UMONS
University of Mons

CyberExcellence
By CyberWal

Deep Learning based LoRa device Identification using Radio Frequency Fingerprinting

CyberExcellence Seminar, 28th Nov, 2023

Aqeel Ahmed
(Ph.D. candidate, CyberExcellence, UMONS)

Supervisor: Prof. Bruno Quoitin, UMONS

Agenda

1. Introduction
2. LoRa Background
3. LoRa Security
4. Radio Frequency Fingerprinting Identification
5. Deep Learning for LoRa device identification
6. Conclusion: Challenges and Future Work

1

Introduction

Introduction

- Low Power Wide Area Networks (LPWANs)
- LoRaWAN is LPWAN standard
- Key-enabler for many IoT applications
- LoRaWAN is built on top of **LoRa physical layer**
- LoRa (Long-Range)
  - Long-range communication **(more than 10 km in rural areas)**
  - low power consumption
  - Low-cost hardware
  - Robust against noise and interference
- LoRa operates in free ISM band i.e **EU868MHz, US915MHz**

Figure 1: LoRaWAN applications

Source: Microchip

2 LoRa Background

LoRa Background

- LoRa is based *on* ***Chirp Spread Spectrum (CSS)*** modulation scheme
- The frequency of a chirp signal increases **(upchirp)** and decrease (**downchirp**) over the time
- A LoRa symbol represents one or more encoded data bits.
- Number of bits/symbol are decided by spreading factor (SF)
- Encoding 7 raw bits on the symbol, means SF is 7.
- The SF range: **SF7 – SF12**
- A LoRa symbol can take any value in the range {0, 1,2, ….2^SF-1}.
- A modulated LoRa symbol can be denoted as follows:

$$c(nT_s + kT) = \frac{1}{\sqrt{2^{SF}}} e^{j2\pi \left(\frac{(S(nT_s)+k) \bmod 2^{SF}}{2^{SF}} \right) kT \frac{BW}{2^{SF}}}$$

- BW range: 125KHz, 250KHz and 500 kHz
- Code Rate CR = $\frac{4}{4+CR}$, CR: 1,2,3,4

SF	Chirps / Symbol	SNR	Airtime [a]	Bitrate
7	128	−7.5	56.5 ms	5469 bps
8	256	−10	−103 ms	3125 bps
9	512	−12.5	185.3 ms	1758 bps
10	1024	−15	371 ms	977 bps
11	2048	−17.5	741 ms	537 bps
12	4096	−20	1318.9 ms	293 bps

[a] 20 bytes per packet and Code Rate = 4/5.

SF Vs Bitrate [1]

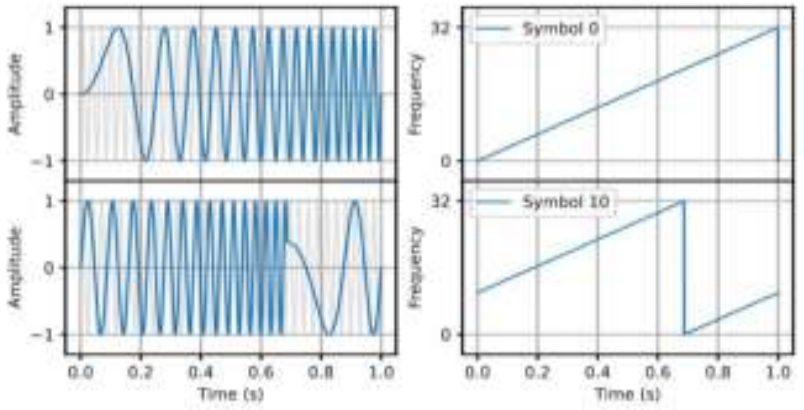


Fig:2 LoRa CSS modulated symbols

3

LoRa Security

LoRa Security

- LoRaWAN provides end-to-end security using cryptographic method i.e AES-128.
- Despite the security features of LoRaWAN, LoRa devices are susceptible to security attacks
- Vulnerable to:
  - ✓ **Impersonation**
  - ✓ **Jamming**
  - ✓ **Wormhole**
  - ✓ **Replay**
  - ✓ **DoS**

