Federated Learning as an Enabler for Collaborative Security between not Fully-Trusting Distributed Parties

2022/11/15 — C&ESAR

Léo LAVAUR (Chaire CNI) & Benjamin COSTE (Airbus Cyber)

Marc-Oliver PAHL, Yann BUSNEL, Fabien AUTREL

speaker series
PhD schools
MOOCs

TABLE OF CONTENT

1. A COLLABORATIVE SECURITY APPROACH

Context, motivation and research question

CONTEXT: SECURITY MONITORING

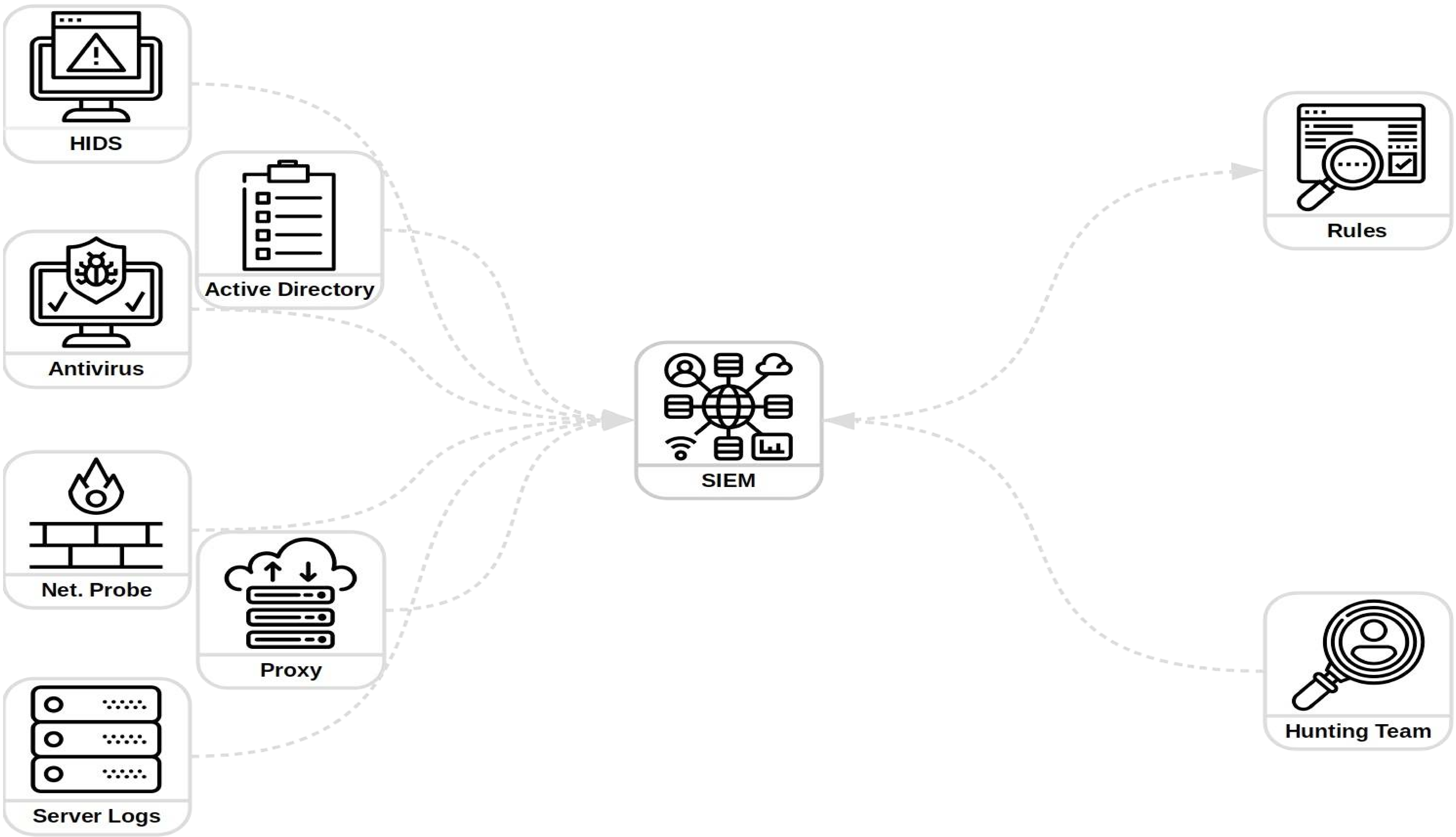


Figure 1: Security monitoring and data collection

Icons by Flaticon

SECURITY MONITORING: A DATASCIENCE PERSPECTIVE

Datascience can help hunting workflow automation [19]

- Structuring data processing allows automating some hunts;
- Clustering to reduce the number of alerts to process manually;
- Anomaly detection to prioritize investigations and limit the time needed to fine tune detection conditions.

Does it scale?

COLLABORATIVE DETECTION PROBLEM

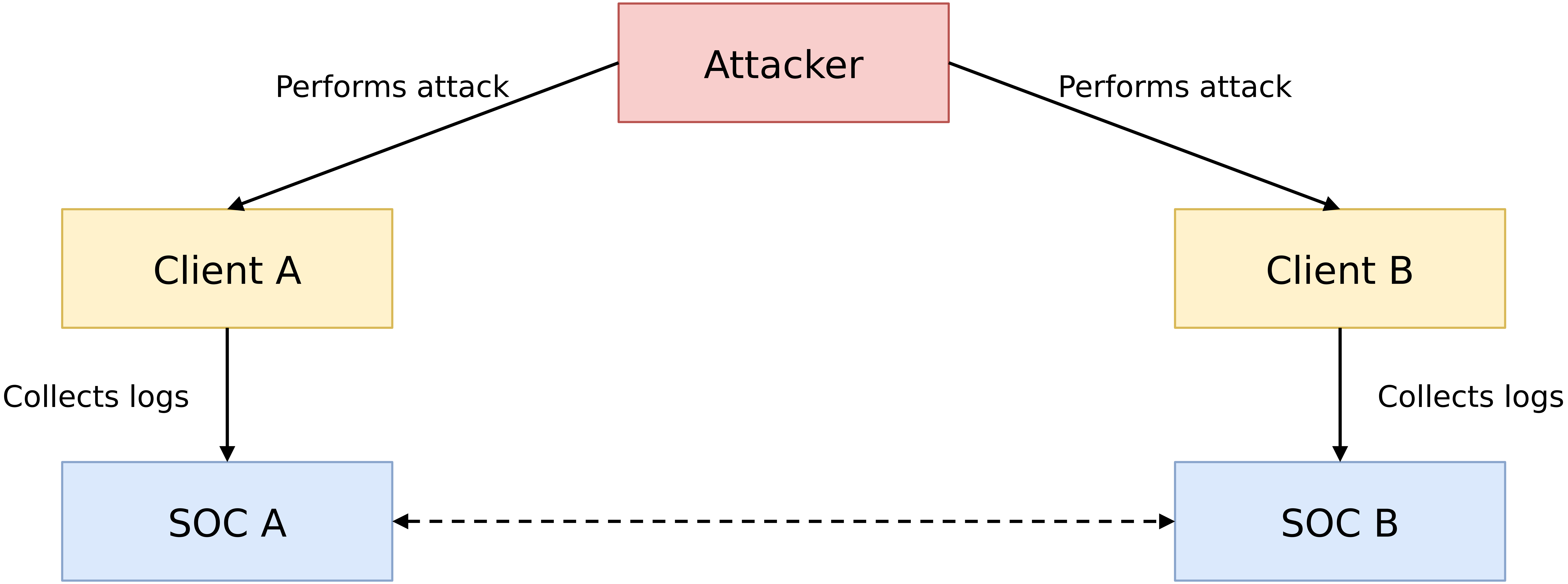


Figure 2: Collaboration in intrusion detection

How to collaborate?
How to ensure trust?

COLLABORATIVE DETECTION: A TRUST PROBLEM

SOCs are hunting for the same incident…

- Attackers reuses malwares and attack patterns
- Clients may use same apps
  - Widely used apps (e.g., Office Pack, SAP)
  - Domain specific apps (e.g., hospitals, bank)

But:

- Datasets must not be shared due to sensitivity (GDPR, IP, National Regulation)
- SOCs may use AI with distinct approaches
  - Different skillsets (datascience vs cybersecurity)
  - Different performances
  - Different training datasets (paid CTI feeds, different past incidents, different malwares)

SECURITY MONITORING: A DATASCIENCE PERSPECTIVE

Datascience can help hunting workflow automation [19]

- Structuring data processing allows automating some hunts;
- Clustering to reduce the number of alerts to process manually;
- Anomaly detection to prioritize investigations and limit the time needed to fine tune detection conditions.

Limitations

- Each monitored system has its own monitoring tools and risks;
- Analysts have limited datascience knowledge and datascientists have limited cybersecurity knwoledge;
- Centralisation of security logs might faces to confidentiality requirements.

RESEARCH QUESTION

Collaborating and sharing information is hard (privacy, security, availability...) [1]-[3]

R.Q: How to federate knowledge between non-trusting parties?

- What data should organizations collect locally?
- What part of that of that data should organization share with each other?
- How to share data between organizations (models, algorithms, sharing strategies)?

