CoLearn: Enabling Federated Learning in MUD compliant IoT Edge Networks

Angelo Feraudo[1], Poonam Yadav[2], Vadim Safronov[3], Diana Andreea Popescu[3], Richard Mortier[3], Shiqiang Wang[4], Paolo Bellavista[1], Jon Crowcroft[3]

[1]University of Bologna, Italy, [2]University of York, UK

[3]University of Cambridge, UK, [4]IBM Research, USA

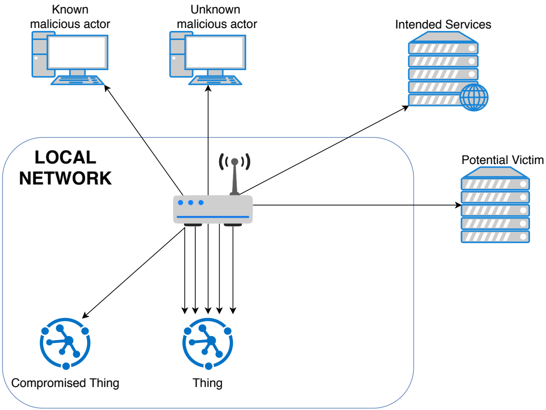
International Workshop on Edge Systems, Analytics and Networking (EdgeSys 2020)

Co-located with EuroSys 2020

Agenda

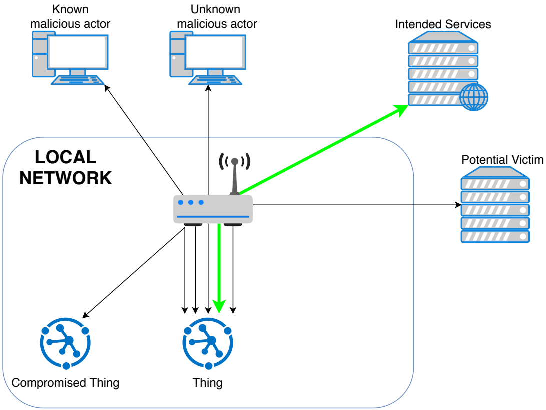
Motivations: Internet of Things

IoT devices are **resource-constrained** and **highly heterogeneous** in both underlying system capability and statistical network behaviour, and are widely distributed



Motivations: Internet of Things

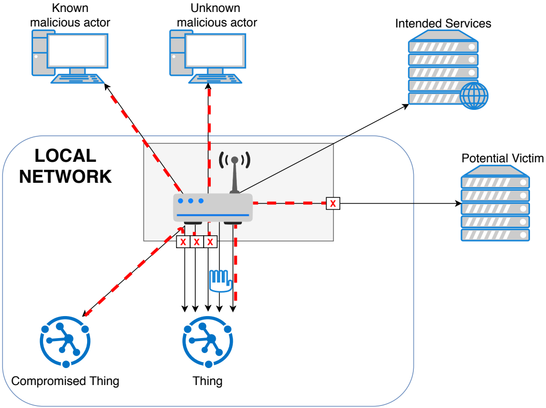
IoT devices are **resource-constrained** and **highly heterogeneous** in both underlying system capability and statistical network behaviour, and are widely distributed



Motivations: Initial Goal

Initial Goal

Improving security systems in IoT environments by preserving privacy of generated data



Agenda

Contributions

With this work we **provide**:

CoLearn an **infrastructure** that aims to create **safe deployment conditions for IoT devices**

With this work we **demonstrate**:

- an **asynchronous participation mechanism** for IoT devices in machine learning model training using a publish/subscribe architecture
- a **mechanism for reducing the attack surface in Federated Learning architecture**
- a **trade-off** between communication bandwidth usage, training time and device temperature

Agenda

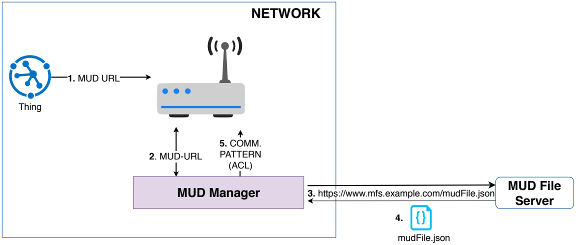
Manufacturer Usage Description Specification

IoT devices are able to **signal** to the network which **functionalities need** to properly work

The MUD standard **restricts** and **limits** traffic end-points and rates in and out of IoT devices

```
"to-device-policy": {
  "access-lists": {
  "access-list": [
   {
     "name":"mud-96898-v4to"
   },
   {
     "name": "mud-96898-v6to
        "
   }
  ]}
 }
```

```
"acl": [
{
 "name":"mud-96898-v4to",
 "type": "ipv4-acl-type",
 "aces": {
 "ace": [
 {
   "name": "cl0-todev",
   "matches": {
     "ipv4": {
       "ietf-acldns:src-
          dnsname": "cloud-
          service.example.
          com"
     }
   },
     "actions": {
       "forwarding": "accept
          "
     }
 }]}
}
```



