



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

15.07.2024

Stefan Richard Saxer

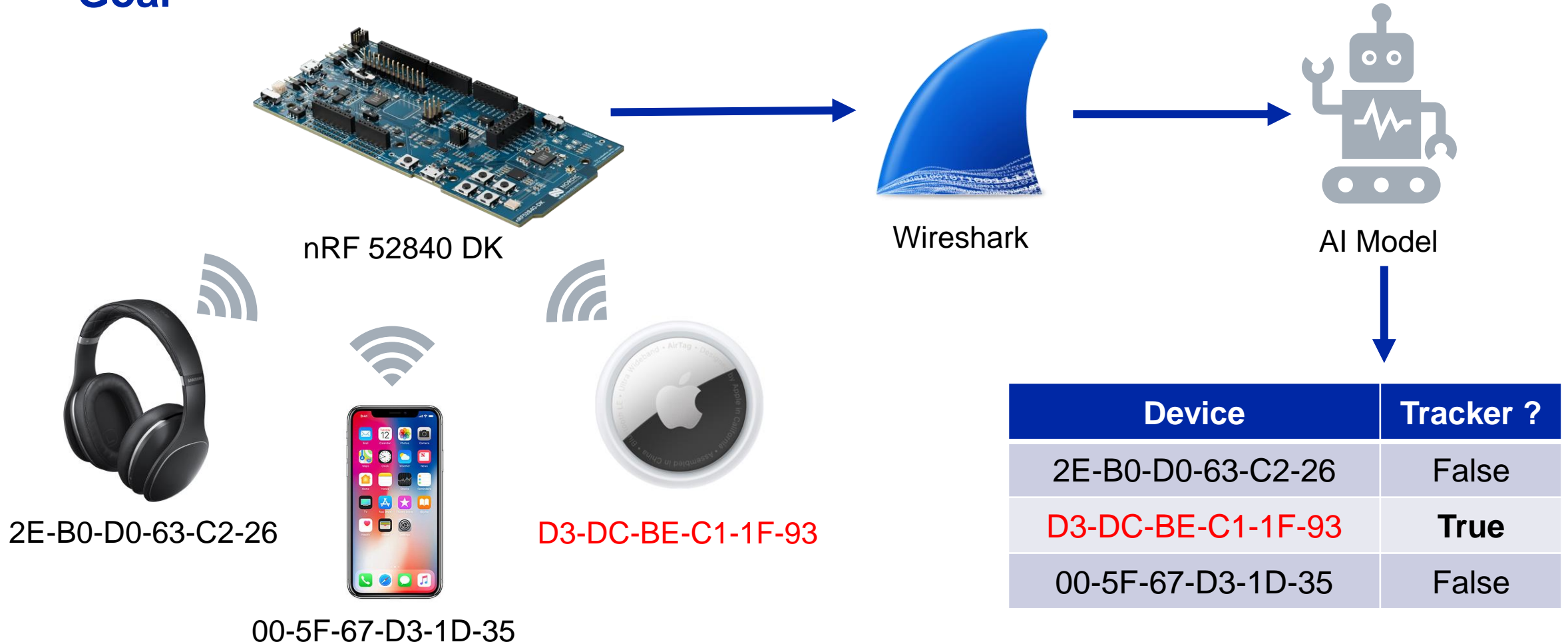


**University of
Zurich** ^{UZH}

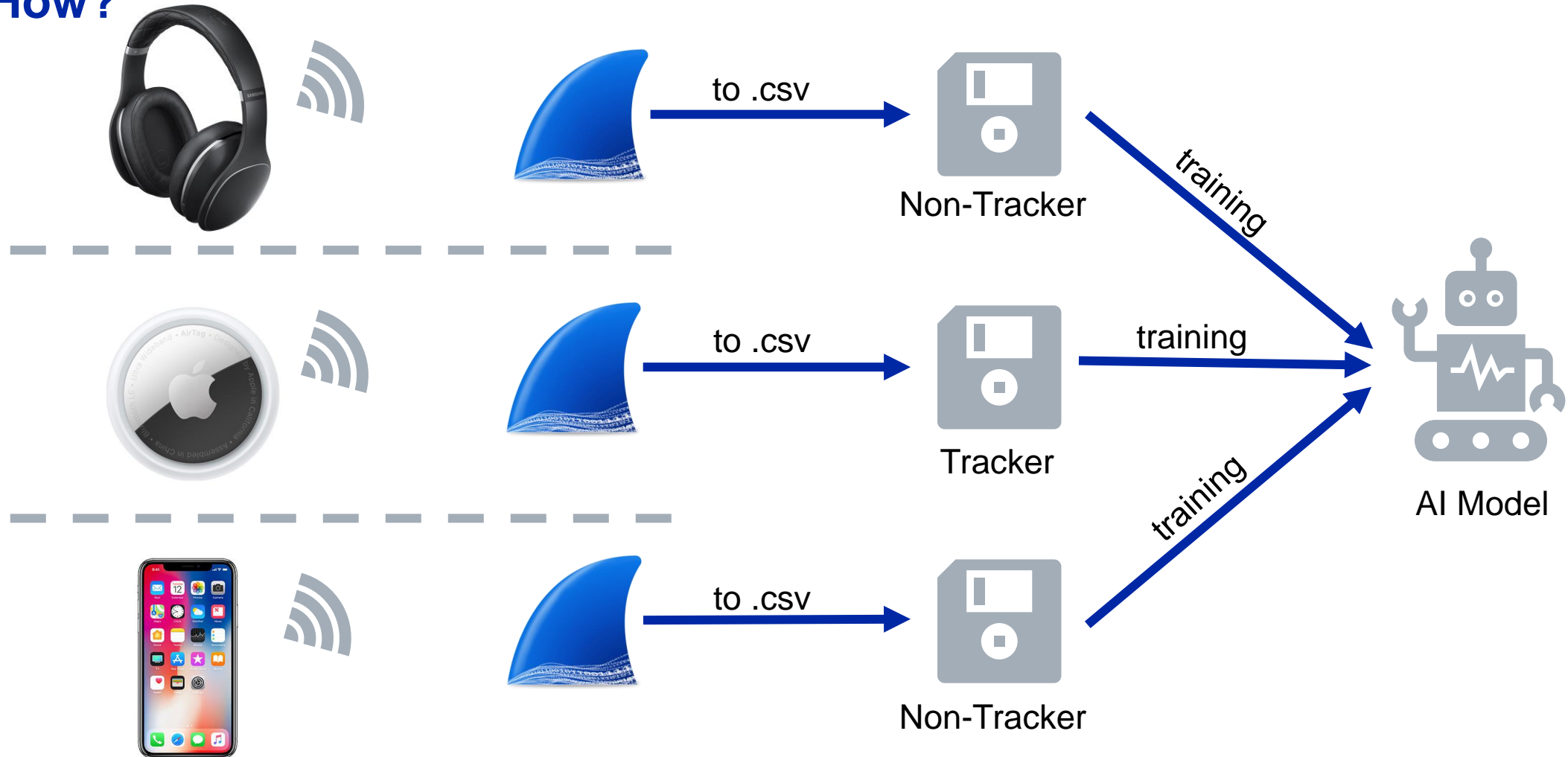
Department of Informatics - CSG

Recap from Midterm

Goal



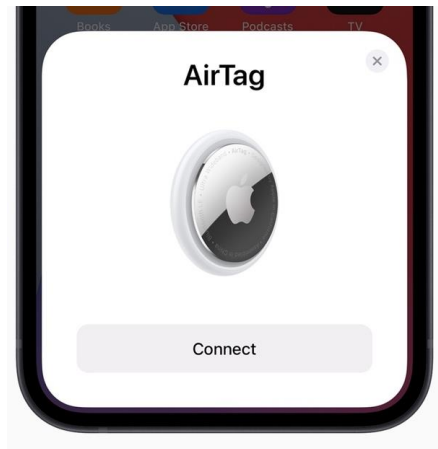
How?



States of Trackers



powerless



unpaired



nearby owner Device



lost

Low Threat



High Threat

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

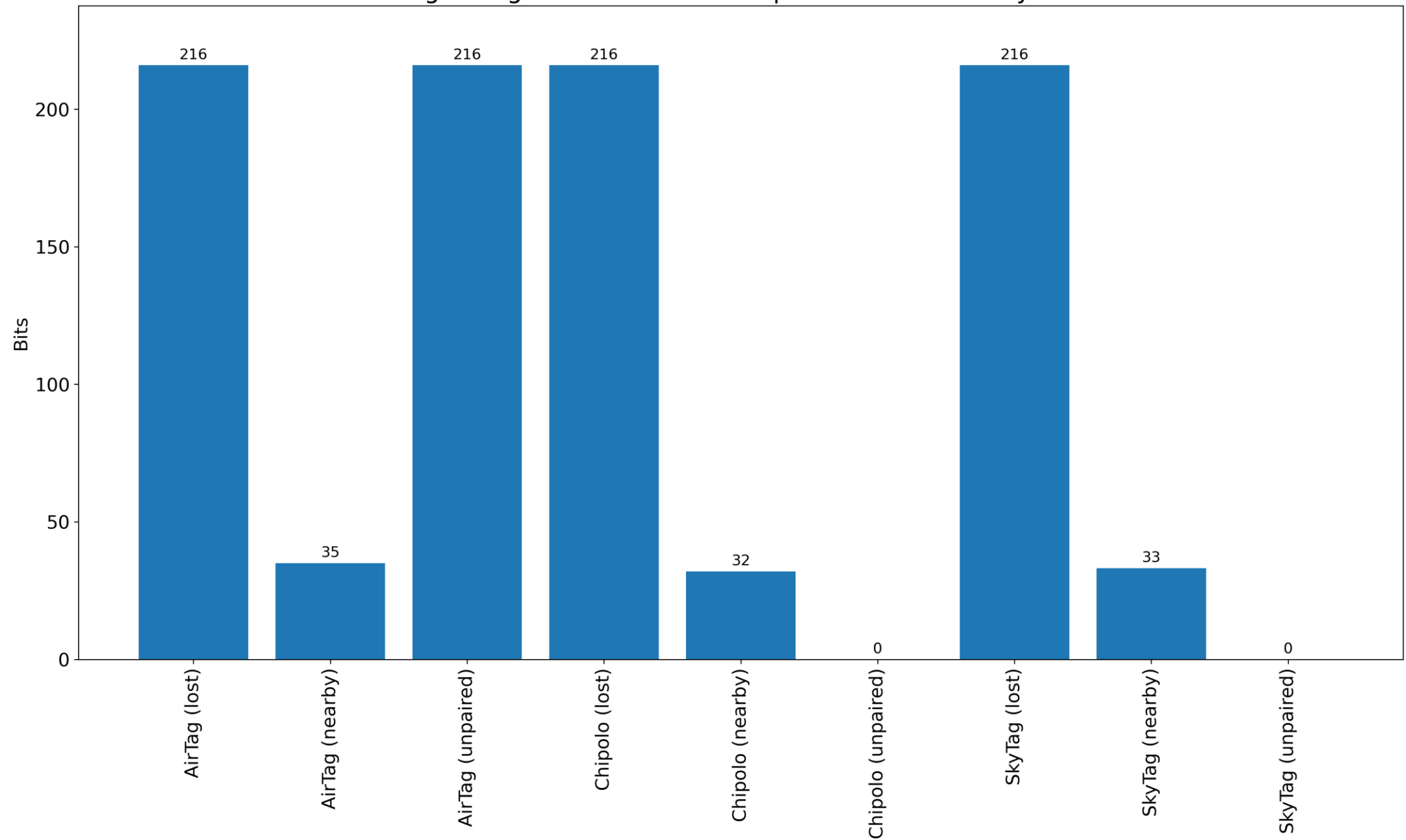
 - Type: Manufacturer Specific (0xff)

