermeasures to Safeguard DFL Systems

- Robust Data Protection
  - Employing cryptographic methods and differential privacy
- Robust Aggregation Mechanisms
  - Proactive creation and direct deployment among fresh network entrants
- Additional Security Layers
  - Incorporation of anomaly and model misbehavior detection
  - Blockchain technology ensures resilient decentralization and inter-participant reliability
  - **Moving Target Defense (MTD)** for dynamically altering attack surfaces



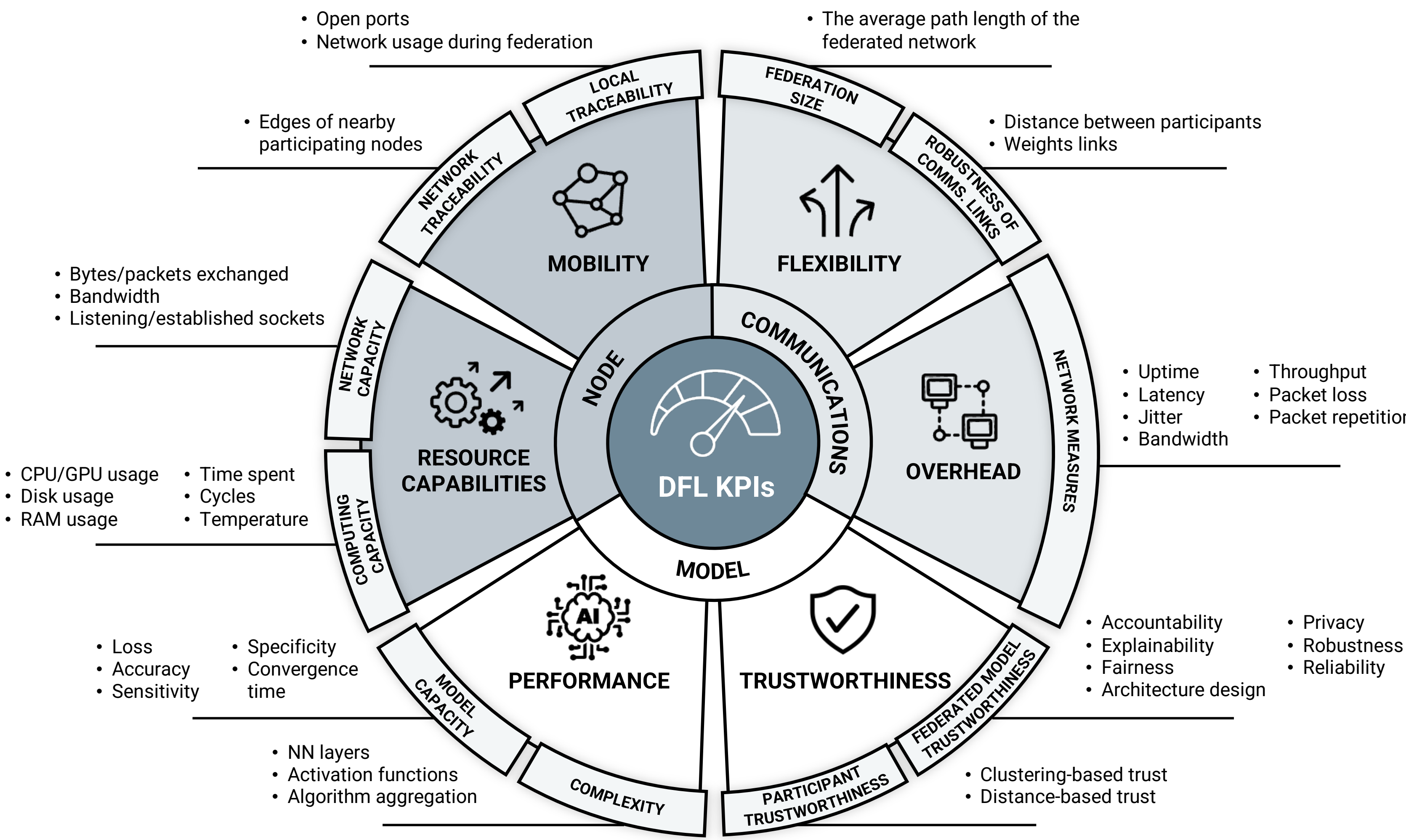


Key Performance Indicators



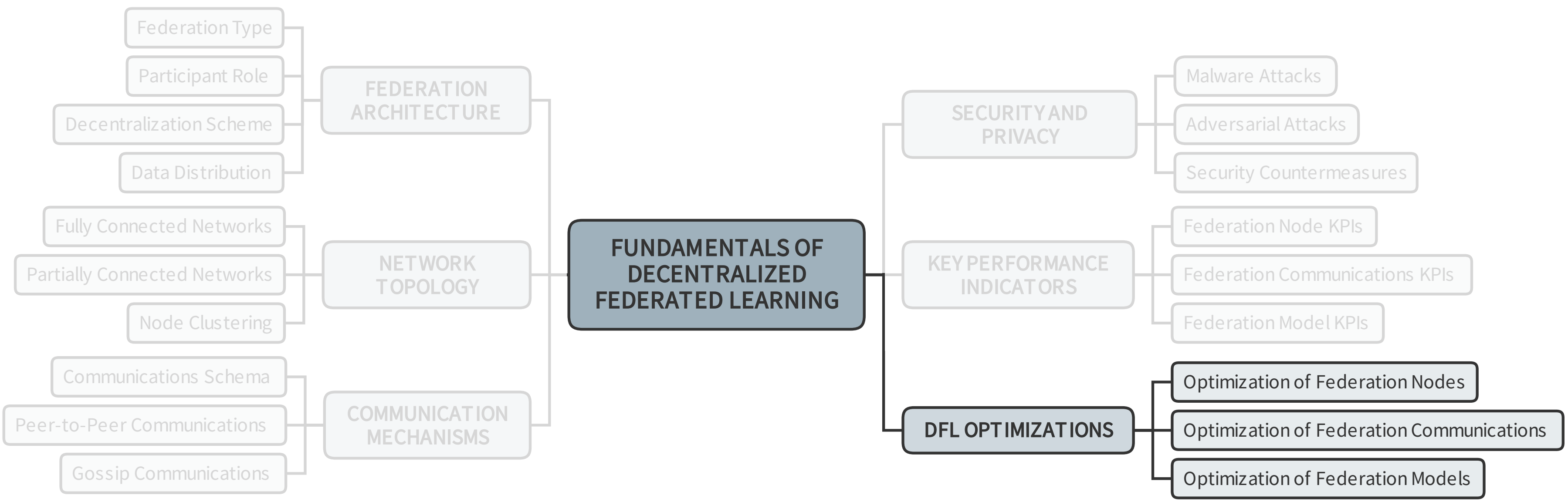
Fundamentals | Key Performance Indicators

