Dataset Generation and Feature Extraction for High-Traffic Environment Personal Tracker Identification (Midterm)

23.05.2023

Stefan Richard Saxer
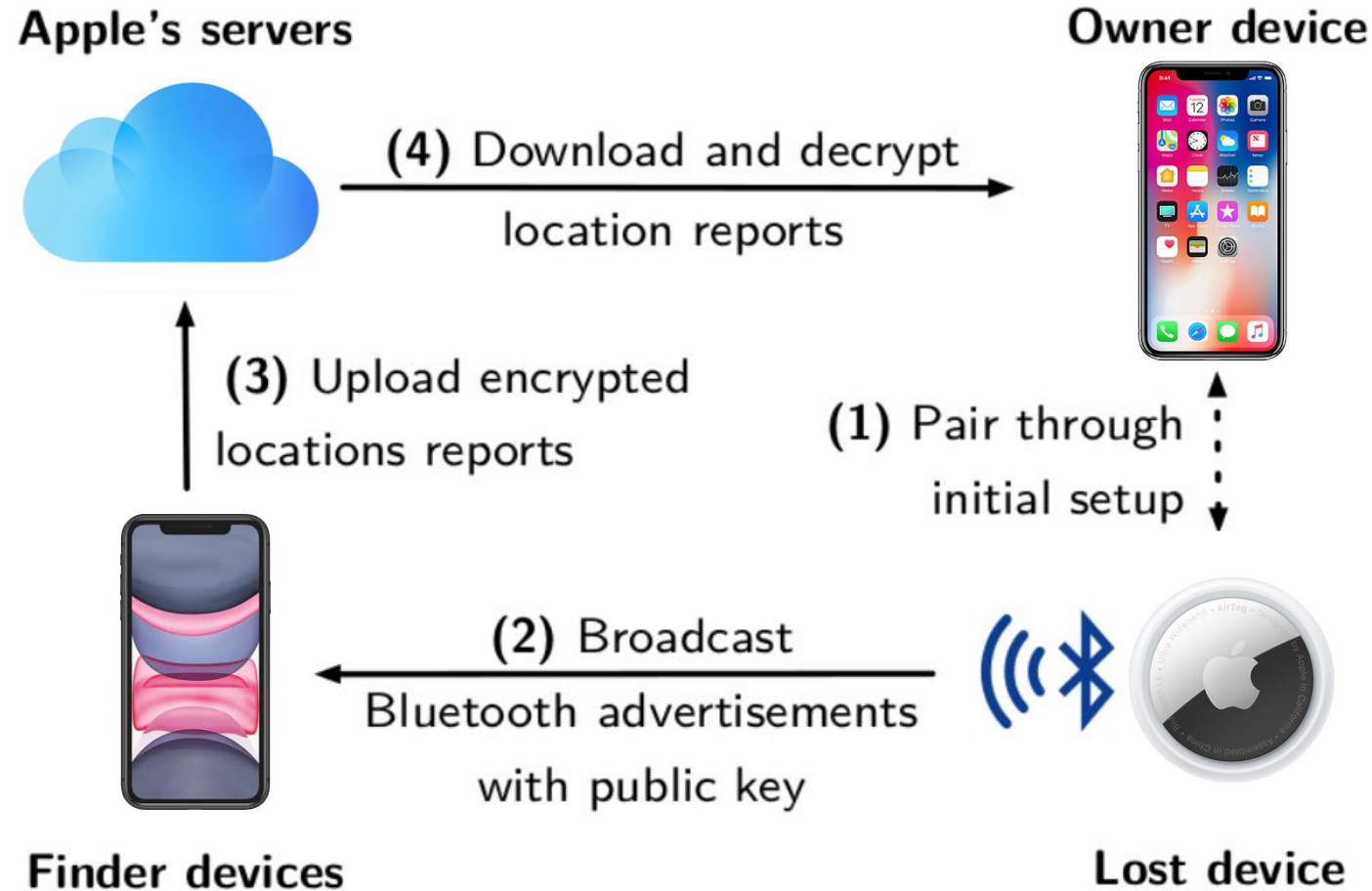
Introduction

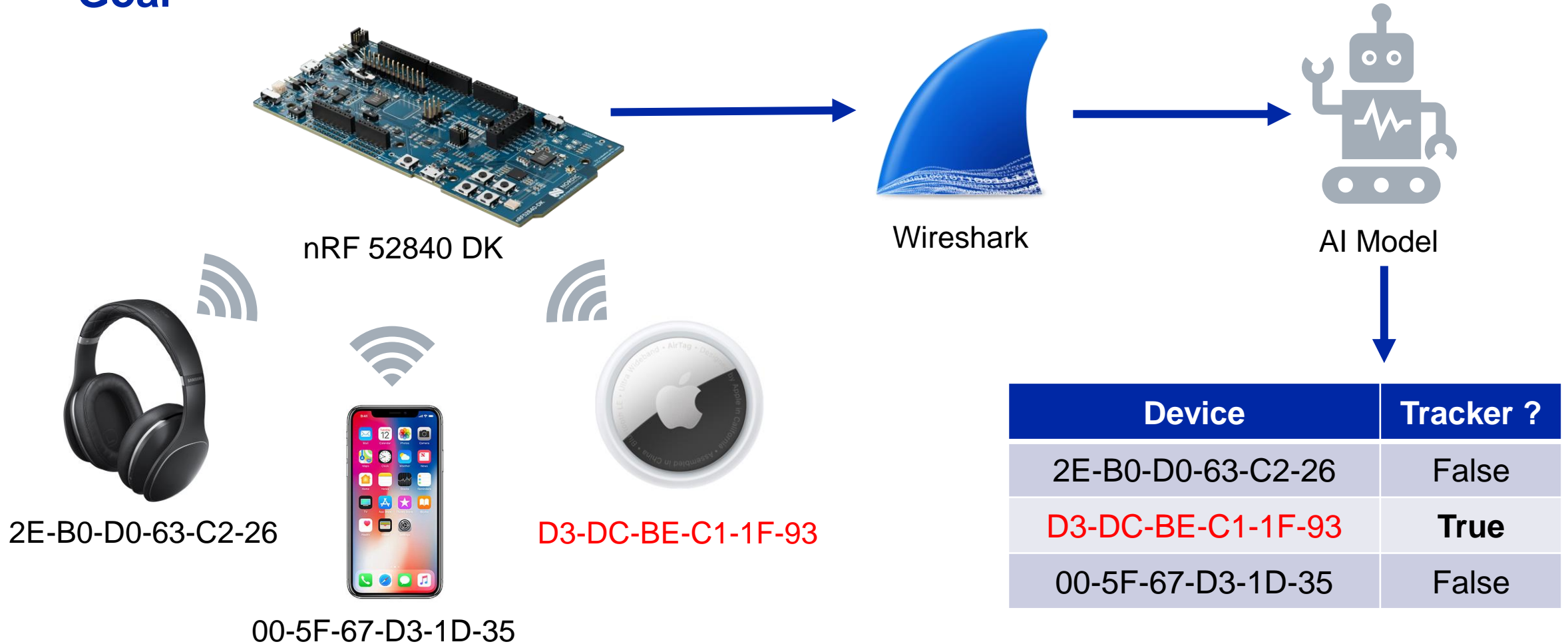# BLE Tracking Devices (Find My)

Goal

nRF 52840 DK

Wireshark

AI Model

2E-B0-D0-63-C2-26

00-5F-67-D3-1D-35

D3-DC-BE-C1-1F-93

Device	Tracker ?
2E-B0-D0-63-C2-26	False
D3-DC-BE-C1-1F-93	**True**
00-5F-67-D3-1D-35	False

How ?

