IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

A Cross-evaluation approach for Reputation-aware Model Weighting

Filtering Contributions in Federated Learning for Intrusion Detection

Journées non-thématiques du GDR RSD, Lyon, 27/01/2023

Authors :
Léo Lavaur
Pierre-Marie Lechevalier

Supervisors :
Fabien Autrel
Yann Busnel
Hélène Le Bouder
Romaric Ludinard
Marc-Oliver Pahl
Géraldine Texier

CHAIR OF CYBER CNI
Critical National Infrastructures

Context: Intrusion Detection

- ➔ Different families: misuse detection, anomaly detection, specification-based…
- ➔ Machine learning (ML) and deep learning (DL) often used for their performance;
  - Eg., auto-encoder (AE) can be used for anomaly detection.
- ➔ DL need a lot of data to be efficient, training them locally is a challenge;
  - Eg., for AE, anything not known is an anomaly → higher false-positive rate.

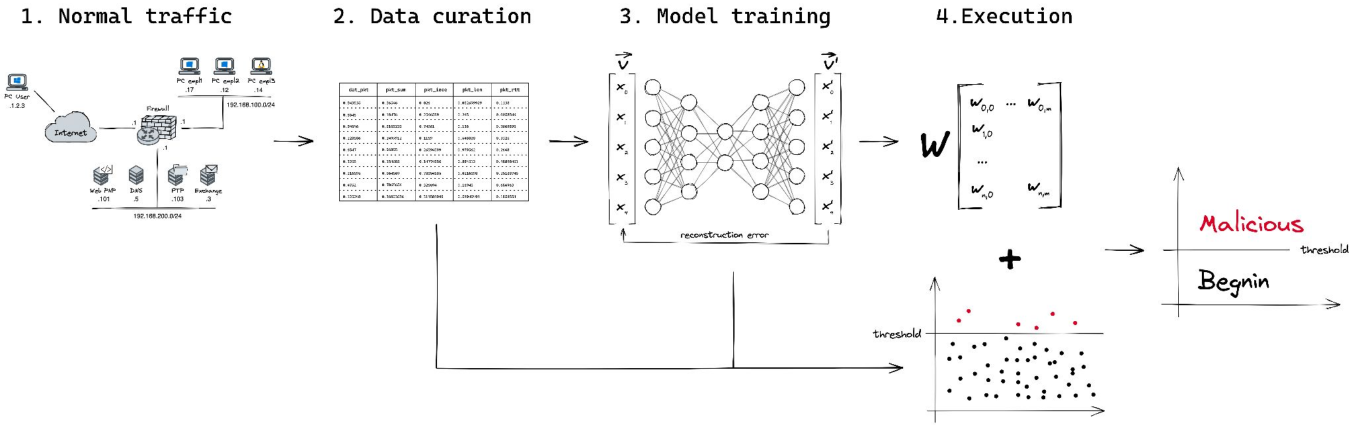


Fig 1: Typical AE workflow for IDS

Collaborative Intrusion Detection

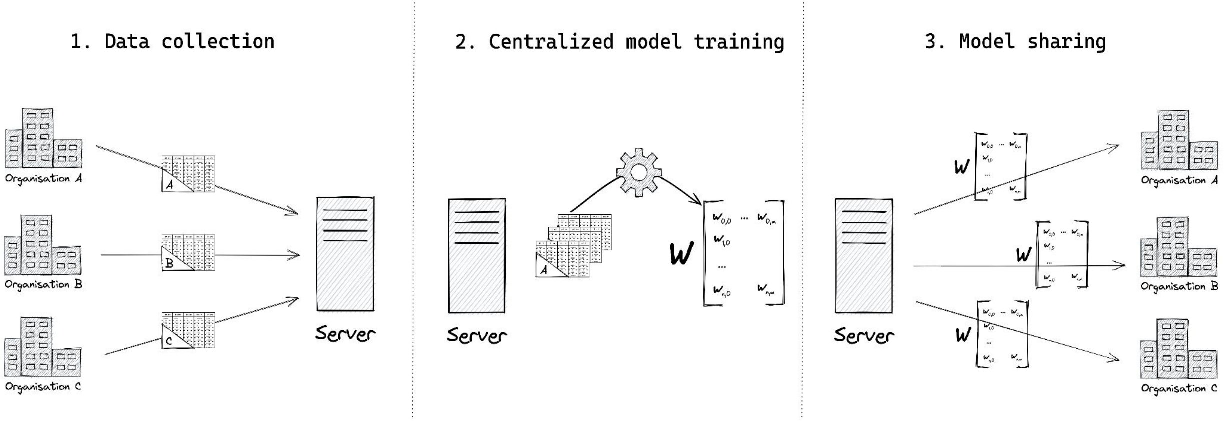


Fig 2: Typical CIDS (Collaborative Intrusion Detection System) workflow

Objective

➔ Consolidate normal behavior modeling by sharing knowledge with other participants;

Challenges

➔ Security & Privacy – *eg.* revealing internals, poisoning, trust [1];
➔ Availability – *eg.* single point of failure in centralized systems [2];
➔ Resources – *eg.* high bandwidth consumption when sharing data [3];

[1] C. Fung et al. "Trust Management for Host-Based Collaborative Intrusion Detection." In Managing Large-Scale Service Deployment, 2008.
[2] S. Rathore, et al., "BlockSecIoT-Net: Blockchain-based decentralized security architecture for IoT network," *Journal of Network and Computer Applications*, 2019
[3] B. McMahan, et al., "Communication-efficient learning of deep networks from decentralized data", *20th International conference on artificial intelligence and statistics*, 2017