UNIVERSIDAD DE MURCIA



FUNDAMENTALS

DFL Optimizations



FUNDAMENTALS

Fundamentals I DFL Optimizations

UNIVERSIDAD DE MURCIA

❑ Node Optimizations

Unlock the full potential of each node in the federation

- Optimize node selection during federation
- Enhance aggregation algorithm

❑ Communications Optimizations

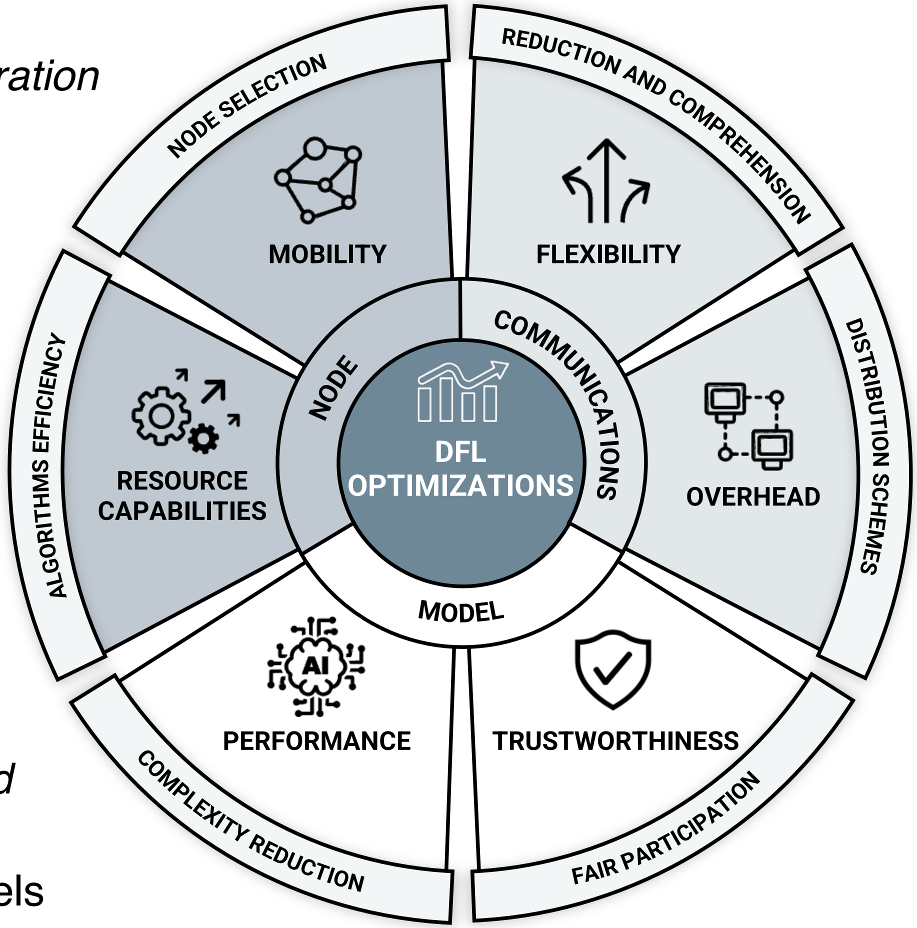
Ensure smooth, efficient communication across the network

- Parameters reduction and comprehension
- Improve distribution schemes

❑ Model Optimizations

Achieve superior federated models with balanced node contribution

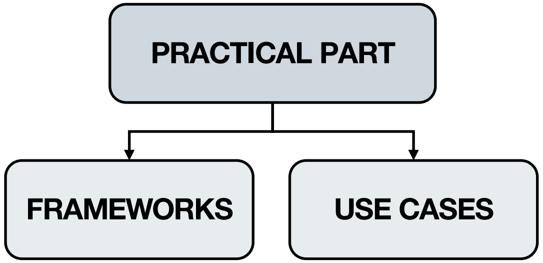
- Complexity reduction in the federated models
- Fair participation and trustworthiness



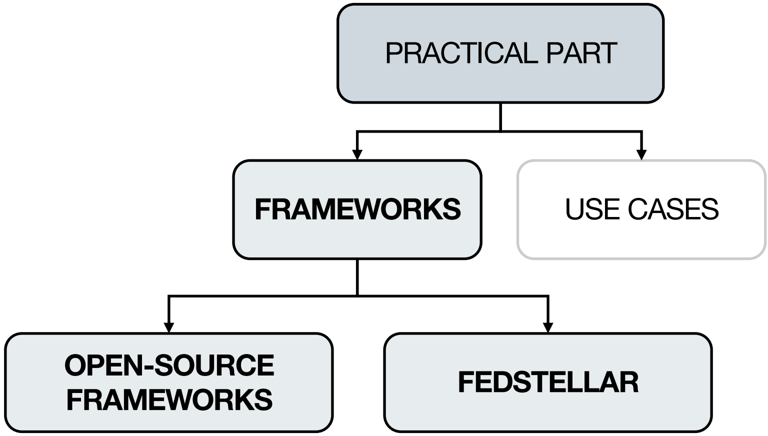
FUNDAMENTALS

TUTORIAL – PART II

Frameworks, Applications, and Use Cases



Frameworks


PRACTICAL PART
FRAMEWORKS
USE CASES
OPEN-SOURCE FRAMEWORKS
FEDSTELLAR

Open-Source Frameworks

UNIVERSIDAD DE MURCIA

Analysis of **+15 solutions**

Mature Frameworks

- ❑ Examples: Tensorflow Federated, PySyft, FederatedScope
- ❑ Characteristics
  - Supported by large companies
  - Offers robust building blocks and deployment on multiple machines
  - Limited focus on adversaries and privacy mechanisms

Incipient Frameworks

- ❑ Examples: BrainTorrent, IPLS, TrustFed, **Fedstellar**
- ❑ Characteristics
  - Aimed at specific applications like medical imaging
  - Emphasizes P2P communications
  - Utilizes emerging technologies like Blockchain and IPFS

Reference	OS	Federation Architecture Participant Type	 Aggregator Node	Aggregation Algorithms	Communication Protocol	Security and Privacy	Data Type	Scenario	Benchmarking
TFF [143]	Linux MacOS	Cross-silo	Centralized Decentralized	Median FedAvg FedProx	gRPC	✓	Time series Images	Simulation	✓
PySyft [144]	Windows Linux MacOS Mobile	Cross-silo Cross-device	Centralized Decentralized	FedAvg	Websockets	✓	Images	Simulation Real	✓
SecureBoost [56]	Linux MacOS	Cross-silo	Centralized Decentralized	FedAvg GBDT	gRPC	✓	Time series	Simulation	✗
FederatedScope [145]	Windows Linux MacOS Mobile	Cross-silo Cross-device	Centralized	FedAvg FedOpt	gRPC	✓	Time series Images	Simulation Real	✓
FedML [146]	Linux MacOS	Cross-silo	Centralized Decentralized	FedAvg FedOpt FedNova	gRPC MPI MQTT	✓	Time series Images	Simulation Real	✓
LEAF [147]	Linux MacOS	Cross-silo	Centralized Decentralized	-	-	-	Time series Images	Simulation	✓
BrainTorrent [148]	Windows Linux MacOS	Cross-device	Decentralized	FedAvg	N/S	✗	Time series	Simulation Real	✗
Scatterbrained [43]	Windows Linux MacOS	Cross-device	Centralized Decentralized	FedAvg	ZeroMQ	✗	Time series Images	Simulation	✗
IPLS [149]	Windows Linux MacOS	Cross-device	Decentralized	FedAvg	P2P	✓	Time series Images	Simulation Real	✓
TrustFed [142]	Windows Linux MacOS	Cross-device	Centralized Decentralized	FedAvg	P2P	✓	Time series Images	Simulation	✗
FLoBC [150]	Windows Linux MacOS	Cross-device	Decentralized	FedAvg	HTTP (REST API)	✗	Time series Images	Simulation Real	✓
BLADE-FL [151]	Windows Linux	Cross-device	Decentralized	Custom algorithm	P2P	✓	Images	Simulation	✓
DISCO [15]	Windows Linux MacOS Mobile	Cross-device	Centralized Decentralized	Custom FedAvg	peer.js	✓	Time series Images	Simulation	✗
CMFL [152]	Windows Linux MacOS	Cross-device	Decentralized	Median Trimmed Mean Krum Multi-Krum	P2P	✓	Time series Images	Simulation	✓
DeFL [40]	Windows Linux MacOS	Cross-device	Decentralized	Custom algorithm	N/S	✗	Time series Images	Simulation Real	✓
FL-SEC [12]	Windows Linux MacOS	Cross-device	Decentralized	FedAvg	N/S	✓	Time series	Simulation	✓
DisPFL [38]	Windows Linux MacOS	Cross-device	Decentralized	FedAvg Ditto FOMO Sub-FedAvg	P2P	✗	Time series Images	Simulation Real	✗
GossipFL [70]	Windows Linux MacOS	Cross-device	Decentralized	FedAvg S-FedAvg D-PSGD CHOCO-SGD	P2P	✗	Images	Simulation	✓
Fedstellar [153]	Windows Linux MacOS	Cross-silo Cross-device	Centralized Decentralized Semi-Decentralized	FedAvg Krum TrimmedMean Median	P2P HTTP (REST API)	✓	Time series Images	Simulation Real	✓