to .csv

Non-Tracker

training

to .csv

Tracker

training

to .csv

Non-Tracker

training

AI Model
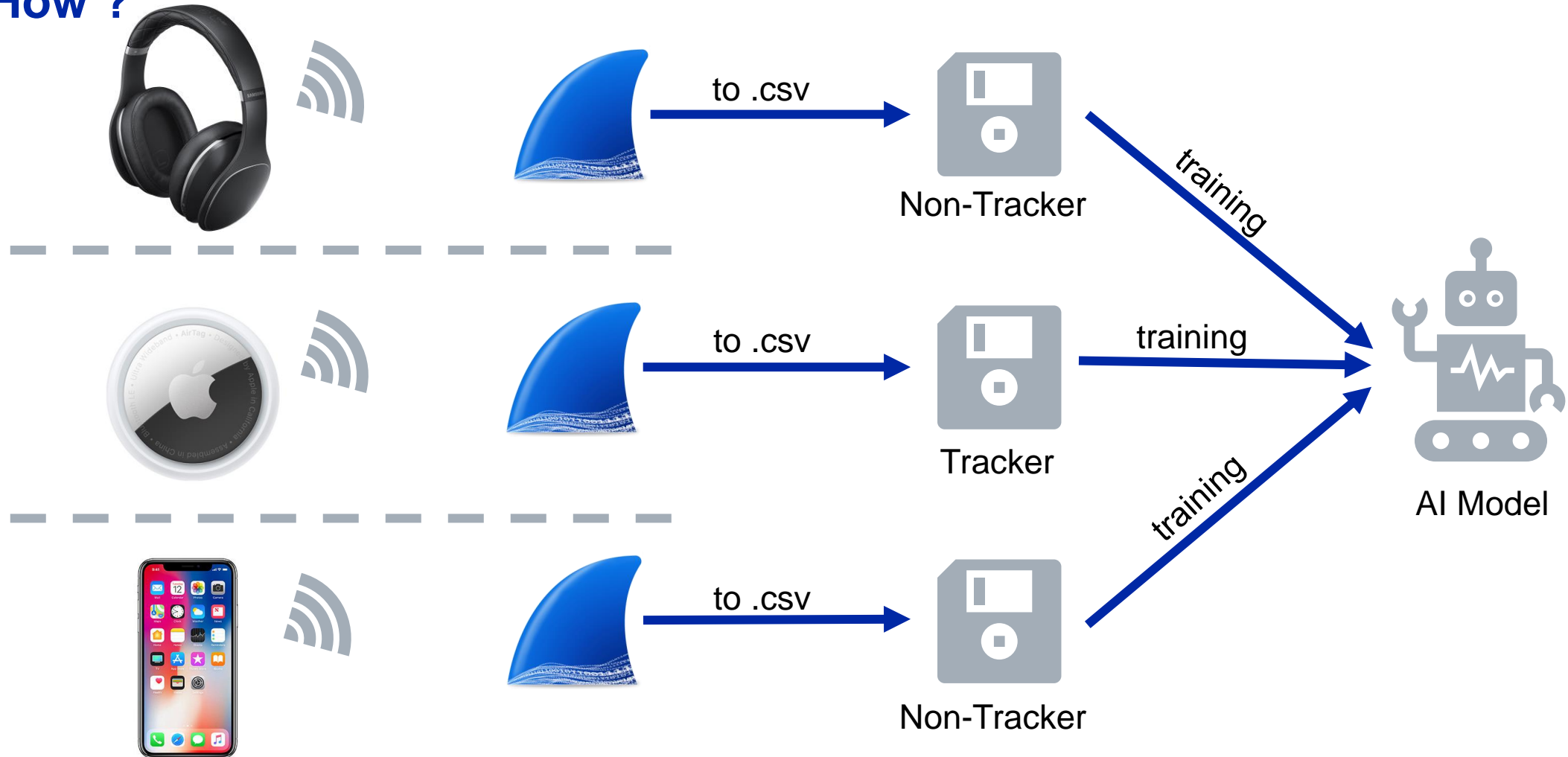
My Workbench @ Home
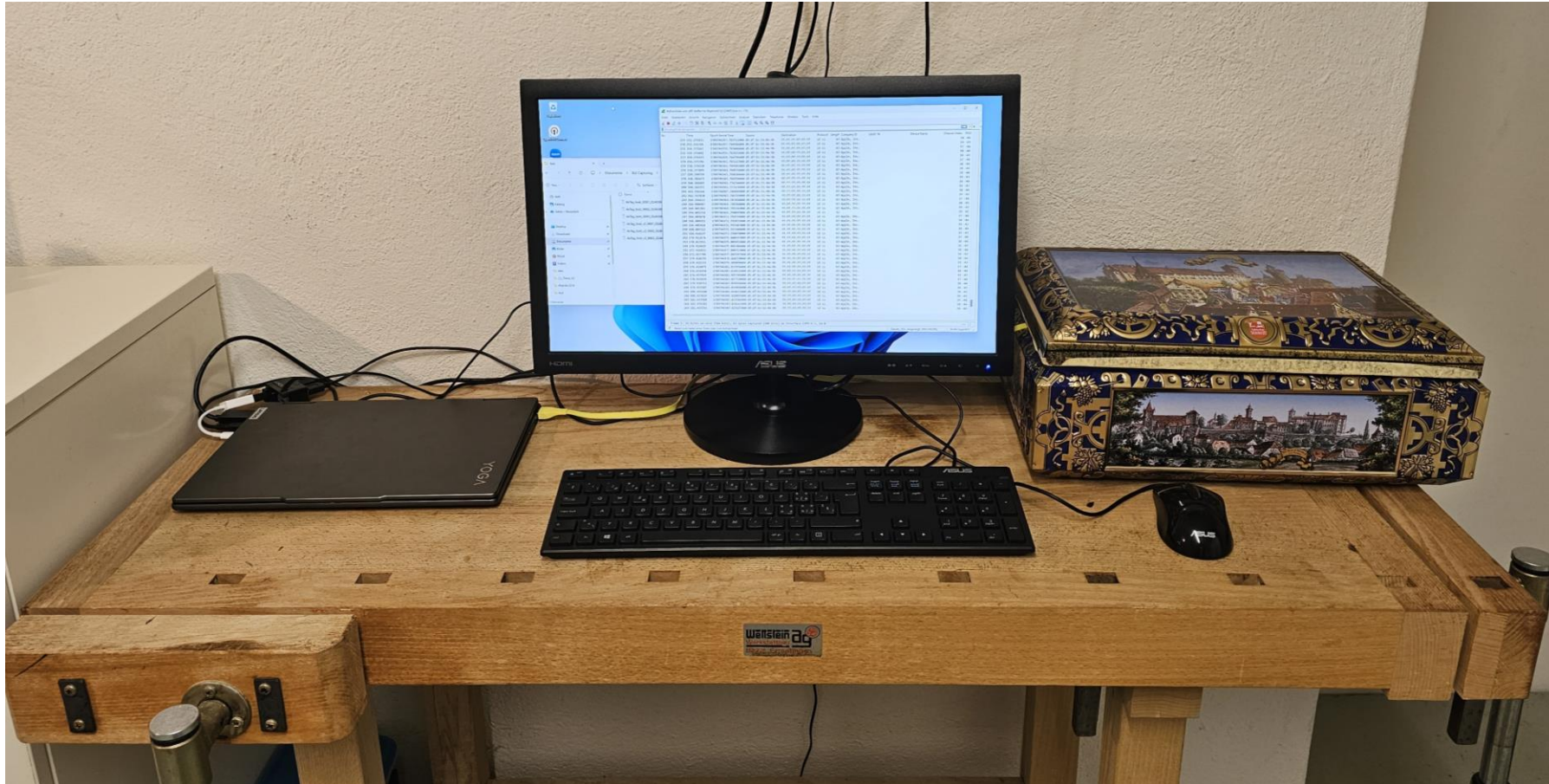
Faraday Cage

How ?
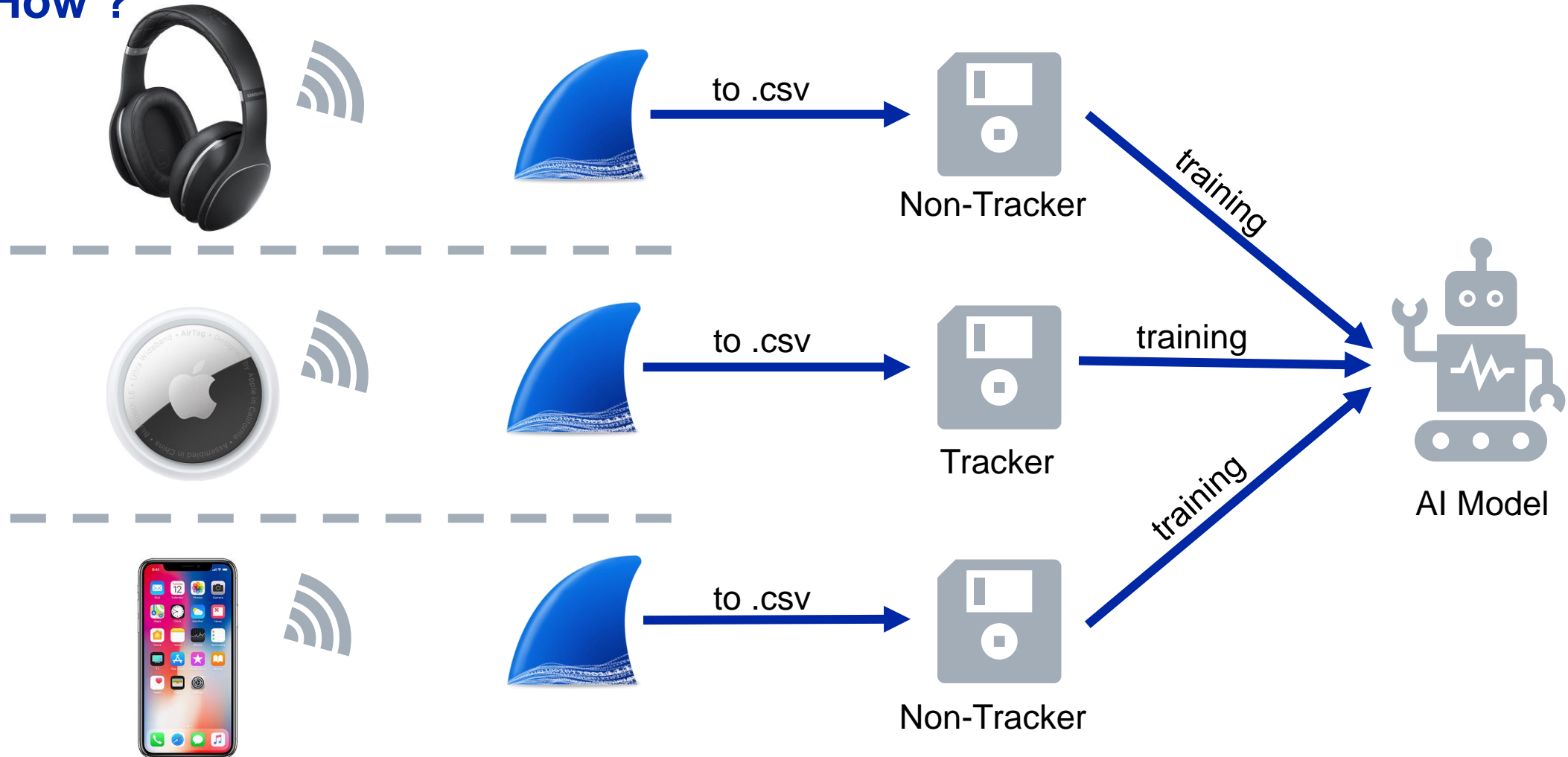
BLE Packet Format
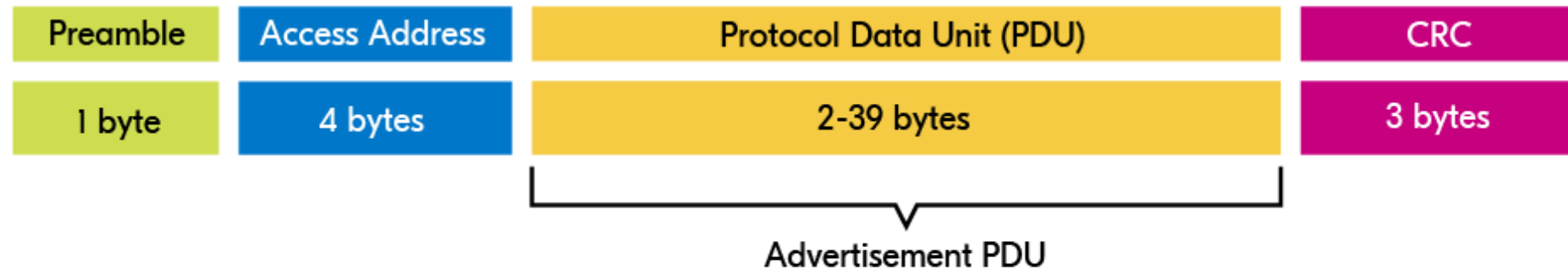
BLE Advertisement Payload Types