Federated Learning as a Collaborative Learning System

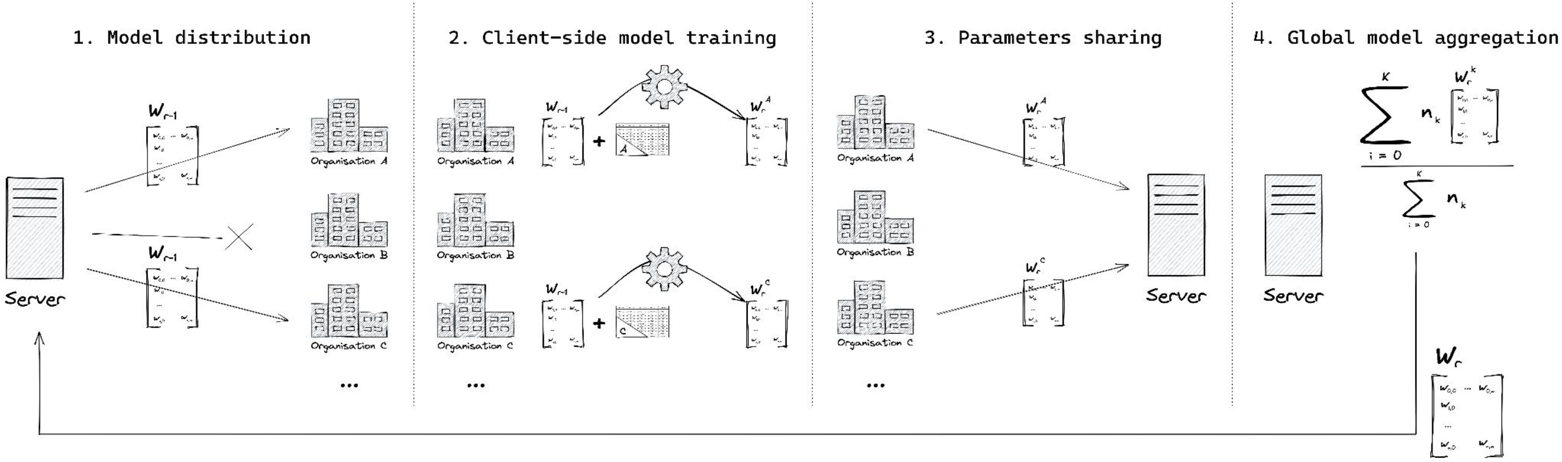


Fig 3: Typical Horizontal Federated Learning workflow for CIDS

Challenges [4]

➔ Heterogeneity – unsuitable global aggregation when participants are too different.

➔ Trust – assessing peer contributions.

[4] L. Lavaur, et al., "The Evolution of Federated Learning-Based Intrusion Detection and Mitigation: A Survey", *IEEE Transactions on Network and Service Management*, 2022

Our approach

Objective: Mitigate the impact of *bad* contributions to the local models;

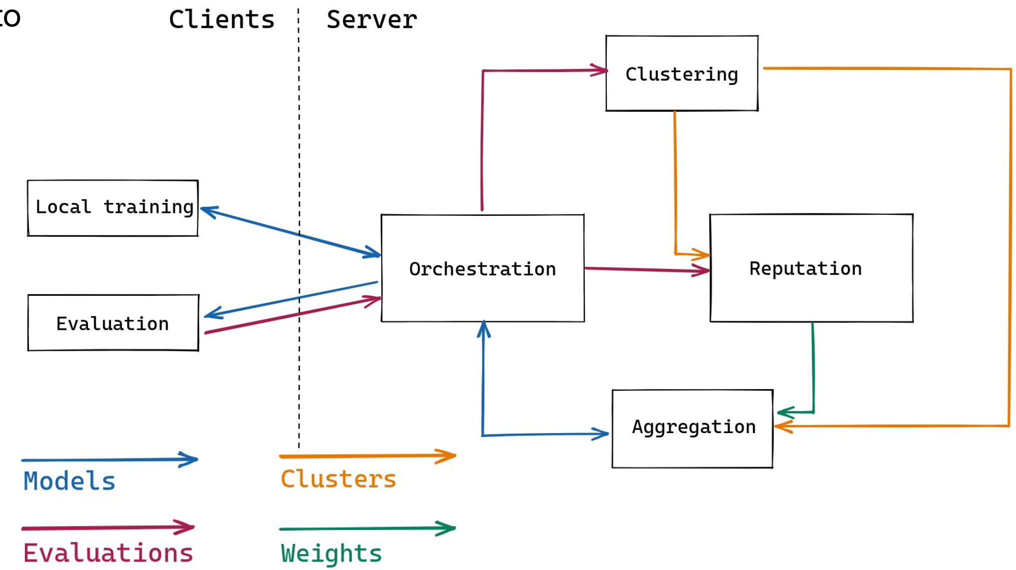


Fig 4: Proposed architecture

Our approach

Objective: Mitigate the impact of *bad* contributions to the local models.

➔ How to evaluate models in highly heterogeneous settings?

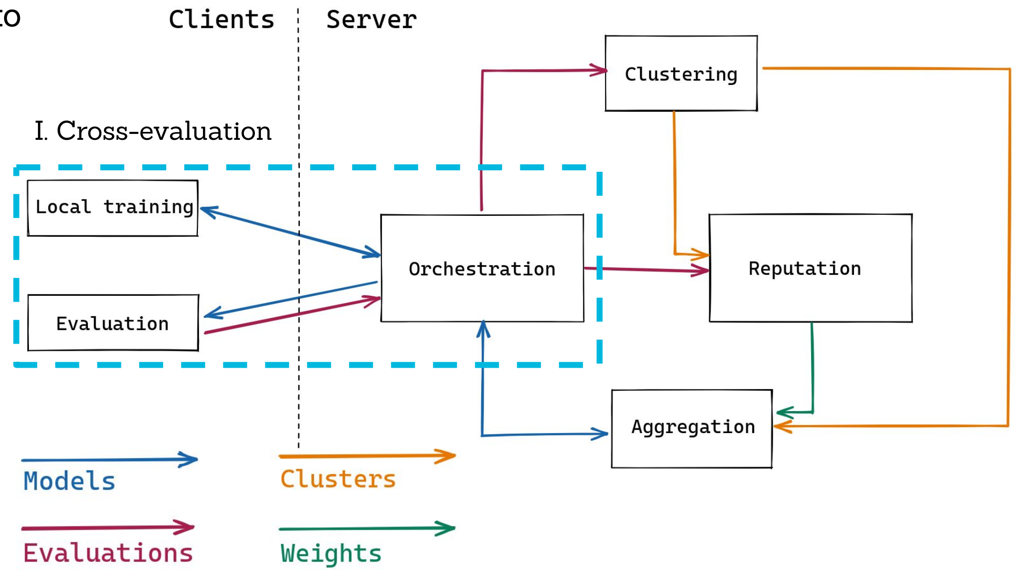


Fig 4: Proposed architecture

Our approach

Objective: Mitigate the impact of *bad* contributions to the local models;