OPEN-SOURCE FRAMEWORKS

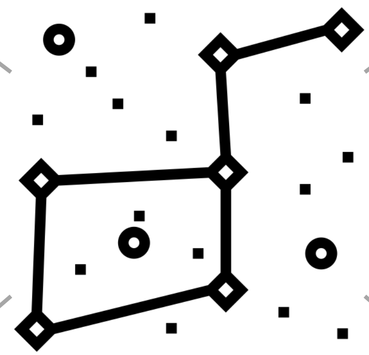
Fedstellar | Overview

***Fedstellar:** A Platform for Decentralized Federated Learning*

01 **Generation of Topologies**
Wide variety of network topologies for information exchange between participants

02 **Federation Architecture**
Type of decentralization of the federation: *decentralized, semi-decentralized or centralized*

03 **ML/DL Customization**
Definition of models and datasets adapted to the federation and to the task to be solved by the participants



04 **Deployment Scenario**
Deployment of nodes in simulated environments, physical devices or remote servers, adapted to needs and resources

05 **Real-time Monitoring**
Monitoring and management of node and network performance to optimize the federation or deployment process

06 **Documentation**
Extensive documentation on the installation and use of the platform

[1]Martínez Beltrán, et al. (2023). Fedstellar: A Platform for Decentralized Federated Learning. arXiv preprint arXiv:2306.09750.

FEDSTELLAR

Fedstellar | Goal

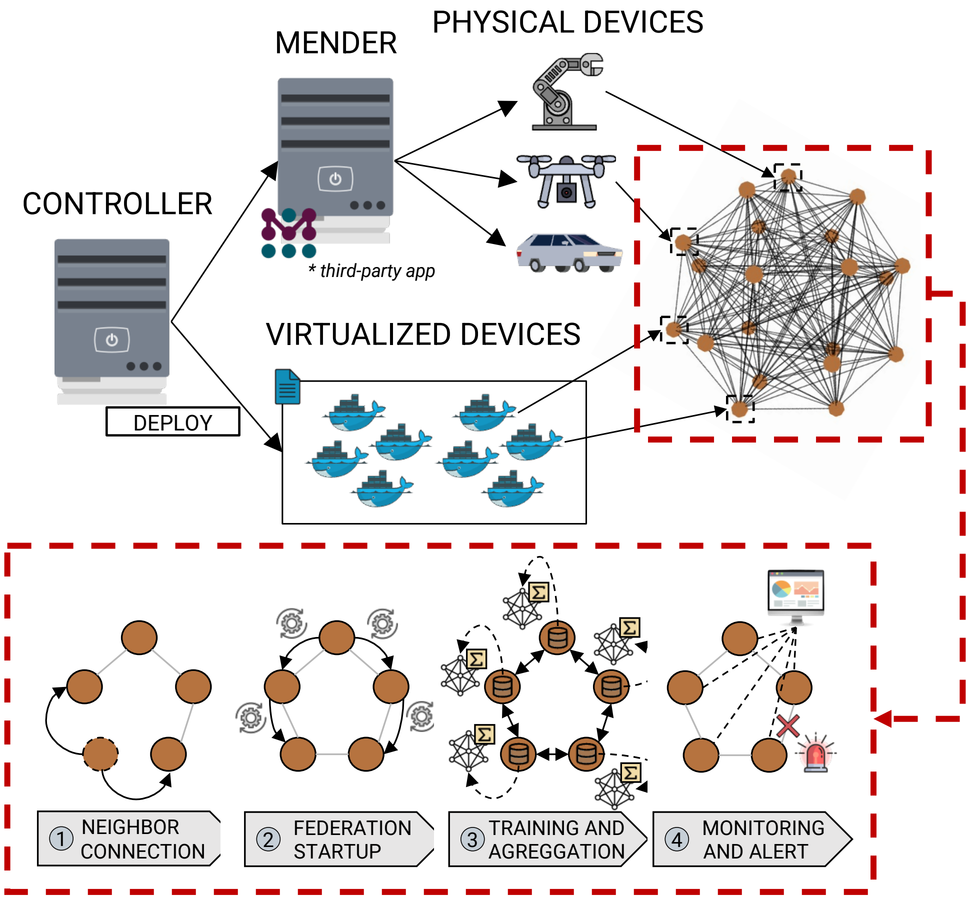
UNIVERSIDAD DE MURCIA

Fedstellar is an **innovative platform** that facilitates the training of FL models in a decentralized fashion.

- ❑ Deployment **of physical and virtual devices**
- ❑ **Topology generation**
- ❑ Provisioning of **federated functionality**
- ❑ Federation **monitoring and management**

Each device performs the following process

1. Creates a **communication link** with its immediate neighbors
2. Broadcasts a message outlining the **federation definition**
3. **Model training, decentralized aggregation, and asynchronous exchange** of model parameters
4. **Assess and report** on any disruptions