– **Manufacturer Specific Data**

  – Company ID (Ex: "Apple, Inc.")

  – Data: hex Number

– **Service Data**

  – UUID (Ex: "Tile, Inc.")

  – Data: hex Number

– **Device Name**

  – Name (Ex: "JBL Headphone")

– …

AirTag Example Packet

```
˅ Bluetooth Low Energy Link Layer
      Access Address: 0x8e89bed6
  ˃   Packet Header: 0x2560 (PDU Type: ADV_IND, ChSel: #2, TxAdd: Random)
      Advertising Address: fd:e5:68:1f:9f:9c
  ˅   Advertising Data
      ˅  Manufacturer Specific
            Length: 30
            Type: Manufacturer Specific (0xff)
            Company ID: Apple, Inc. (0x004c)
         ˃  Data: 121910bc3904307fa14c973cb6aefd08ba26426c0a15fb2a4c022c
      CRC: 0x0f0801
```

Tile Tracker Example Packet

```
Bluetooth Low Energy Link Layer
    Access Address: 0x8e89bed6
>   Packet Header: 0x1b60 (PDU Type: ADV_IND, ChSel: #2, TxAdd: Random)
    Advertising Address: db:7e:1b:05:af:8e
v   Advertising Data
  v   Flags
        Length: 2
        Type: Flags (0x01)
        000. .... = Reserved: 0x0
        ...0 .... = Simultaneous LE and BR/EDR to Same Device Capable (Host): false (0x0)
        .... 0... = Simultaneous LE and BR/EDR to Same Device Capable (Controller): false (0x0)
        .... .1.. = BR/EDR Not Supported: true (0x1)
        .... ..1. = LE General Discoverable Mode: true (0x1)
        .... ...0 = LE Limited Discoverable Mode: false (0x0)
  v   Service Data - 16 bit UUID
        Length: 13
        Type: Service Data - 16 bit UUID (0x16)
        UUID 16: Tile, Inc. (0xfeed)
        Service Data: 02008f4d8cee199f3295
  v   16-bit Service Class UUIDs
        Length: 3
        Type: 16-bit Service Class UUIDs (0x03)
        UUID 16: Tile, Inc. (0xfeed)
    CRC: 0xaa2815
```

The BLE Tracker Market in Switzerland

| Air Tag (Apple) | ONE (Chipolo) | SkyTag (4Smarts) | SmartTag (Samsung) | Tile |

Find My Network

Dataset and Training Devices

other Device
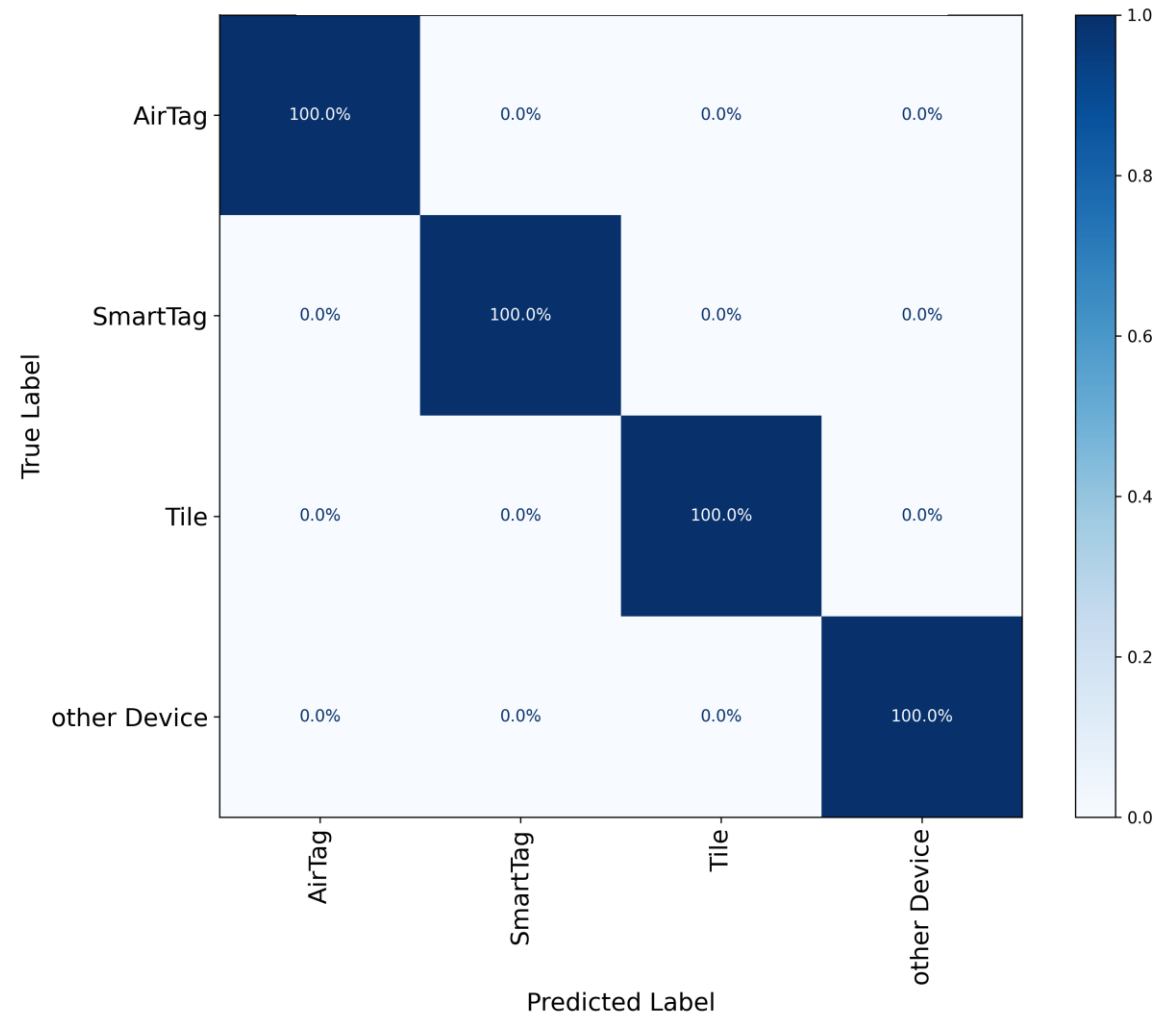

AirTag

Tile

SmartTag

Feature Extraction

Company ID «Apple»	UUID «Samsung»	UUID «Tile»	Label
1	0	0	AirTag
0	**1**	0	SmartTag
0	0	**1**	Tile
0	0	0	other Device

Evaluation of AI Model

"AI Model"

```python
def predictLabel(row):
    if bool(row['UUID Tile']):
        return 'Tile'
    elif bool(row['UUID Samsung']):
        return 'SmartTag'
    elif bool(row['Company ID Apple']):
        return 'AirTag'
    else:
        return 'other Device'
```