- ➔ How to evaluate models in highly heterogeneous settings?
- ➔ How to set aside dissimilar participants?

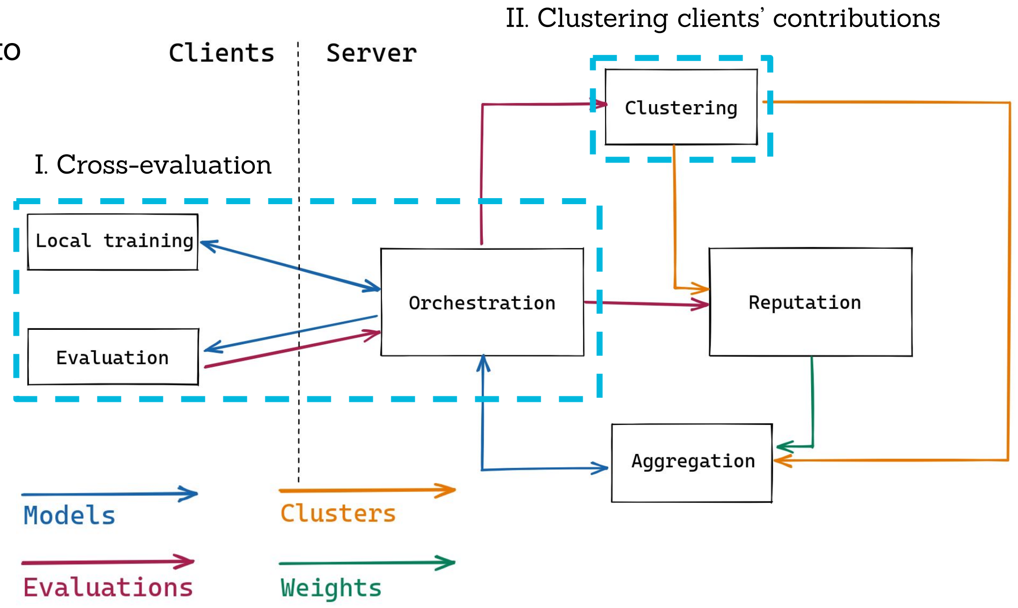


Fig 4: Proposed architecture

Our approach

Objective: Mitigate the impact of *bad* contributions to the local models;

- ➔ How to evaluate models in highly heterogeneous settings?
- ➔ How to set aside dissimilar participants?
- ➔ How to identify and discard similar but negative behaviors?

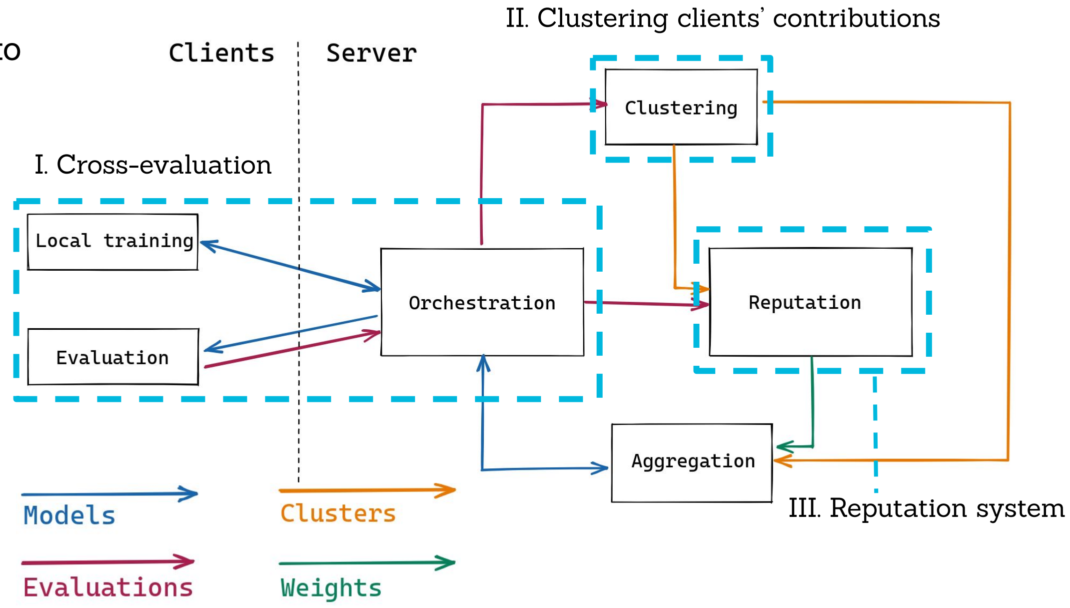
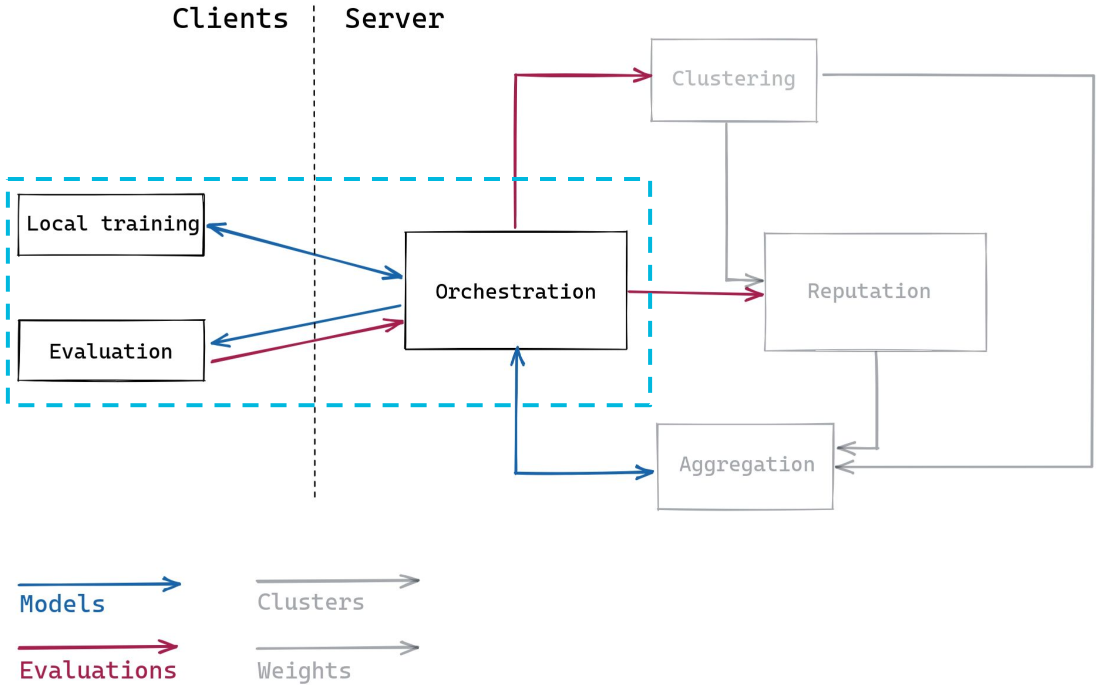