Icons by Flaticon

2. AN OVERVIEW OF FEDERATED LEARNING

Topic definition, literature review and open issues

RELEVANCE OF FEDERATED LEARNING

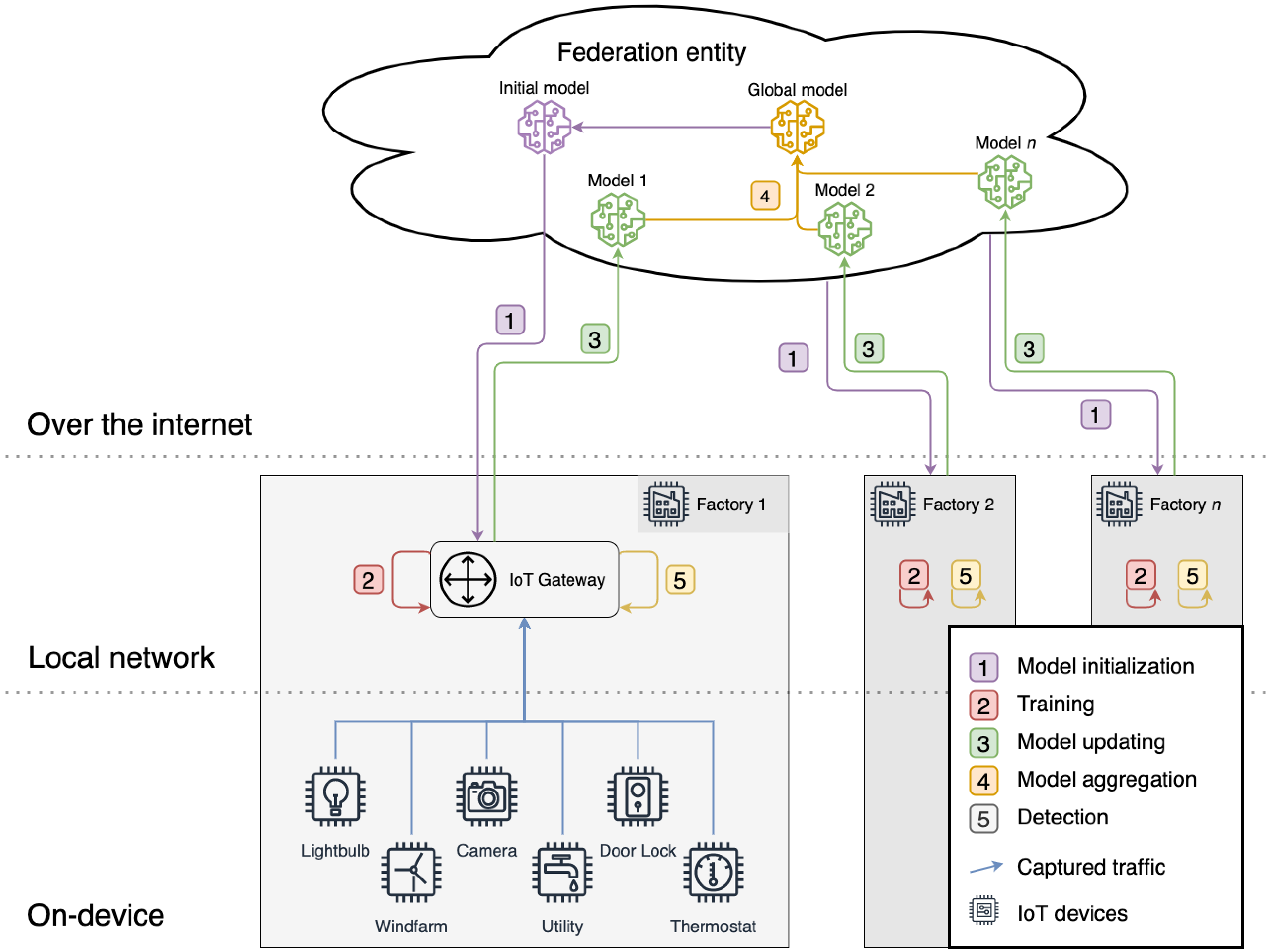


Figure 3: FL for intrusion detection, an application to Industrial IoT [4] — © IEEE 2022

Local operation

- Independent of the server for detection
- Faster and lower bandwidth consumption

Collaboration

- More data to train on
- Shares models not data (++ privacy)

OVERVIEW

"The Evolution of FL-based intrusion detection and mitigation: a Survey" [1] [4]

- Systematic Literature Review
- Four contributions
  - Quantitative and qualitative structured analyses
  - Reference architecture
  - Taxonomy
  - Open issues and research directions

RQs answered by the survey

- How are FIDSs used in different domains?
- What are the differences between FIDS architectures?
- What is the state of the art of FIDSs?

1. submitted Nov. 2021, accepted May 2022, published Jun. 2022

QUANTITATIVE OVERVIEW

- "Trending topic" since ~2018-2019
  - exponential: more than doubled since the realization of the survey
- Very heterogeneous venues
- Heterogeneous community