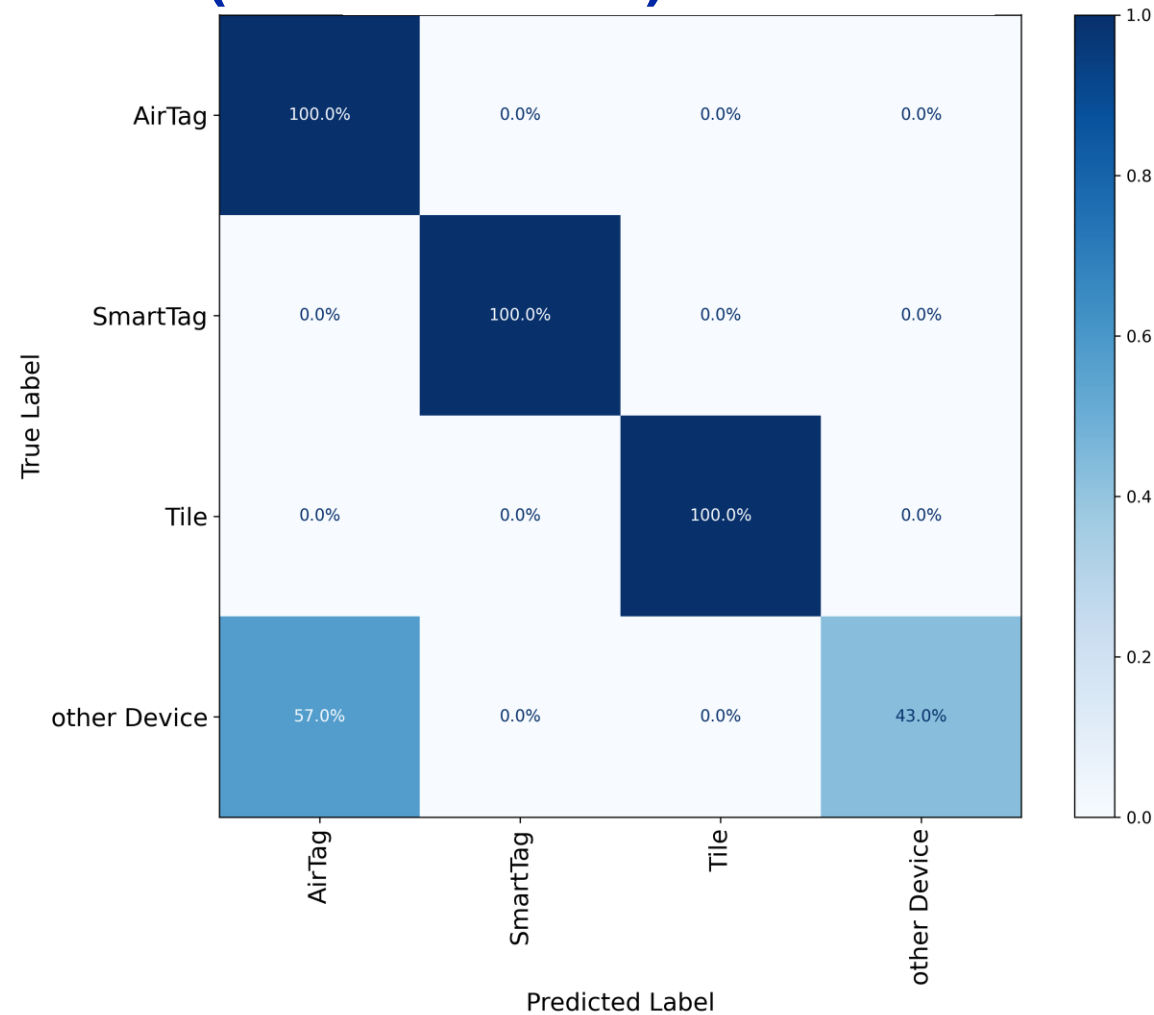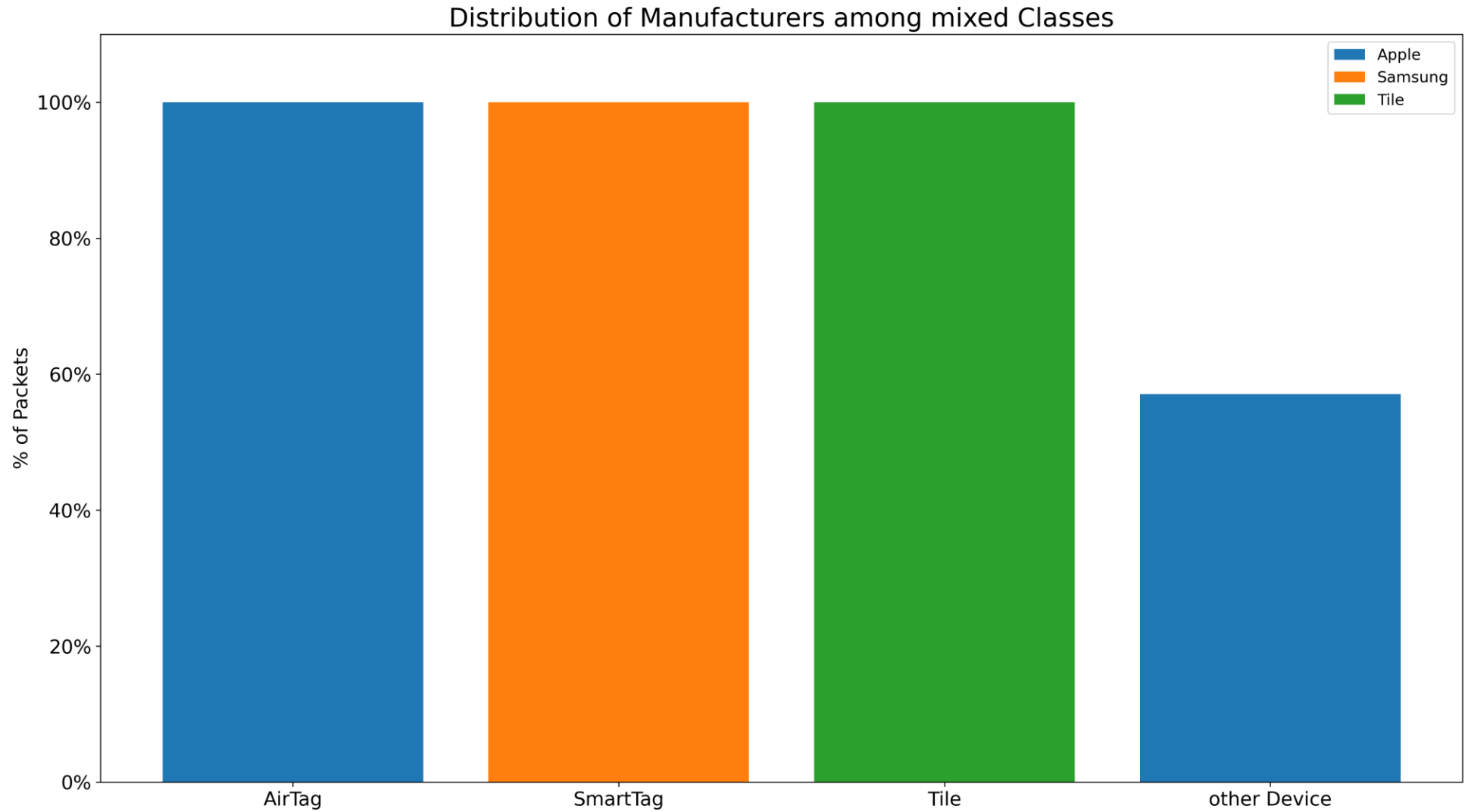# More Training Devices

other Device

AirTag

Tile

SmartTag

Evaluation of Model (Decision Tree)

Distribution of Manufacturers among mixed Classes

States of Trackers ?
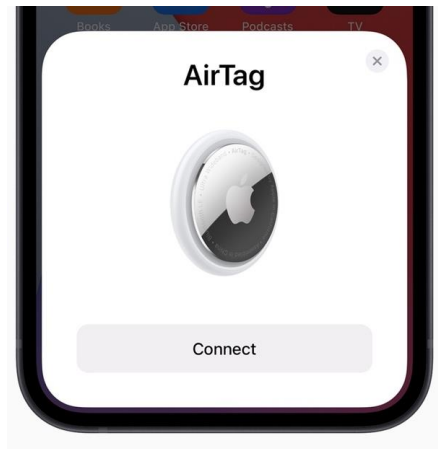
| powerless | unpaired | nearby owner Device | lost |

Low Threat ➞ **High Threat**

Trackers in the "lost" State
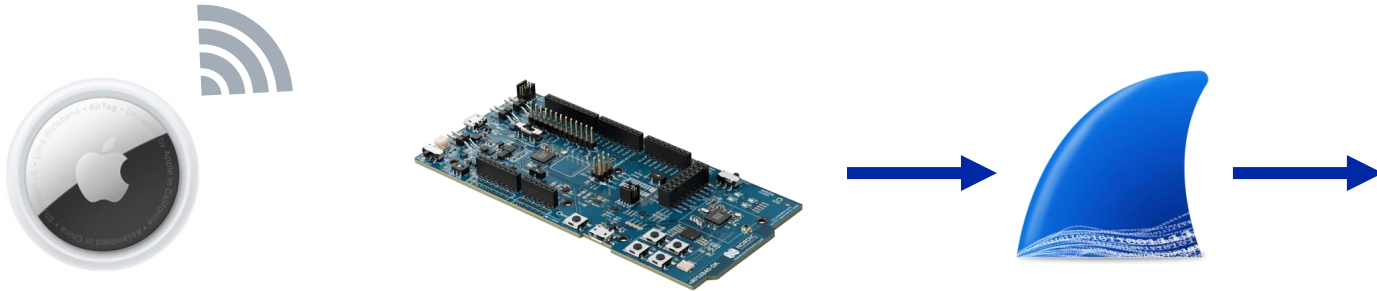
Packet Nr.	Source
1	2E-B0-D0-63-C2-26
2	2E-B0-D0-63-C2-26
3	2E-B0-D0-63-C2-26
…	…
4'638	D3-DC-BE-C1-1F-93
…	…