Fig 4: Proposed architecture

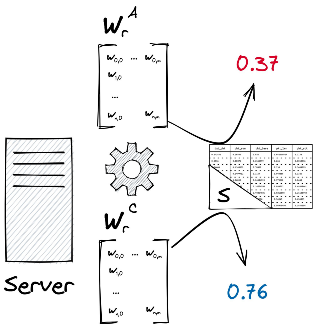
I. Assessing Contributions with Cross-Evaluation



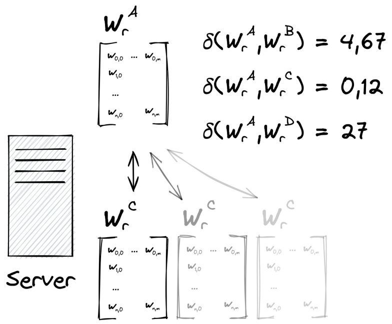
IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Methods for filtering contributions

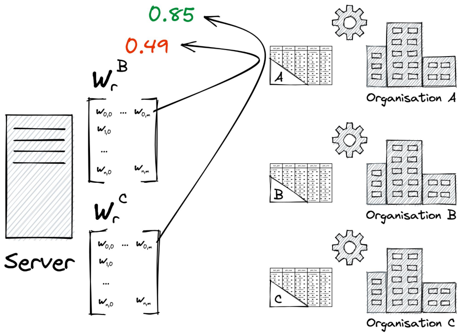
Server-side evaluation [5]



Server-side model comparison [6]



Client-side evaluation [7]



- ➔ only applicable in IID settings
- ➔ single source of truth

- ➔ less related to client data
- ➔ more appropriated for high-dimensional features

- ➔ high cost in cross-device settings

[5] J. Zhou, et al., "A Differentially Private Federated Learning Model against Poisoning Attacks in Edge Computing", 2022

[6] C. Briggs, et al., "Federated learning with hierarchical clustering of local updates to improve training on non-IID data", 2020

[7] L. Zhao, et al., "Shielding Collaborative Learning: Mitigating Poisoning Attacks through Client-Side Detection", 2020

Cross-evaluation workflow

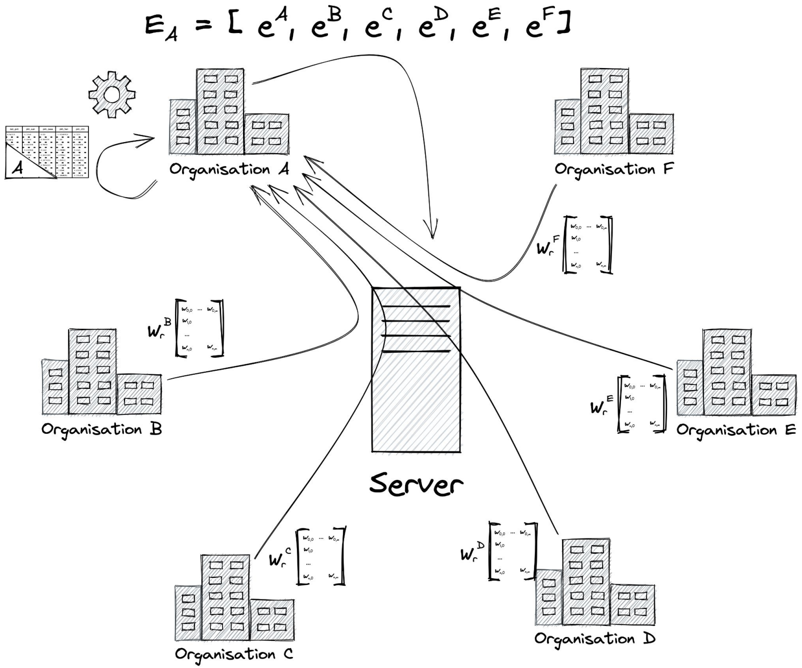


Fig 5: Cross-evaluation

Advantages

- ➔ Central server doesn't need prior knowledge.
- ➔ Evaluates how each model fits the data (*eg*., accuracy).
- ➔ Exhaustive overview of the entire system at round *r*.
- ➔ Keeps the subjectivity of the evaluations.