 - Company ID: Apple, Inc. (0x004c)

 - > Data: 121910bc3904307fa14c973cb6aefd08ba26426c0a15fb2a4c022c

 - CRC: 0x0f0801

Average Length of Manufacturer specific Data - FindMy Tracker





**University of
Zurich** ^{UZH}

Department of Informatics - CSG

Modeling

Classes of BLE Trackers



Air Tag (Apple)



ONE (Chipolo)



SkyTag (4Smarts)



SmartTag (Samsung)



Tile



Find My Trackers

States of BLE Trackers



Air Tag (Apple)



ONE (Chipolo)



SkyTag (4Smarts)



SmartTag (Samsung)



Tile

unpaired / nearby / lost

nearby / lost

lost

Apple iDevices



One app to find it all.





iPhone Find My (offline) Example Packet

- ✓ Bluetooth Low Energy Link Layer

 - Access Address: 0x8e89bed6

 - > Packet Header: 0x2542 (PDU Type: ADV_NONCONN_IND, TxAdd: Random)

 - Advertising Address: ea:54:39:35:b4:52

- ✓ Advertising Data

 - ✓ Manufacturer Specific

 - Length: 30

 - Type: Manufacturer Specific (0xff)

 - Company ID: Apple, Inc. (0x004c)

 - > Data: 121900f67e333bb0b4cb1b8d736af371736bb658f258e0a0070300

 - CRC: 0x715c47





iPhone Example Packet

✓ Bluetooth Low Energy Link Layer

Access Address: 0x8e89bed6

➢ Packet Header: 0x1742 (PDU Type: ADV_NONCONN_IND, TxAdd: Random)

Advertising Address: 4c:55:37:a4:37:d7

✓ Advertising Data

✓ Flags

Length: 2

Type: Flags (0x01)

000. = Reserved: 0x0

...1 = Simultaneous LE and BR/EDR to Same Device Capable (Host): true (0x1)

.... 1... = Simultaneous LE and BR/EDR to Same Device Capable (Controller): true (0x1)

.... .0.. = BR/EDR Not Supported: false (0x0)

.... ..1. = LE General Discoverable Mode: true (0x1)

.... ...0 = LE Limited Discoverable Mode: false (0x0)

✓ Manufacturer Specific

Length: 13

Type: Manufacturer Specific (0xff)

Company ID: Apple, Inc. (0x004c)

➢ Data: 160800be40aecaa21edc

CRC: 0x5da8f8



Class Labels for Apple Devices (iDevices)



- **iDevice FindMy (online)**
- **iDevice FindMy (offline)**
- **iDevice**

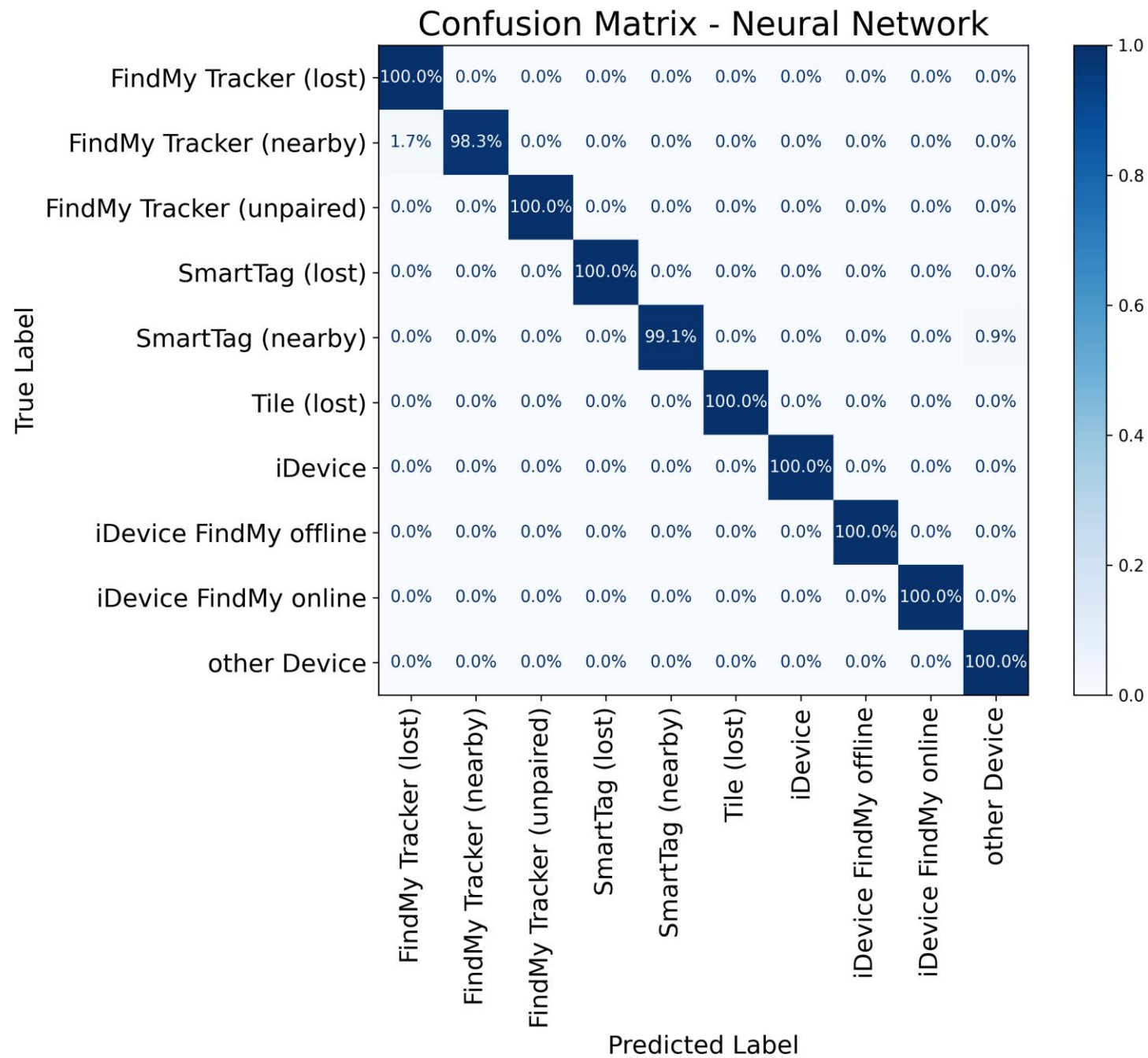


Example: Class Labels for Apple Devices (iDevices)

| Packet Nr. | Source | Continuity Type | Label |
|------------|-------------------|-----------------|-------------------------|
| 1 | 2E-B0-D0-63-C2-26 | 0x16 | iDevice |
| 2 | C9-2D-8B-7F-9B-A6 | 0x10 | iDevice |
| 3 | 00-5F-67-D3-1D-35 | 0x12 | iDevice FindMy (online) |
| 4 | 2E-B0-D0-63-C2-26 | 0x16 | iDevice |
| 5 | C9-2D-8B-7F-9B-A6 | 0x10 | iDevice |
| 6 | 00-5F-67-D3-1D-35 | 0x12 | iDevice FindMy (online) |
| 7 | 2E-B0-D0-63-C2-26 | 0x16 | iDevice |
| 8 | 2E-B0-D0-63-C2-26 | 0x16 | iDevice |
| ... | ... | | ... |