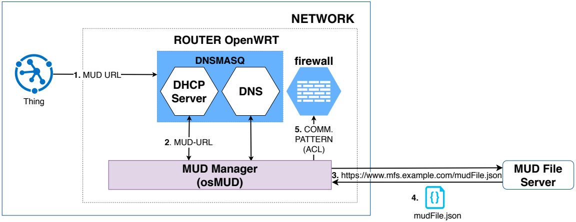
Manufacturer Usage Description Specification

IoT devices are able to **signal** to the network which **functionalities need** to properly work

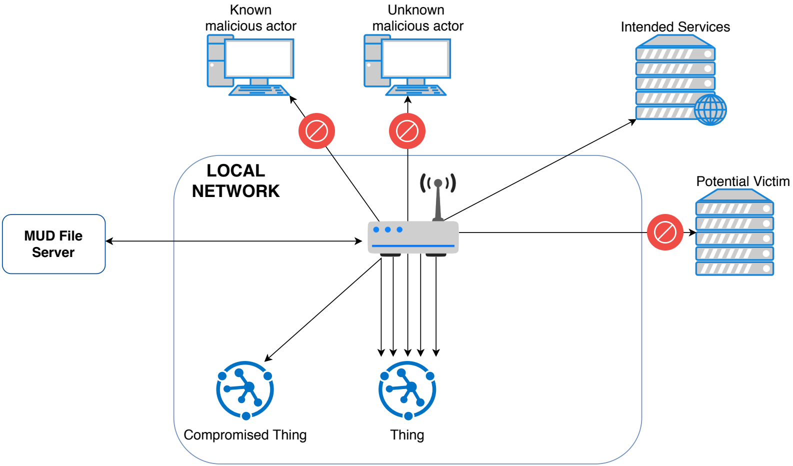
The MUD standard **restricts** and **limits** traffic end-points and rates in and out of IoT devices

```
"to-device-policy": {
  "access-lists": {
  "access-list": [
    {
      "name":"mud-96898-v4to"
    },
    {
      "name": "mud-96898-v6to
        "
    }
  ]}
}
```

```
"acl": [
{
 "name":"mud-96898-v4to",
 "type": "ipv4-acl-type",
 "aces": {
 "ace": [
 {
   "name": "cl0-todev",
   "matches": {
     "ipv4": {
       "ietf-acldns:src-
          dnsname": "cloud-
          service.example.
          com"
     }
   },
     "actions": {
       "forwarding": "accept
          "
     }
 }]}
}
```



MUD compliant Network



MUD Problems

Problem 1

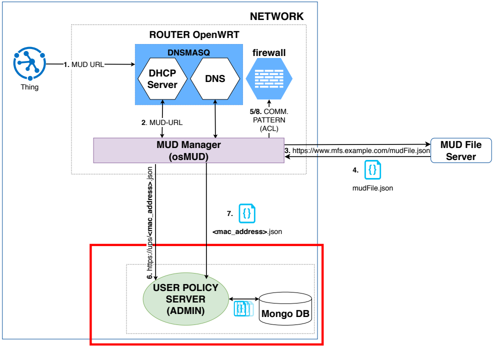
MUD rules could be not sufficient, even if all devices are MUD compliant: **individual users may have their deployment setup which may require specific rules**

Problem 2

Manufacturers **are not able to define rules for IoT devices that behave as general purpose devices** (Alexa, Google Home, smartphone etc.), and users as well

User Policy Server (UPS)

It provides the opportunity to an administrator/end-user to **interact with MUD components through a user-friendly interface**, thus allowing to **define rules suitable for the network in which MUD is deployed**[1]



The administrator rules are defined through specific MUD Files (UPS Files).

[1] SoK: Beyond IoT MUD Deployments – Challenges and Future Directions, *https://arxiv.org/abs/2004.08003*

Agenda

Motivations

Contributions

Manufacturer Usage Description (RFC 8520)

Federated Learning

CoLearn

Conclusion and Future Works

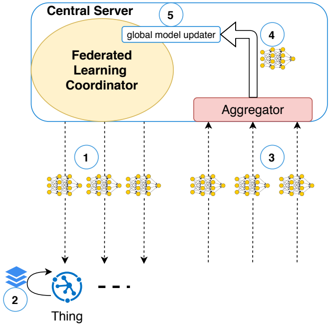
Federated Learning: Overview

"bringing the code to the data, instead of the data to the code"

