Aeronautics Institute of Technology
Laboratory of Command and Control and Cyber-security (C2DC)

Improving detection of scanning attacks on heterogeneous networks with Federated Learning

Gustavo de Carvalho Bertoli Lourenço Alves Pereira Junior Osamu Saotome

Agenda

- Motivation
- Dataset
- Data Preprocessing
- Federated Learning Setup
- Results
- Conclusions

Motivation

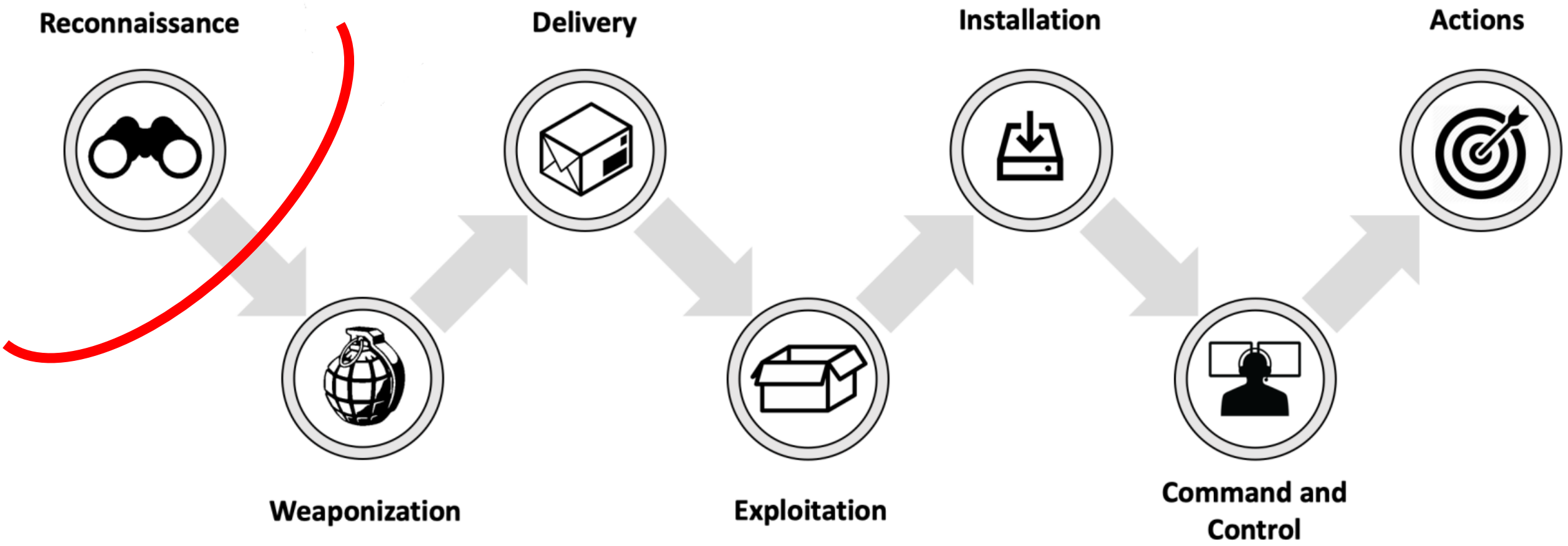
- The amount of network traffic requires automated ways for monitoring threats

- Machine learning-based Network Intrusion Detection Systems are a well-known research field