Other Devices







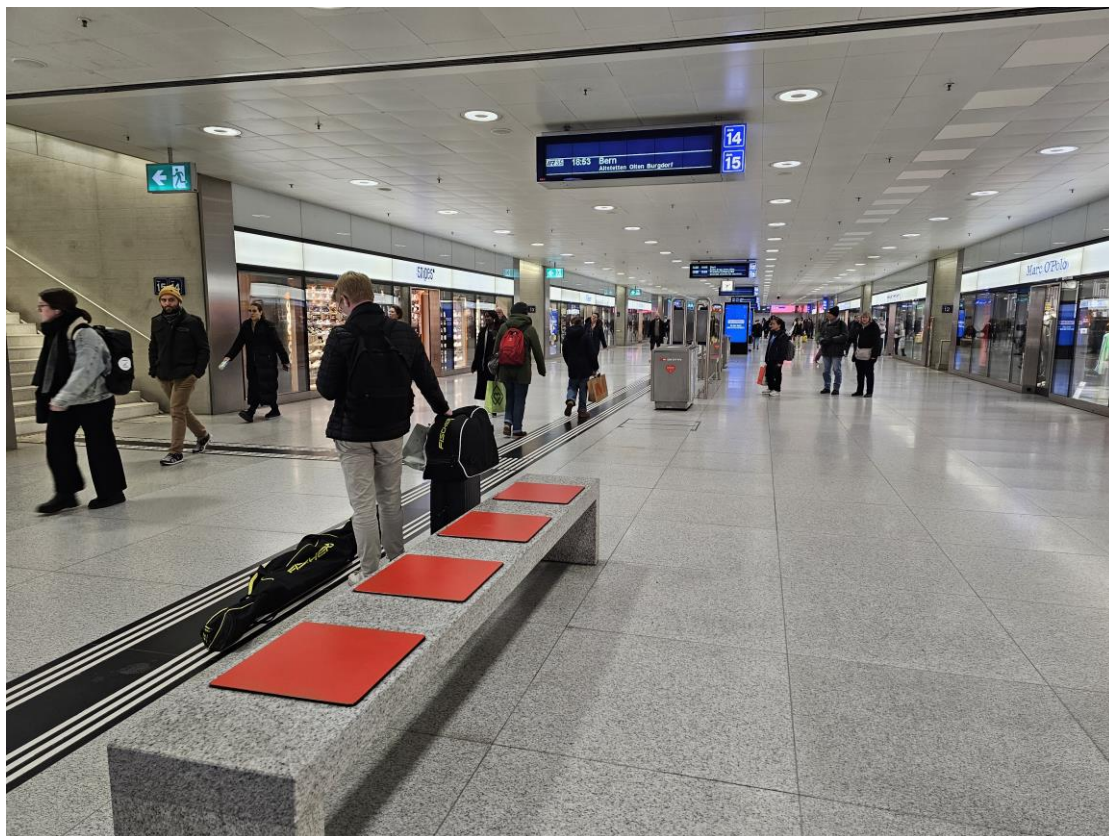
**University of
Zurich** ^{UZH}

Department of Informatics - CSG

Inference



Capturing BLE Packets at Zürich HB



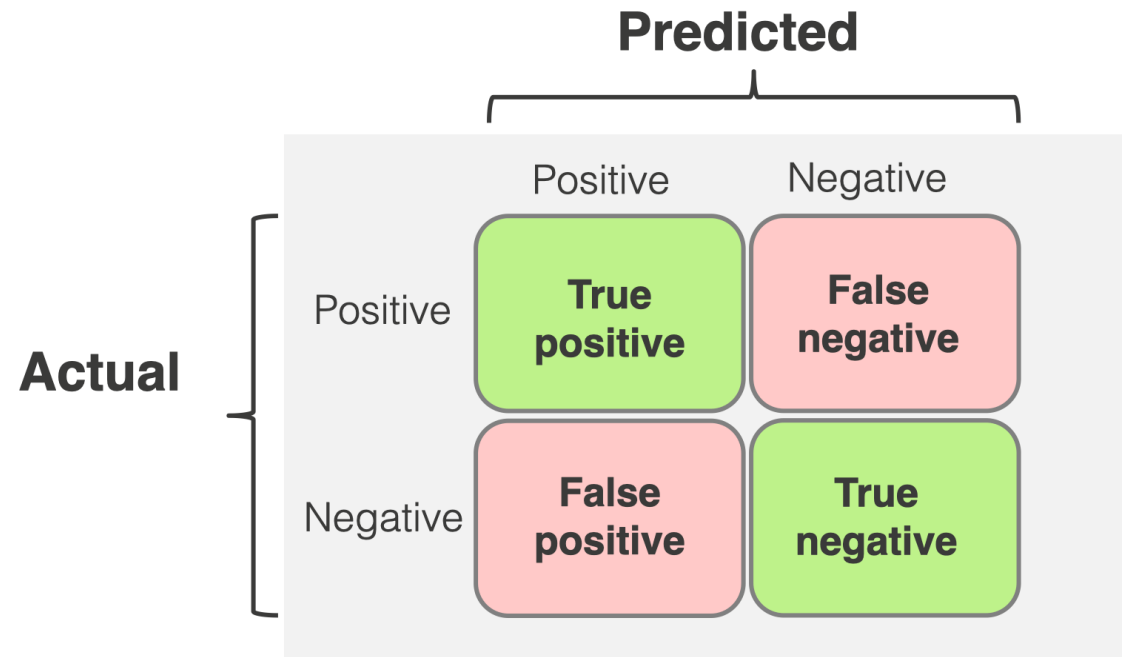


One Source Address multiple Labels?

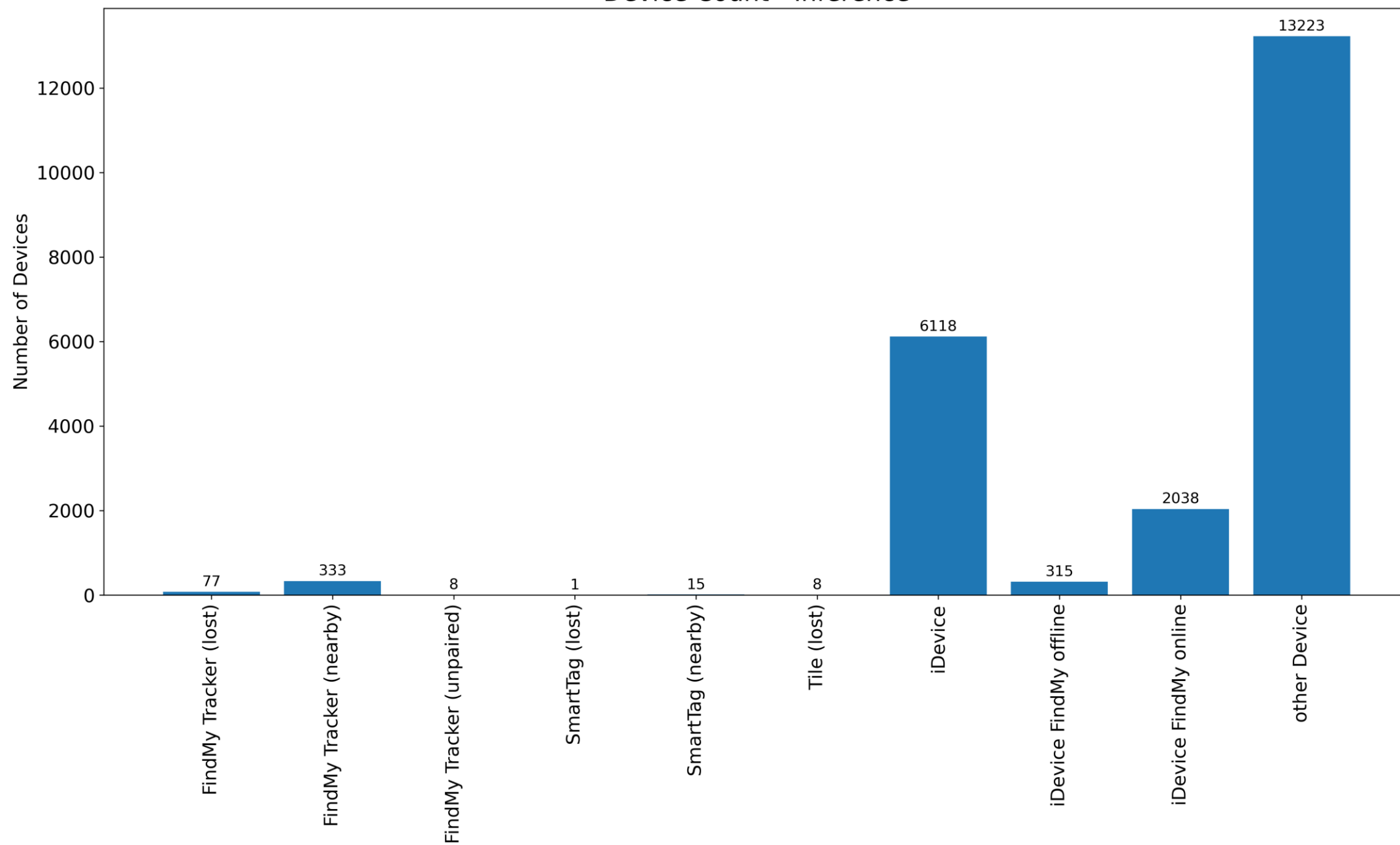
| Packet Nr. | Source | Predicted Label | Soft Max Confidence |
|------------|-------------------|-----------------|---------------------|
| 1 | 2E-B0-D0-63-C2-26 | SmartTag | 36% |
| 2 | 2E-B0-D0-63-C2-26 | AirTag | 92% |
| 3 | 2E-B0-D0-63-C2-26 | AirTag | 100% |
| 4 | 2E-B0-D0-63-C2-26 | SmartTag | 42% |
| 5 | 2E-B0-D0-63-C2-26 | AirTag | 97% |
| 6 | 2E-B0-D0-63-C2-26 | AirTag | 100% |
| 7 | 2E-B0-D0-63-C2-26 | AirTag | 95% |
| 8 | 2E-B0-D0-63-C2-26 | AirTag | 98% |