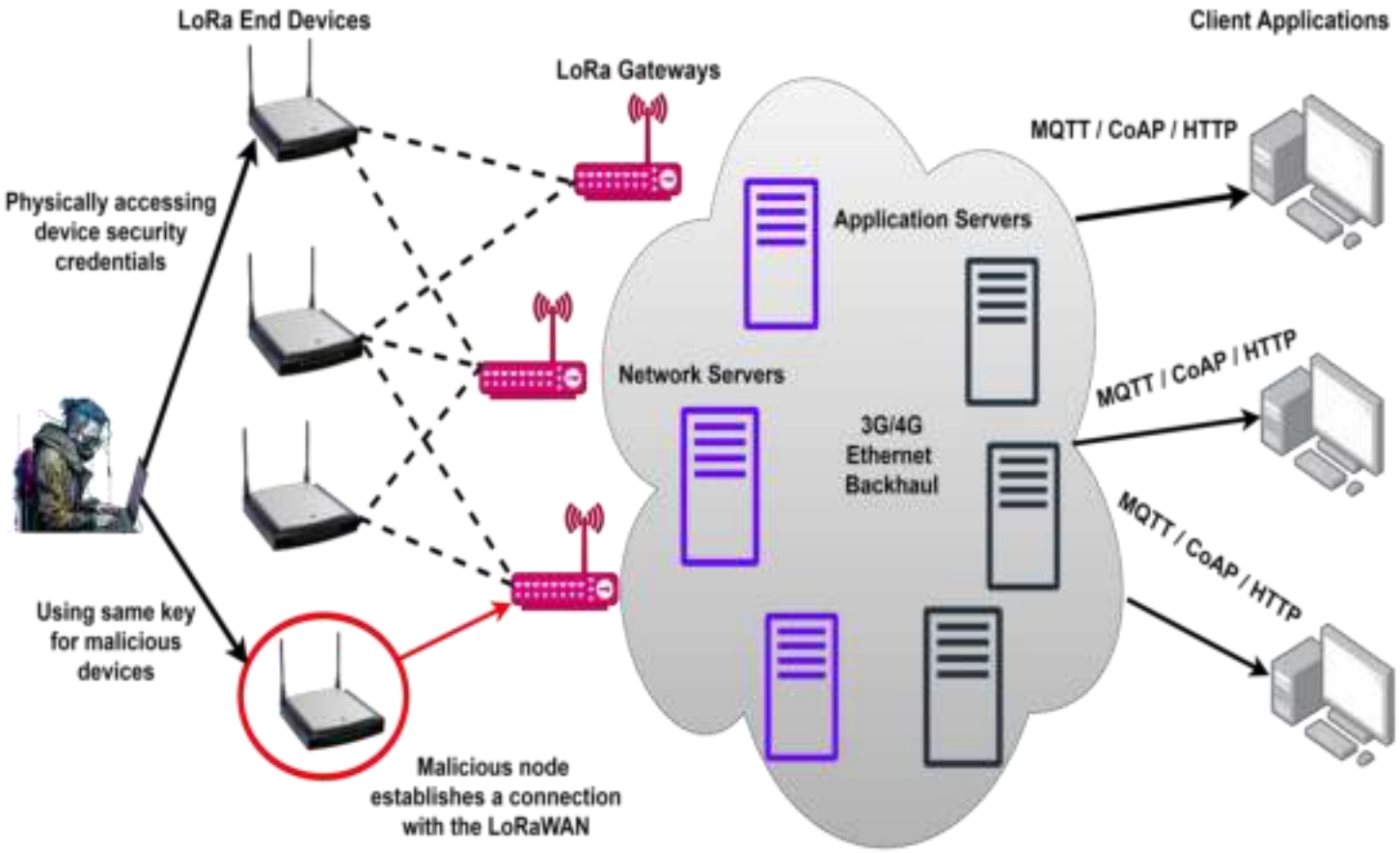


Fig 3: LoRaWAN impersonation attack scenario

Ref: E. Aras, G. S. Ramachandran, P. Lawrence, and D. Hughes, "**Exploring the security vulnerabilities of LoRa**," in 2017 3rd IEEE International Conference on Cybernetics (CYBCONF), pp. 1–6, IEEE, 2017

X. Yang, E. Karampatzakis, C. Doerr and F. Kuipers, "**Security Vulnerabilities in LoRaWAN**," 2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI), Orlando, FL, USA, 2018, pp. 129-140, doi: 10.1109/IoTDI.2018.00022.