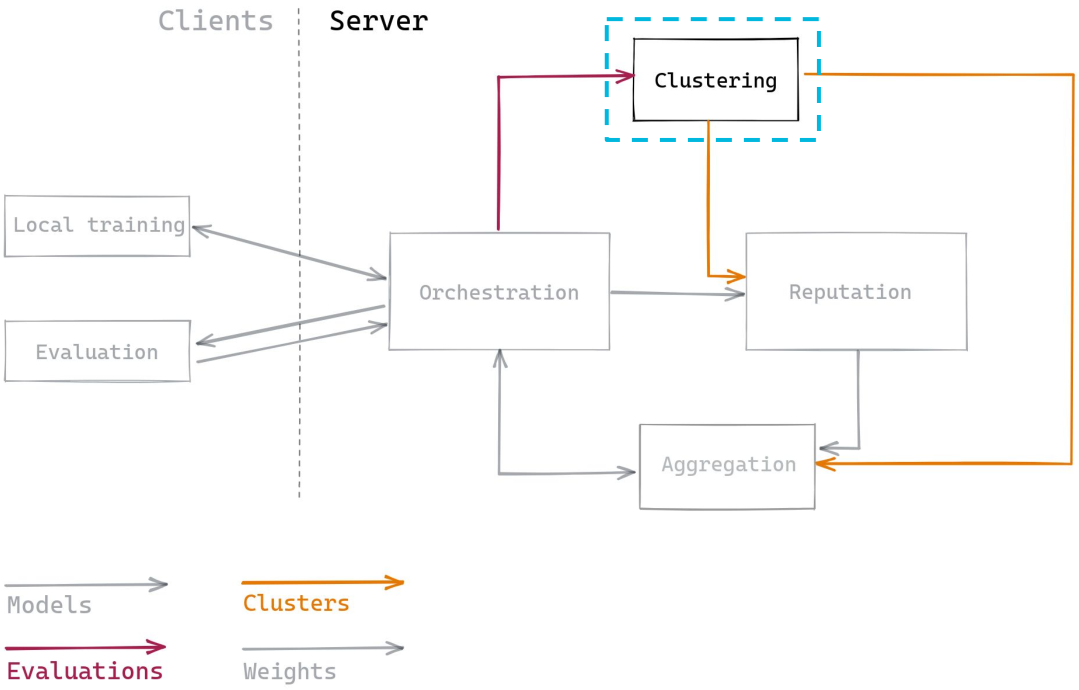
Drawbacks

- ➔ Expansive in communication and computation.
- ➔ Doesn't scale well.

But…

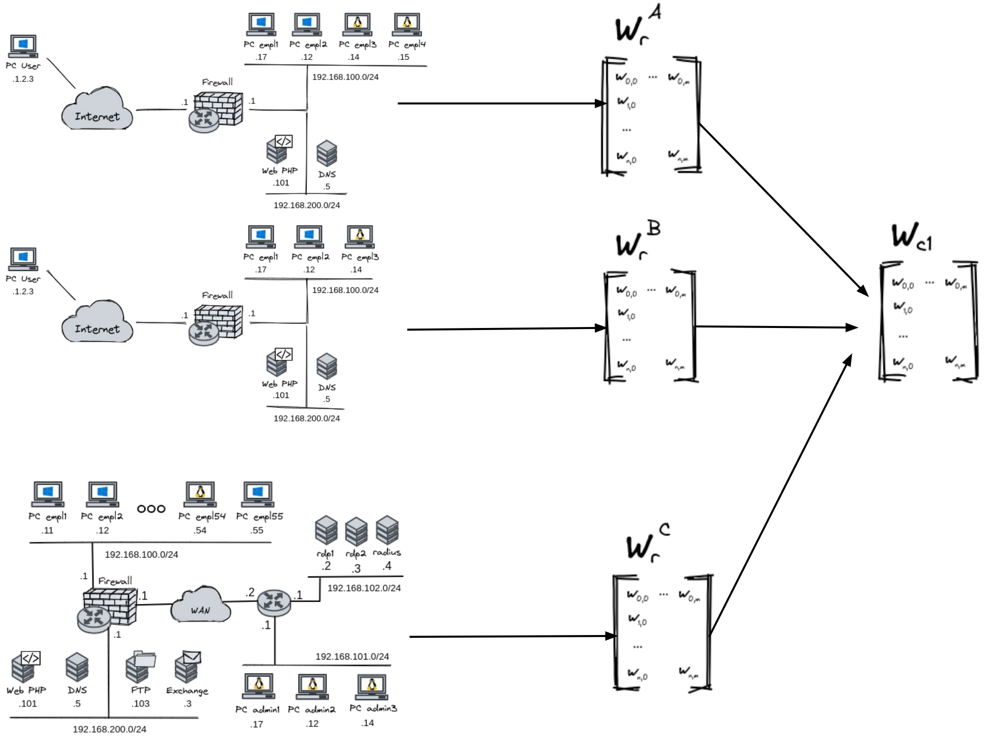
- ➔ Cross-silo: Few clients, with reasonable computing capacity.
- ➔ Slow workflow: long time between rounds.