AirTag

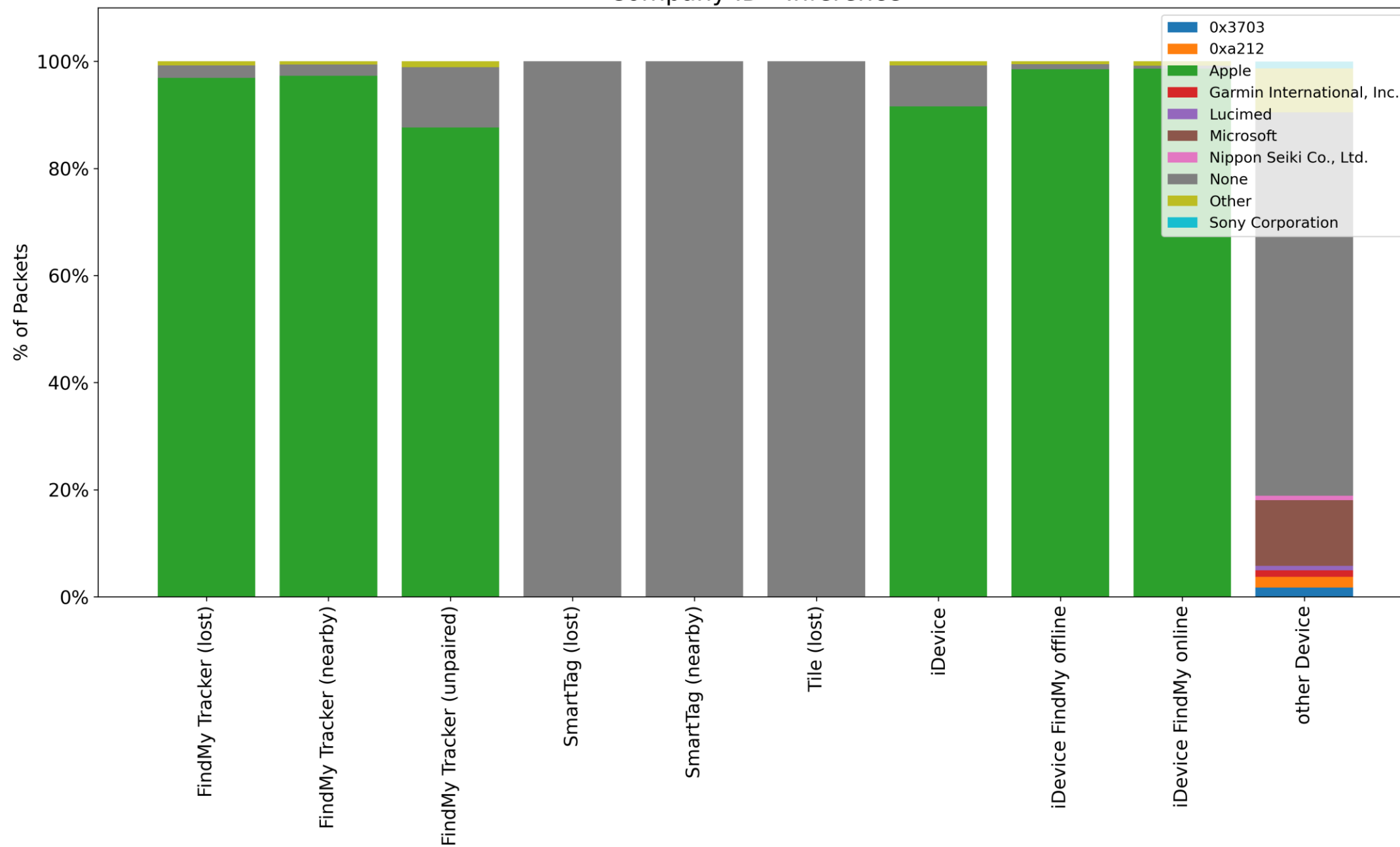
Evaluation?