FEDSTELLAR

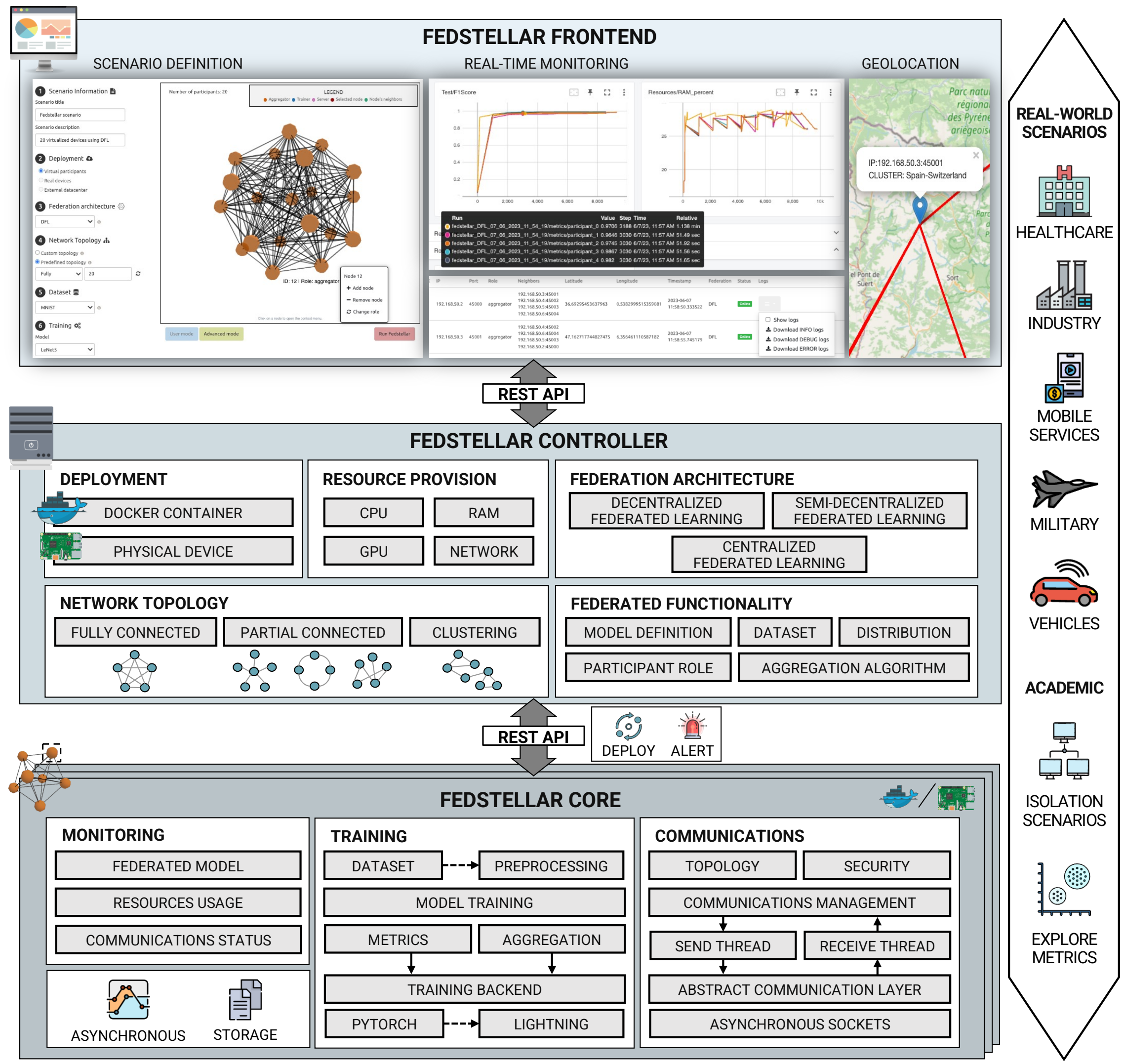
Fedstellar | Architecture

UNIVERSIDAD DE MURCIA

- **Frontend** provides high-level functionality for easy and fast deployment of federations
  - **Deployment** of federated topologies
  - **Monitoring** of metrics from the federation
- **Controller** serves as the orchestration of the platform
- **Core** provides the basic functionality for the execution of federations

SCAN ME

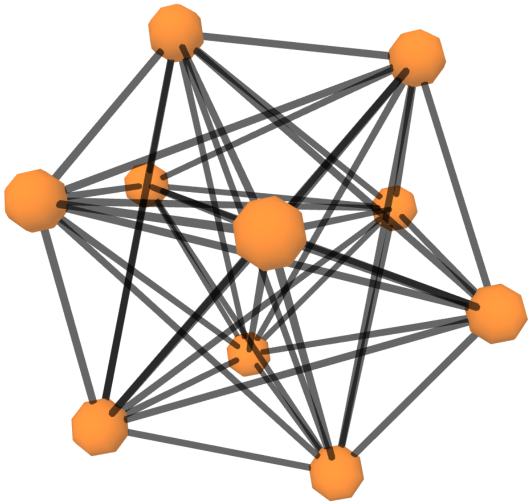
Fedstellar: A Platform for Decentralized Federated Learning
Expert Systems With Applications
https://arxiv.org/abs/2306.09750



FEDSTELLAR

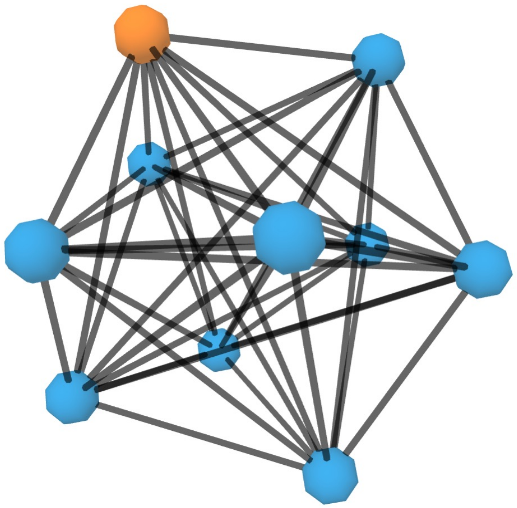
UNIVERSIDAD DE MURCIA

Decentralized Federated Learning (DFL)



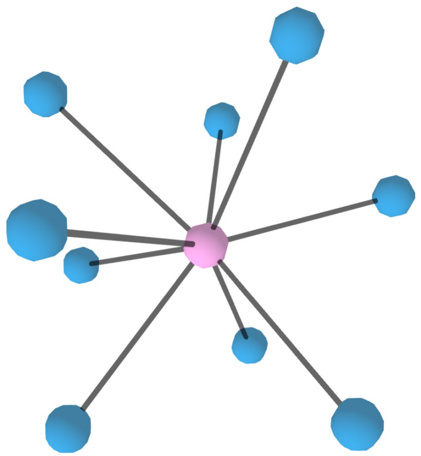
- **Fully connected** topology
- **All nodes aggregate** model parameters

Semi-Decentralized Federated Learning (SDFL)



- **Fully connected** topology
- **A different node aggregates** the model parameters in each federation round

Centralized Federated Learning (CFL)



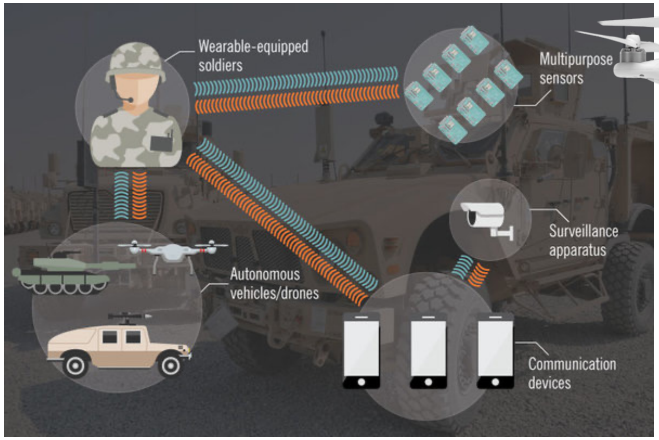
- **Star** topology
- **A server node** that **aggregates** the model parameters from the rest of the network