4 Radio Frequency Fingerprinting Identification

Radio Frequency Fingerprinting

- RF fingerprinting is a physical-layer security method
- Aims to develop a unique RF fingerprint for a wireless device

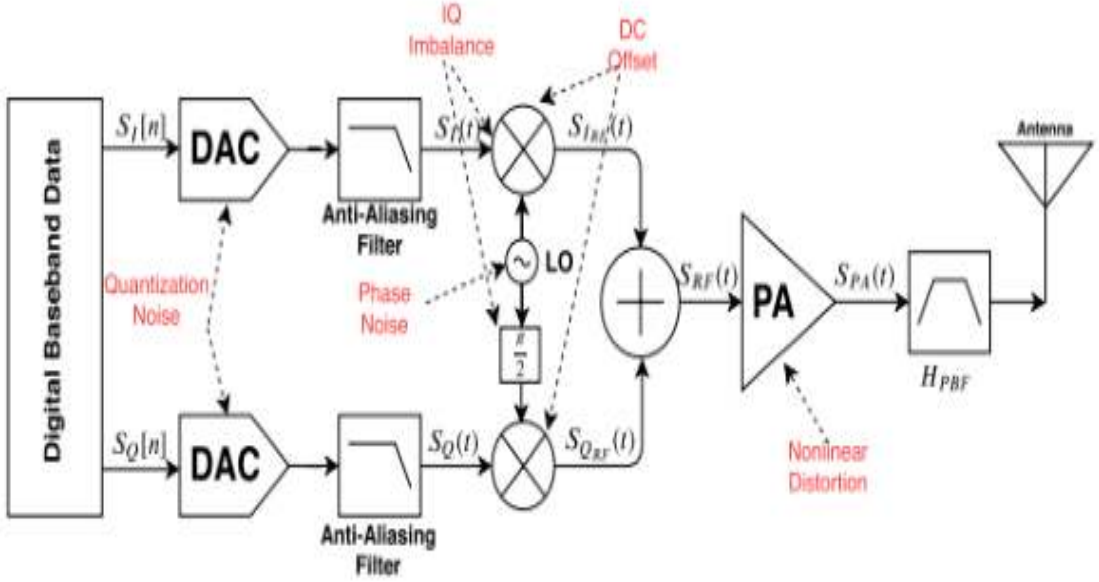


Fig 4: A typical transceiver with various RF impairments [2]

Why (RFFI) for LoRa device identification?

- Existing security method: (1) **hard on resources** (2) **vulnerable to impersonation**
- RFFI relies on device specific features caused by oscillator, amplifiers and mixers
- DC Offset, Carrier Frequency Offset, Phase Noise, IQ imbalance etc
- Unique fingerprint and hard to temper
- RFFI does not impose any additional power consumption on device [3]