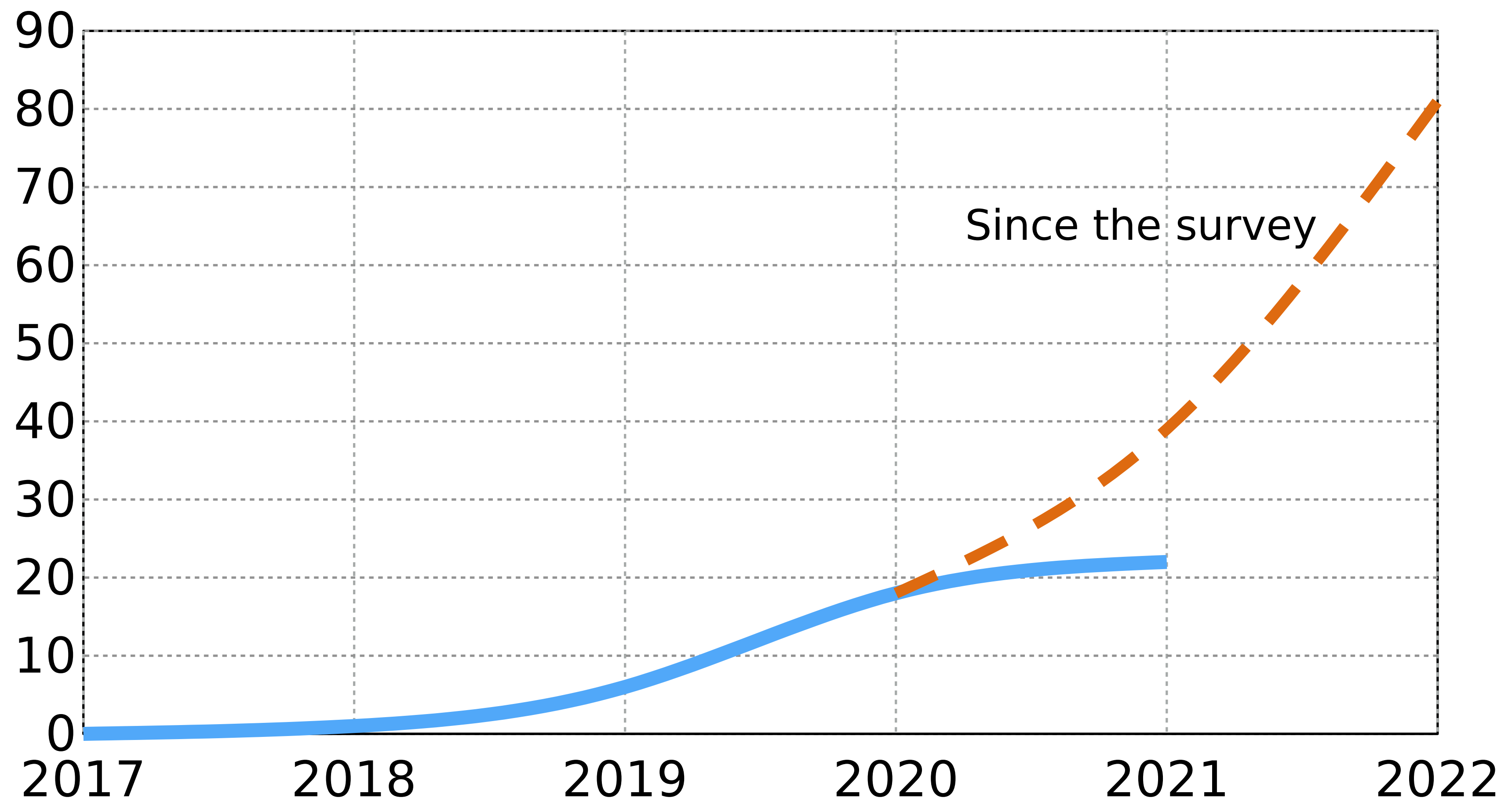


Figure 4: Evolution of FIDSs — data from [4] © IEEE 2022

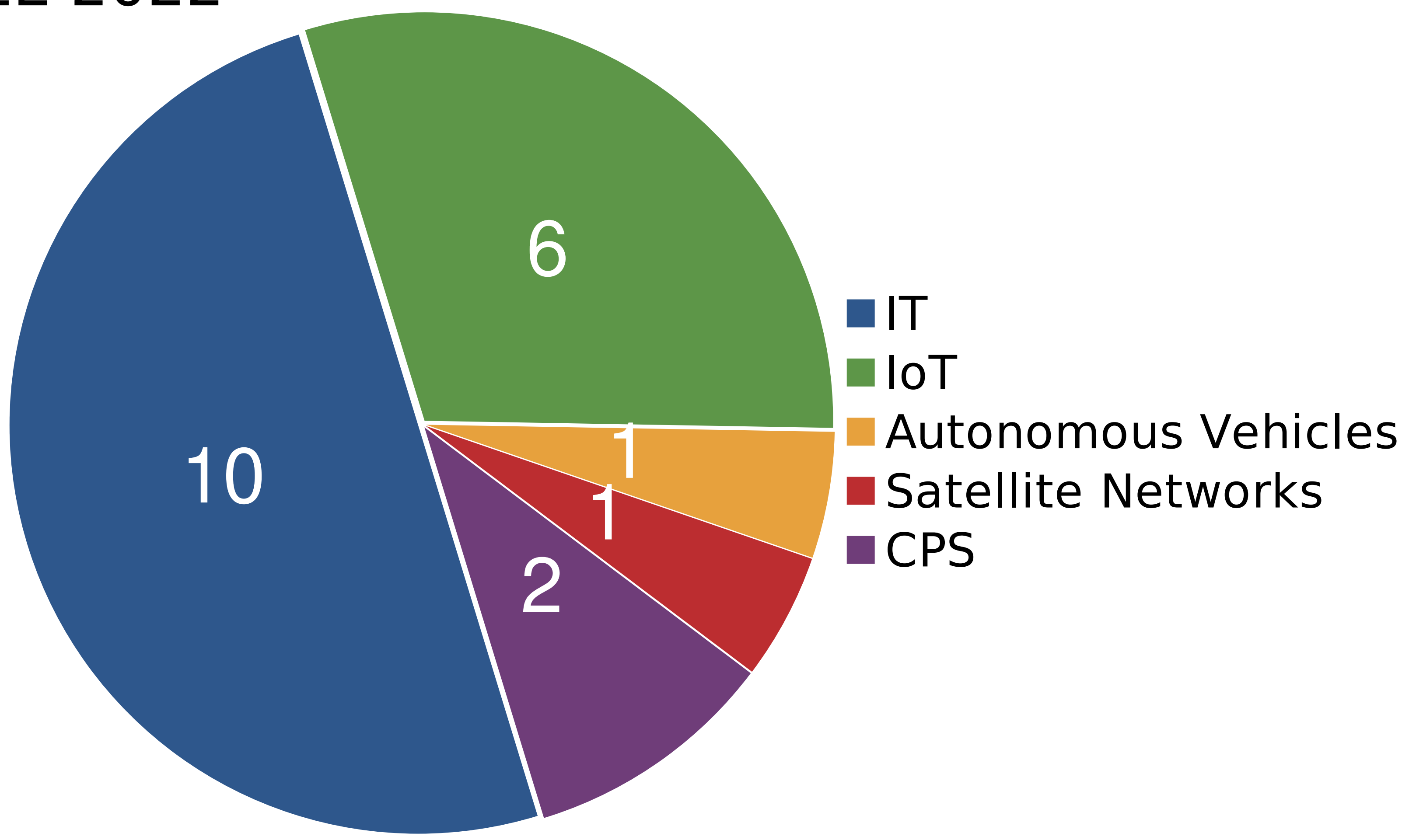


Figure 5: Publications by domain — data from [4] © IEEE 2022

QUALITATIVE OVERVIEW

	Ref	Horizontal FL	Vertical FL	Federated Transfer Learning	Federated MTL	Federated Mimic Learning	Differential privacy	Homomorphic encryption	Secret Sharing	Personalized models	Available dataset	Internet of Things	Information Technologies	Cyber Physical Systems	Autonomous Vehicles	Satellite-terrestrial networks	Distributed Ledgers	Defense/Mitigation	Software Defined Networking	Training location	Data type	Dataset	Local Algorithm	Federation Algorithm
2018	Pahl *et al.* [9]	●	○	○	○	○	○	○	○	●	●	●	○	○	○	○	○	○	○	Device	IoT network traffic (middleware)	Generated	BIRCH K-means	Parameter addition
2019	Rathore *et al.* [8]	●	○	○	○	○	○	○	○	●	●	○	●	○	○	○	●	●	●	Edge-controller (SDN)	Network traffic (SDN)	NSL-KDD [55]	ANN	Vector concatenation
2019	Schneble *et al.* [10]	●	○	○	○	○	○	○	○	○	●	○	○	●	○	○	●	○	○	Gateway	Sensor values	MIMIC [78]	MLP	Weight and biases average
2019	Nguyen, Marchal, *et al.* [11]	●	○	○	○	○	○	○	○	●	○	●	○	○	○	○	○	○	○	Gateway	IoT network traffic	Generated	GRU	FedAvg
2019	Zhao *et al.* [12]	○	○	○	●	○	○	○	○	●	◐	○	●	○	○	○	○	○	○	Gateway	Network traffic (encrypted)	CICIDS2017 [79] ISCXVPN2016 [80] ISCXTor2016 [81]	FC (shared layers) → FC	Weight and biases average
2019	Cetin *et al.* [13]	●	○	○	○	○	○	○	○	○	●	○	●	○	○	○	○	○	○	Gateway	Network traffic (WIFI)	AWID [56]	SAE	FedAvg
2020	Li, Wu, *et al.* [14]	●	○	○	○	○	○	●	○	○	●	○	○	●	○	○	○	○	○	Gateway	MODBUS	CPS dataset [82]	CNN-GRU → MLP	Homomorphic parameter addition
2020	Chen, Zhang, *et al.* [15]	●	○	○	○	○	○	○	○	○	●	○	●	○	○	○	○	○	○	Gateway	Network traffic	KDD 99 [54]	DAGMM	Parameter addition
2020	Zhang, Lu, *et al.* [16]	●	○	○	○	○	○	○	○	○	○	●	○	○	○	○	●	○	○	Gateway	Sensor values	Generated	ANN	CDW_FedAvg
2020	Fan *et al.* [17]	○	○	●	○	○	○	○	○	●	◐	●	○	○	○	○	○	○	○	Gateway (MEC)	IoT network traffic	CICIDS2017 [79] NSL-KDD [55] Generated	CNN	Parameter aggregation
2020	Rahman *et al.* [18]	●	○	○	○	○	○	○	○	○	●	○	●	○	○	○	○	○	○	Device	IoT Network traffic	NSL-KDD [55]	ANN	FedAvg
2020	Sun, Ochiai, *et al.* [19]	●	○	○	○	○	○	○	○	●	○	○	●	○	○	○	○	○	○	Gateway	Network traffic	LAN-Security Monitoring Project [83]	CNN	Parameter aggregation
2020	Al-Marri *et al.* [20]	○	○	○	○	●	○	○	○	○	●	○	●	○	○	○	○	○	○	Gateway	Network traffic	NSL-KDD [55]	ANN	FedAvg
2020	Kim, Cai, *et al.* [21]	●	○	○	○	○	○	○	○	○	●	○	●	○	○	○	○	○	○	Gateway	Network traffic	NSL-KDD [55]	MLP	FedAvg
2020	Qin, Poularakis, *et al.* [22]	●	○	○	○	○	○	○	○	●	●	○	●	○	○	○	○	○	●	Gateway (SDN)	Network traffic	CICIDS2017 [79] ISCX Botnet 2014 [84]	BNN	SignSGD
2020	Chen, Lv, *et al.* [23]	●	○	○	○	○	○	○	○	○	●	○	●	○	○	○	○	○	●	Gateway	Network traffic	CICIDS2017 [79] KDD 99 [54] WSN-DS [85]	GRU-SVM	FedAGRU
2020	Hei *et al.* [24]	●	○	○	○	○	○	○	○	○	●	○	●	○	○	○	●	○	○	Device	Network traffic	KDD 99 [54]	MLP	FedAvg
2020	Li, Zhou, *et al.* [25]	●	○	○	○	○	○	●	○	○	○	○	●	○	○	●	○	○	○	Gateway	Network traffic	Generated	CNN	Homomorphic parameter addition
2021	Liu *et al.* [26]	●	○	○	○	○	●	○	●	○	●	○	○	○	●	○	○	○	○	Device	Network traffic	KDD 99 [54]	MLP	Parameter aggregation
2021	Popoola *et al.* [27]	●	○	○	○	○	○	○	○	○	●	●	○	○	○	○	●	○	○	Gateway	IoT Network traffic	Bot-IoT [86] N-BaIoT [87]	ANN	FedAvg
2021	Qin and Kondo [28]	●	○	○	○	○	○	○	○	●	●	○	●	○	○	○	○	○	○	Device	Network traffic	NSL-KDD [55]	ONLAD [89] (ELM + AE)	FedAvg
2021	Sun, Esaki, *et al.* [29]	●	○	○	○	○	○	○	○	●	○	○	●	○	○	○	○	○	○	Gateway	Network traffic	LAN-Security Monitoring Project [83]	CNN	Parameter aggregation
				FL type				Protection		Training				Use case				Features						

Figure 6: Comparative overview of selected works [4] — © IEEE 2022

Key points:

- Mostly horizontal FL settings
- Often cross-silo, training on dedicated devices
- Mainly NIDS IT datasets
- Often NNs
- Few sophisticated aggregation algorithms

OPEN ISSUES

and research directions

1. TRANSFERABILITY

Transfer knowledge between models from heterogeneous client.

- Train multiple variations of the same models [13];
- Transfer knowledge between use cases or environments [12];
- Finding trade-off between specialization and generalization/federation [7], [14].

2. SECURITY AND TRUST

Preventing FIDS to represent a threat.

- Improve model-poisoning detection [14];
- Use reputation systems to deal with untrusted participants [15];
- Protect aggregation with HE, MPC, or differential privacy [16];

3. DATASET REPRESENTATIVITY

Providing datasets that fit real-world situations.

- Provide datasets generated in federated settings (heterogeneous participants);
- Evaluate knowledge transfer (new behaviors learned by peers).

OPEN ISSUES

and research directions

4. MODEL PERFORMANCE

Improving detection in regards of usual metrics (accuracy, precision, recall, …).

- Use GANs as a training input [5];
- Study the impact of hyper- [6] and meta-parameters on detection rate;
- Behavior modeling (protocol-mining, periodicity-mining, manual feature selection) [7]-[8].

5. MODEL CONVERGENCE

Preventing FIDS models to diverge

- Considering aggregation as an optimisation problem [14];
- Weighting mechanisms to improve the convergence [15];.

6. ADAPTABILITY AND SCALABILITY

Dealing with high client volume and constrained environments.