Motivation > Federated Learning

- Data privacy
- Efficiency
- Capability to generalize for multiple environments

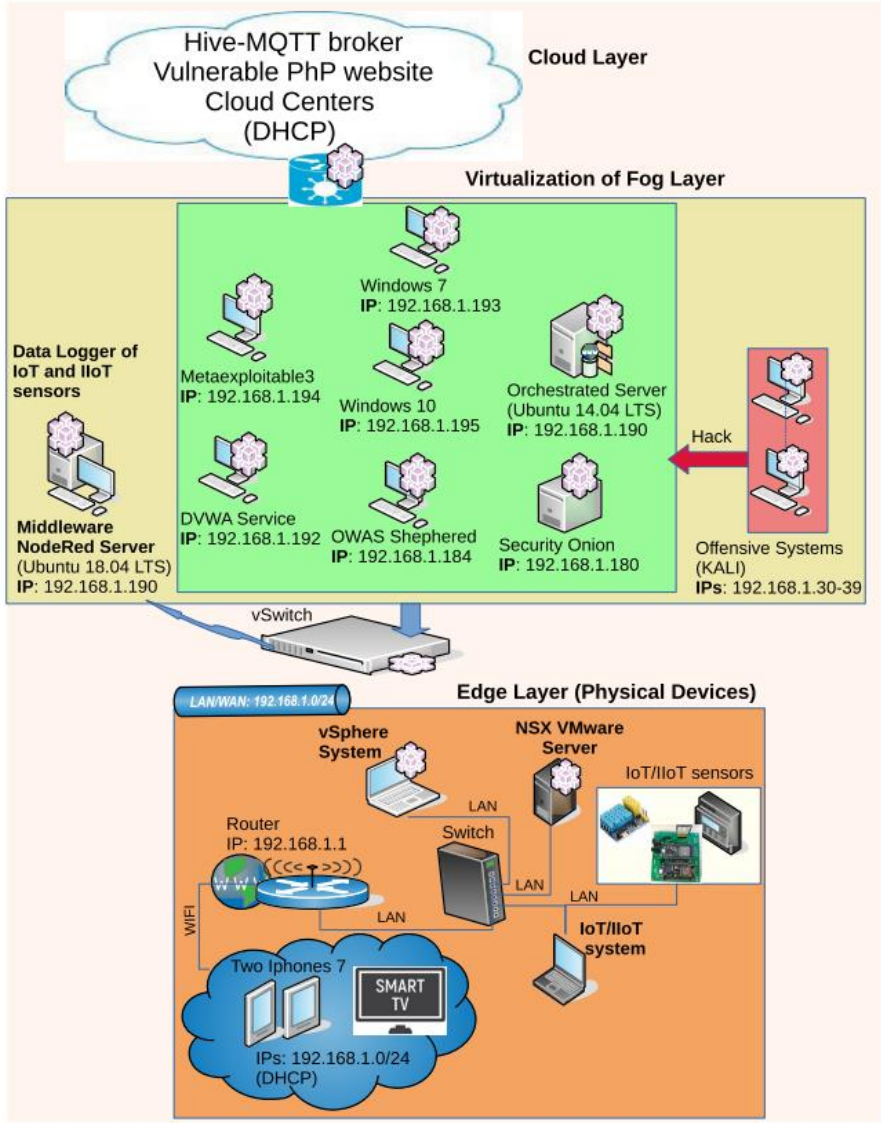
Motivation > Scanning Attacks



Dataset

TON_IoT dataset

- Recent dataset
- Heterogeneous network
- Multitude of attacks
- NetFlow Features