5

Deep Learning based LoRa RFFI System

Deep learning based LoRa RFFI System

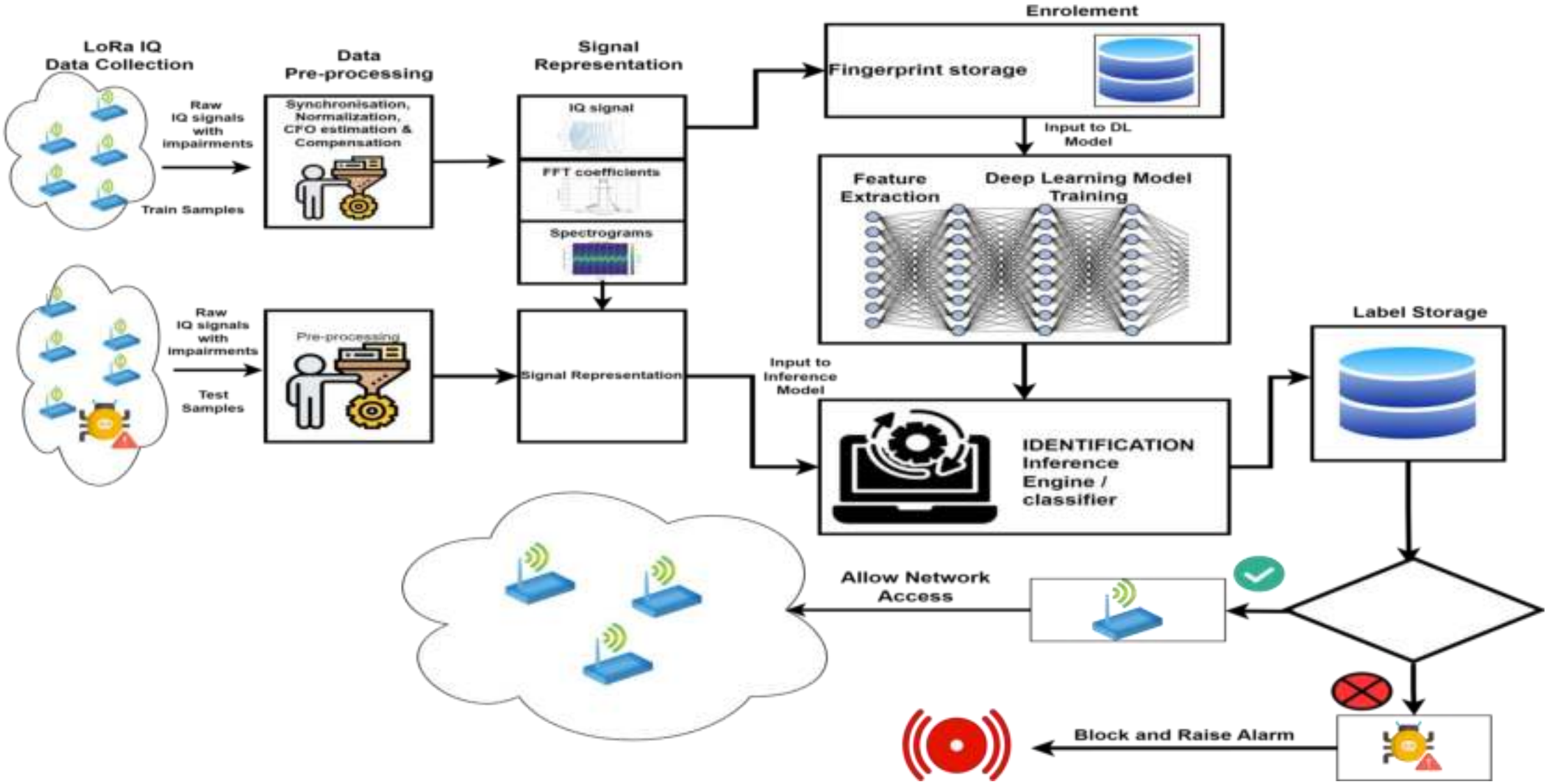


Fig 5: DL based LoRa RFFI System

Related Work

Reference	Description	Features	Accuracy
ELMAGHBUB et al .[2]	Create various test bed scenarios using DUT for analysing the impact. Use CNN for prediction. Also release dataset publicly.	Pycom LoRa devices ,IQ, FFT, Polar coordinates of IQ; amplitude and phase	65%- 90%
Shen et al.[4]	CFO estimation and compensation using real LoRa devices. Use CNN, MLP, LSTM for device classification.	Pycom LoPy LoRa devices,IQ, FFT, Spectrogram as input signals.	98.11% CNN hybrid
Robyns et al.[5]	MLP, SVM, & CNN models implemented to identify various LoRa devices with different chipsets	IQ samples as input	59%- 99%
Shen et al [6]	CFO and Phase noise impairments studied using channel independent spectrogram. CNN extractor + KNN classifier used	Pycom LoPy & FiPy, Dragino, IQ signals converted into spectrogram	> 95% for same case. Varies under channel conditions
Elmaghbub et al.[7]	CNN based DL model for device identification using in-band and out-of-band distortion.	IQ samples as input	96%