Trackers in the "nearby" State
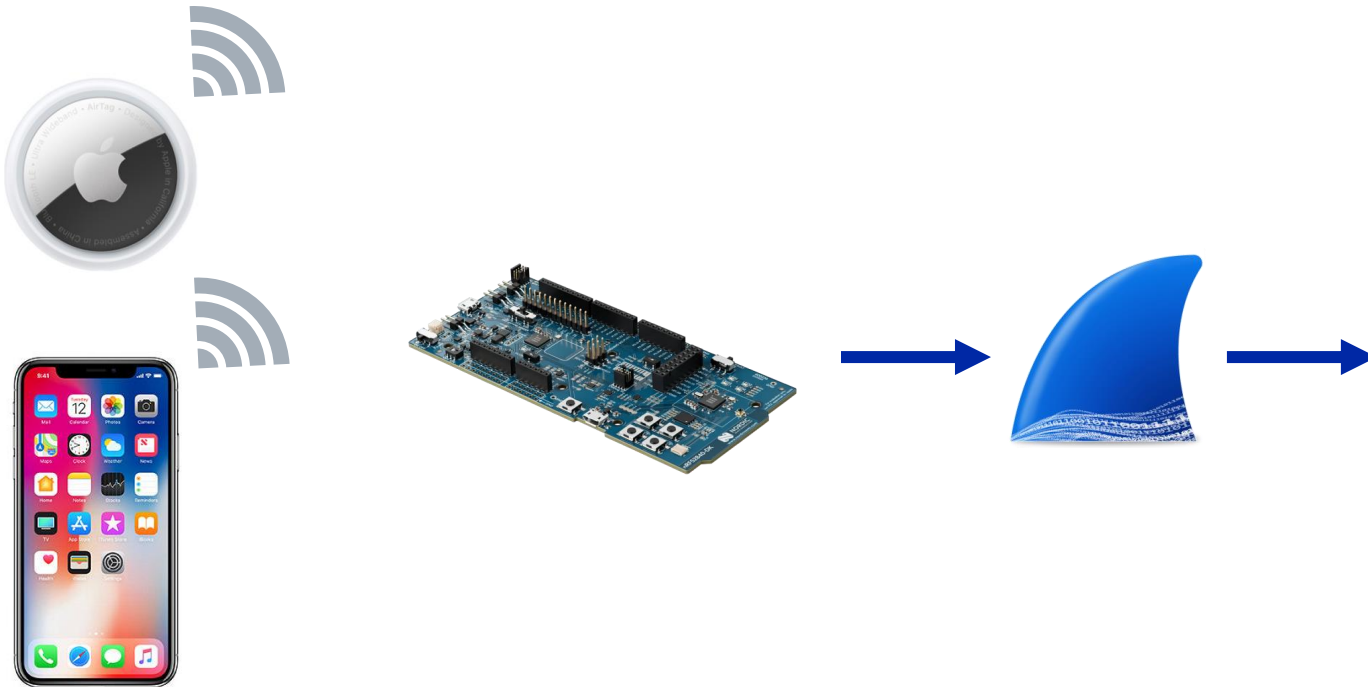
Packet Nr.	Source	Label
1	2E-B0-D0-63-C2-26	?
2	00-5F-67-D3-1D-35	?
3	00-5F-67-D3-1D-35	?
4	2E-B0-D0-63-C2-26	?
…	…	…
631'253	D3-DC-BE-C1-1F-93	?

Trackers in the "nearby" State

Packet Nr.	Source	Label
1	2E-B0-D0-63-C2-26	AirTag
2	2E-B0-D0-63-C2-26	AirTag

Trackers in the "nearby" State

Packet Nr.	Source	Label
1	2E-B0-D0-63-C2-26	AirTag
2	2E-B0-D0-63-C2-26	AirTag
3	00-5F-67-D3-1D-35	?
4	2E-B0-D0-63-C2-26	?
5	2E-B0-D0-63-C2-26	?
6	00-5F-67-D3-1D-35	?
7	00-5F-67-D3-1D-35	?

Trackers in the "nearby" State

Packet Nr.	Source	Label
1	2E-B0-D0-63-C2-26	AirTag
2	2E-B0-D0-63-C2-26	AirTag
3	00-5F-67-D3-1D-35	iPhone
4	2E-B0-D0-63-C2-26	AirTag
5	2E-B0-D0-63-C2-26	AirTag
6	00-5F-67-D3-1D-35	iPhone
7	00-5F-67-D3-1D-35	iPhone

Trackers in the "nearby" State

Packet Nr.	Source	Label
1	2E-B0-D0-63-C2-26	AirTag
2	2E-B0-D0-63-C2-26	AirTag
3	00-5F-67-D3-1D-35	iPhone
4	2E-B0-D0-63-C2-26	AirTag
5	2E-B0-D0-63-C2-26	AirTag
6	00-5F-67-D3-1D-35	iPhone
7	00-5F-67-D3-1D-35	iPhone
…	…	…
10'272	2E-B0-D0-63-C2-26	AirTag

Trackers in the "nearby" State

Packet Nr.	Source	Label
1	2E-B0-D0-63-C2-26	AirTag
2	2E-B0-D0-63-C2-26	AirTag
3	00-5F-67-D3-1D-35	iPhone
4	2E-B0-D0-63-C2-26	AirTag
5	2E-B0-D0-63-C2-26	AirTag
6	00-5F-67-D3-1D-35	iPhone
7	00-5F-67-D3-1D-35	iPhone
…	…	…
10'272	2E-B0-D0-63-C2-26	AirTag
10'273	C9-2D-8B-7F-9B-A6	?