This approach allows to do **model learning on edge-devices**, while **keeping all the training data on them**

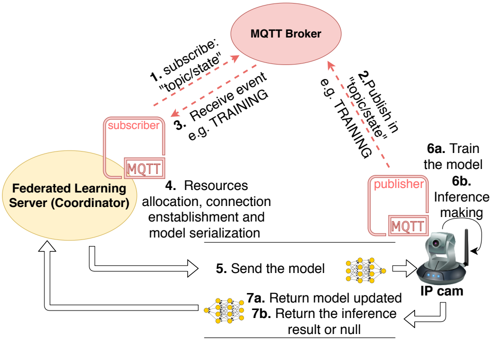
Implementation problems:

1. **Model distribution**
2. **Device's state communication**
3. **Training requests management**
4. Model cryptography (?)



Federated Learning: Our approach

- **Model distribution**
  → **PySyft framework** that employs **WebSockets** to communicate the global model to Federated Learning participants and is built on top of **PyTorch**
- **Device's state communication**
  → Pattern **publish/subscribe** implemented through **MQTT**
  → Three states: **TRAINING**, **INFERENCE**, **NOT_READY**



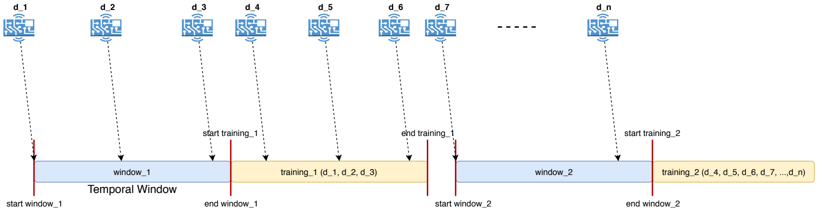
Federated Learning: Our approach

- Training requests management

→We introduce the **temporal window** concept, in which the Coordinator **waits and collects training requests**.

→The devices can **remove** or **drop** themselves from the Coordinator's devices list

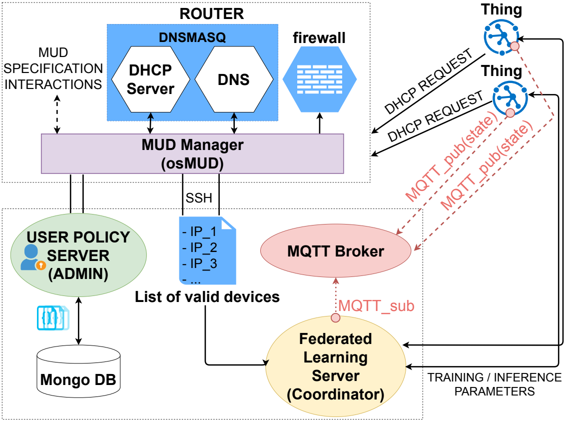
→ Useful to define **lower bound threshold**, **upper bound threshold** and **device selection criteria**



Agenda

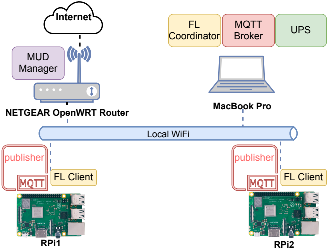
CoLearn: MUD and FL together

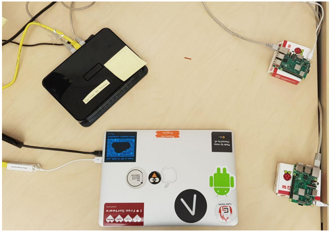
- Introduction of an **entity** hosting UPS and FL Coordinator
- **Device filtering**: only MUD compliant devices can participate in the Federated Learning Protocol



Deployment

- **Router**: NETGEAR WNDR 3700v2
- **Machine hosting UPS and Coordinator**: MacBook Pro Intel Core i5 e 8 GB RAM
- **Edge devices**: two Raspberry Pi 3B+ running FL clients and supporting the Python environment needed for PySyft.
- **Data-set**: Bot-IoT Dataset[2]
- **Computational model**: Feed-Forward neural network (2 hidden layers, one with 50 neurons and the other with 30 neurons, an input size of 10)



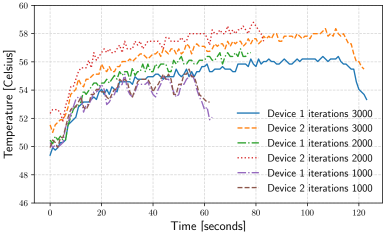


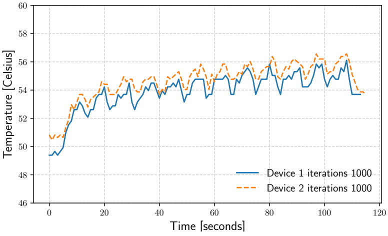
[2] *https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/bot_iot.php*