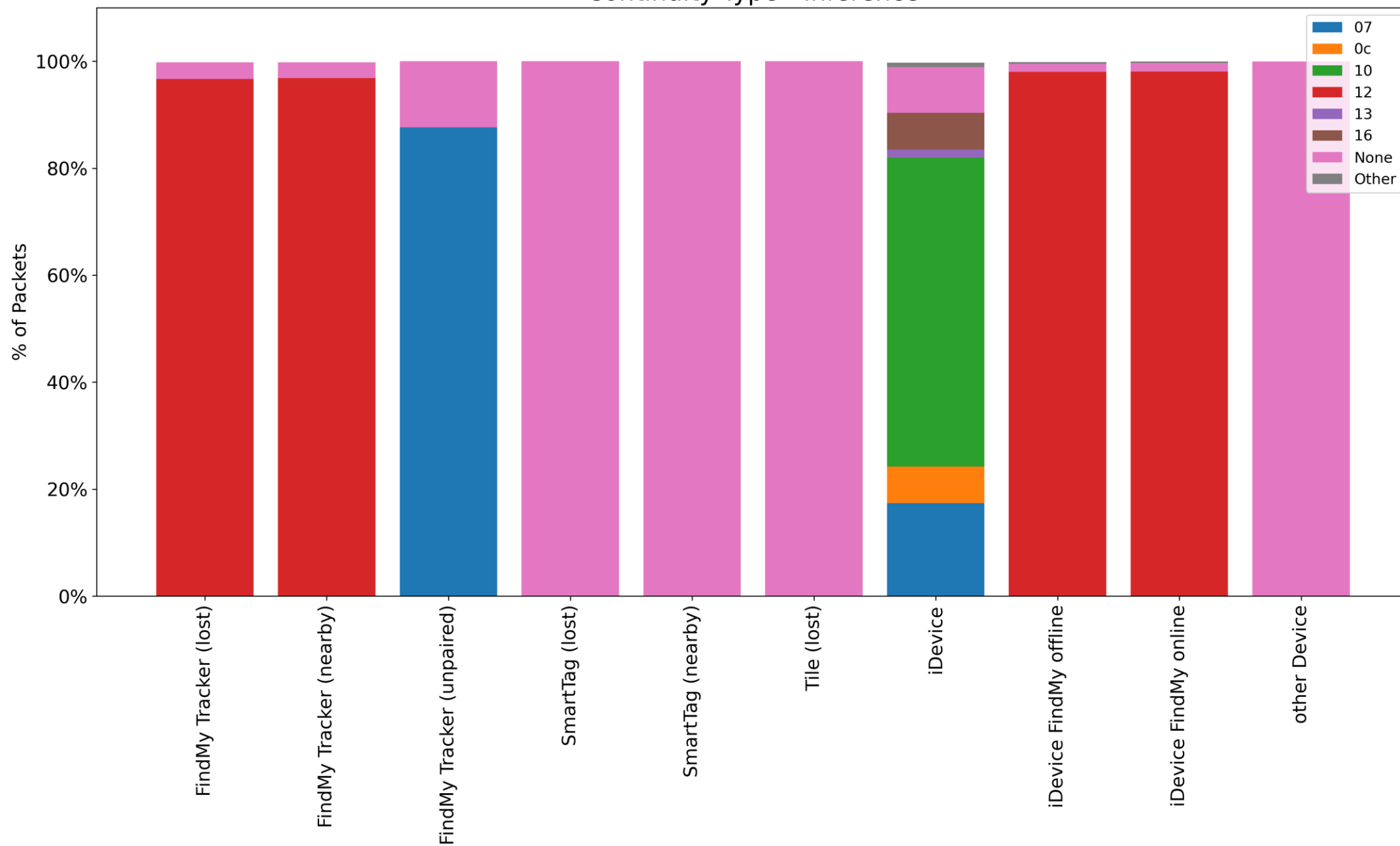
Device Count - Inference



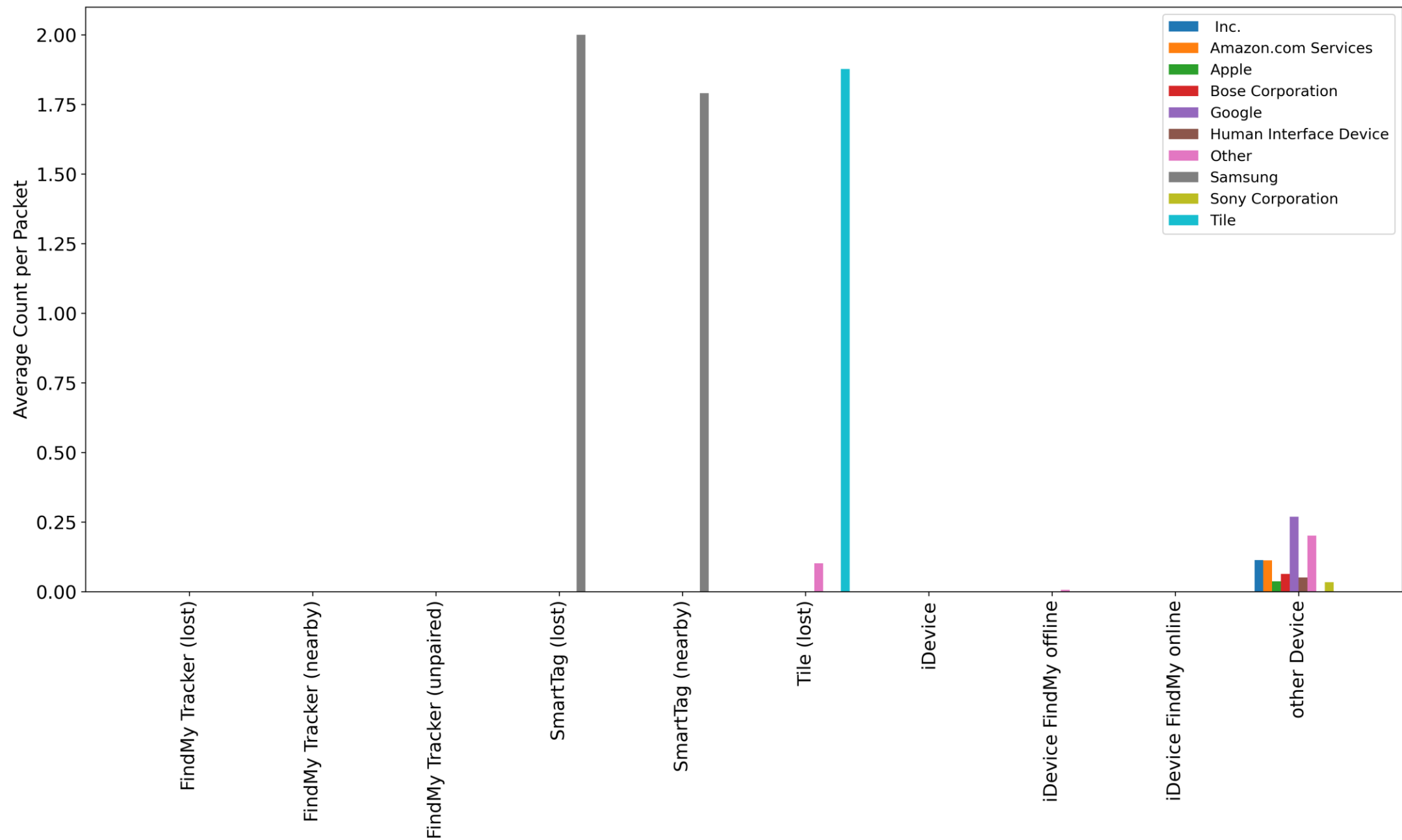
Company ID - Inference



Continuity Type - Inference



UUIDs - Inference





**University of
Zurich** ^{UZH}

Department of Informatics - CSG

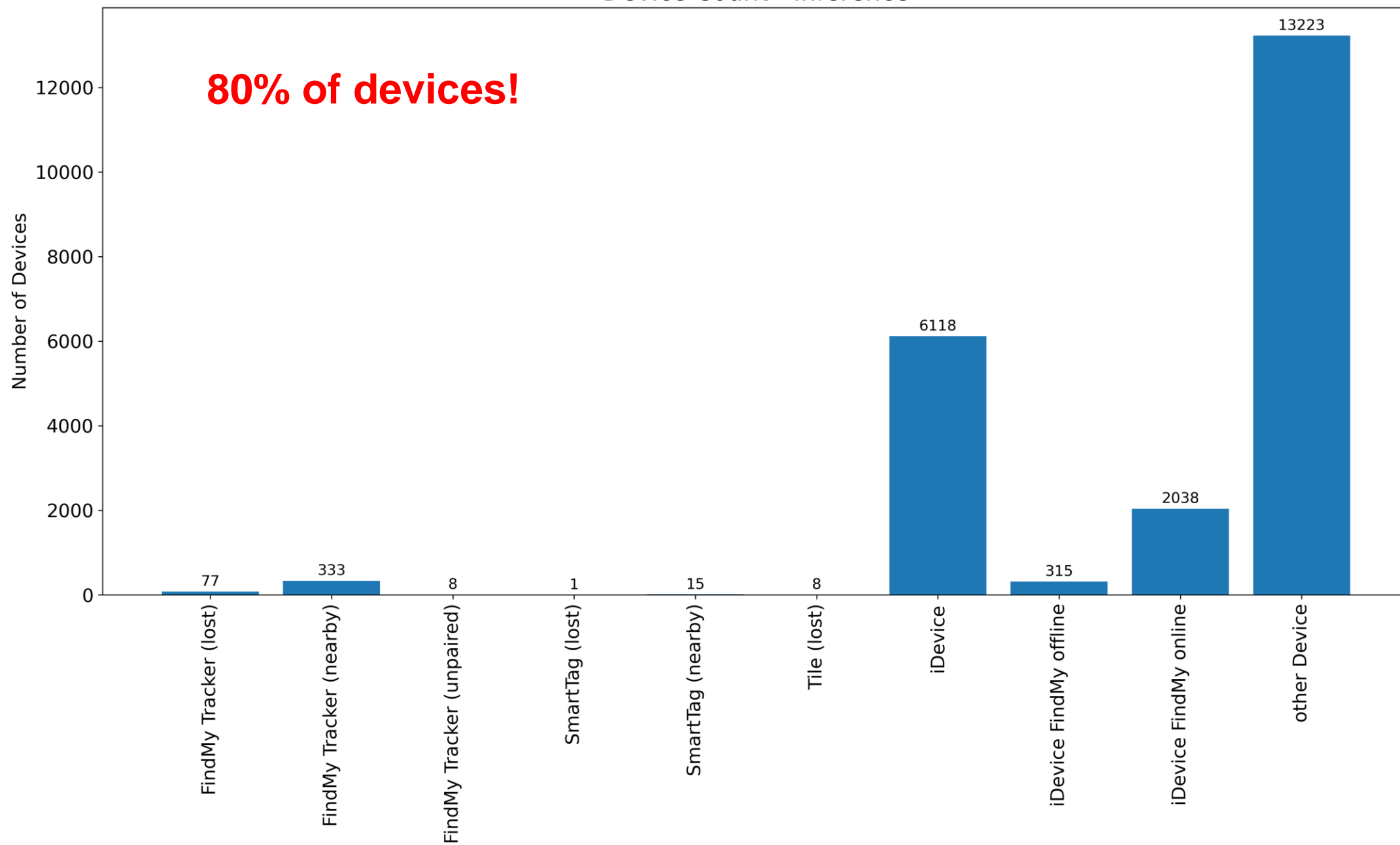
What's next?

One Source Address multiple Labels?

| Packet Nr. | Source | Predicted Label | Soft Max Confidence |
|------------|-------------------|-----------------|---------------------|
| 1 | 2E-B0-D0-63-C2-26 | SmartTag | 36% |
| 2 | 2E-B0-D0-63-C2-26 | AirTag | 92% |
| 3 | 2E-B0-D0-63-C2-26 | AirTag | 100% |
| 4 | 2E-B0-D0-63-C2-26 | SmartTag | 42% |
| 5 | 2E-B0-D0-63-C2-26 | AirTag | 97% |
| 6 | 2E-B0-D0-63-C2-26 | AirTag | 100% |
| 7 | 2E-B0-D0-63-C2-26 | AirTag | 95% |
| 8 | 2E-B0-D0-63-C2-26 | AirTag | 98% |