FEDSTELLAR

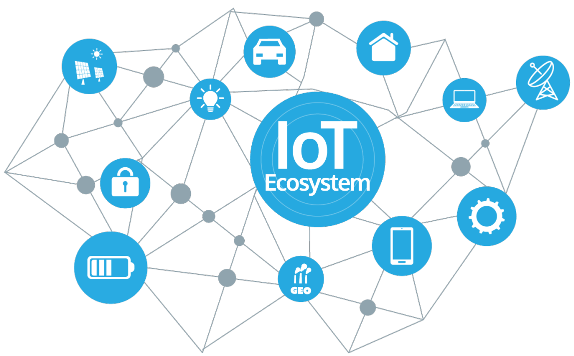
Application Scenarios for DFL

UNIVERSIDAD DE MURCIA

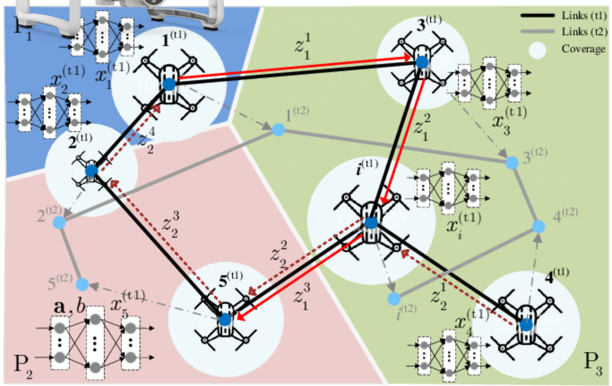
Military 🪖



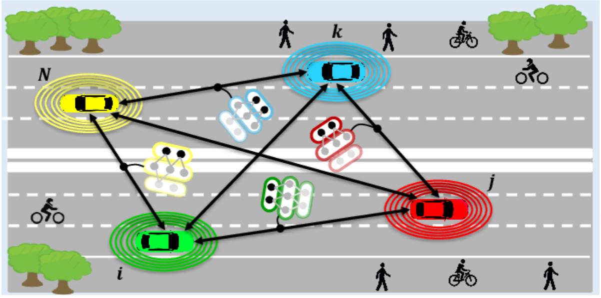
IoT devices 💻



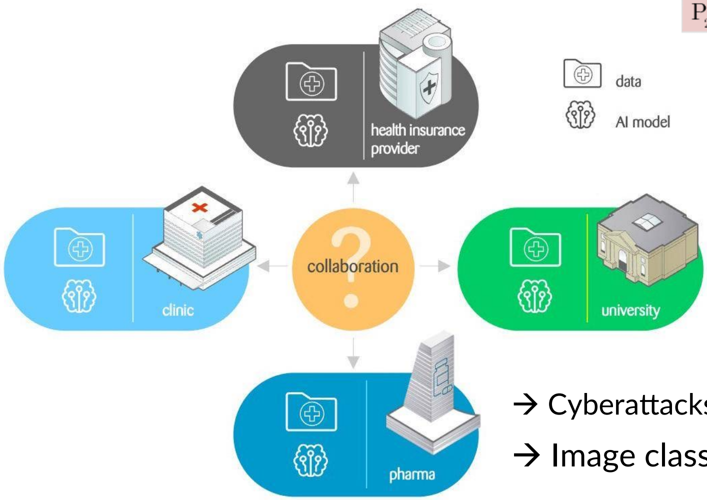
Drones ✈️



Vehicles 🚙

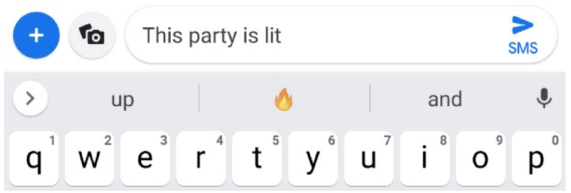


Healthcare 🩺 I Industry 4.0 🏭



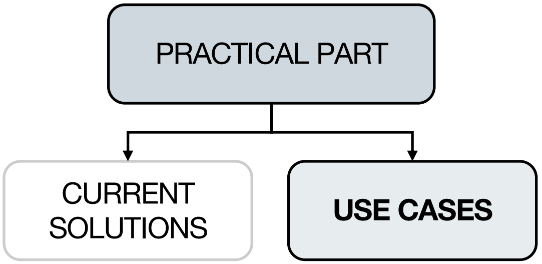
- → Cyberattacks detection
- → Image classification

Mobile services 📲



REAL-WORLD APPLICATTIONS

Use Cases



Use Case 1

Simulated deployments using docker containers and well-known distributed datasets

URL: https://federatedlearning.inf.um.es
User: demo / Password: DFL-ECAI-2023

USE CASES

Use Case 2

Decentralized Federated Learning to detect anomalies produced by malware affecting Raspberry Pis

Use Case 2 (I)

URL: https://federatedlearning.inf.um.es
User: demo / Password: DFL-ECAI-2023

UNIVERSIDAD DE MURCIA

***Fedstellar: A Platform for Decentralized Federated Learning*[1]**

Measure federated models, resource usage, and network of all participants

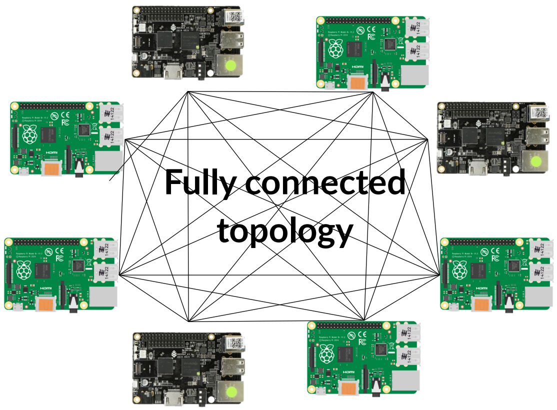
❑ Virtualized scenario

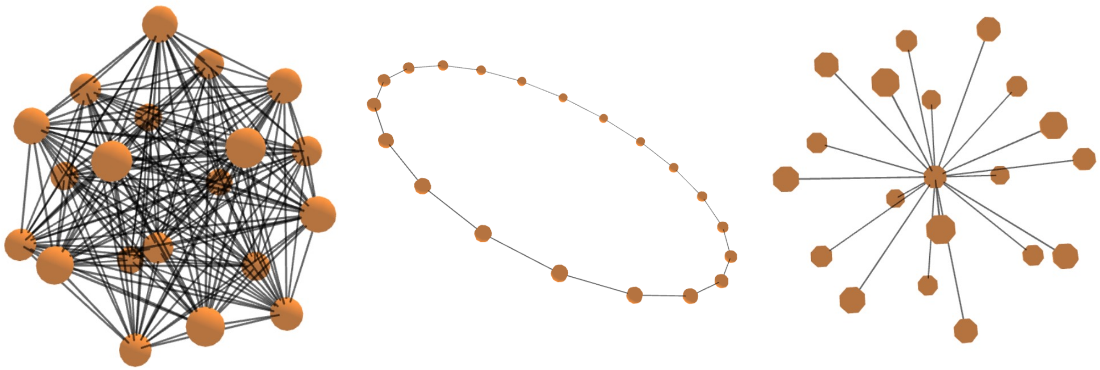
- 20 docker containers for image classification
- Using MNIST and CIFAR10 (Non-IID data)

❑ Physical scenario

- 8 single-board devices for detecting cyberattacks
- Using a dataset of syscalls

Characteristic	Physical Scenario	Virtualized Scenario
Participant Characteristics	8 (5 Raspberry Pi 4 / 3 Rock64)	20 (docker containers)
Dataset	Syscalls (Huertas Celdrán et al., 2023b)	MNIST (Deng, 2012) / CIFAR-10 (Krizhevsky, 2009)
Federated Model	Autoencoder	LeNet5 / MobileNet
Network Topology	Fully connected	Fully connected Star, Ring
Federation Architecture	DFL	DFL, SDFL, CFL