CoLearn evaluations

Experiments performed: Temperature monitoring, Bandwidth monitoring, Training Loss, Training Time

- The **number of iterations** influences the **temperature** of the components involved

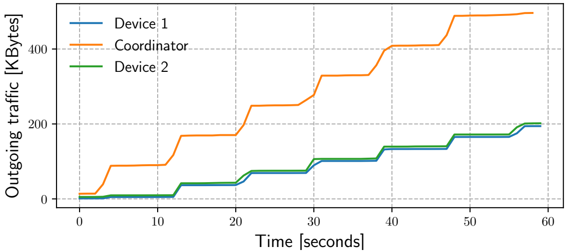




CoLearn evaluations

Experiments performed: Temperature monitoring, Bandwidth monitoring, Training Loss, Training Time

- The **number of iterations** influences the **temperature** of the components involved
- **Total outgoing traffic**, as expected, is strictly correlated to the **model dimension**, **number of rounds**, and the **number of devices** involved



CoLearn evaluations

Experiments performed: Temperature monitoring, Bandwidth monitoring, Training Loss, Training Time

- The **number of iterations** influences the **temperature** of the components involved
- **Total outgoing traffic**, as expected, is strictly correlated to the **model dimension**, **number of rounds**, and the **number of devices** involved
- As expected, the **total training time** increases with total number of iterations

	Iterations	Rounds	Total iterations	Training time (s)	Training loss
1	1000	3	3000	26.868	0.001814
2	1000	6	6000	53.148	0.001068
3	1000	12	12000	105.921	0.000863
4	2000	3	6000	38.378	0.00107
5	2000	6	12000	76.139	0.000877
6	3000	3	9000	56.467	0.000957
7	3000	6	18000	112.247	0.000852

CoLearn evaluations

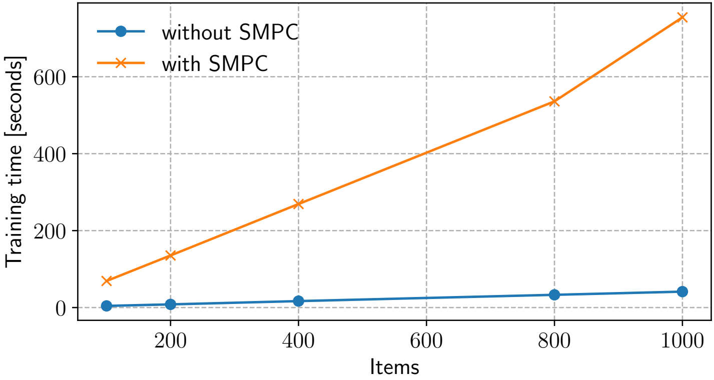
Experiments performed: Temperature monitoring, Bandwidth monitoring, Training Loss, Training Time

- The **number of iterations** influences the **temperature** of the components involved
- **Total outgoing traffic**, as expected, is strictly correlated to the **model dimension**, **number of rounds**, and the **number of devices** involved
- As expected, the **total training time** increases with total number of iterations

	Iterations	Rounds	Total iterations	Training time (s)	Training loss
1	1000	3	3000	26.868	0.001814
2	1000	6	6000	53.148	0.001068
3	1000	12	12000	105.921	0.000863
4	2000	3	6000	38.378	0.00107
5	2000	6	12000	76.139	0.000877
6	3000	3	9000	56.467	0.000957
7	3000	6	18000	112.247	0.000852

Impact of Privacy Preservation via SMPC

Secure Multi-Party Computation: it replaces the key concept with party concept



Agenda

Conclusions

In summary we provided:

- a **user-friendly interface** able to interact with MUD components
- infrastructure **Federated Learning based** able to interact with real devices
- a direction to optimise the Federated Learning trade-off
- infrastructure that can use and train **anomaly detection models** and ready for **Transfer Learning**
- to the best of our knowledge, the first deployment **hosting both MUD and FL**

Discussion and Future Works

In the **current CoLearn** deployment:

- we assumed that edge devices (RPis) do not fail in the training phases and during their activity of traffic eavesdropping.
- we did not focus on IoT device identification and authentication, which is vital for both MUD-compliant networks and FL architecture

Future CoLearn deployment could include:

- Extension of YANG-based MUD file by adding a field containing structure and weights of a model
- Improving of UPS functionalities
- Adaptive temporal window sizing

Q&A

Thanks!

Questions?

For more info, please contact:
Angelo Feraudo <angelo.feraudo@studio.unibo.it>
(<aferaudo34@gmail.com>)
Dr Poonam Yadav <poonam.yadav@york.ac.uk>
(@pooyadav)