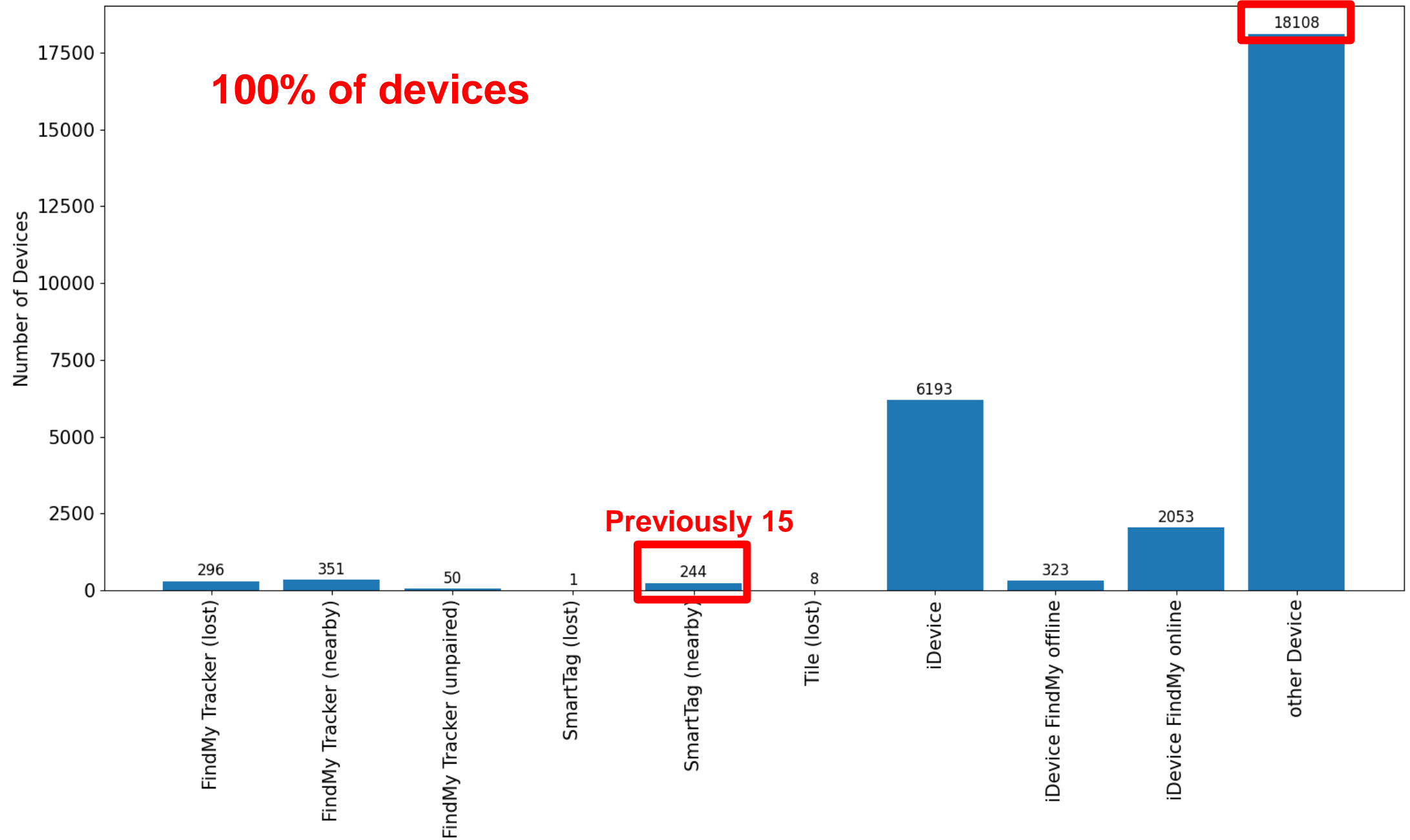
~~AirTag (97%)~~
< 98%

Device Count - Inference

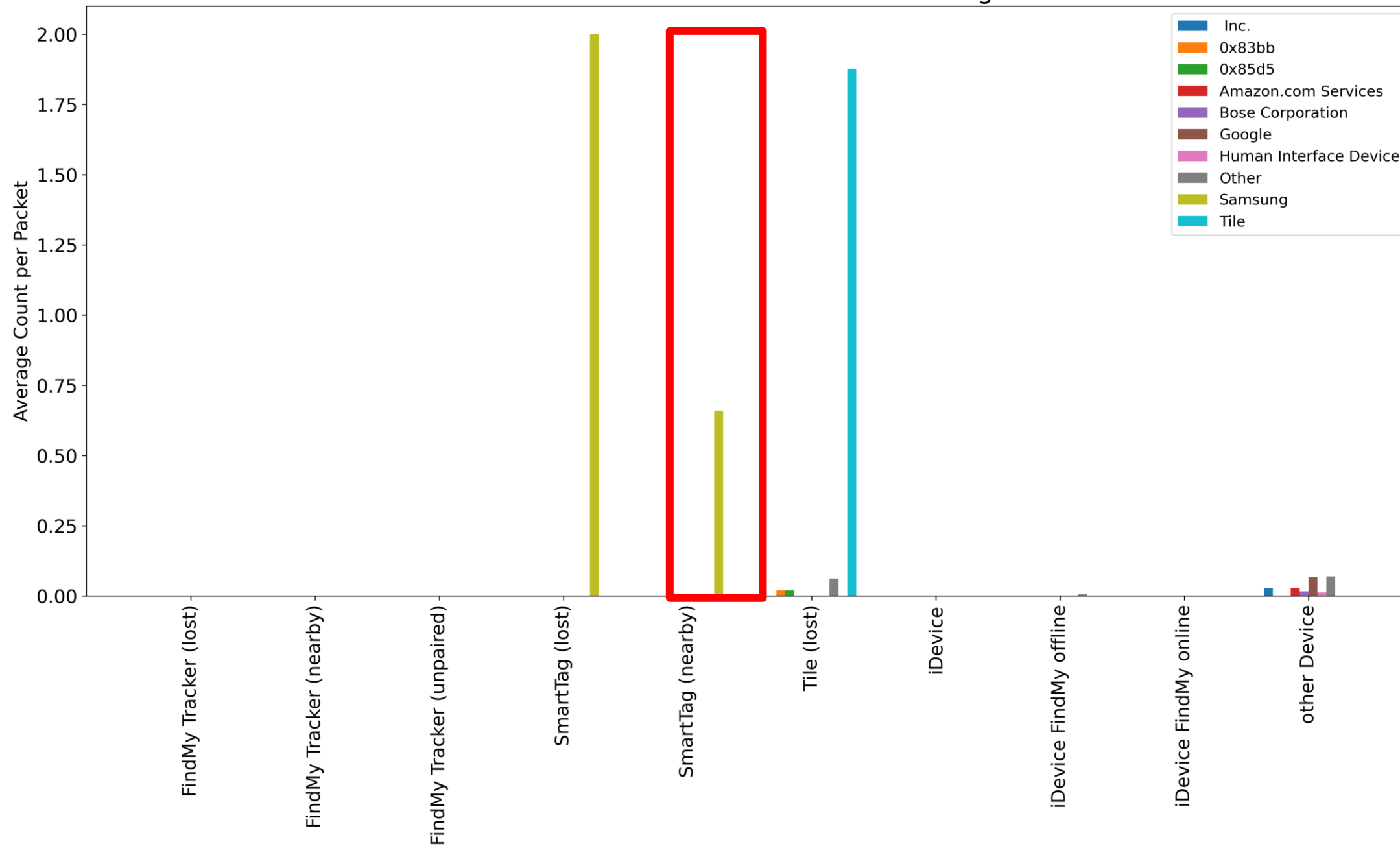


Device Count - Inference

Previously 13'223



UUIDs - Inference without Thresholding





We need (a lot) more Training Data!

Capture more BLE Packets?