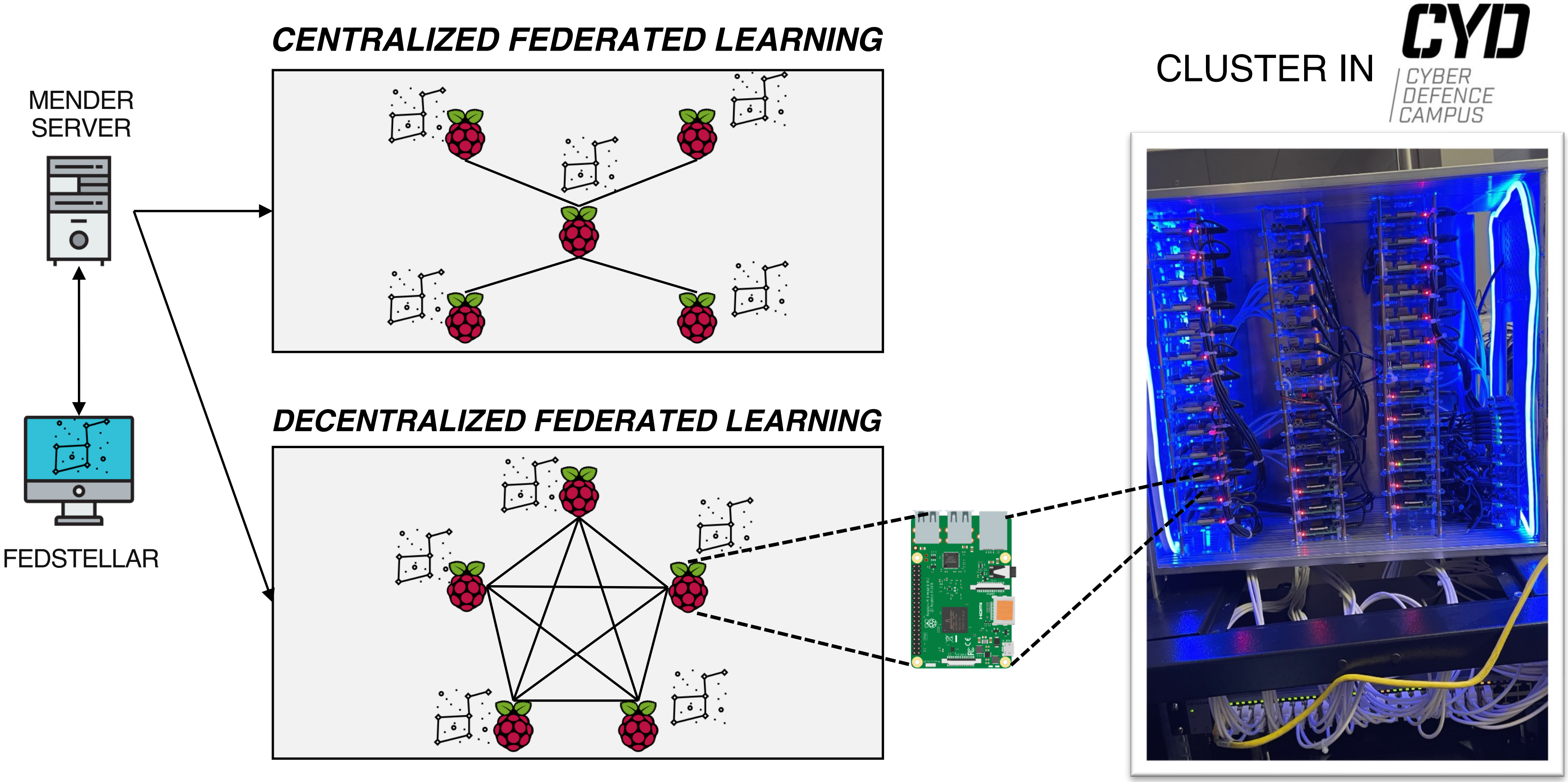


[1]Martínez Beltrán, et al. (2023). Fedstellar: A Platform for Decentralized Federated Learning. arXiv preprint arXiv:2306.09750.

USE CASES

Use Case 2 (II)

URL: https://federatedlearning.inf.um.es
User: demo / Password: DFL-ECAI-2023



[1]Martínez Beltrán, et al. (2023). Fedstellar: A Platform for Decentralized Federated Learning. arXiv preprint arXiv:2306.09750.

Use Case 2 (III)

URL: https://federatedlearning.inf.um.es
User: demo / Password: DFL-ECAI-2023

UNIVERSIDAD DE MURCIA

❑ Physical scenario

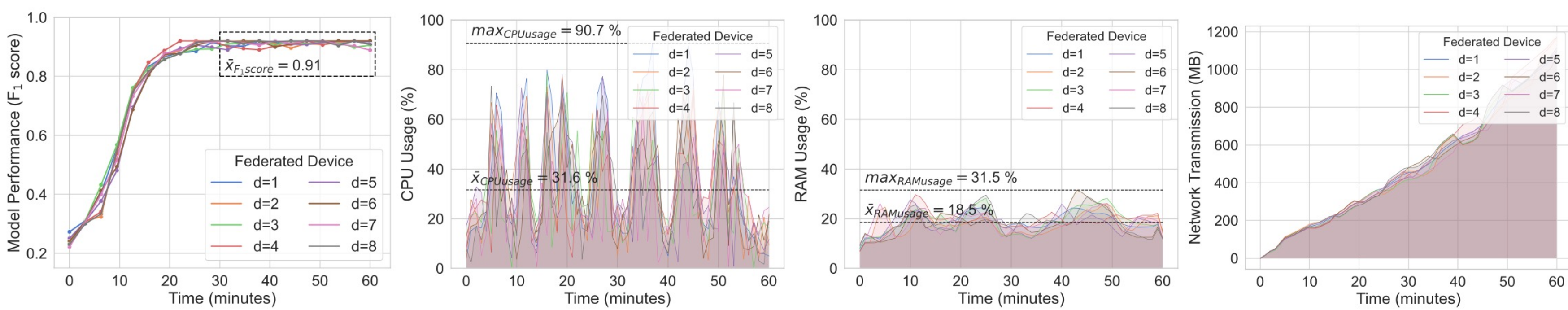
- Fedstellar achieved an **F1 score of 91%**
- **CPU usage of 31.6%** during federation
- **RAM usage of 18.5%** during federation

❑ Virtualized scenario

- Fedstellar obtained an **F1 score of 98% using DFL** and **97.3% using SDFL** with MNIST
- **Reduction of the training time** for model convergence by 32% compared to centralized architectures

Federation Architecture	Network Topology	Model (F_1 *score*)	CPU (%)	RAM (%)	Network (MB)	Time* (*min.*)
DFL	Fully	0.987 ±0.009	78 ±15 %	29 ±6 %	≈ 1243 MB	≈ 28
	Star	0.955 ±0.012	72 ±13 %	28 ±5 %	≈ 1165 MB	≈ 35
	Ring	0.917 ±0.019	70 ±14 %	26 ±4 %	≈ 1089 MB	≈ 41
SDFL	Fully	0.973 ±0.015	69 ±12 %	28 ±5 %	≈ 1148 MB	≈ 32
	Star	0.938 ±0.020	66 ±11 %	27 ±4 %	≈ 1065 MB	≈ 38
	Ring	0.901 ±0.027	64 ±13 %	25 ±4 %	≈ 1023 MB	≈ 45
CFL	Star	0.992 ±0.010	58 ±10 %	26 ±3 %	≈ 985 MB	≈ 40

* Overall time to reach model F_1 *score* ≥ 90 %



(a) Federated models (F_1 *score*) (b) CPU usage (%) (c) RAM usage (%) (d) Network usage (MB)

[1]Martínez Beltrán, et al. (2023). Fedstellar: A Platform for Decentralized Federated Learning. arXiv preprint arXiv:2306.09750.

Use Case 2 (IV)

URL: https://federatedlearning.inf.um.es
User: demo / Password: DFL-ECAI-2023

UNIVERSIDAD DE MURCIA

Mitigating Communications Threats in Decentralized Federated Learning through Moving Target Defense[1]

❑ Motivation

- DFL poses **various types of sensitive information** to federation risks, including network topology, participants' roles, and communication patterns
- Presence of **communication-based attacks:** disrupting the model aggregation process and cause security breaches or privacy infringements

❑ Contributions

- Create a **threat model** that identifies sensitive information susceptible to eavesdropping, Man-in-the-Middle, and eclipse attacks
- Design and implement an **eclipse attack**
- Develop a **security module** enabling encryption and proactive defense using MTD
- Three security configurations were assessed
  - **No security** baseline
  - A configuration with **encryption**
  - A configuration integrating the **encryption and MTD**

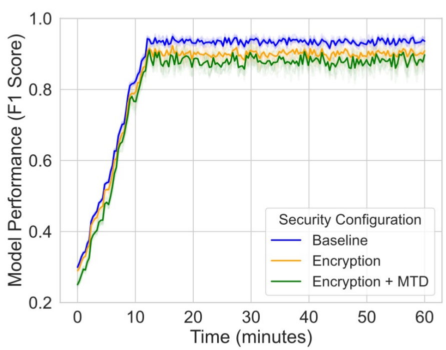
Characteristic	Description
DFL Platform	Fedstellar [18]
Federation Architecture	DFL
Participants	5 Raspberry Pi 4 3 Rock64
Network Topology	Random
Federated Model	LeNet5
Dataset	MNIST [19]
Security Configuration	1 Baseline 2 Encryption 3 Encryption and MTD
Attack	Eclipse attack: • One external attacker • One target participant

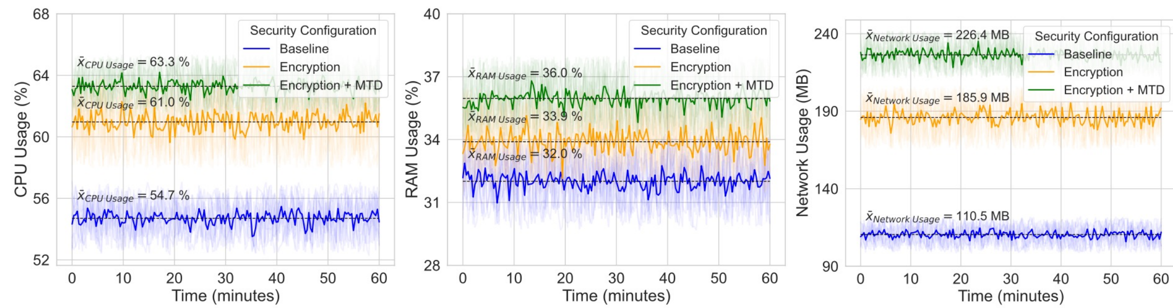
[1]Martínez Beltrán, et al. (2023). Mitigating Communications Threats in Decentralized Federated Learning through Moving Target Defense. arXiv preprint arXiv:2307.11730.