ca. 10 Min of labelled data

Semi-Unsupervised Labelling
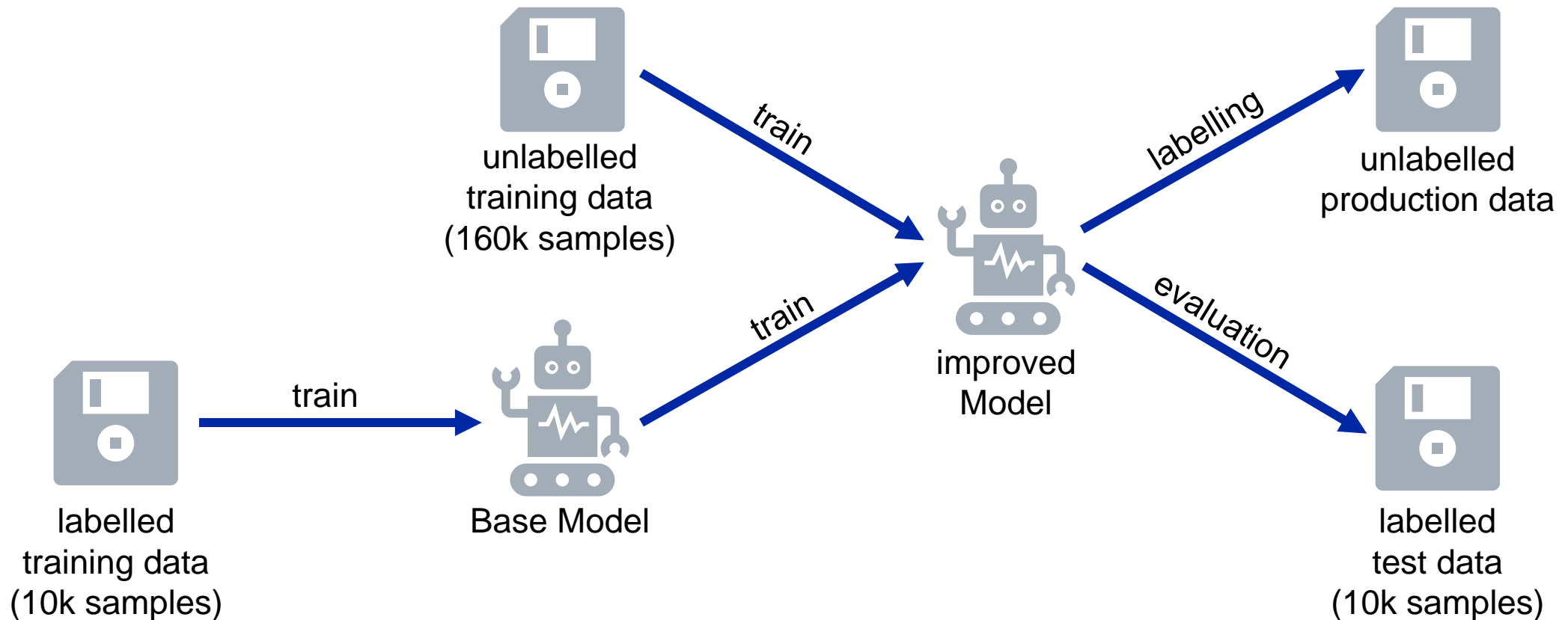
Semi-Unsupervised Labelling
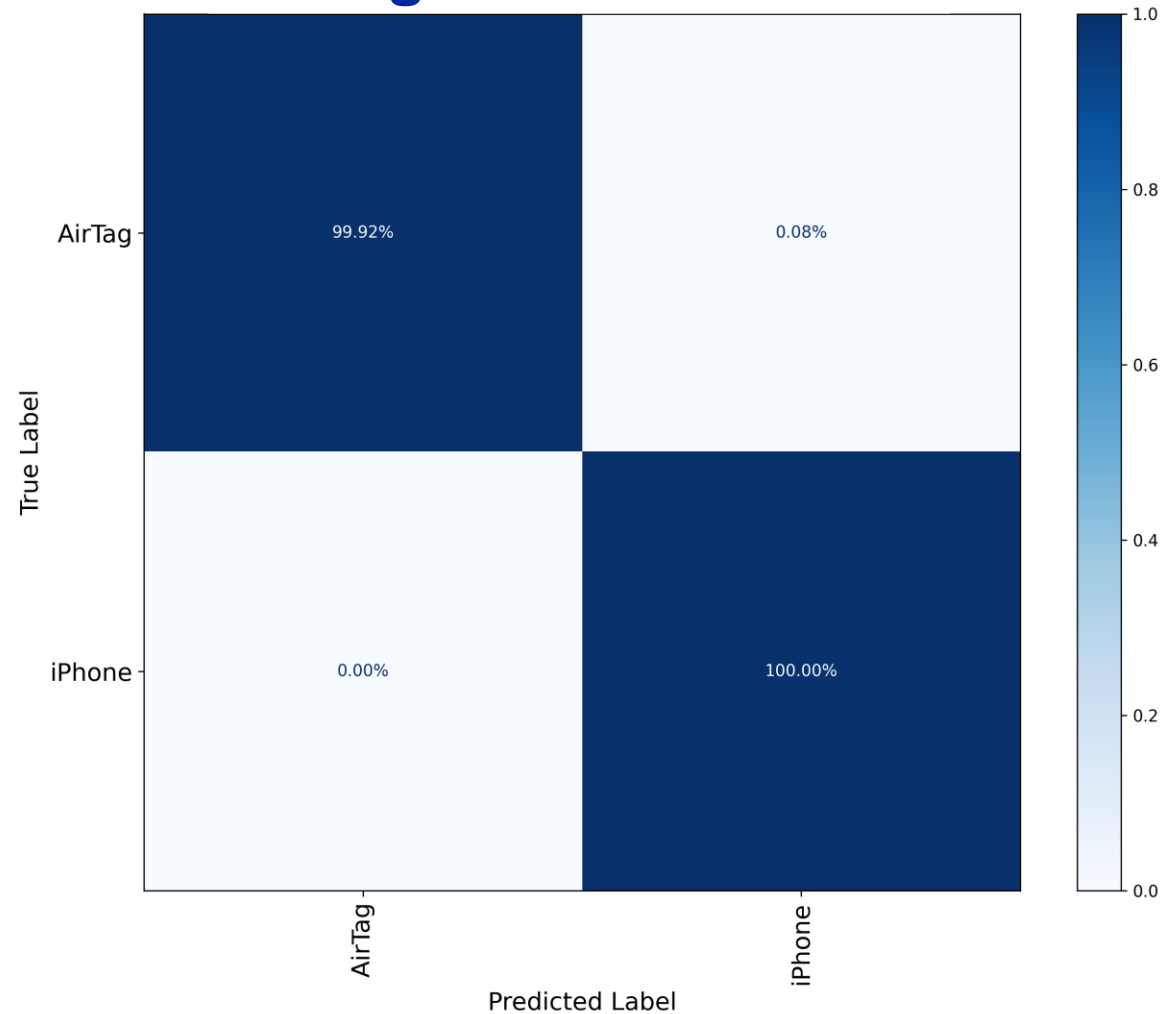
Data Pre-Processing Pipeline
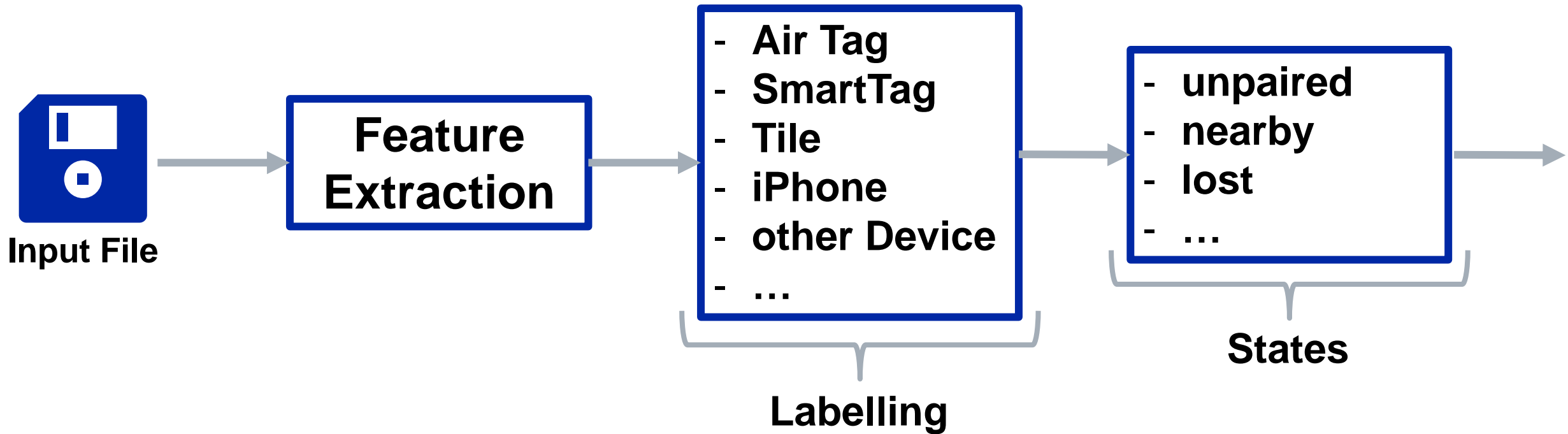
Task-Group-Framework

UML Class Diagram: Task-Group-Framework
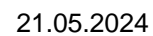
```
Analysis Pipeline
    10 Pre Processing: Pre Processing
            10 Select and order columns
            15 Company ID and UUID
                    10 Fill Company ID with None
                    20 Fill UUID with None
                    30 Replace Company IDs
                    40 Replace Company UUIDs
            20 MS Data Processing
                    10 Fill MS Data with empty String
                    20 Length of MS Data
                    30 Continuity Type
                    40 Drop MS Data Column
            25 Service Data Processing
                    10 Fill Service Data with empty String
                    20 Length of Service Data
                    30 Samsung Type
                    40 Drop Service Data Column
            30 Fill Numeric NA with 0
            40 Fill String NA with None
            50 Broadcast
            60 Datetime conversion
            70 Malformed Packet
            80 Order DataFrame
            90 Convert object type to string
    20 Dummy Processing: Dummy Processing
            20 Dummies Protocol
            30 Dummies Channel
            40 Dummies AD Type
            50 Dummies Company
            55 Dummies UUID
            70 Dummies PDU Type
            80 Dummies Continuity Type
            90 Dummies SmartTag Type
    30 Labeling: Labeling
            10 Labeling auto: Labeling auto
                    Label AirTag and iPhone: Label AirTag and iPhone
                    Label Chipolo and iPhone: Label Chipolo and iPhone
                    Label SkyTag and iPhone: Label SkyTag and iPhone
                    Label Tile and iPhone: Label Tile and iPhone
                    Label AirPod and iPhone: Label AirPod and iPhone
                    Label other Device and SmartTag: Label SmartTag and other Device
            20 Labeling manual: Labeling manual
                    Label AirTag: Label AirTag
                    Label Chipolo: Label Chipolo
                    Label SkyTag: Label SkyTag
                    Label iPhone: Label iPhone
                    Label iPad: Label iPad
                    Label MacBook: Label MacBook
                    Label AirPod: Label AirPod
                    Label SmartTag: Label SmartTag
                    Label Tile: Label Tile
                    Label other Device: Label other Device
    40 States: States
            10 State Column
            20 States iDevices: States iDevices
                    Continuity
                    online: State online
                    offline: State offline
            30 States Trackers: States Tracker
                    lost: State lost
                    nearby: State nearby
                    unpaired: State unpaired
                    searching: State searching
            Collapse State Column
    80 Modeling: Modeling
            0 Drop Column Continuity
            10 Drop Columns
```