Methodology

Dataset Collection (Test bed)

Figure 6: LoRa transceivers and HACKRF SDR

- Transmitter: Three Pycom LoPy 4
- Receiver: HackRF SDR
- Distance between Tx & Rx: ~ 1m
- GNU Radio Software
- Fc (868.1MHz), SF (7), BW (125 kHz), Sampling Rate (8 MHz)
- Decimation factor of 8
- Environment : Indoor (with some obstacles)
- Time interval between two signals ~100 ms
- Training Frames: 3000
- Test Frames: 3000

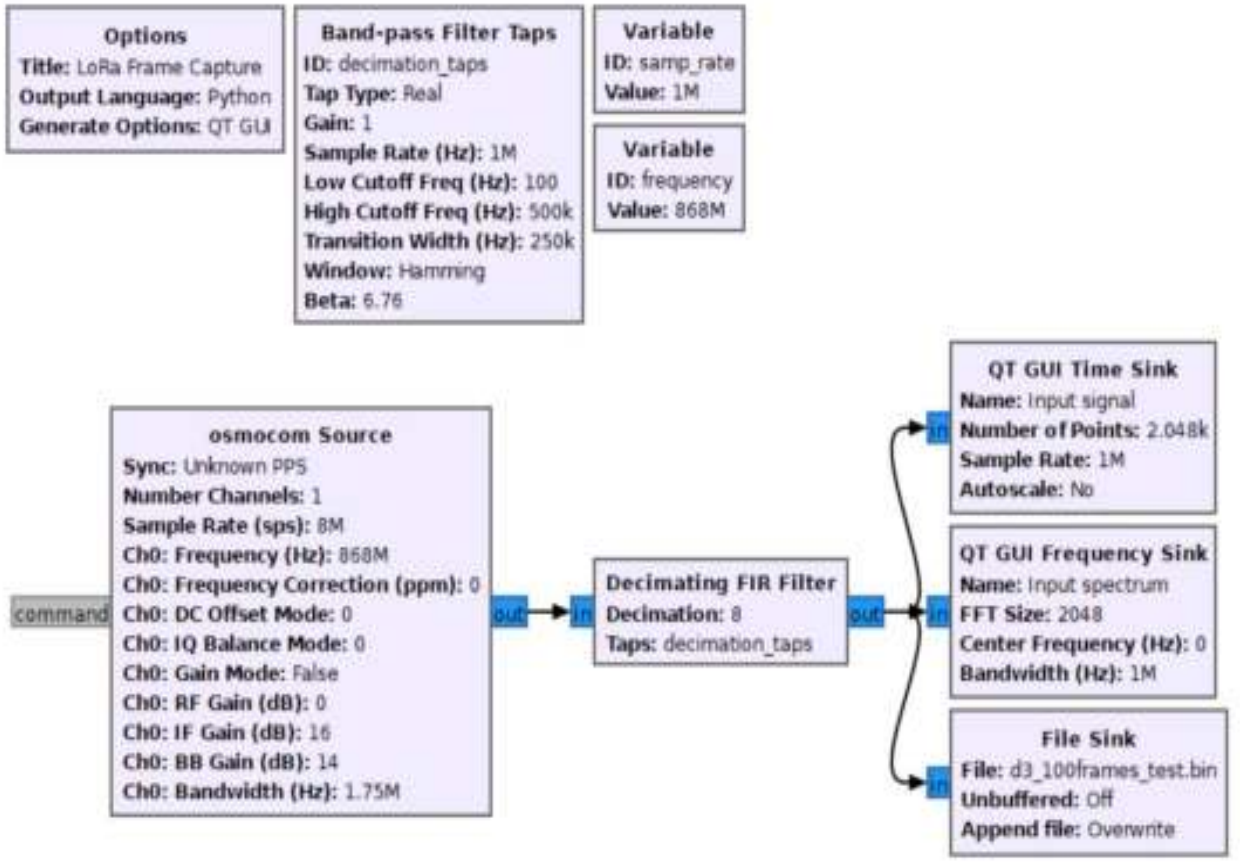


Figure 7: GNU radio flowgraph for processing IQ signals

Methodology (Cont..)