URL: https://federatedlearning.inf.um.es
User: demo / Password: DFL-ECAI-2023

UNIVERSIDAD DE MURCIA

- An extensive experimental evaluation used a **real-world topology** with diverse connections and participants
- The **MNIST dataset** and eclipse attacks were utilized for the evaluation





(a) CPU usage (%) (b) RAM usage (%) (c) Network usage (MB)

- An average **F1 score of 93%** was reached, peaking at 97% without security measures
- Secure configurations using MTD shows a minor increase in CPU usage and network traffic and a slight rise in RAM

Table 5: Security Settings, Protection, and Performance in DFL

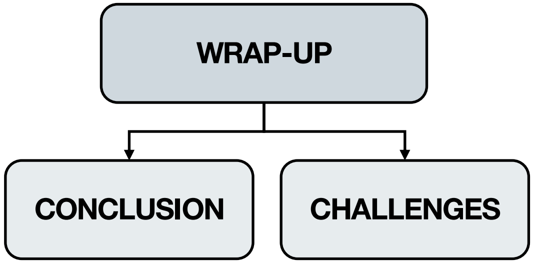
Security Configuration	Information Protected	Performance Metrics			
		F1 Score	CPU	RAM	Network
Baseline (No security)	N/A	97%	54.6% ±1.8%	31.9% ±2.3%	110.2 MB ±12 MB
Encryption	• Model Parameters • Roles • Communication Patterns	94%	60.9% ±3.7%	33.8% ±2.41%	185.2 MB ±21 MB
Encryption + MTD	• Model Parameters • Roles • Communication Patterns • Topology • Activity Periods	92.5%	63.2% ±3.5%	35.9% ±1.5%	226 MB ±15 MB

[1]Martínez Beltrán, et al. (2023). Mitigating Communications Threats in Decentralized Federated Learning through Moving Target Defense. arXiv preprint arXiv:2307.11730.

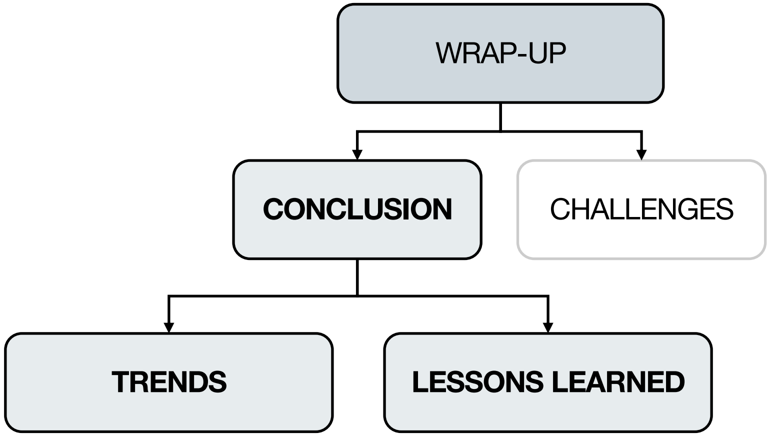
USE CASES

TUTORIAL – PART III

Conclusions and Challenges



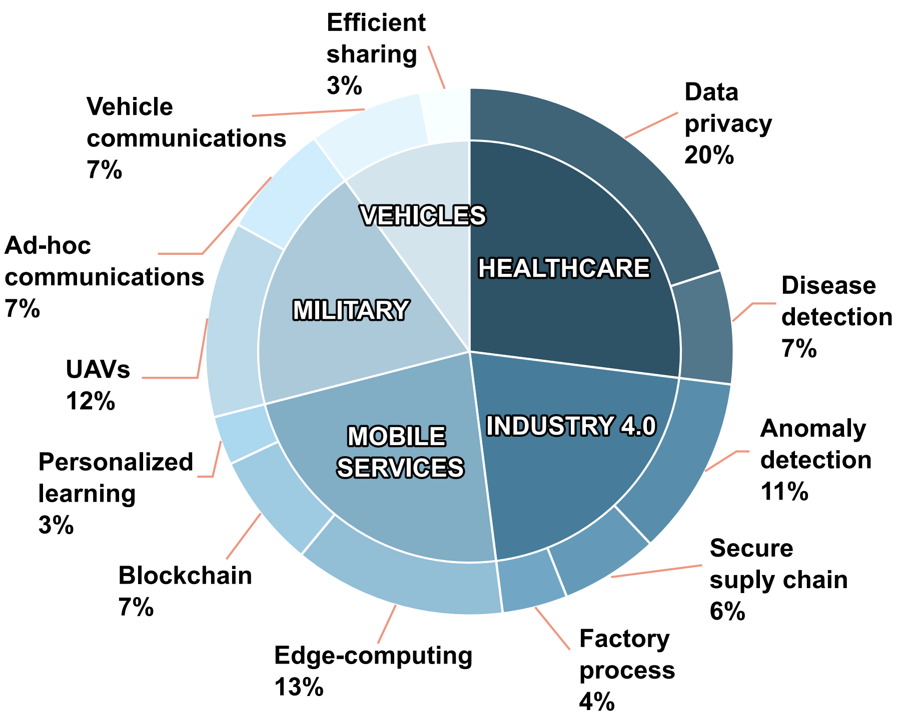
Conclusions



Current Trends

UNIVERSIDAD DE MURCIA

*Explore the breadth of **DFL applications**, from healthcare to vehicles, demonstrating its adaptability and vital role across **application scenarios***



- **Healthcare**
  - Widely used in electronic medical records, medical imaging, disease detection, and collaborative drug discovery
- **Industry 4.0**
  - Reducing costs and enhancing operational efficiency
- **Mobile Services**
  - Augments personalization and privacy, refining user experiences in edge devices
- **Military**
  - Prominent in UAV deployment, ensuring enhanced security and operational excellence
- **Vehicles**
  - Facilitates anomaly detection and improved communication systems, ensuring safer and more efficient vehicle operations

CONCLUSIONS

Current Trends

UNIVERSIDAD DE MURCIA

Unearth the fundamentals of DFL, focusing on architecture, communication, and optimization

Federation Architecture

- Central to DFL, ensuring robust and effective communication between diverse network nodes

Network Topology

- Predominantly fully connected, offering versatility and simplicity in DFL scenarios (used in about 50%)

Communication Mechanisms

- Over 65% of solutions focus on enhanced, streamlined communication, particularly in healthcare and mobile services

Security and Privacy

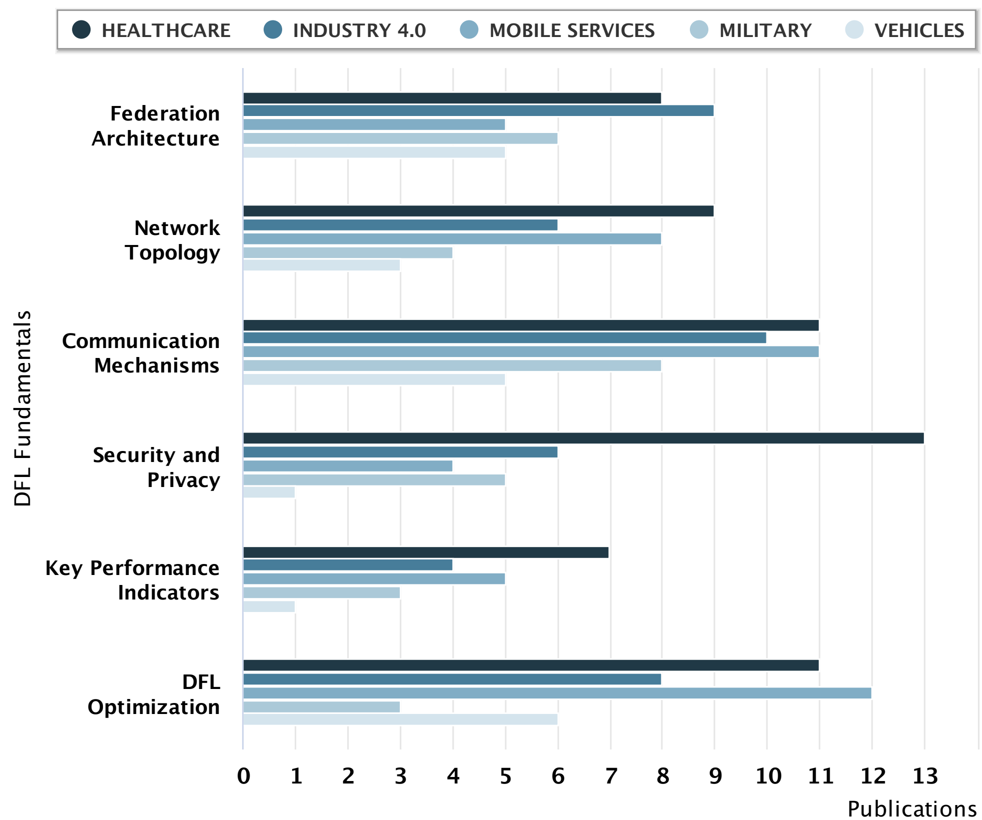
- Paramount in DFL, ensuring user and data protection across all application scenarios