II. Fighting Heterogeneity with Clustering



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Merging heterogeneous contributions



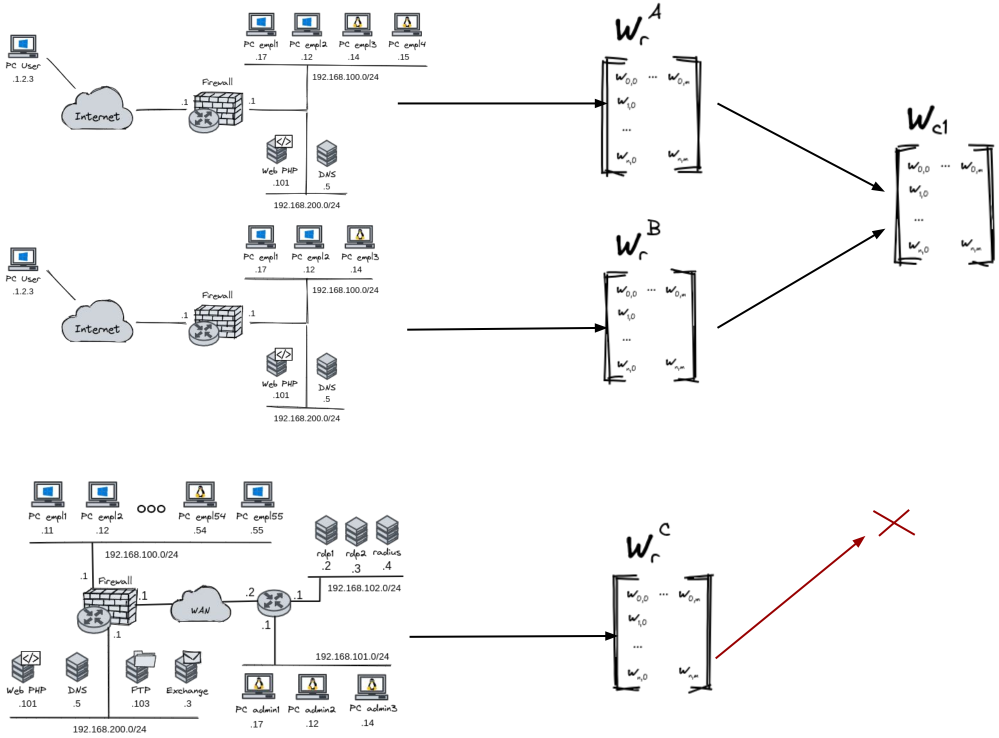
Global model can either:

➔ lose accuracy by trying to fit all participants [8];
➔ dismiss some participants [9].

[8] Cai, et al. "Cluster-Based Federated Learning Framework for Intrusion Detection." In 2022 IEEE 13th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP)

[9] Blanchard, et al. "Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent." Advances in Neural Information Processing Systems 30 2017.

Merging heterogeneous contributions



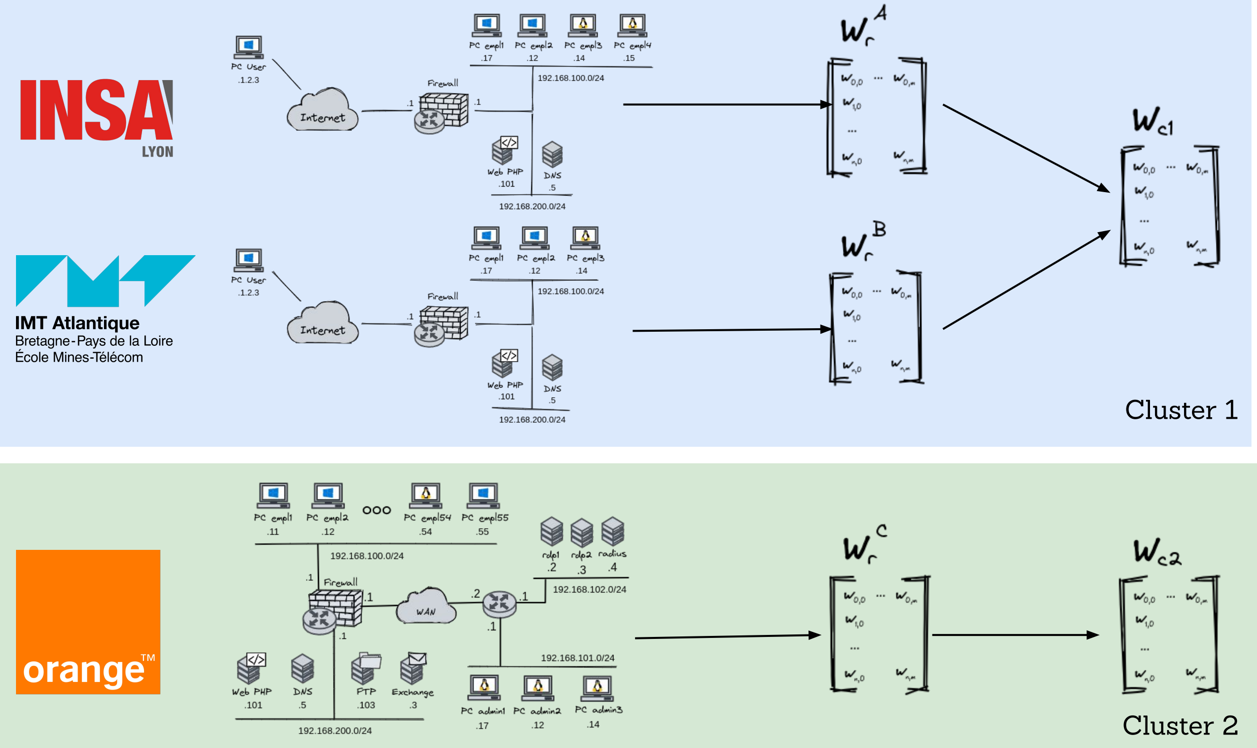
Global model can either:

➔ lose accuracy by trying to fit all participants [8];
➔ dismiss some participants [9].

[8] Cai, et al. "Cluster-Based Federated Learning Framework for Intrusion Detection." In 2022 IEEE 13th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP)

[9] Blanchard, et al. "Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent." Advances in Neural Information Processing Systems 30 2017.

Merging heterogeneous contributions



Clustering goal:

- ➔ regroup similar participants together;
- ➔ create an aggregated model per cluster.