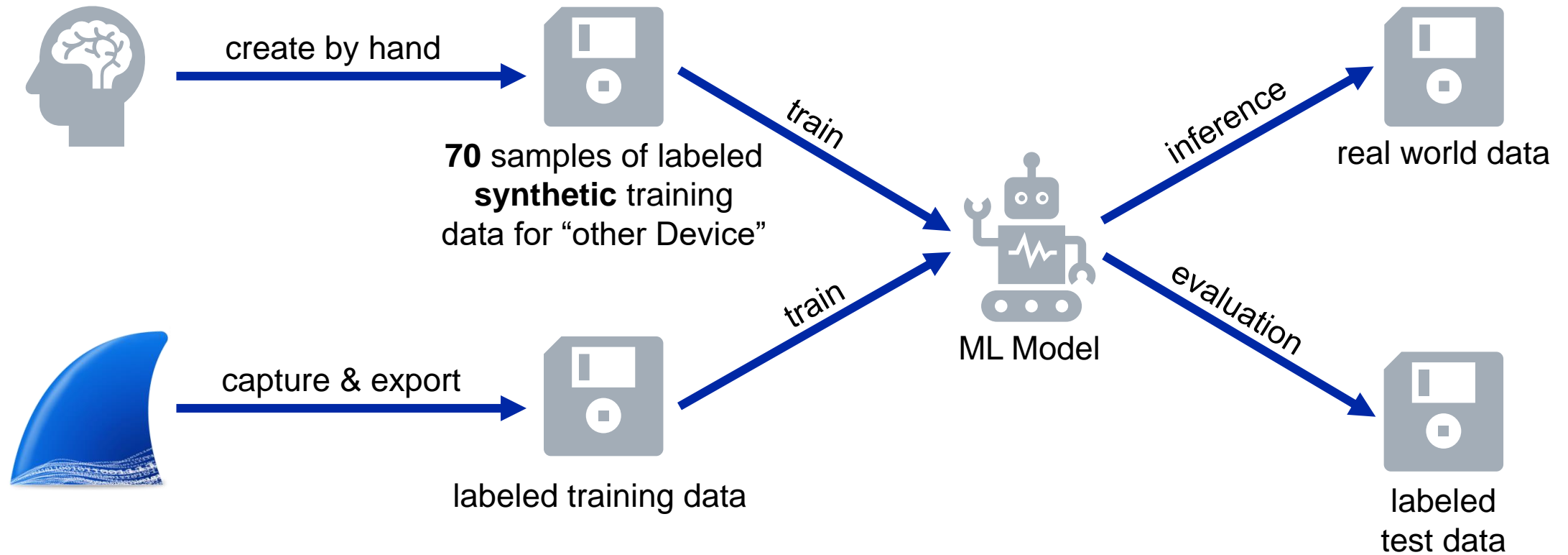


**University of
Zurich** ^{UZH}

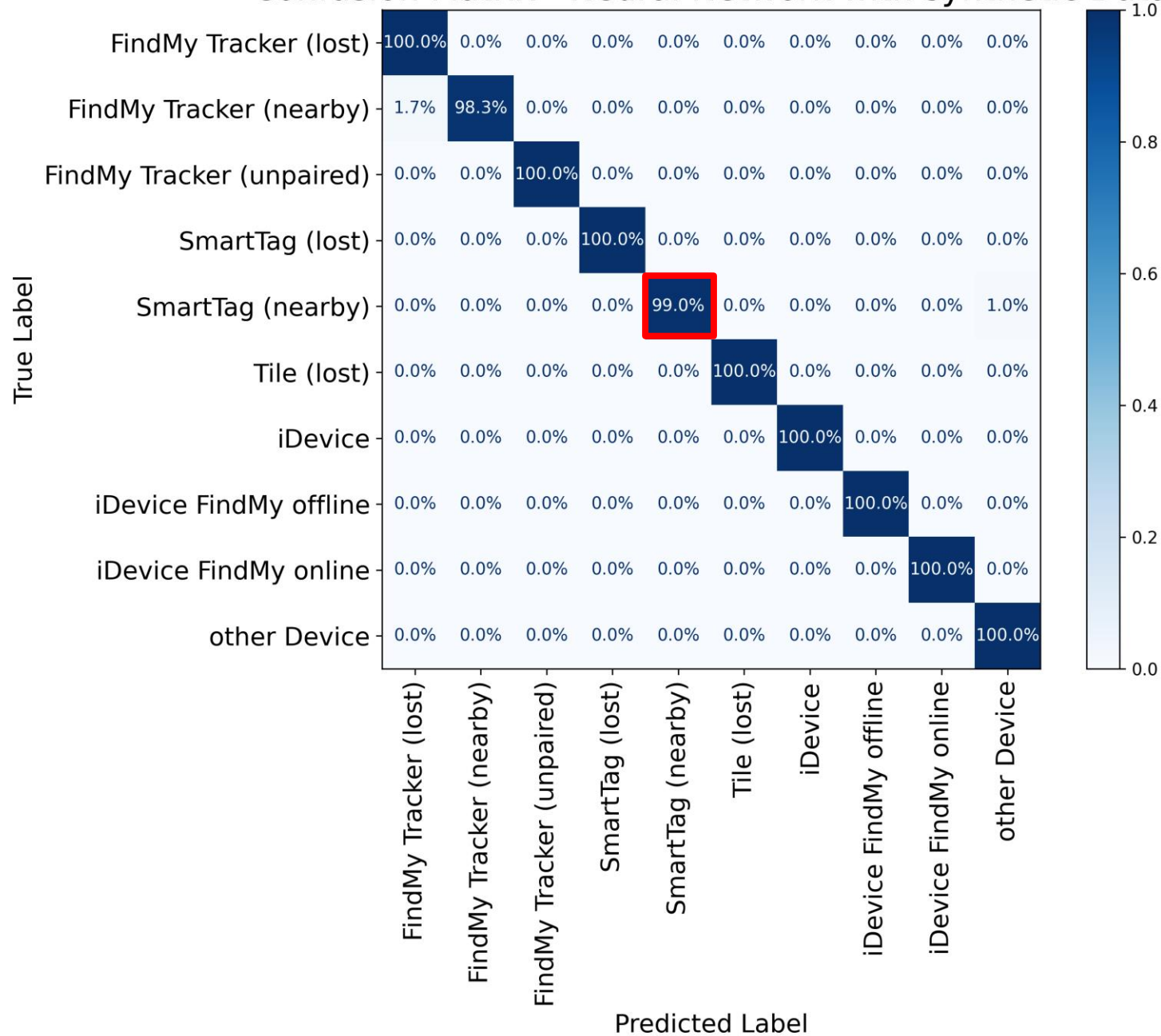
Department of Informatics - CSG

Synthetic Data Generation for BLE Packets

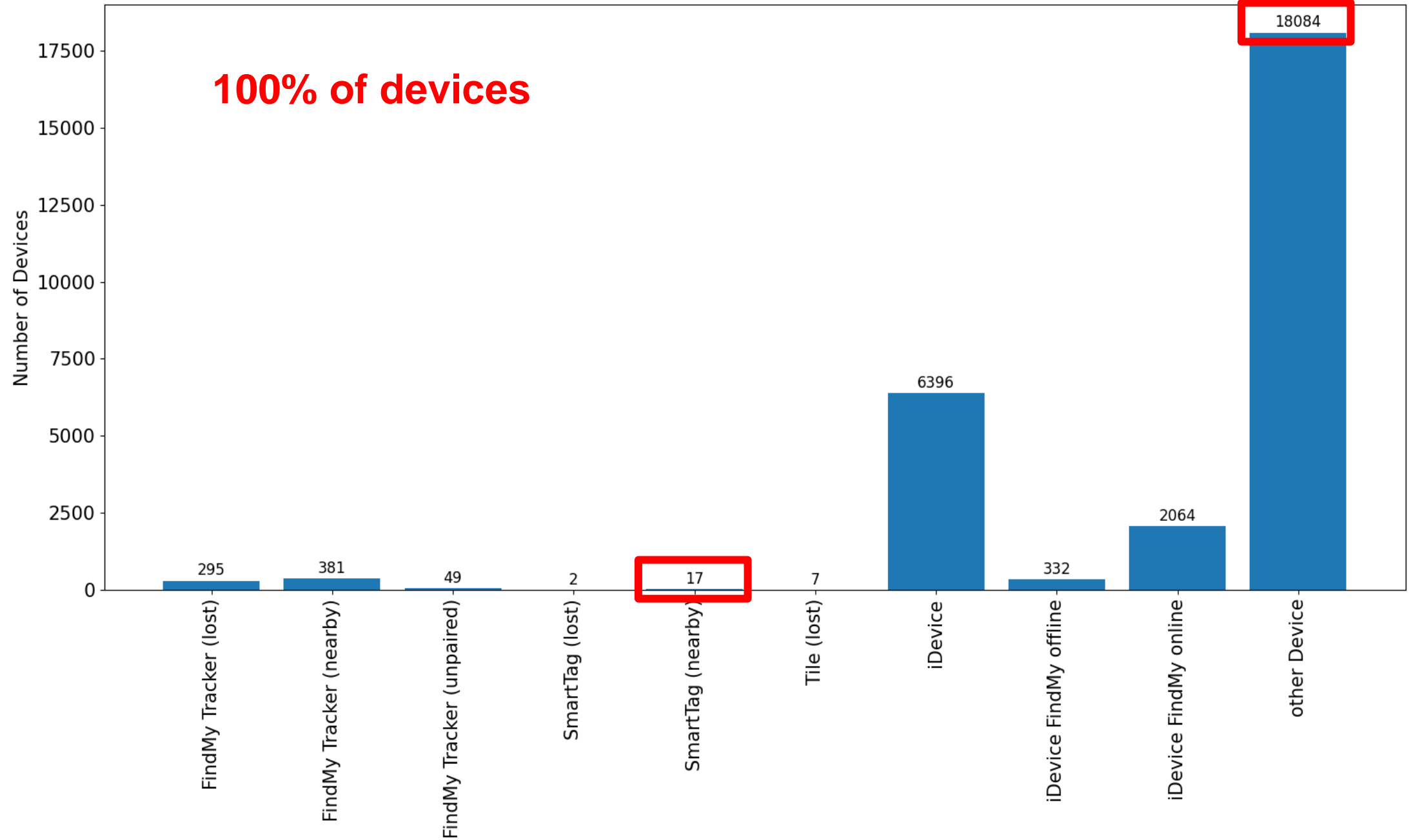
Training Machine Learning Models with synthetic Data



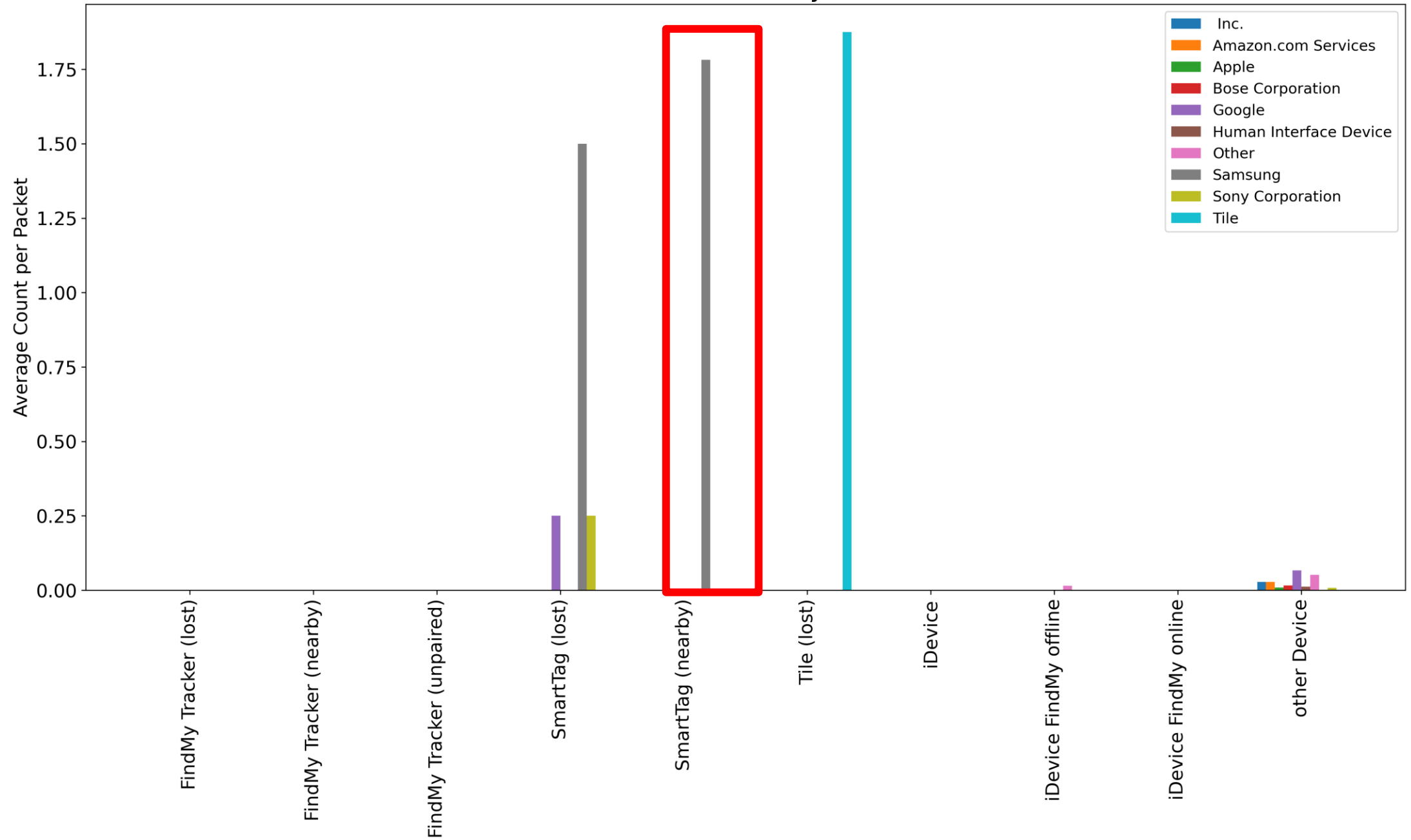
Confusion Matrix - Neural Network with synthetic Data