- Deal with constrained environments (compressed updates, fewer rounds) [10]-[11];
- Provide update strategies to keep with the evolution of attacks [10].

7. SELF-DEFENSE AND SELF-HEALING

Providing reaction, resilience, and sharing counter-measures.

- Provide automated or assisted mitigation strategy [9];
- Study the application of FL to improve mitigation;

MOTIVATION

Transferability, **adaptability**, and **trust** are identified open issues in the research community.

- #1 and #2 are due to the differences between clients in cross-silo settings like intrusion detection.
  - Organizations may process very different data and still require collaboration, thus producing very different models.
- Trust is particularly important in collaborative security context.
- Existing datasets for intrusion detection are created for a local-detection use case.

RESEARCH QUESTIONS

RQ 1. How to federate data from heterogeneous sources?

RQ 2. How to trust participants and evaluate their performance?

RQ 3. How to weight each contribution for aggregation?

RQ 4. How to evaluate FIDSs?

3. EXPERIMENTS AND FUTURE WORKS

Addressed issues and contributions

USE CASE: Collaborative NIDS in IT Networks

IT NETWORKS

Leverage NIDS capabilities to detect distributed threats in realistic IT networks.

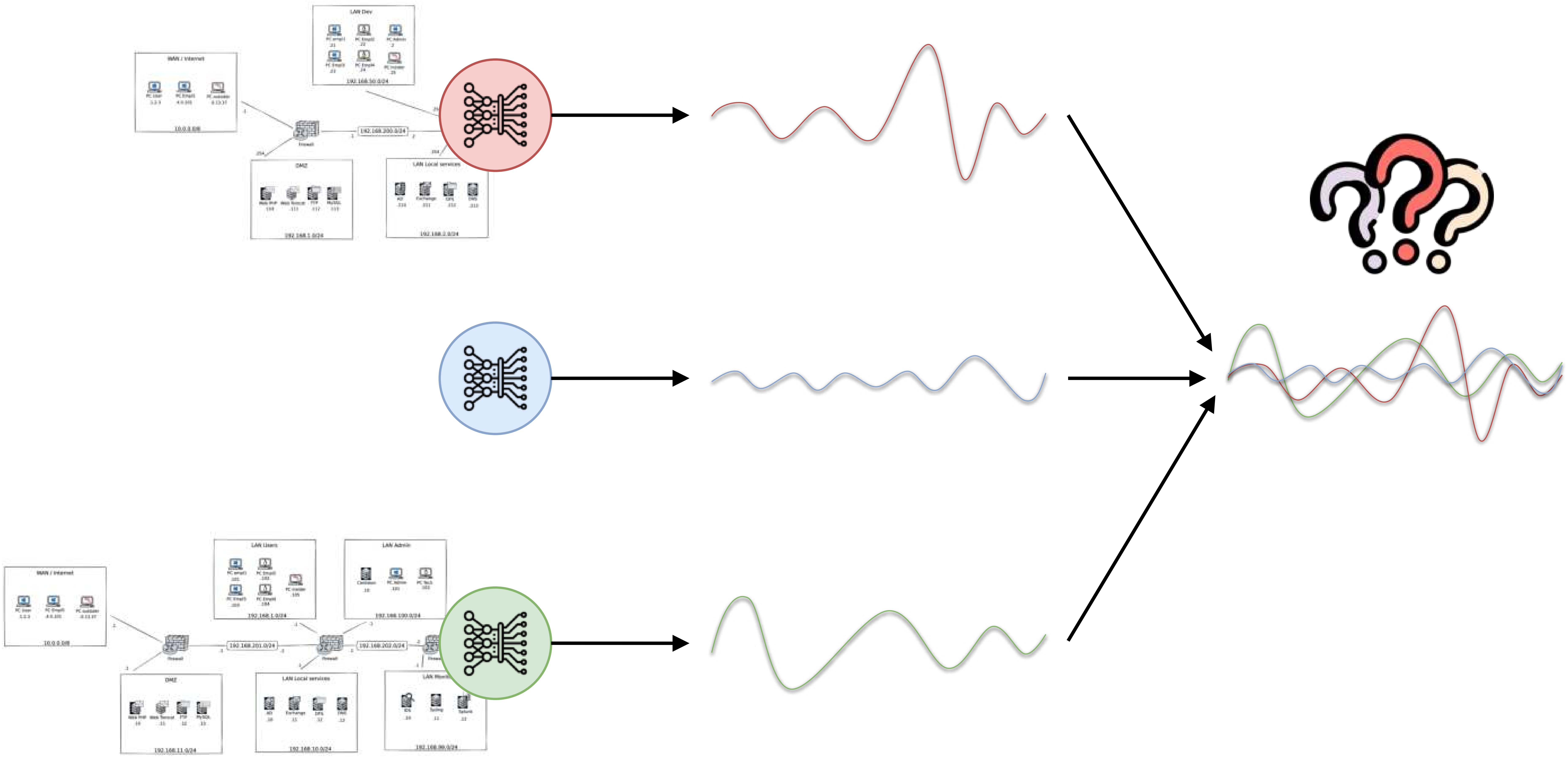
Relevance of the use case

- "Easy" to build and to experiment on.
- A lot of existing works, allows comparison with related works.
- Virtualization enables reproducibility and modularity in experimentations.

Different heterogeneities

- ~~Organizations may use different models for detection~~.
- Organizations may have differences in their training data and environments.