Notebook on GitHub

Midterm Presentation (23.05.2024)

```python
In [1]: import sys
        sys.path.insert(0, '..')

        import pandas as pd
        import numpy as np
        from src import Task, TaskGroup, Flag
        from src.pre_processing.interfaces.executor_interface import ExecutorInterface

        import warnings
        warnings.filterwarnings("ignore")
```

```python
In [2]: matrix = pd.DataFrame(np.identity(5, dtype=int), columns=[0, 1, 2, 3, 4])

        def plot(df):
            print(df.to_string(index=False, header=False))

        plot(matrix)
```

```
1 0 0 0 0
0 1 0 0 0
0 0 1 0 0
0 0 0 1 0
0 0 0 0 1
```

```python
In [3]: result = 3 * matrix + 1
        plot(result)
```

```
4 1 1 1 1
1 4 1 1 1
1 1 4 1 1
1 1 1 4 1
1 1 1 1 4
```

Analysis of BLE Capture

BLE Device Capture

Average Length of Manufacturer specific Data - FindMy Tracker
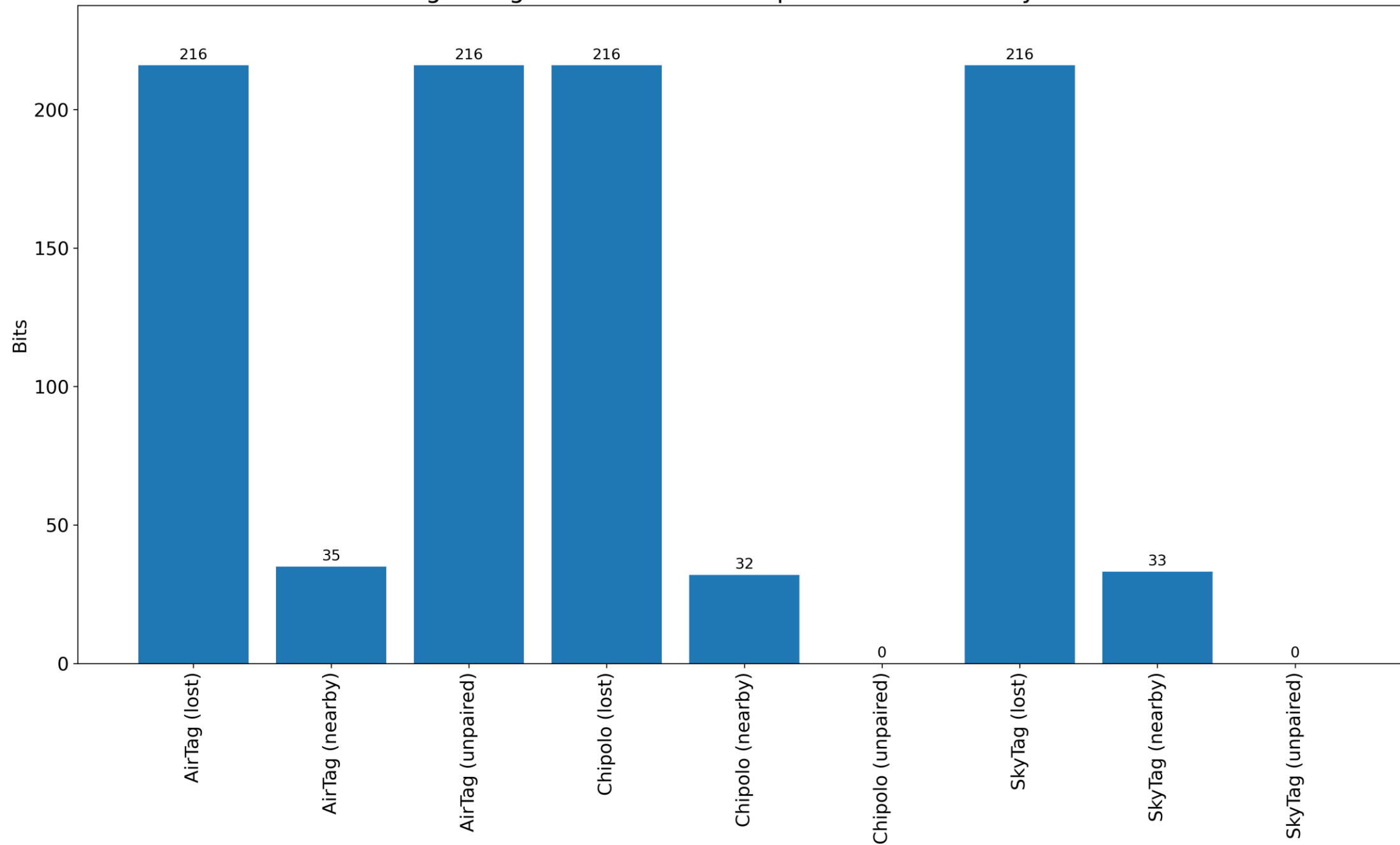
Advertisement Data Type - FindMy Tracker
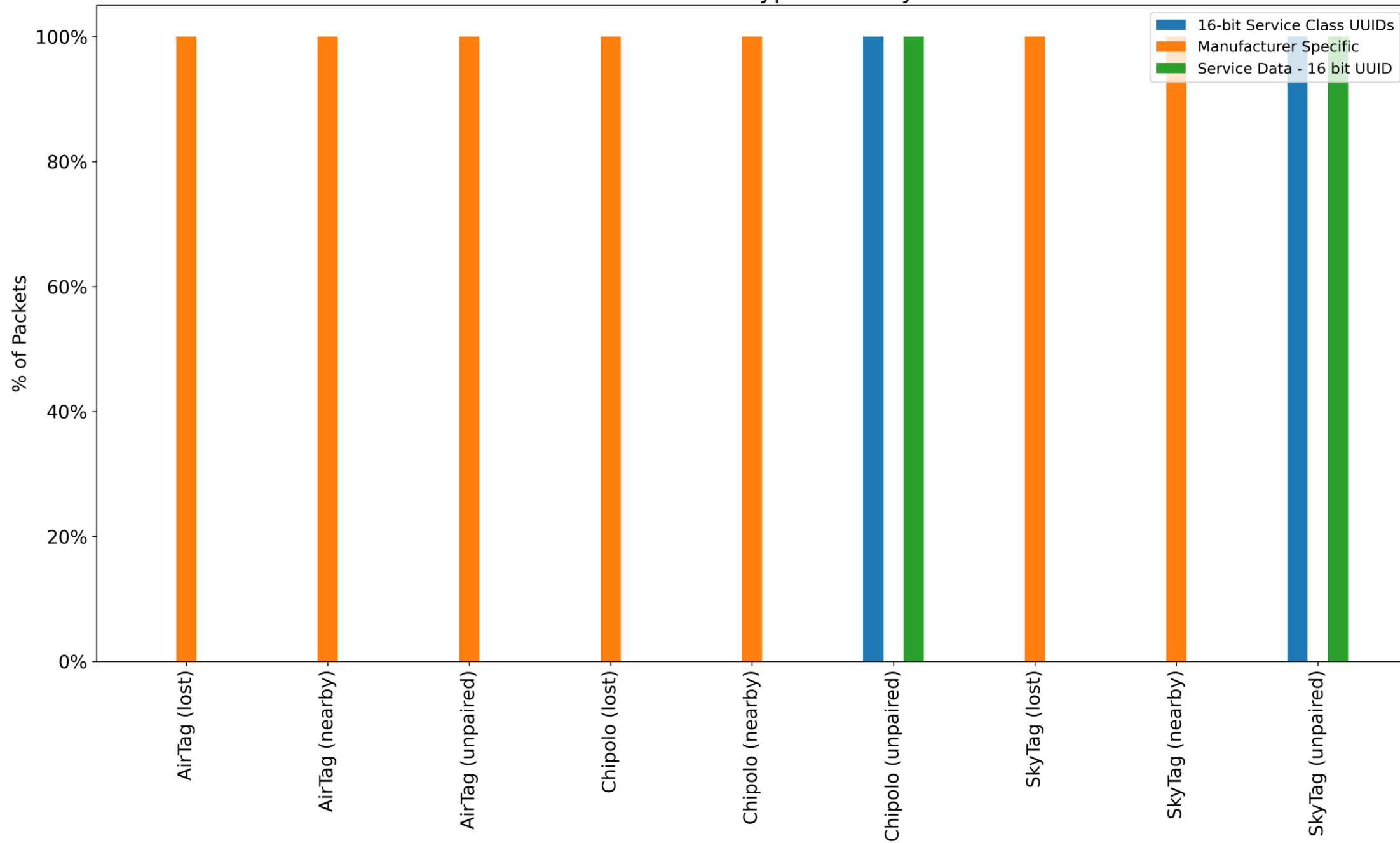
AirTag Example Packet

```
∨ Bluetooth Low Energy Link Layer
    Access Address: 0x8e89bed6
  > Packet Header: 0x2560 (PDU Type: ADV_IND, ChSel: #2, TxAdd: Random)
    Advertising Address: fd:e5:68:1f:9f:9c
  ∨ Advertising Data
    ∨ Manufacturer Specific
        Length: 30
        Type: Manufacturer Specific (0xff)
        Company ID: Apple, Inc. (0x004c)
      > Data: 121910bc3904307fa14c973cb6aefd08ba26426c0a15fb2a4c022c
    CRC: 0x0f0801
```