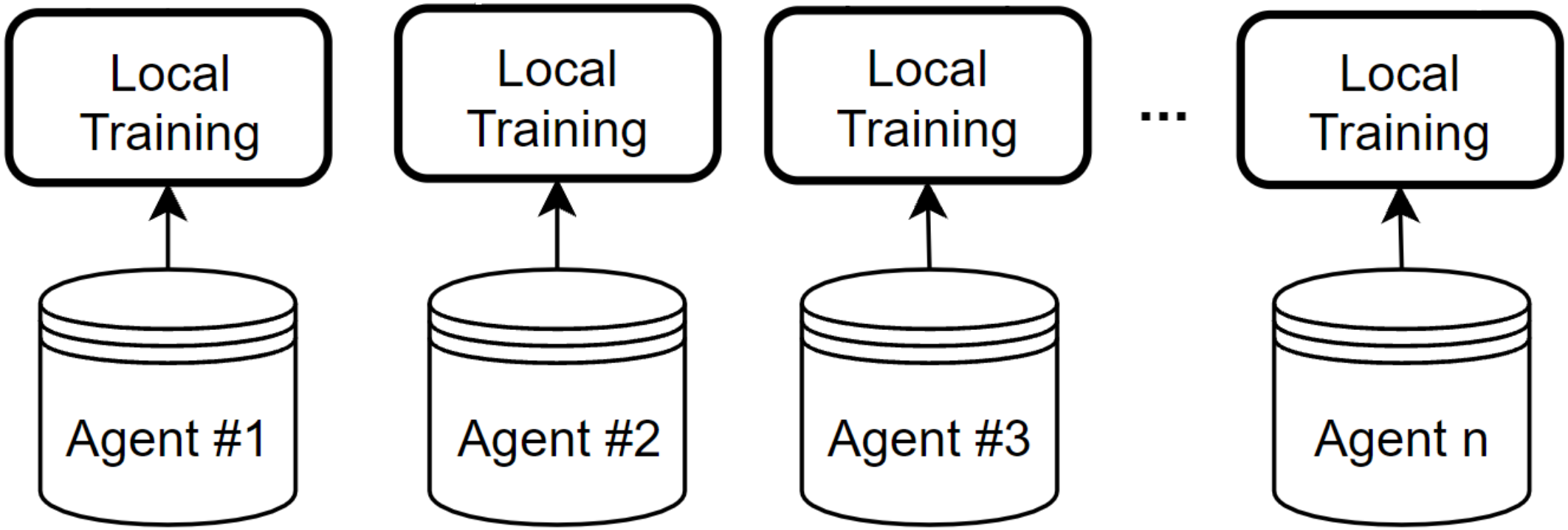


Source: Alsaedi et al, TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems (IEEE Access 2020)

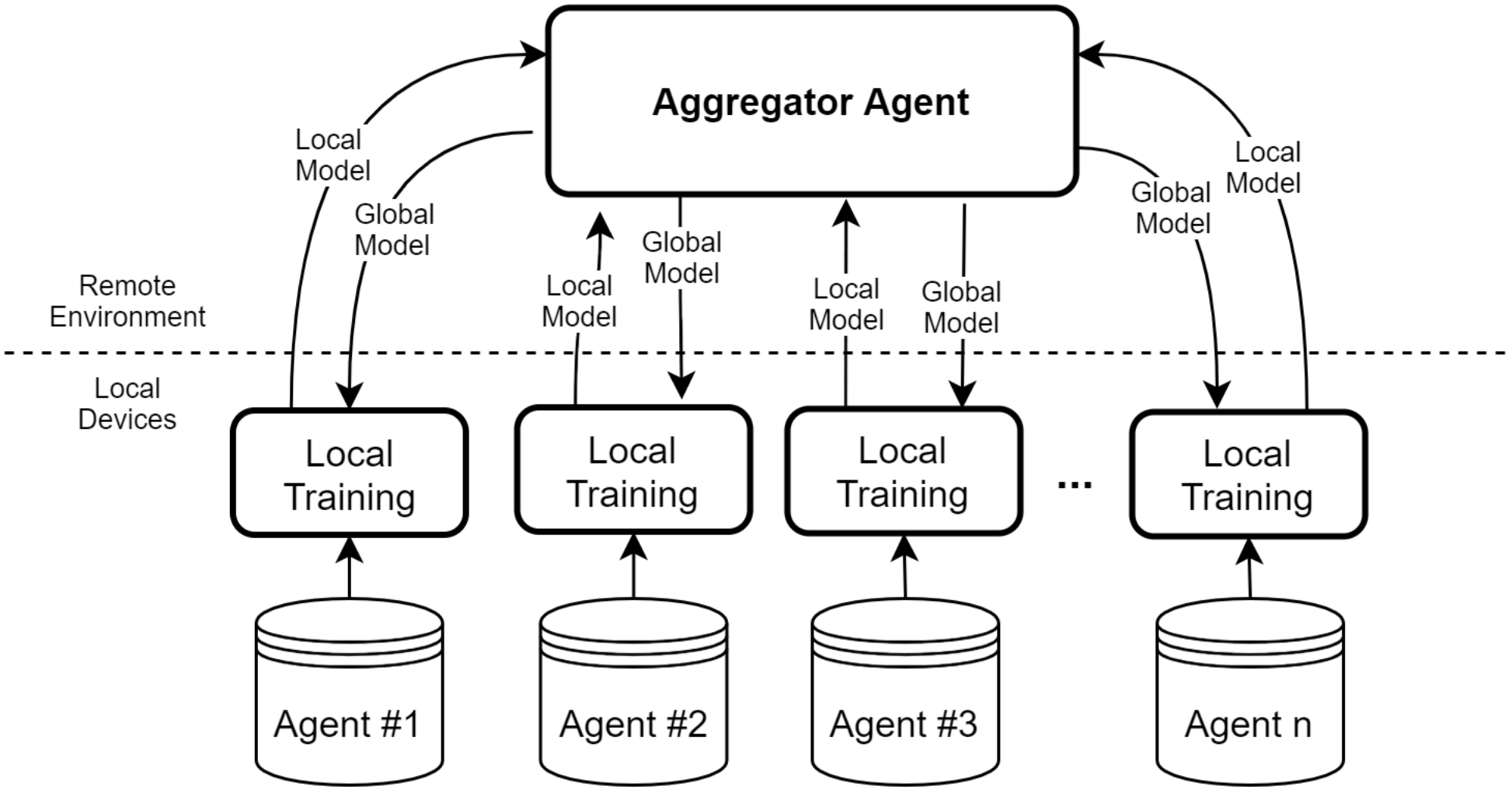
Data Preprocessing

- Keep just scanning and benign traffic (5.7GB → 3.3GB)
- Keep just targets with more than 3.000 traces, reduces to 229 targets
- Use of domain knowledge to keep targets with fair amount of benign traffic (229 → 13)
- These 13 agents results in 13 data silos