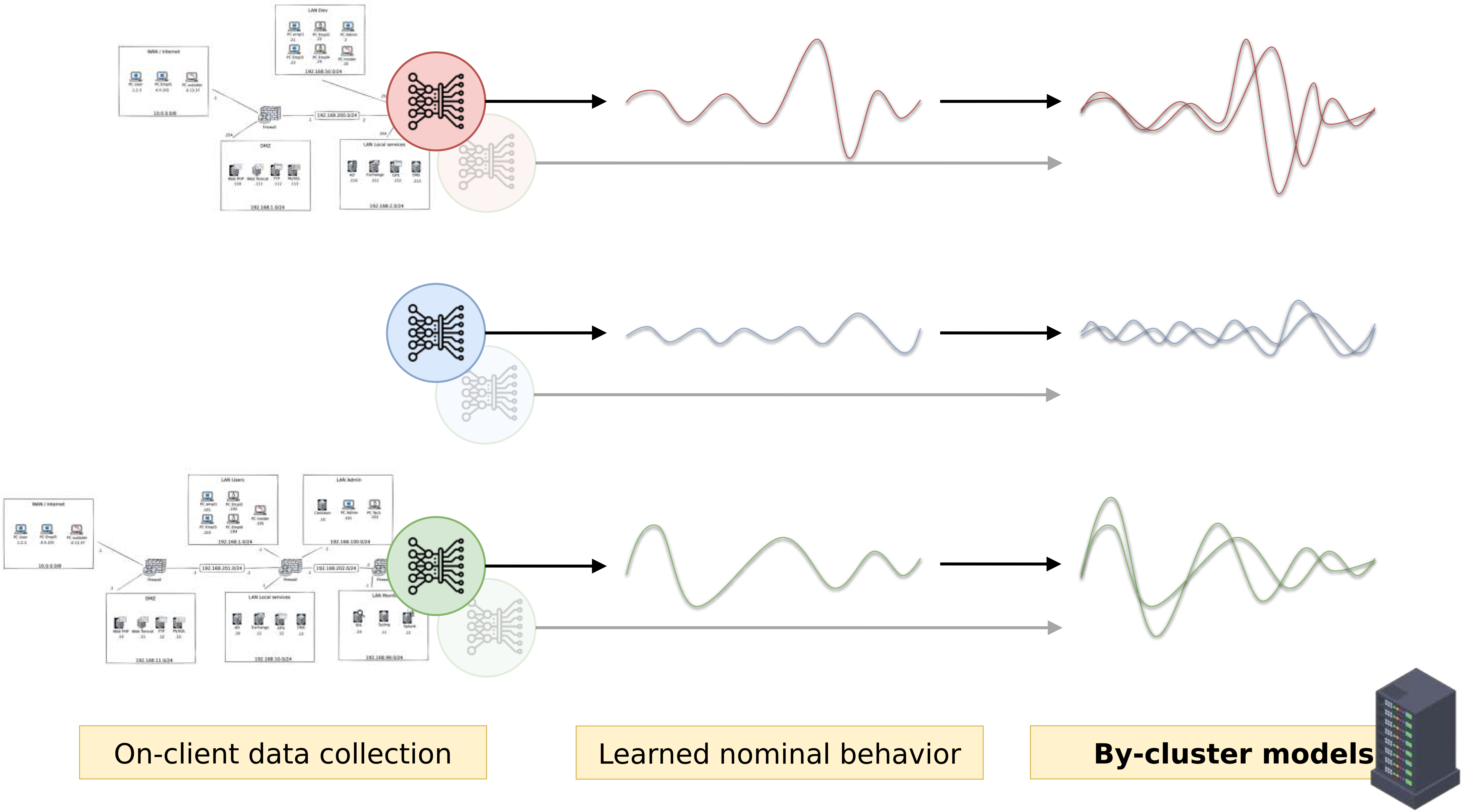
DEALING WITH HETEROGENEITY



On-client data collection | Learned nominal behavior | **Aggregated model**

Figure 7a: hetereogeneity in FIDS

DEALING WITH HETEROGENEITY



On-client data collection

Learned nominal behavior

By-cluster models

Figure 7b: hetereogeneity in FIDS

DEALING WITH TRUST

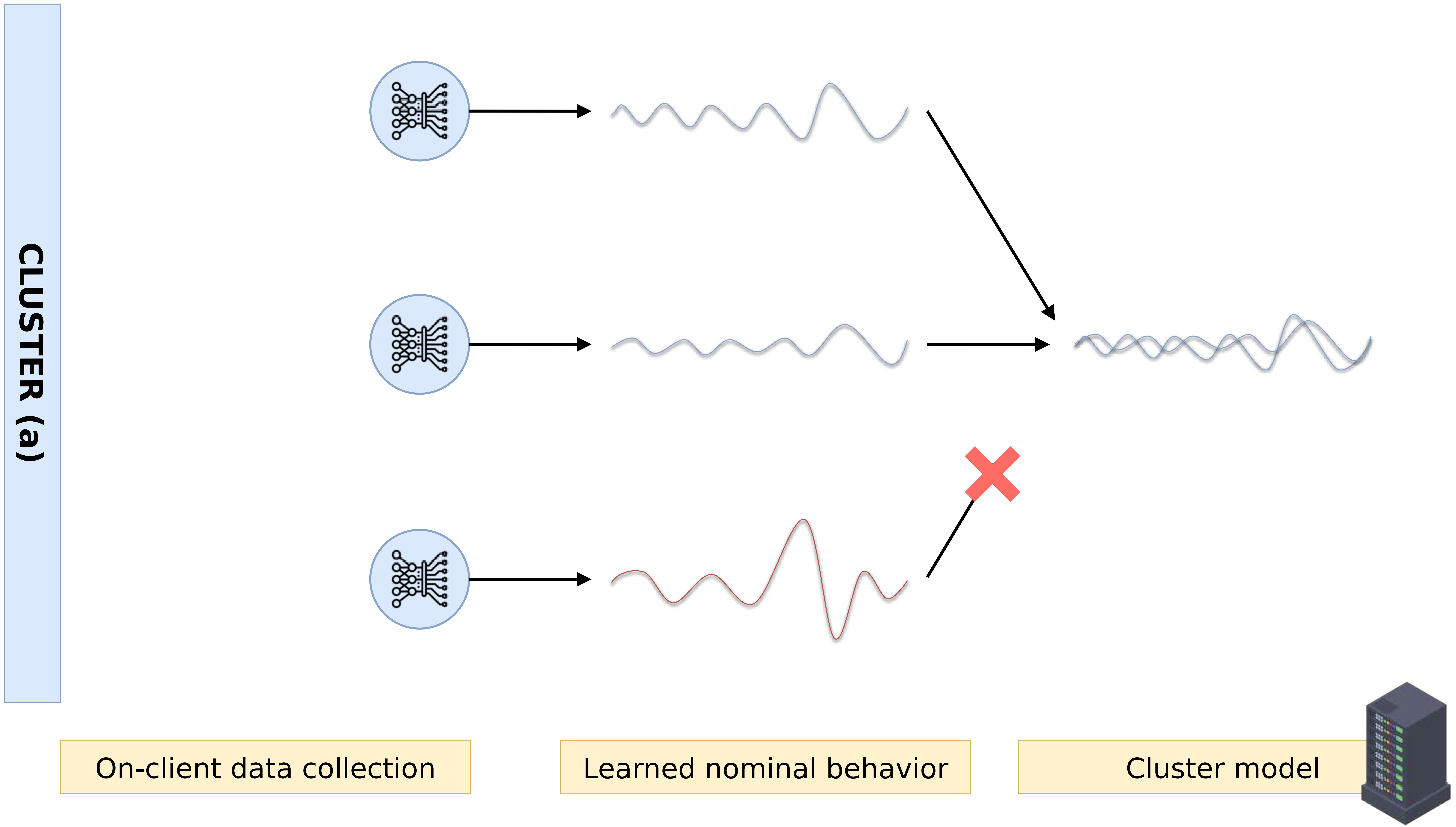


Figure 8a: trust in FIDS

DEALING WITH TRUST: reputation system

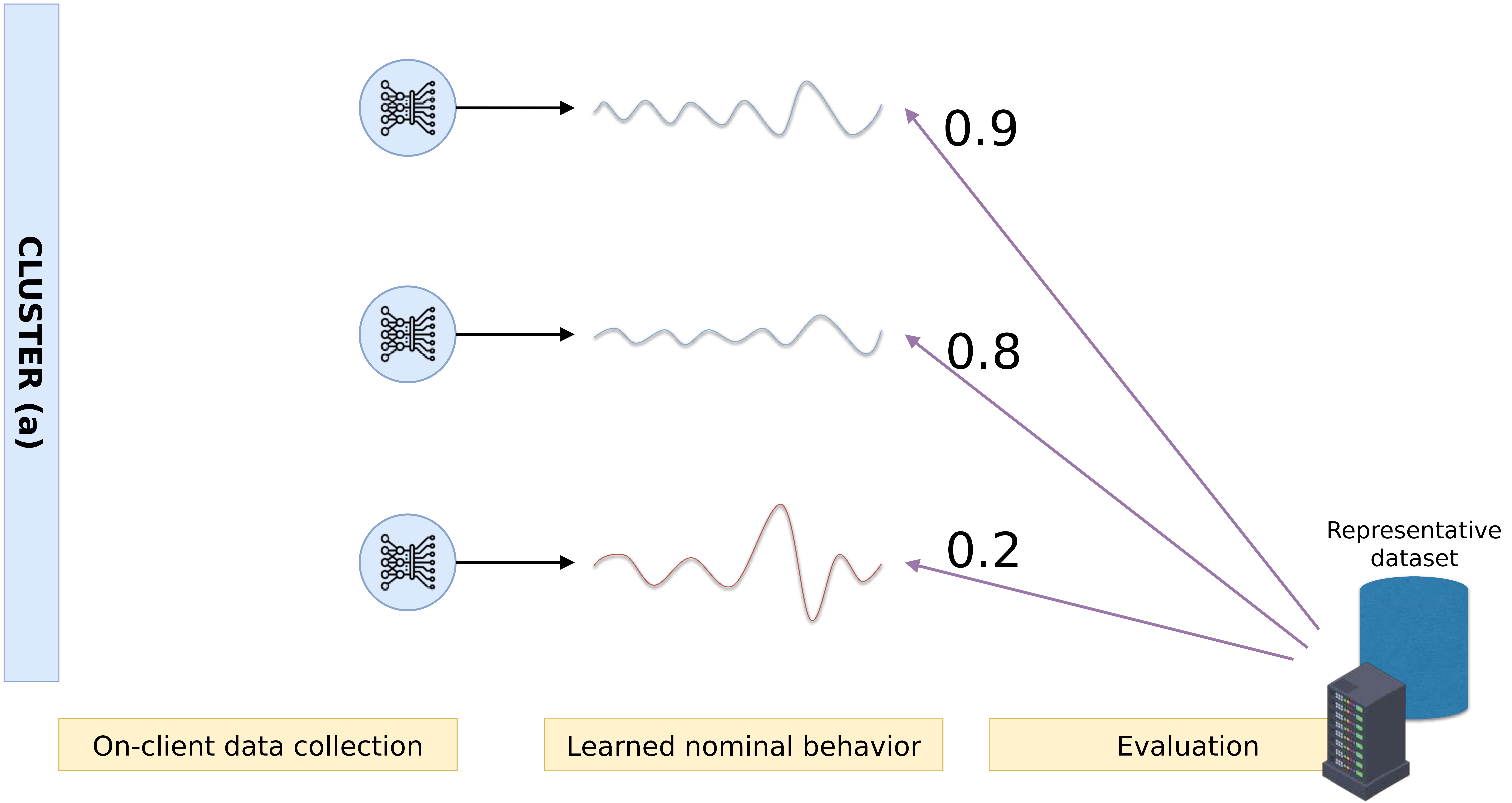


Figure 8b: trust in FIDS

DEALING WITH TRUST: reputation system

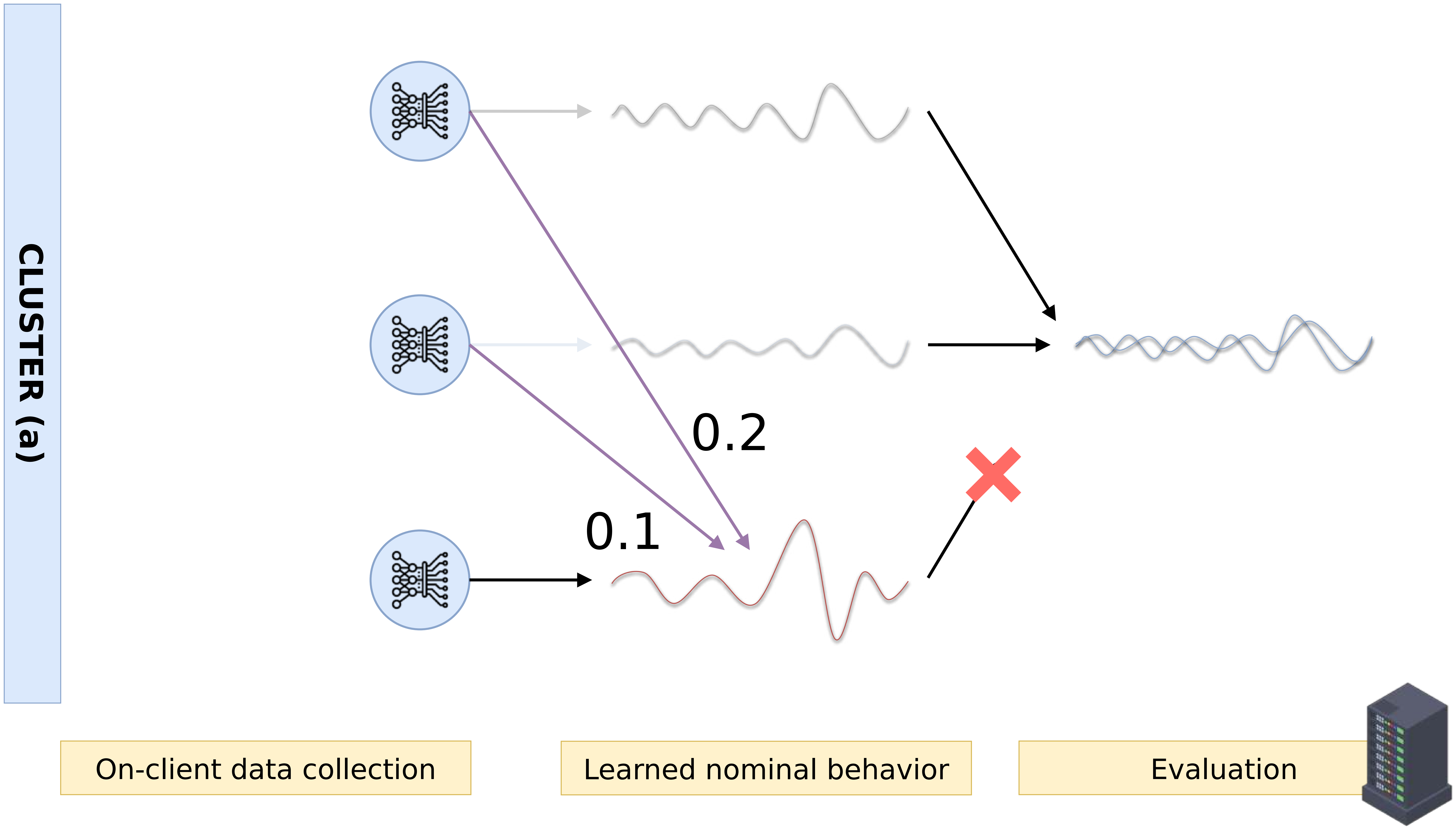


Figure 8c: trust in FIDS

TRUST-FIDS[1]

- **Aim:** tackle heterogeneity and lack of trust in FL-based collaboration. (RQ1-3)

- **Means:**
  - use clustering to group clients by data-similarity
  - use reputation to iteratively build trust between clients

- **How:** introduce cross-evaluation between clients, which provides feedbacks on how each client views the other models.

ollaboration project with another PhD student at IMT Atlantique who focus on reputation systems.

GENERATING DATA

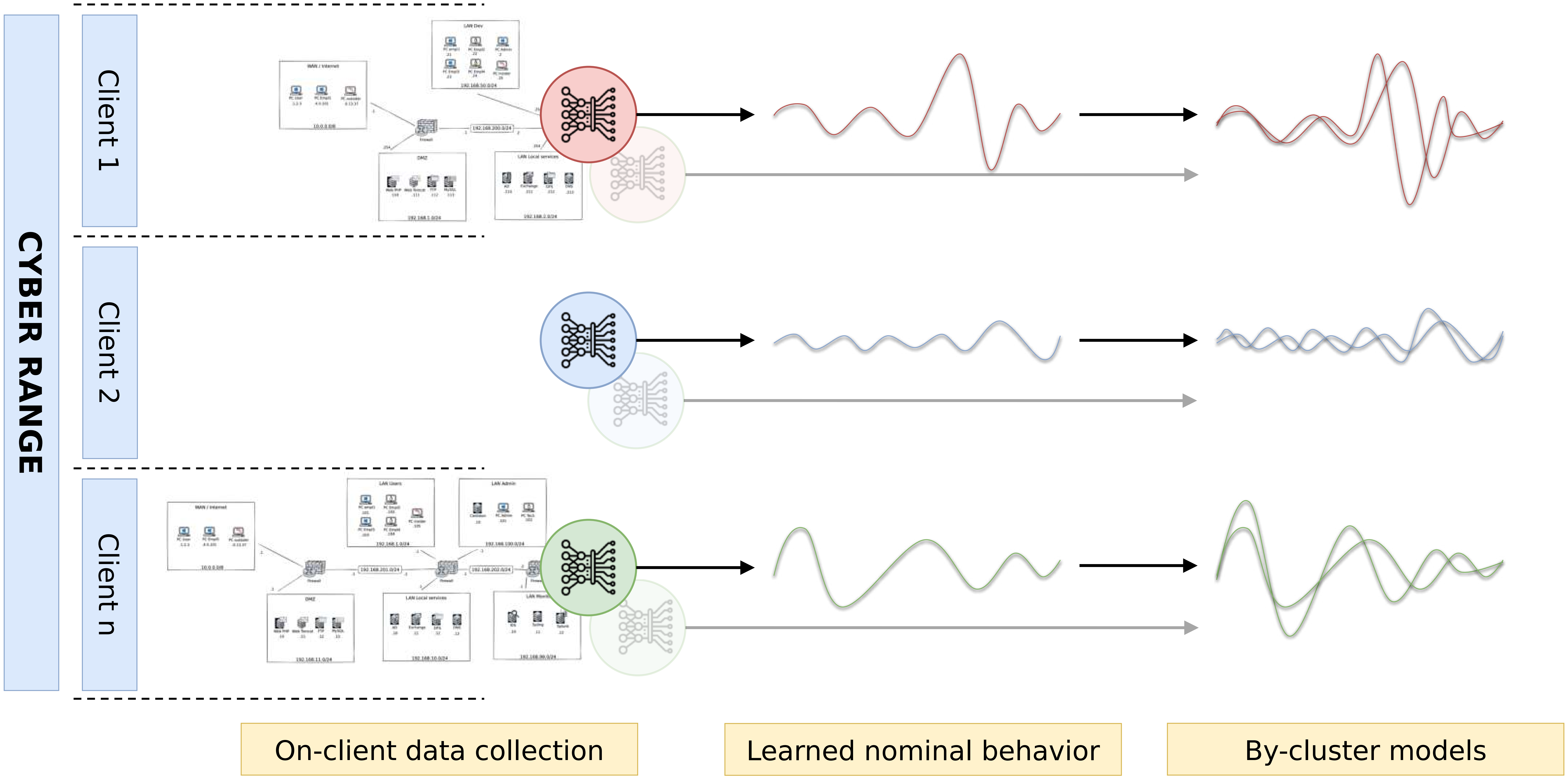


Figure 9: evaluation in FIDSs

FedITN

- **Aim:** provide tools dedicated to evaluate FIDSs and other collaborative IDSs (RQ1, RQ4)
  - performance against heterogeneity;
  - knowledge transfer between clients;
  - model adaptability;
  - generation capability;

- **Means:**
  - a new dataset with four network topologies
  - evaluation baselines and tools for reproducibility

#. CONCLUSION

Outcomes and perspectives

CONCLUSION

- **Federated Learning for Collaborative IDSs:**
  - Addresses actual problems from the industry (e.g., SOC collaboration);
  - Focus on heterogeneity and trust;
  - Emphasis on evaluation, reproducibility, and sound experiments.

- **Other research directions:**