Existing clustering approaches for federated learning

Clustering data source

- ➔ Participants models [10].
- ➔ Cross evaluation results.

Clustering for federated learning

- ➔ Dynamic split and merge [11].
- ➔ Hierarchical clustering [10].

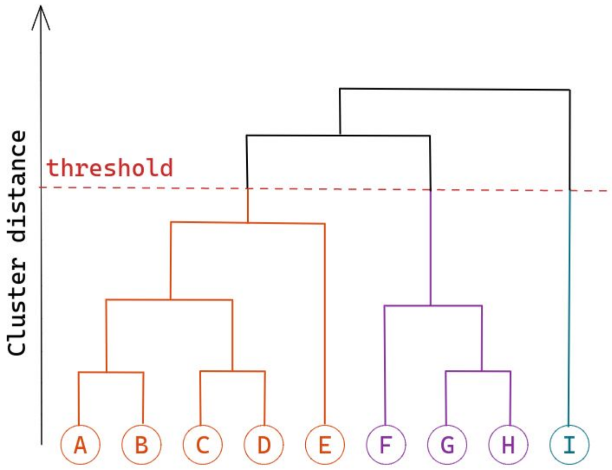
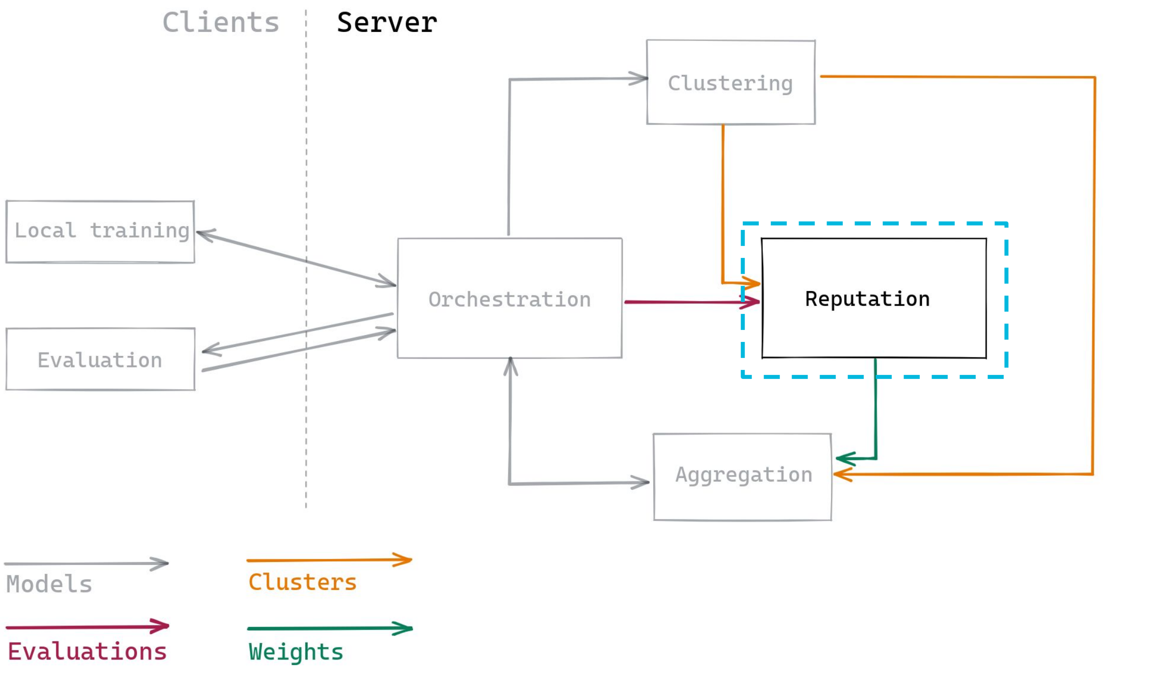


Fig 6: Hierarchical clustering

[10] Briggs, et al. "Federated Learning with Hierarchical Clustering of Local Updates to Improve Training on Non-IID Data." In 2020 International Joint Conference on Neural Networks (IJCNN),2020
[11] Chen, et al. "Zero Knowledge Clustering Based Adversarial Mitigation in Heterogeneous Federated Learning." IEEE Transactions on Network Science and Engineering 8, no. 2, 2021

III. Ensuring Quality Contributions with Reputation



IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Motivation for reputation

Objectives reminder:

- ➔ weight participants contributions;
- ➔ detect change that occur over time [12].

Definition [13]

- ➔ **Long-lived entities** that inspire an expectation of **future interaction**;
- ➔ **Capture and distribution of feedback** about current interactions (such information must **be visible** in the future); and
- ➔ Use of feedback to guide trust decisions.

[12] Karimireddy, et al. "Learning from History for Byzantine Robust Optimization." In Proceedings of the 38th International Conference on Machine Learning, PMLR, 2021.
[13] Resnick, Paul, et al. "Reputation systems." Communications of the ACM 43.12 (2000): 45-48.

Evaluation weighting

Similarity

- ➔ Cluster outliers shouldn't have too much impact on evaluation.
- ➔ Ponderate client evaluation using their similarity to other cluster members [14].

Historical considerations

- ➔ No specific constraints.
- ➔ Exponential decay: older results fade away.