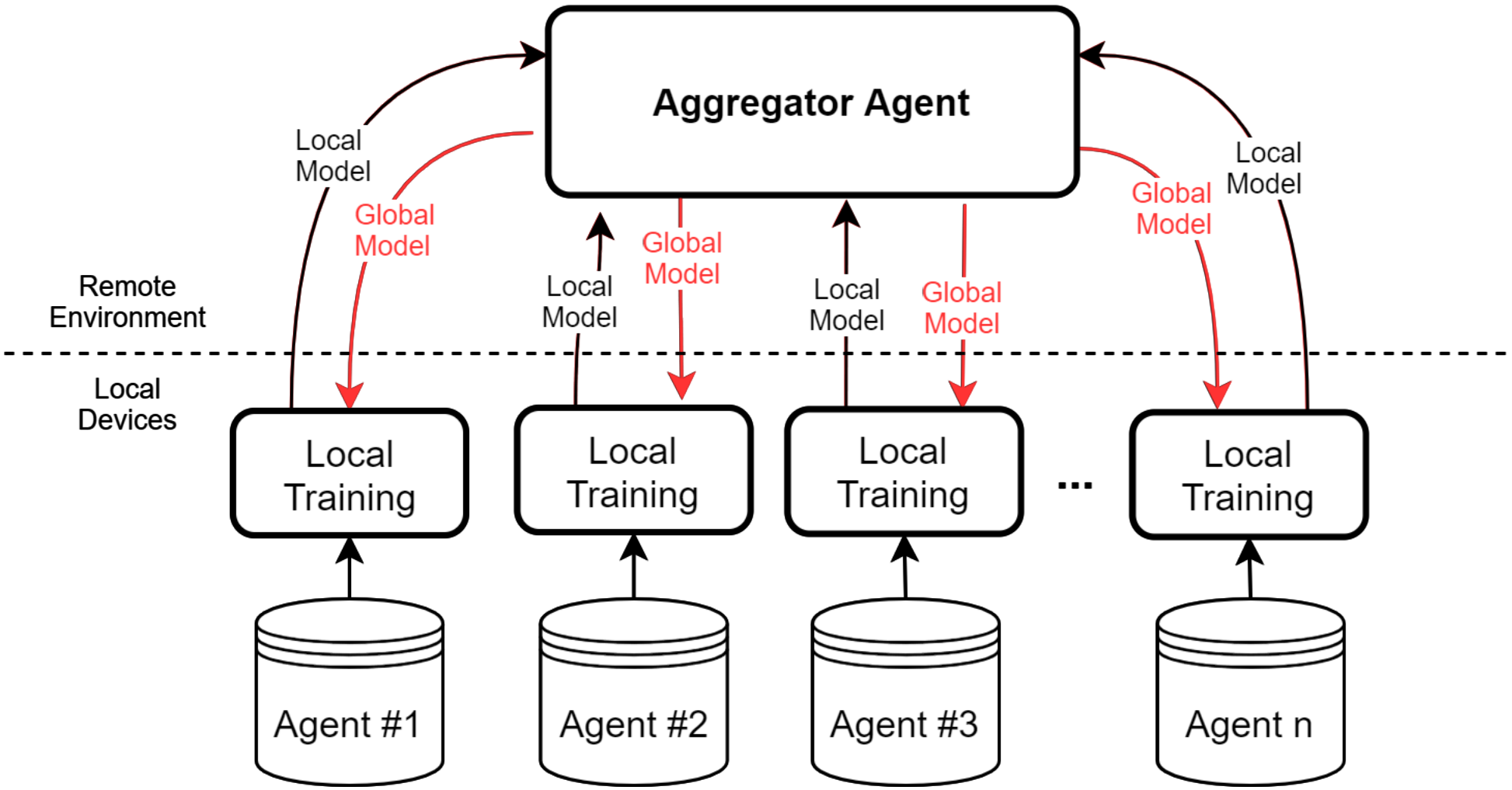
Federated Learning Setup



Federated Learning Setup



Federated Learning Setup



Federated Learning Setup

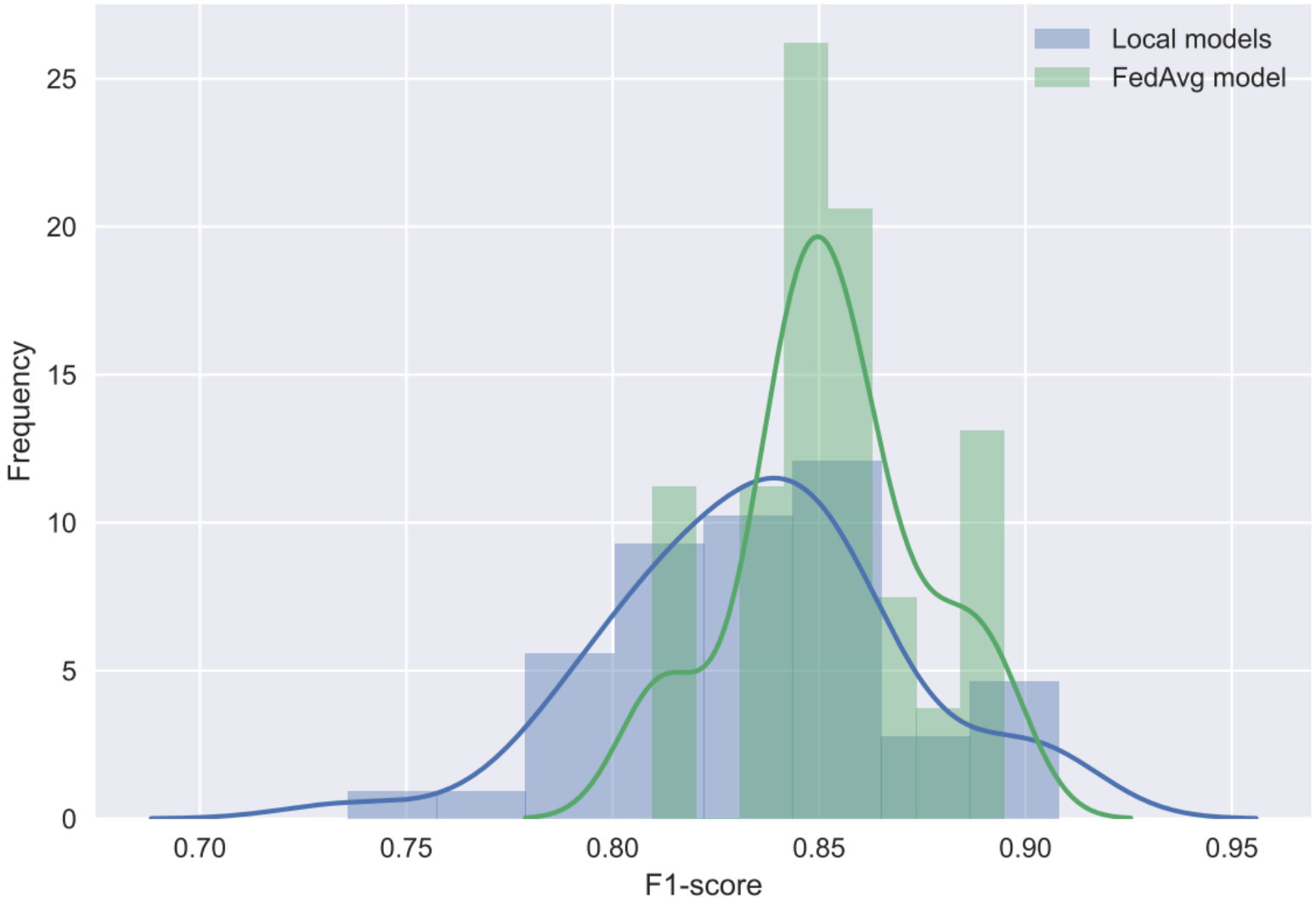
- We use the logistic regression through the stochastic gradient descent sharing the weights of each agent
- We focus on the F1-score metric to address the imbalanceness of the dataset
- Using the FedAvg strategy weighting by the imbalanceness of each participant

Results

The reported results were obtained with the following conditions:

- 50 rounds
- 10 epochs
- 100 examples as batch size
- Learning Rate of 0.15
- Using all 13 agents / data silos