  - scalability, model selection, …

- **Prospective vision:**
  - Opt-in and open collaboration;
  - Federation of models of all kind;
  - *Magic* collaboration.

QUESTIONS ?

PUBLICATIONS

Journals and International conferences

[a] L. Lavaur, M. -O. Pahl, Y. Busnel and F. Autrel, “The Evolution of Federated Learning-based Intrusion Detection and Mitigation: a Survey,” in IEEE Transactions on Network and Service Management (TNSM), 2022, doi: 10.1109/TNSM.2022.3177512.

National and local venues

[b] L. Lavaur, B. Costé, M. -O. Pahl, Y. Busnel, and F. Autrel, “Federated learning as enabler for collaboration between not fully-trusting distributed parties,” in 29th Computer & Electronics Security Application Rendezvous: Ensuring Trust in a Decentralized World (C&ESAR 2022), 2022

REFERENCES

[1] N. Neshenko, E. Bou-Harb, J. Crichigno, G. Kaddoum, and N. Ghani, "Demystifying IoT Security: An Exhaustive Survey on IoT Vulnerabilities and a First Empirical Look on Internet-Scale IoT Exploitations," *IEEE Communications Surveys & Tutorials*, 2019

[2] T. D. Wagner, K. Mahbub, E. Palomar, and A. E. Abdallah, "Cyber threat intelligence sharing: Survey and research directions", *Computers & Security*, 2019.

[3] W. Tounsi and H. Rais, "A survey on technical threat intelligence in the age of sophisticated cyber attacks," *Computers & Security,* 2018

[4] L. Lavaur, M.-O. Pahl, Y. Busnel, and F. Autrel, "The Evolution of Federated Learning-based Intrusion Detection and Mitigation: a Survey," *IEEE Trans. On Network and Services Management, Special Issue on Advances in Network Security Management,* 2022

[5] W. Schneble and G. Thamilarasu, "Attack detection using federated learning in medical cyber-physical systems," *International Conference on Computer Communications and Networks*, 2019.