Dirichlet distribution [15,16]

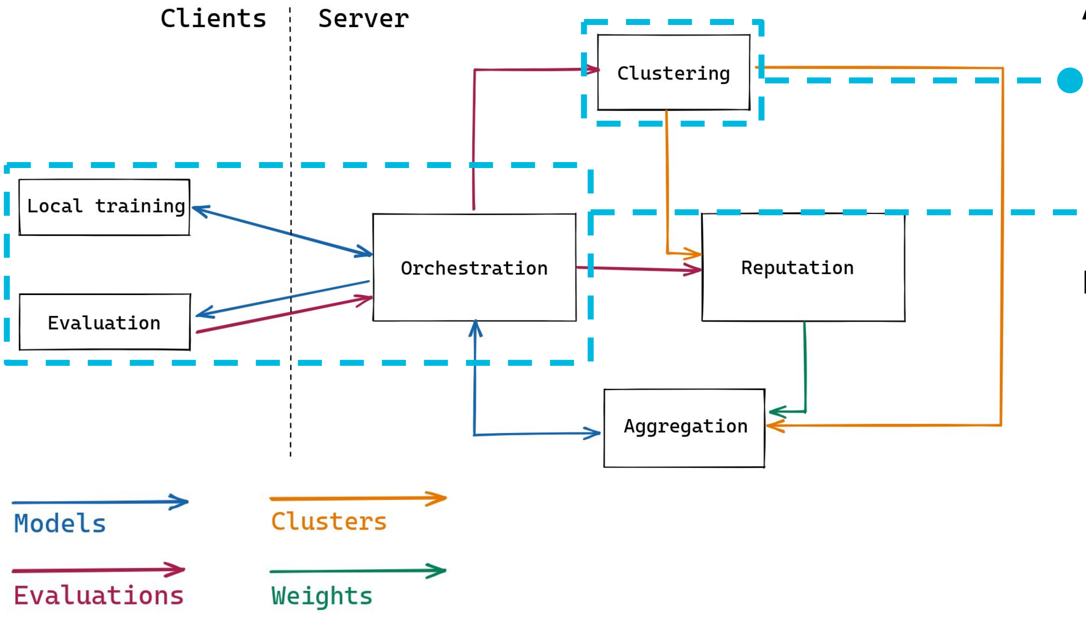
- ➔ Multinomial distribution.
- ➔ Allow discretization of cross evaluation results.

[14] Li Xiong, et al. "PeerTrust: Supporting Reputation-Based Trust for Peer-to-Peer Electronic Communities." IEEE Transactions on Knowledge and Data Engineering 16, no. 07, 2004

[15] Josang, et al. "Dirichlet Reputation Systems." In The Second International Conference on Availability, Reliability and Security (ARES'07) 2007.

[16] Fung, et al. "Dirichlet-Based Trust Management for Effective Collaborative Intrusion Detection Networks." IEEE Transactions on Network and Service Management 8, no. 2, 2011

Conclusion



Achievements

➔ Preliminary clustering validation:
- done at the end of the first round of FL;
- 8/10 clients are in the correct cluster;
- empirical demonstration of cross-evaluation.

Future works

➔ Chain the functional blocks:
- implement clustering and cross-evaluation; into the Flower framework;
- test the reputation system.

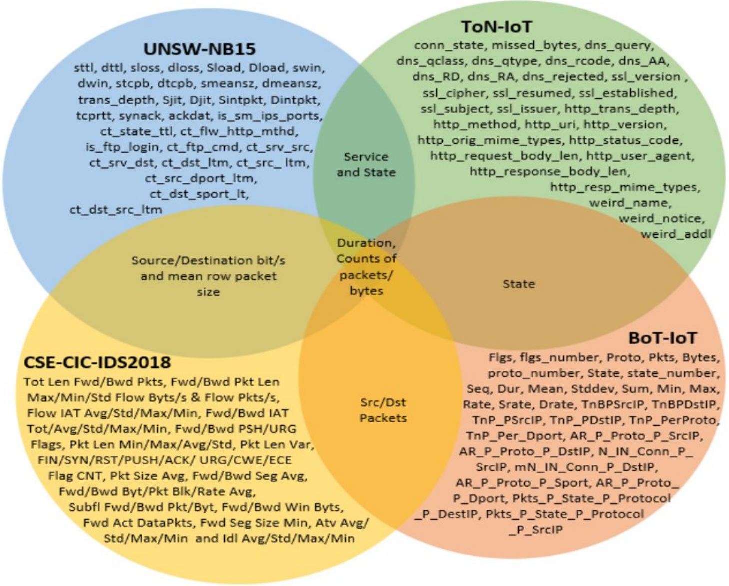
➔ Extensive evaluation:
- of each atomic block;
- of the chained system.

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

CHAIR OF CYBERCNI
Critical National Infrastructures

Annex 1: Datasets and experimental platform

- "standardized IDS datasets" [17] (UNSW-NB15, BoT-IoT, ToN-IoT, and CSE-CIC-IDS2018)
  - 4 datasets = 4 use cases where clients are distributed in the use cases
  - normalized features among all datasets
- Flower FL Framework (https://flower.dev)



[17] M. Sarhan, S. Layeghy, and M. Portmann, "Towards a Standard Feature Set for Network Intrusion Detection System Datasets," arXiv.org, 2021

Annex 2: Preliminary clustering results

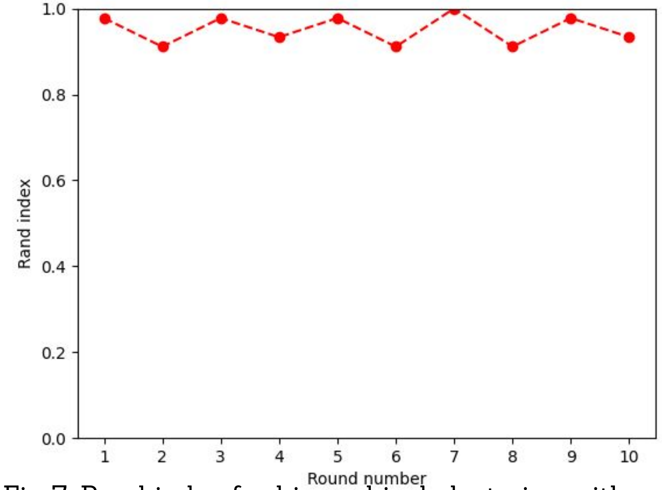


Fig 7: Rand index for hierarchical clustering with mean cluster interdistance as threshold

$$RI = \frac{TP + TN}{TP + FP + FN + TN}$$