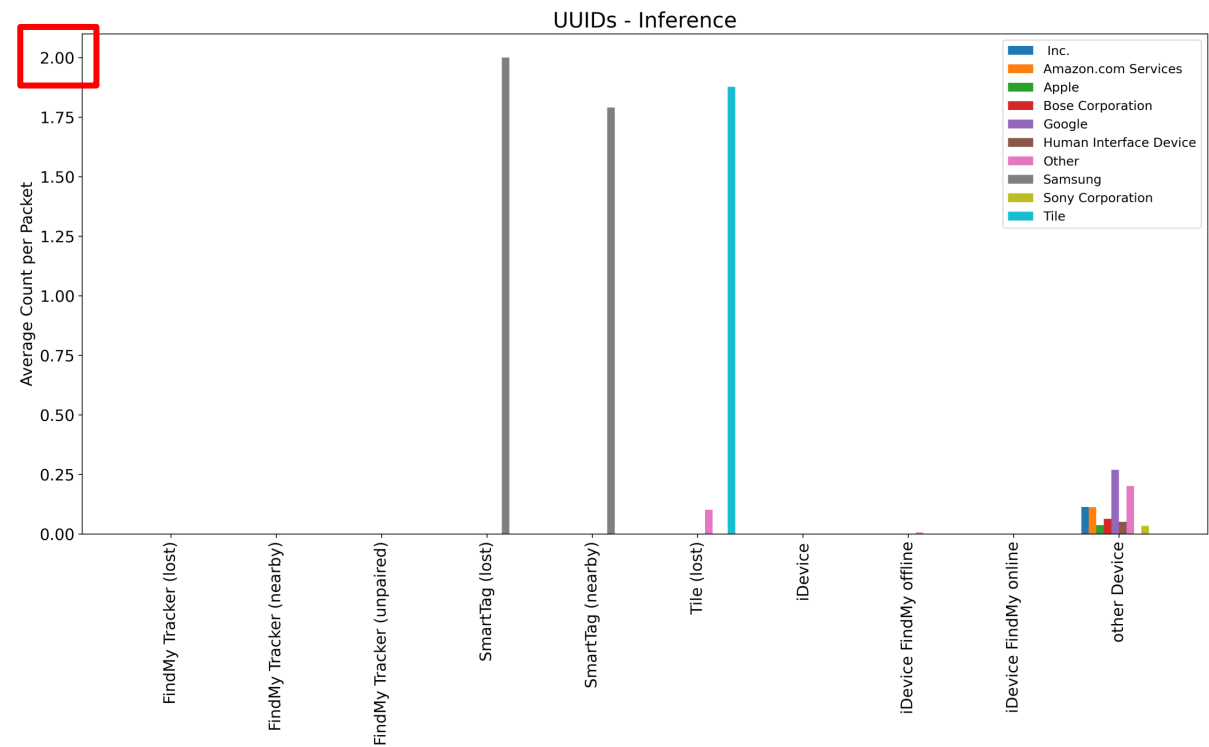
Device Count - Inference



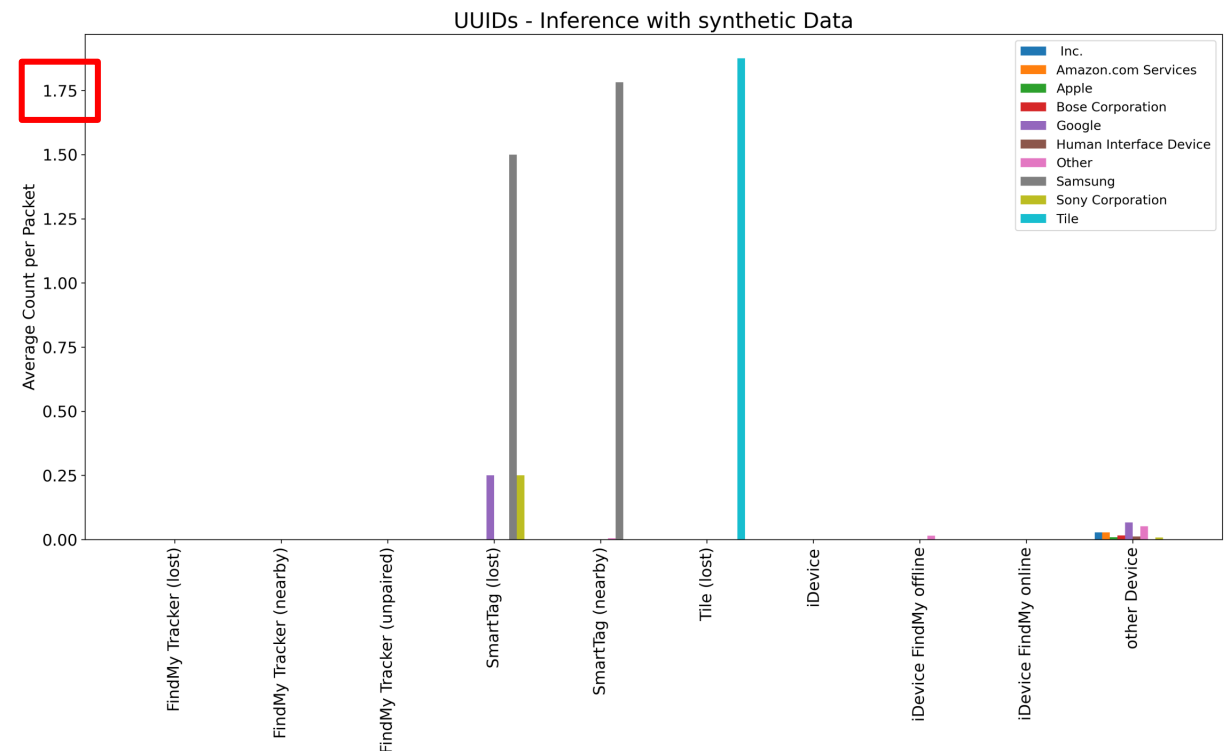
UUIDs - Inference with synthetic Data



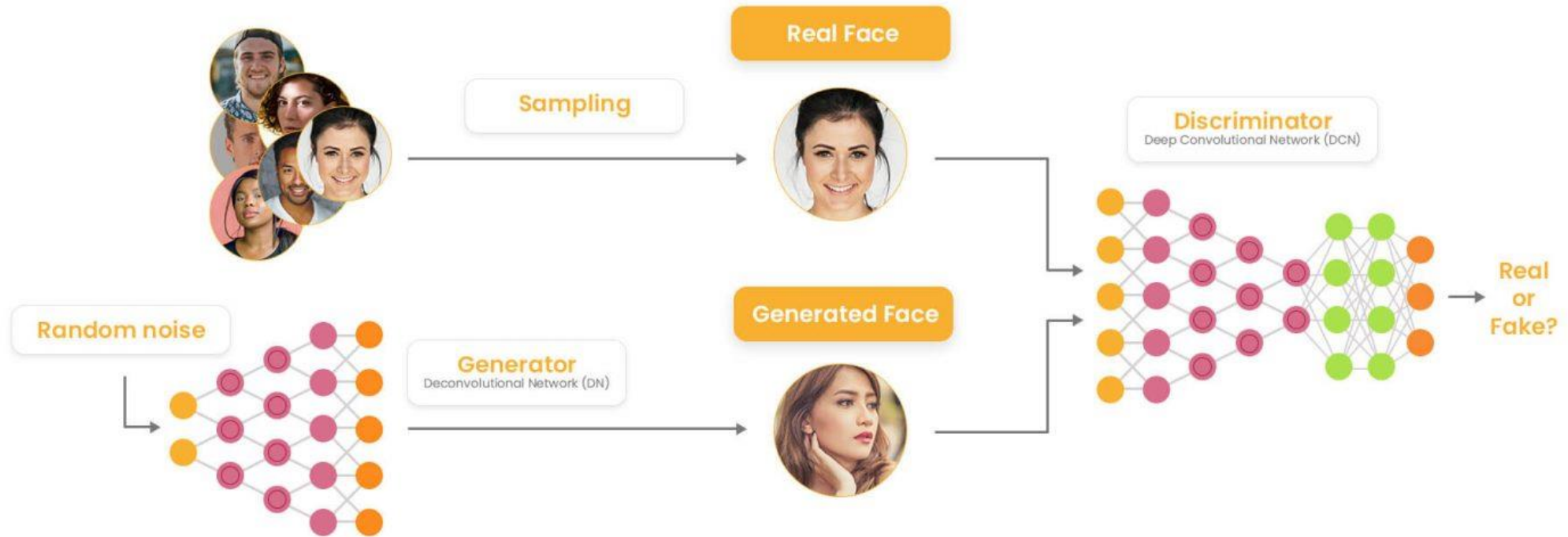
Thresholding **without** synthetic Data



100% of devices **with** synthetic Data

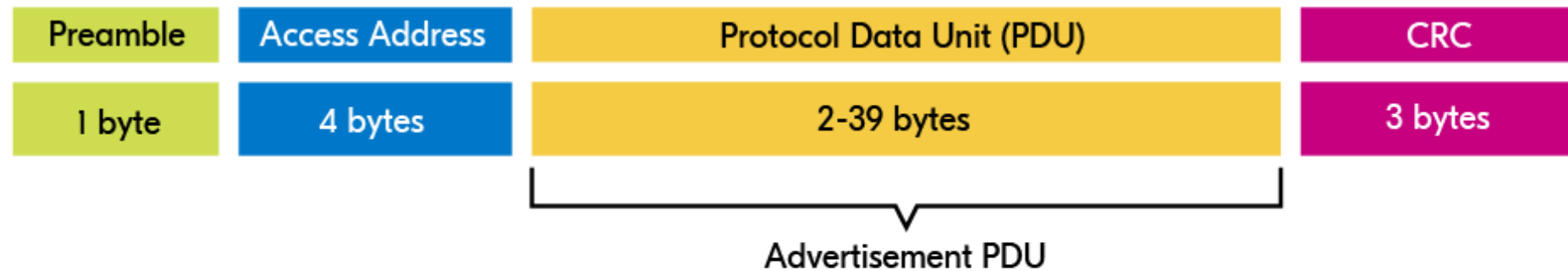


Data Generation with Deep Neural Networks (GAN)?



BLE Packet Format

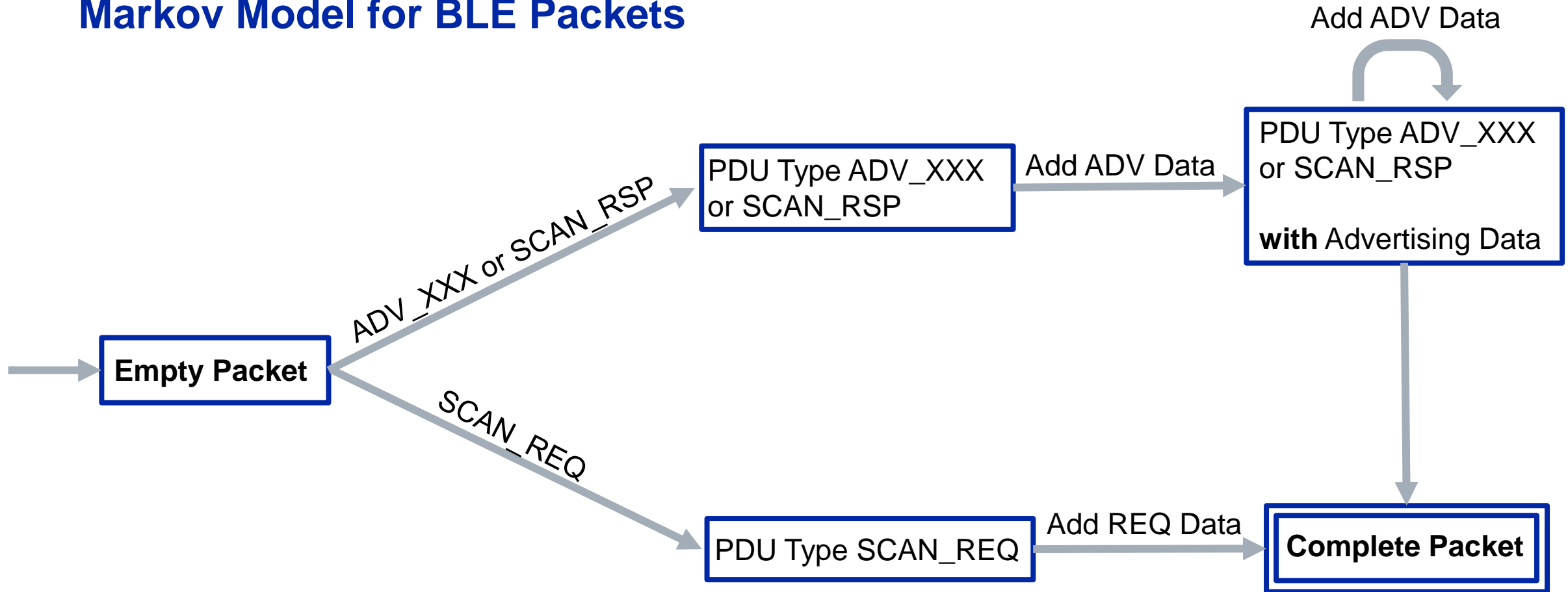
Bluetooth LE Packet

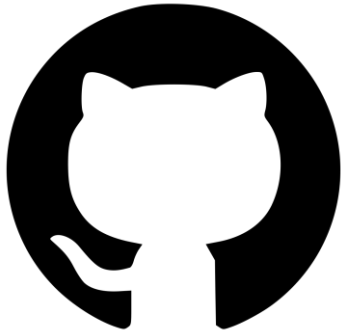


Advertisement PDU



Markov Model for BLE Packets






 Bachelor-Thesis-Stefan-Richard-Saxer Public













 master  1 Branch  0 Tags




 Go to file 

Add file 

 Code 

About 


| | | | |
|---|----------------------|------------------------|--|
|  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 | |


 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).


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

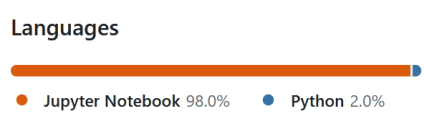
No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)





Questions?

Thank you for your attention and good luck reviewing my thesis!