[6] Y. Sun, H. Ochiai, and H. Esaki, "Intrusion Detection with Segmented Federated Learning for Large-Scale Multiple LANs," *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020

[7] M.-O. Pahl and F. X. Aubet, "All Eyes on You: Distributed Multi-Dimensional IoT Microservice Anomaly Detection," *14th International Conference on Network and Service Management*, 2018

[8] T. D. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan, and A.-R. Sadeghi, "DÏoT: A Federated Self-learning Anomaly Detection System for IoT*," IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, 2019

[9] S. Rathore, B. Wook Kwon, and J. H. Park, "BlockSecIoT-Net: Blockchain-based decentralized security architecture for IoT network," *Journal of Network and Computer Applications*, 2019

REFERENCES

[10] Y. Fan, Y. Li, M. Zhan, H. Cui, and Y. Zhang, “IoTDefender: A Federated Transfer Learning Intrusion Detection Framework for 5G IoT,” in *2020 IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE),* 2020

[11] S. A. Rahman, H. Tout, C. Talhi, and A. Mourad, “Internet of Things Intrusion Detection: Centralized, On-Device, or Federated Learning?” *IEEE Network*, 2020

[12] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, “DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber-Physical Systems,” *IEEE Transactions on Industrial Informatics*, 2020

[13] Y. Chen, J. Zhang, and C. K. Yeo, “Network Anomaly Detection Using Federated Deep Autoencoding Gaussian Mixture Model,” in *Machine Learning for Networking*, 2020

[14] T. D. Nguyen, P. Rieger, H. Yalame, H. Mo¨llering, H. Fereidooni, S. Marchal, M. Miettinen, A. Mirhoseini, A.-R. Sadeghi, T. Schneider, and S. Zeitouni. “FLGUARD: Secure and Private Federated Learning.” , *arXiv*, 2021

[15] W. Zhang, Q. Lu, Q. Yu, Z. Li, Y. Liu, S. K. Lo, S. Chen, X. Xu, and L. Zhu, “Blockchain-based Federated Learning for Device Failure Detection in Industrial IoT,” *IEEE Internet of Things Journal*, 2020

[16] Z. Chen, N. Lv, P. Liu, Y. Fang, K. Chen, and W. Pan, “Intrusion Detection for Wireless Edge Networks Based on Federated Learning,” *IEEE Access,* 2020

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

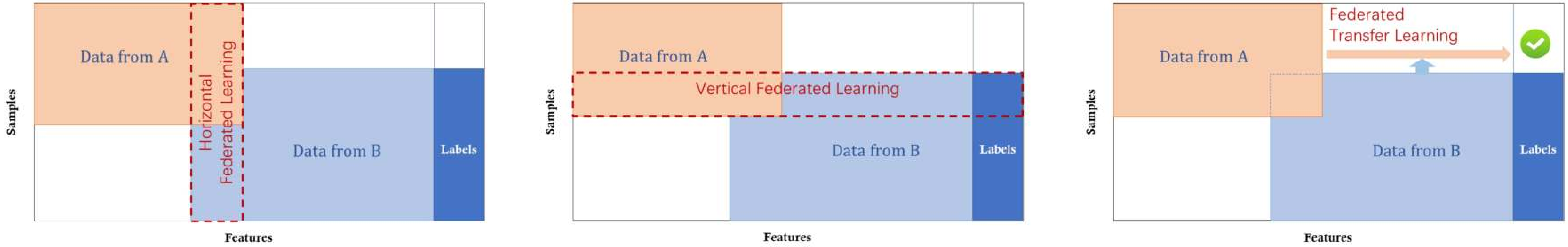
[18] G. Bertoli, L. A. Pereira Junior, A. L. dos Santos, O. Saotome, "Generalizing intrusion detection for heterogeneous networks: A stacked-unsupervised federated learning approach," *arXiv.org*, 2022

REFERENCES

[19] A. Dey, “Science de la donnée en appui des opérations de cybersécurité,” in *Ecole nationale supérieure Mines-Télécom Atlantique Bretagne Pays de la Loire,* thesis manuscript, 2022

[20] Q. Yang, Y. Liu, T. Chen, and Y. Tong, “Federated Machine Learning: Concept and Applications,” in *ACM Transactions on Intelligent Systems and Technology*, 2019

BACKUP SLIDES



(a) Horizontal Federated Learning (b) Vertical Federated Learning (c) Federated Transfer Learning

Figure X: Different settings of FL by Yang et al. [20]

COLLABORATION IN CYBERSECURITY

Collaborating and sharing information to cope with the increase in cyberattacks [1]-[3]

- Privacy risks – *eg.* information disclosure;
- Security risks – *eg.* revealing internals, poisoning;
- Availability – *eg.* single point of failure in centralized systems;
- Resources – *eg.* higher bandwidth consumption when sharing data;
- ...

TRUST-FIDS Methodology

- **Dataset**: "standardized IDS datasets" [17] (UNSW-NB15, BoT-IoT, ToN-IoT, and CSE-CIC-IDS2018)
- **Evaluation:**
  - Comparison with the SoA [18] on the same dataset
  - ***w*** and ***w/o*** clustering
  - ***w*** and ***w/o*** reputation
  - ***w*** and ***w/o*** poisoning attacks / neglecting participants
- **Expected results:**
  - *Clustering* – the dataset is in four parts → four clusters at least
  - *Reputation* – contribution-aware aggregation, detection of neglecting participants
  - faster convergence, better results than without both

TRUST-FIDS Architecture

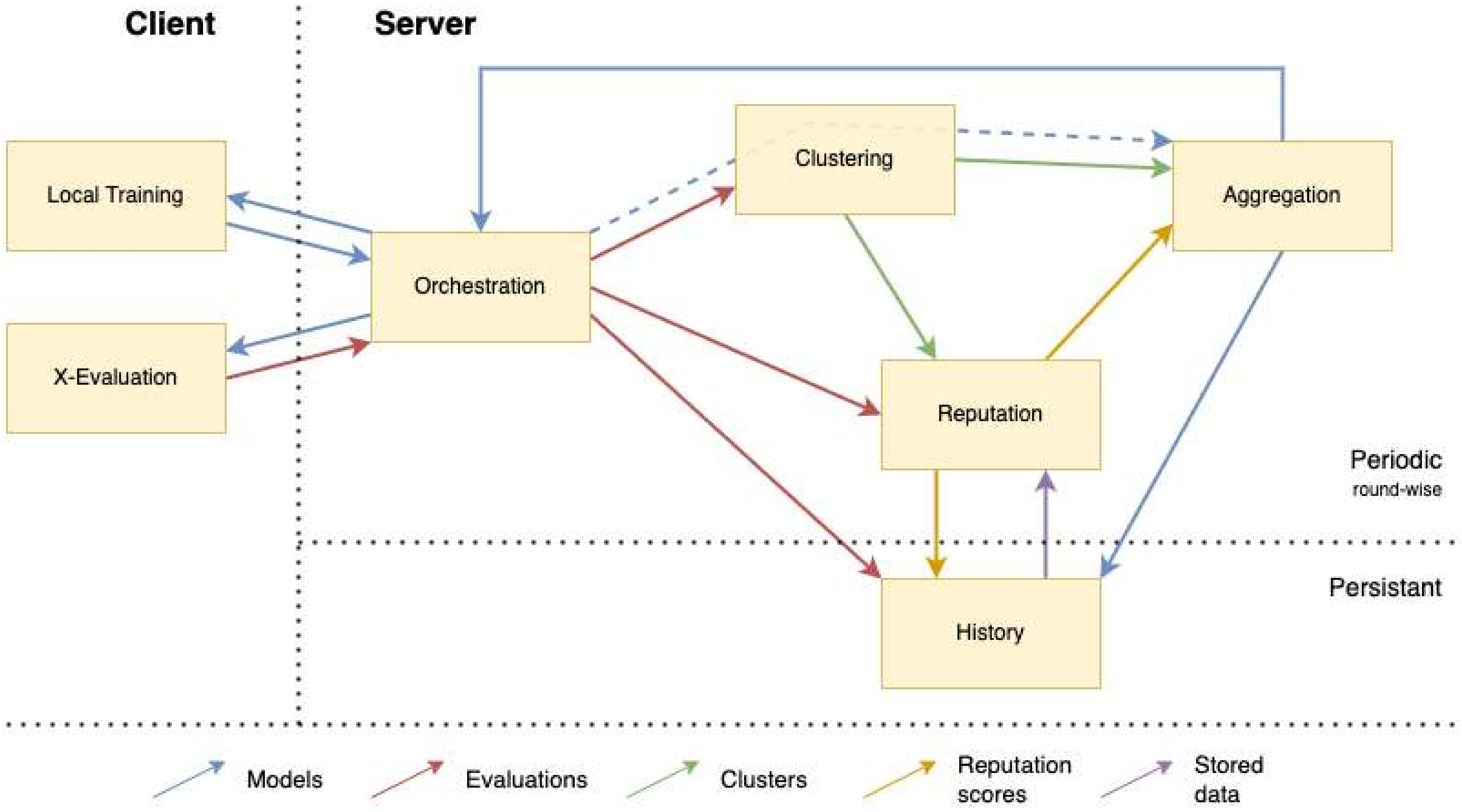


Figure X: Logical architecture of the Trust-FIDS approach

FedITN Methodology

- **Attacks:**
  - ***found in common datasets*** available in cyberrange
  - Implement only what is missing
  - 55 attacks with variations for the underlying services, labelling following the MITRE ATT&CK®
- **Heterogeneity:**
  - Different topologies with different services, architecture (network segmentation), probe location, and *cyber-maturity* (*eg.* firewall rules)
- **Evaluation:**
  - Metric comparison with other datasets (*eg.* NSL-KDD, CIC-IDS-201X, ...);
  - Comparison on SoA [18] approaches with other datasets;
- **Expected results:**
  - Existing approaches focusing on statistical heterogeneity might falter
  - Complexity difference in topology will show if FL can really *transfer* knowledge

FedITN Testbed and topologies

IT networks

Figure X: Airbus CyberRange

Topology 1

Expert topology with good segmentation.