Key Performance Indicators

- Metrics for assessing and enhancing effectiveness and impact

Optimizations

- Refining and improving DFL for optimal performance and efficiency, especially in model parameter exchanges



CONCLUSIONS

Lessons Learned

UNIVERSIDAD DE MURCIA

Delve into the critical lessons learned from extensive DFL research, highlighting existing limitations and areas for further exploration and enhancement

Aggregation Algorithms

- Specific aggregation algorithms for DFL, like FedAvg, are limited in use and often require customization to adapt to diverse federation models

Decentralized Systems

- Limited studies on improving decentralized systems with DFL highlight a need for a deeper analysis of resilience, robustness, and overall security in reducing server dependency

Realistic Federation Benchmarks

- A limited number of solutions provide realistic federation benchmarks. No single benchmark is universally used, and current ones often ignore vital metrics like system efficiency and architecture robustness

Consensus on Frameworks

- No consensus exists in the literature for deploying DFL architectures, with most frameworks adapted to specific validation scenarios. Lack of mature open-source DFL frameworks that are network, node, and data agnostic

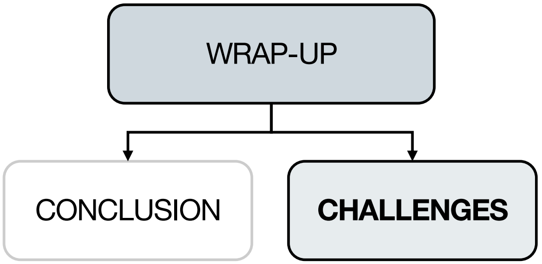
Military and Vehicular Scenarios

- These present complex challenges for deploying DFL solutions, lacking robustness and facing issues like limited bandwidth and high-security requirements

Use of Unsupervised Learning

- A significant gap exists in the literature regarding using unsupervised learning in DFL architectures, with a predominant focus on supervised ML models for classification tasks

Challenges



UNIVERSIDAD DE MURCIA

Confront the pressing challenges in DFL and envision the pathways for future advancements

❑ DFL Fundamentals

→Dynamic participant selection

→Personalized local model learning

→Detecting attacks in DFL scenarios and differentiating based on privacy

❑ DFL Frameworks

→Implement and manage DFL fundamentals and practical application

→Data preprocessing, normalization, and advanced data augmentation techniques

❑ DFL Application Scenarios

→Evaluate DFL in Smart City technologies

→Combine AI, IoT, and DFL for testing tools

Challenge	Future Developments
Fundamentals *[Fund.]*	
Scalability of DFL with increasing participants (!!!)	• Dynamic participant selection • Personalized local model learning
Cybersecurity mechanisms for a secure DFL (!!!)	• Detect attacks in DFL scenarios • Different treatment based on privacy
Trustworthiness among federation participants (!!!)	• Maintain trust policies • Prevent dishonest behavior
Homogeneous node participation (!!)	• Quantization and gradient compression • Use of SDFL
Address participant mobility in DFL scenarios (!!)	• Topology-aware node reconfiguration • Resilient synchronization methods
Study of adversarial attacks (!!)	• Identify the techniques and their impacts • Compare against traditional approaches
Explore the use of Reinforcement Learning (!!)	• Optimize the federated model performance • Improve the selection of participants
DFL standardization efforts (!!)	• Promote comprehensive DFL standards • Involve standard-setting bodies (ISO, IEEE)
5G and 6G technologies for communications (!)	• Network slicing utilization • 5G/6G-integrated edge computing
Frameworks *[Fram.]*	
Modular, scalable, and efficient frameworks (!!!)	• Implement and manage DFL fundamentals • Application in practical scenarios
Heterogeneous datasets in decentralized participants (!!)	• Data preprocessing and normalization • Advanced data augmentation techniques
Dynamic scheduling of federated network (!)	• Adaptable federation architecture • Resilient algorithms
Application scenarios *[Sce.]*	
Exploration of new DFL application scenarios (!!)	• Evaluate DFL in smart city technologies • Combine AI, IoT, and DFL for testing tools

! low importance, !! high importance, !!! critical

CHALLENGES

Tutorial Overview

UNIVERSIDAD DE MURCIA

- ❑ Comprehensive analysis of DFL evolution
- ❑ Outline of fundamentals and comparison with traditional architectures
- ❑ Exploration of frameworks, application scenarios, and challenges

Addressed the following Research Questions:

❑ RQ1. What are the fundamental aspects of DFL?
- Clear delineation from CFL
- Introduction to detailed DFL taxonomy
- Covers architectures, topologies, communication, and security

❑ RQ2. What DFL frameworks exist, and what fundamentals do they provide?
- Discussion on mature and nascent DFL frameworks
- Highlight of DFL's robust foundational elements

❑ RQ3. Which are the main characteristics of the most relevant scenarios of DFL?
- Analysis of key application areas: healthcare, mobile services, Industry 4.0
- Insights into decentralized cross-device architectures

❑ RQ4. What trends, lessons learned, and challenges have emerged in DFL?
- Exploration of advanced federation architectures and topologies
- Detailing limitations and prospective research areas: heterogeneous datasets, cyberattacks, 5G/6G communications

FUNDAMENTALS

Our Research Projects about DFL

UNIVERSIDAD DE MURCIA

DEFENDIS: Decentralized Federated Learning for IOT Device Identification and Security

FEDERAL OFFICE FOR DEFENCE PROCUREMENT ARMASUISSE

INSTITUTO NACIONAL DE CIBERSEGURIDAD

DEFENDER: DEtecting Feasible cybErattacks to iNcrease cybersecurity and cyberDEfence in experimental laboratoRies

INCIBE (Spanish CERT)

ROBUST-6G: smaRt, AutOmated, and ReliaBle SecUrity Service PlaTform for 6G

Horizon Europe Framework Programme (HORIZON)-SNS-2023

To be started in December 2023

More information at https://cyberdatalab.um.es

Cyber Data Lab

ECAI 2023 Tutorial – Kraków, Poland

Thank you for your attention!

Any questions?

SCAN ME

Alberto Huertas Celdrán[1], Enrique Tomás Martínez Beltrán[2], Pedro Miguel Sánchez Sánchez[2], Gérôme Bovet[3], Gregorio Martínez Pérez[2], and Burkhard Stiller[1]

[1]Communication Systems Group CSG, Department of Informatics IfI, University of Zurich UZH, Switzerland

[2]Department of Information and Communications Engineering, University of Murcia, Spain

[3]Cyber-Defence Campus within armasuisse Science & Technology, Thun, Switzerland

UNIVERSIDAD DE MURCIA

Cyber Data Lab

Universität Zürich[UZH]

CSG Communication Systems Group

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

armasuisse
Science and Technology