Continuity Type - FindMy Tracker
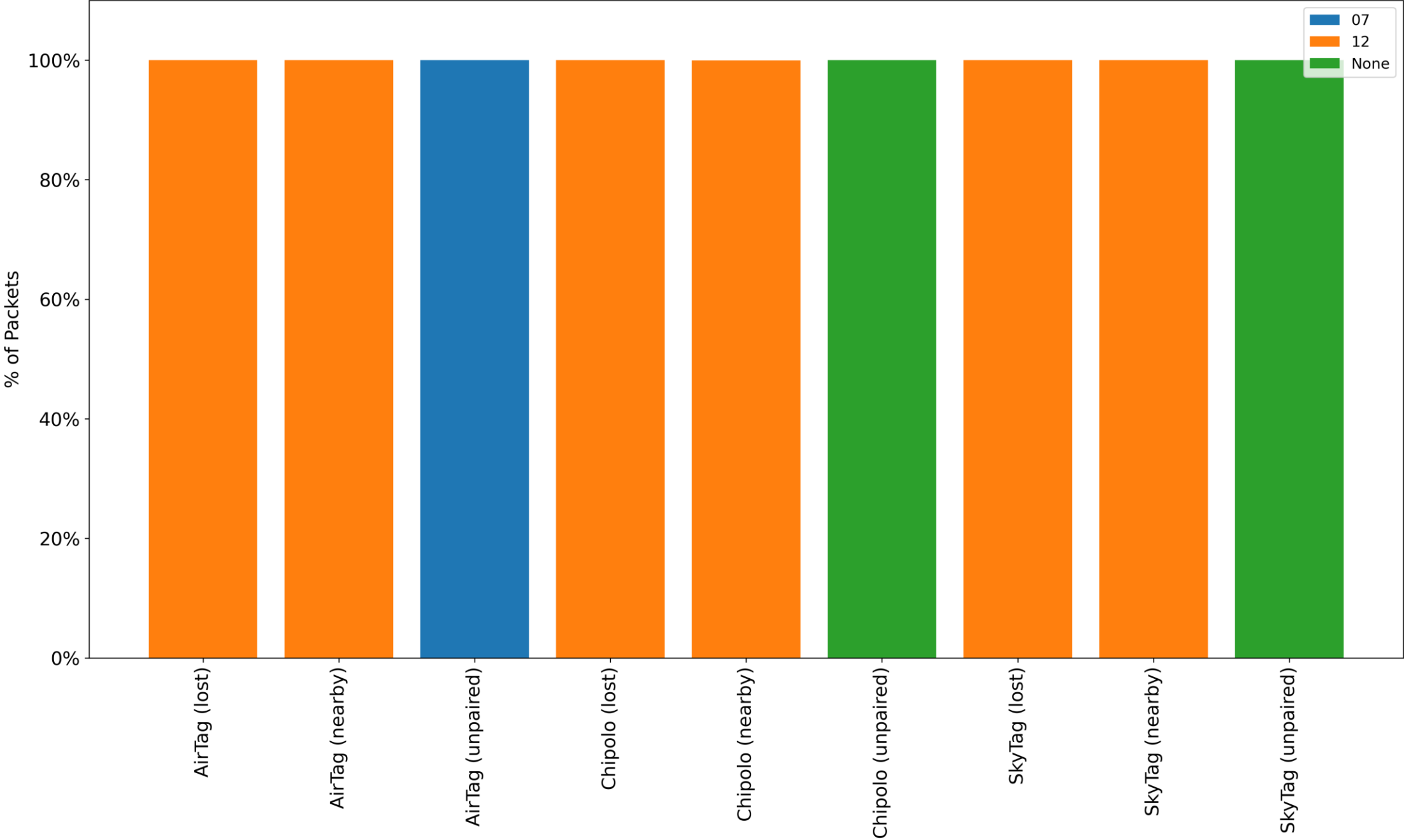
Graph of Packet Rate - AirTag (unpaired) - (AirTag_(unpaired))
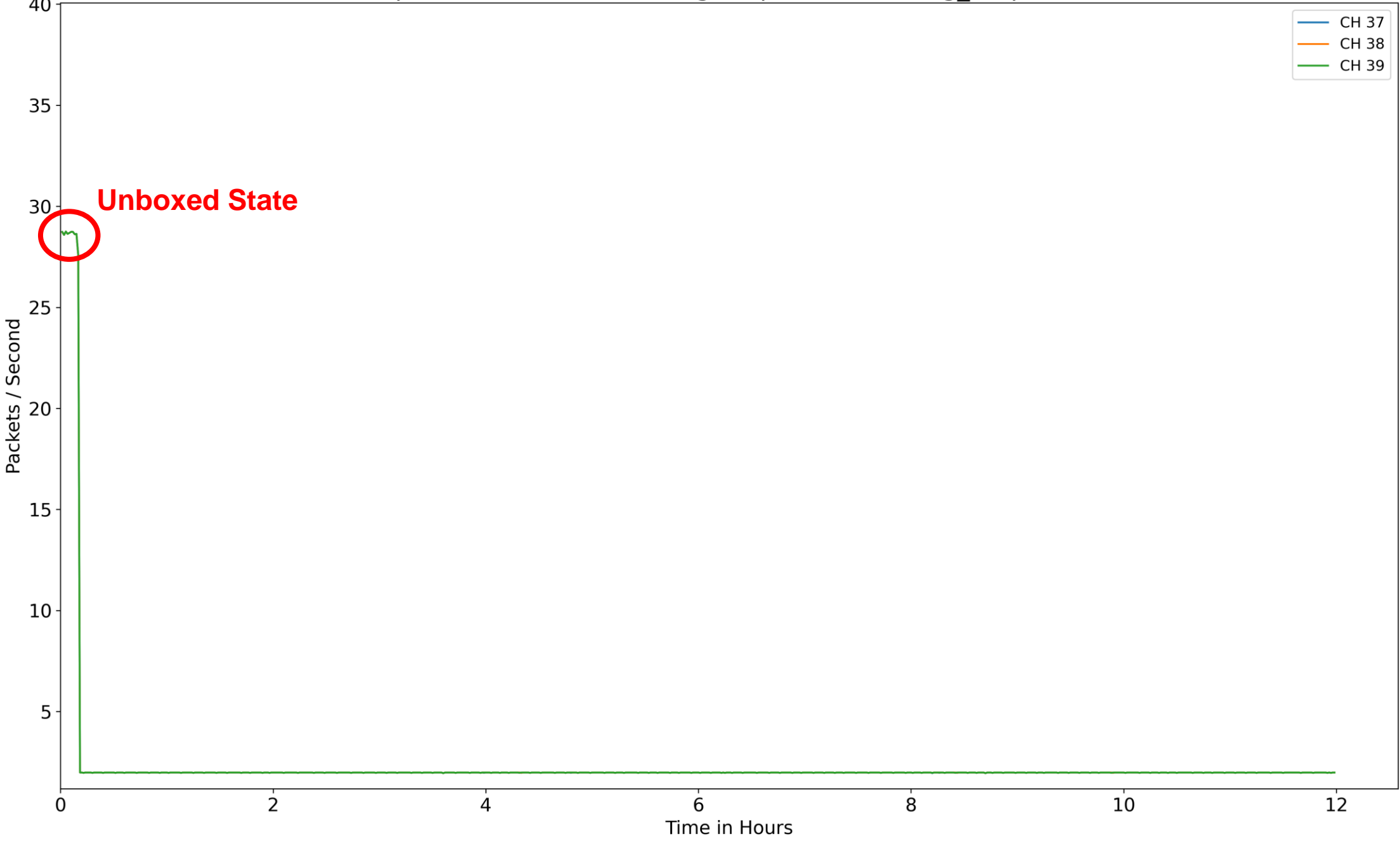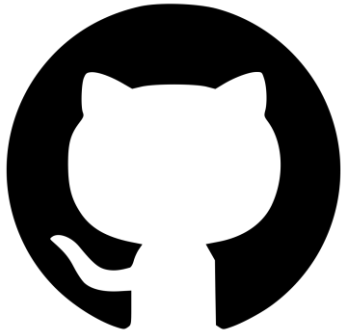
Bachelor-Thesis-Stefan-Richard-Saxer

master ▾   1 Branch   0 Tags

Go to file   t   Add file ▾   <> Code ▾

stsaxe updated introduction   00dceb5 · 13 hours ago   ⟳ 28 Commits

📁 notebooks	updated introduction	13 hours ago
📁 pickle_objects	initial	3 days ago
📁 plots	updated introduction	13 hours ago
📁 slides	updated naming	2 days ago
📁 src	changed imports	3 days ago
📁 tables	updated introduction	13 hours ago
📁 wireshark_profile	updated profile	2 days ago
📄 .gitignore	initial	3 days ago
📄 LICENSE.md	updated LICENSE	2 days ago
📄 README.md	updated readme v10	2 days ago

About

The Bachelor Thesis of Stefan Richard Saxer written in spring of 2024.

📖 Readme

⚖ Apache-2.0 license

∿ Activity

☆ 0 stars

👁 1 watching

⑂ 0 forks

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

● Jupyter Notebook 98.0%   ● Python 2.0%

📖 README   ⚖ Apache-2.0 license   ✏

Dataset Generation and Feature Extraction for High-Traffic Environment Personal Tracker Identification

This is the Bachelor Thesis of Stefan Richard Saxer, written in spring of 2024 with the Communication Systems Group at the University of Zurich (UZH).
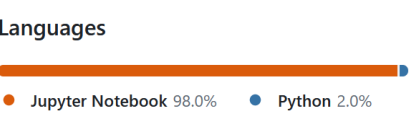
Questions ?

Thank you for your attention.