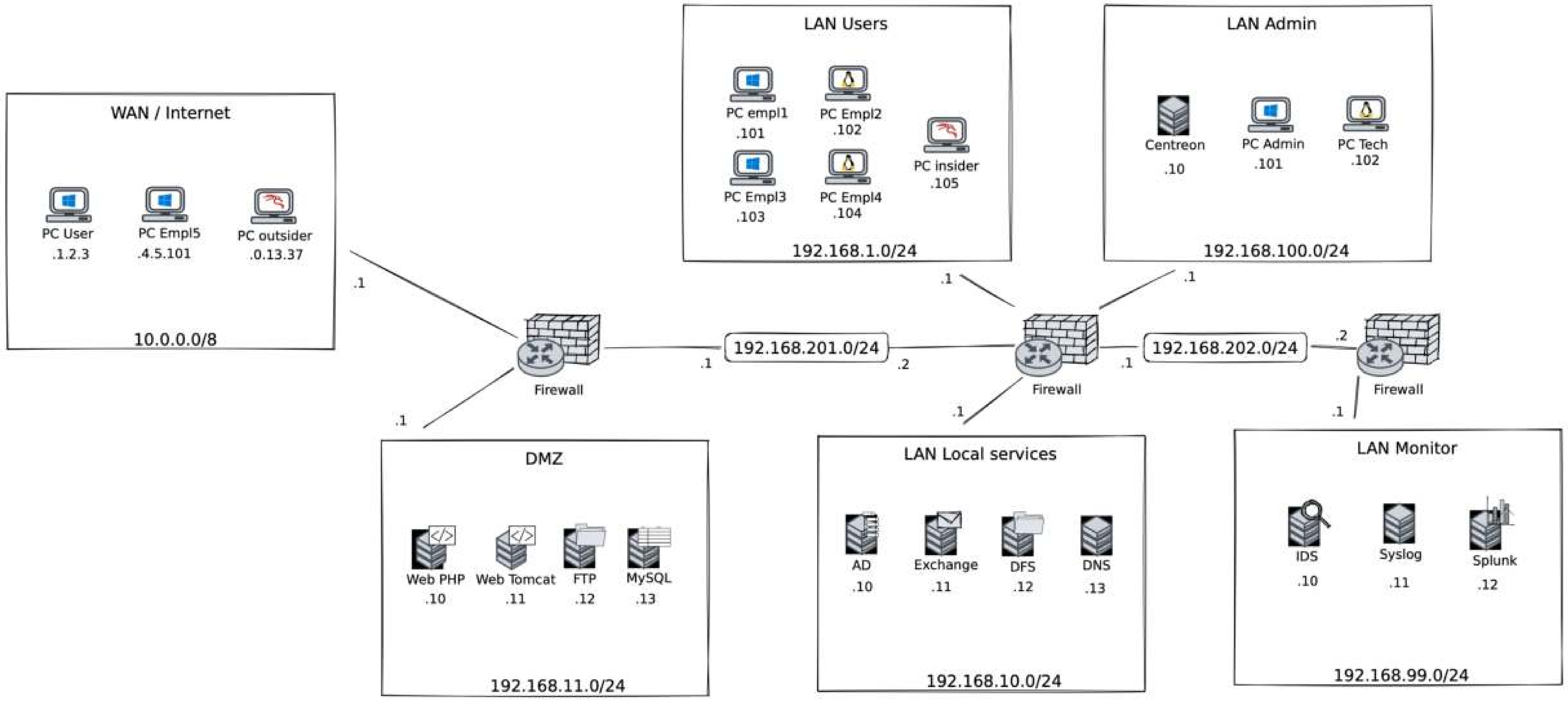


Figure X: Topology 1, modified version of Airbus Cybersecurity's default topologies

FedITN Attacks

Attack	Category	Target	ATT&CK Technique	ATT&CK Tactic
Bruteforce FTP	Bruteforce	FTP Server	Password Guessing (T1110.001)	Credential Access (TA0006)
Bruteforce lopin form	Bruteforce	Web Server w/ login form	Password Guessing (T1110.001)	Credential Access (TA0006)
Bruteforce MySQL	Bruteforce	MySQL server	Password Guessing (T1110.001)	Credential Access (TA0006)
Bruteforce RDP	Bruteforce	Windows Host w/ RDP server	Password Guessing (T1110.001)	Credential Access (TA0006)
Bruteforce SMB	Bruteforce	Windows Host w/ SMB server	Password Guessing (T1110.001)	Credential Access (TA0006)
Bruteforce SSH	Bruteforce	SSH server	Password Guessing (T1110.001)	Credential Access (TA0006)
Bruteforce Telnet	Bruteforce	Telnet server	Password Guessing (T1110.001)	Credential Access (TA0006)
Bruteforce VNC	Bruteforce	VNC server	Password Guessing (T1110.001)	Credential Access (TA0006)
DNS amplification	DoS	Any host	Reflection Amplification (T1498.002)	Impact (TA0040)
ICMP IGMP flood	DoS	Any host	Direct Network Flood (T1498.001)	Impact (TA0040)
PUSH ACK flood	Dos	Any host	Direct Network Flood (T1498.001)	Impact (TA0040)
R.U.D.Y.	DoS	Web Server w/ form	Service Exhaustion Flood (T1499.002)	Impact (TA0040)
slowloris	DoS	Web Server	Service Exhaustion Flood (T1499.002)	Impact (TA0040)
SYN flood	DoS	Any host	OS Exhaustion Flood (T1499.001)	Impact (TA0040)
TCP killer	DoS	Any host	Application or System Exploitation (T1499.004)	Impact (TA0040)
TCP RST flood	DoS	Any host	Direct Network Flood (T1498.001)	Impact (TA0040)
UDP flood	DoS	Any host	Direct Network Flood (T1498.001)	Impact (TA0040)
ZIP bomb	DoS	Any host	ARP Cache Poisoning (T1557.002) Transmitted Data Manipulation (T1565.002) OS Exhaustion Flood (T1499.001)	Credential Access (TA0006) Collection (TA0009) Impact (TA0040)

Figure X: Exemple of considered attacks and according labels

FedITN Experimentation pipeline

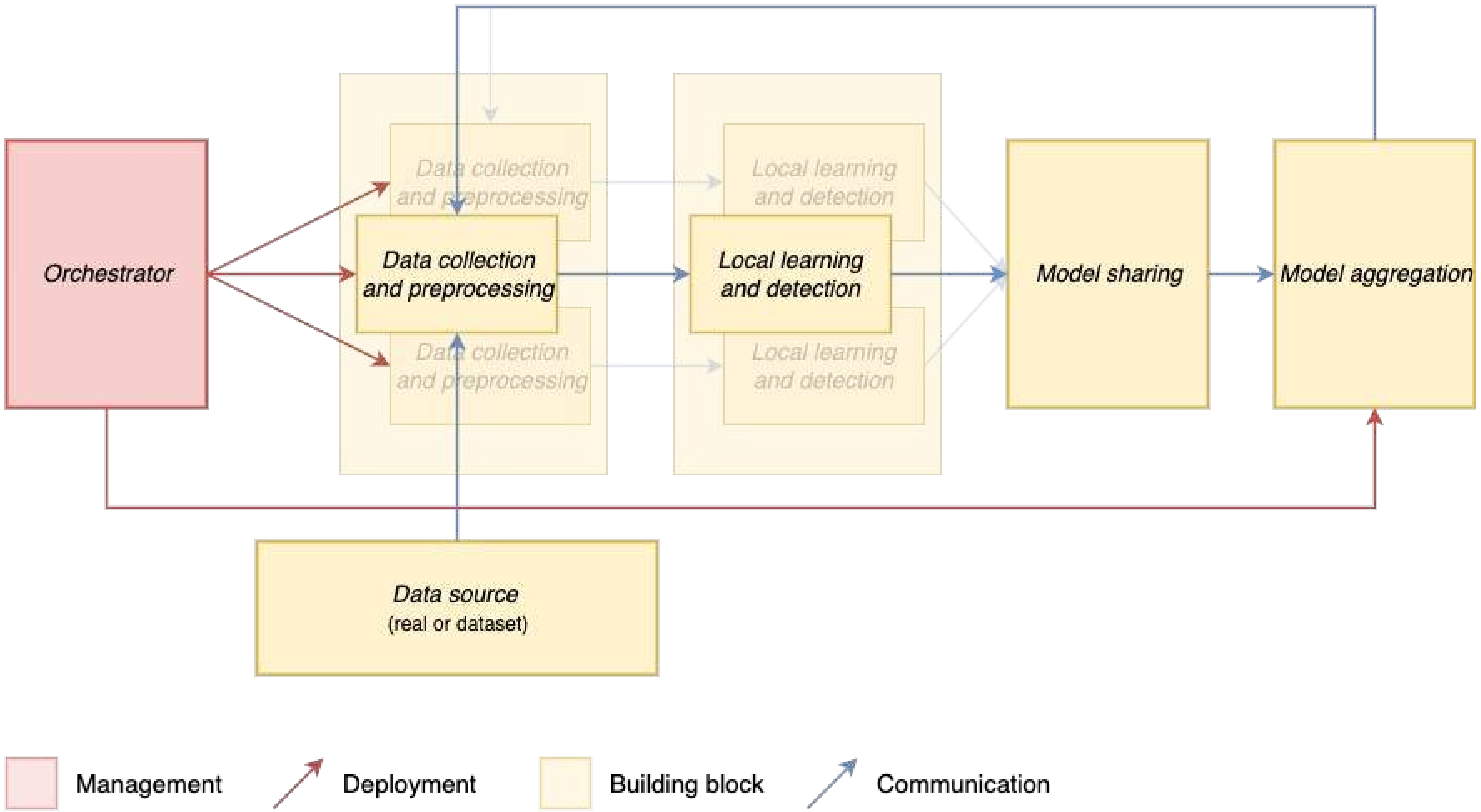


Figure X: FIDS experimentation framework (WIP)

MODEL WEIGHTING
Potential leads

- **data quality**: what is the quality of the data the model has been trained on?
  - Difficult to define
  - Possibly good metric for model weighting
- **cyber-maturity**: which confidence can I put in one participant's data?
  - Arbitrarily attribute maturity to some clients to evaluate the impact on federation
  - Ideally use that to create topologies in the future
- **semantic metadata**: what does this model contain, and what does mine lack?
  - Improve model aggregation information about its content
  - Balance to find with privacy