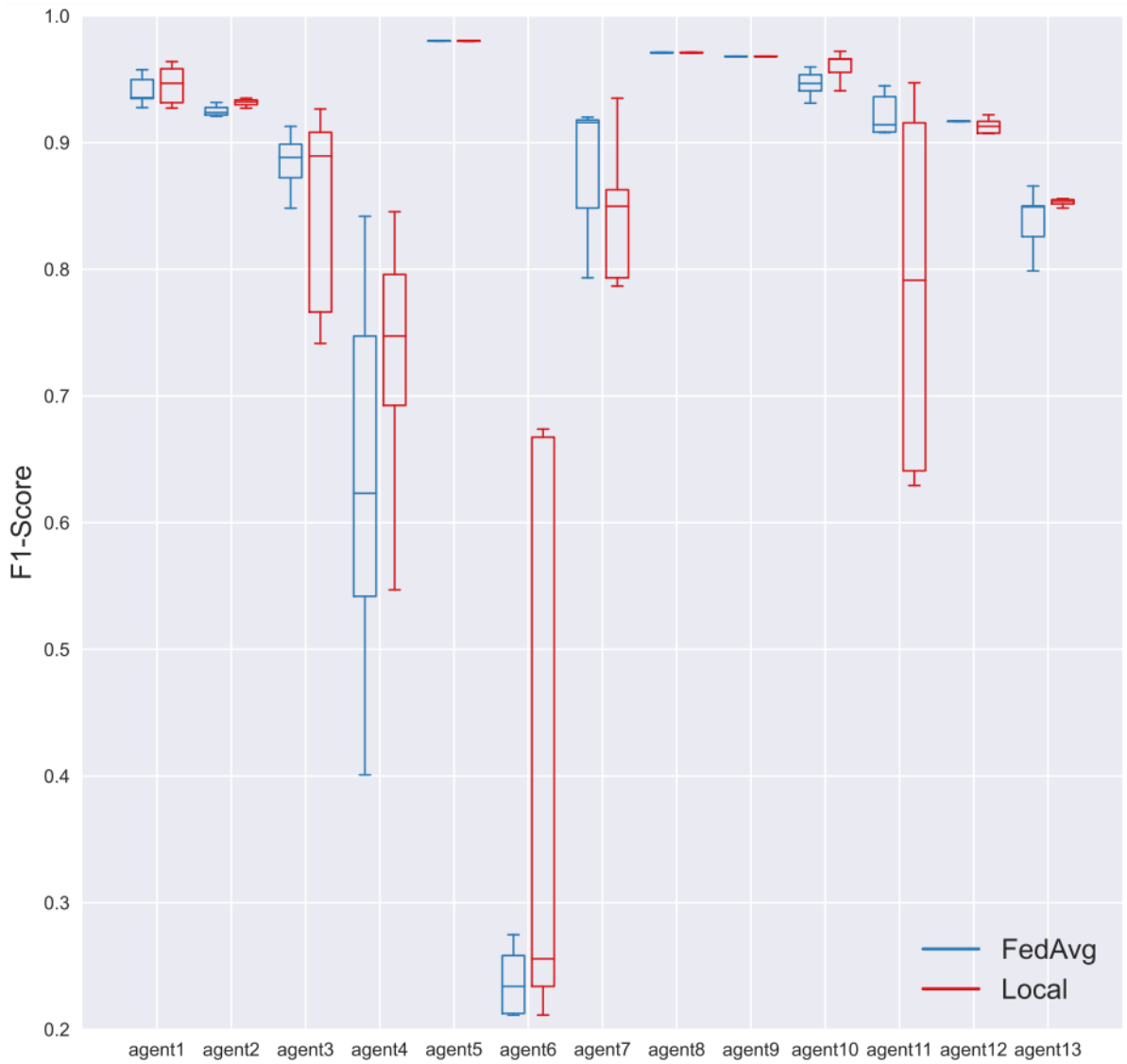
Results



Average F1-Score

- Local models: 0.84
- Federated Learning: 0.85
- Centralized: 0.93

Results



Federated Learning presents less variance performance.

Better learning and attack detection.

Conclusions and Next Steps

- We presented an improved scanning attack detection in comparison with traditional approach
- There are non-IID characteristics that were not explored in this paper
- Further investigates the federated learning application to network intrusion detection using various datasets

Thank you!

Gustavo Bertoli
✉ bertoli [at] ita.br

Repository:
https://github.com/c2dc/wain2021