Data Preprocessing:

- Locating the starting point
- Extracting preamble part of the signal only
- Storing the IQ signals as (complex64 format)
- Signal normalization

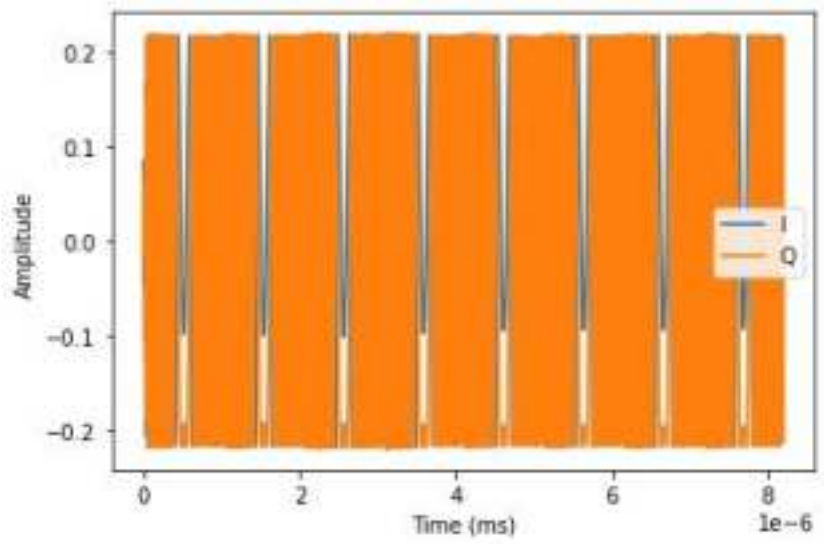


Fig 8: LoRa IQ signal in time domain (preamble)

- Raw signal in complex iq format

```
In [61]: print(data_train)
[[-0.00290618-0.05313189j -0.02153722-0.05378306j -0.0425642 -0.04476689j
  ... -0.39165947-0.45658479j -0.20011072-0.5670293j
  -0.02078168-0.60889018j]
 [ 0.00600547+0.05899644j  0.02393316+0.07105944j  0.04772553+0.07782059j
  ... -0.18533437-0.22635107j -0.16966186-0.23300758j
```

- We convert the signal into sequence of I(real) and Q (imaginary) parts as separate features of length 8192 and provide as input to the DL model.

- We further convert this IQ signal into frequency domain using FFT

- We also use amplitude and phase of the IQ signal as input to the model

$$A(m) = \sqrt{I^2(m) + Q^2(m)}$$

$$\phi(m) = \arctan\left(\frac{Q(m)}{I(m)}\right)$$

Methodology (cont…)

Publicly available Dataset

Figure 9: LoRa transceivers and USRP SDR

- IQ signal data collected under indoor and outdoor environments
- 30 off-the-shelf LoRa devices used as Transmitter
- Pycom LoPy4, Pycom Fipy, Dragino SX1276, mbed SX1261 shield USRP N210 SDR as receiver
- Carrier Frequency: 868.1
- MHz Sampling rate 1M
- BW: 125KHz
- Spreading Factor 7
- Code rate 4/5

Dataset publicly accessible@: **https://ieee-dataport.org/open-access/lorarffidataset**

G. Shen, J. Zhang, A. Marshall, and J. R. Cavallaro**, "Towards Scalable and Channel-Robust Radio Frequency Fingerprint Identification for LoRa**," *IEEE Trans. Inf. Forensics Secur.*, vol. 17, pp. 774–787, 2022, doi: 10.1109/TIFS.2022.3152404

Methodology (cont…)

Implemented Deep Learning Model

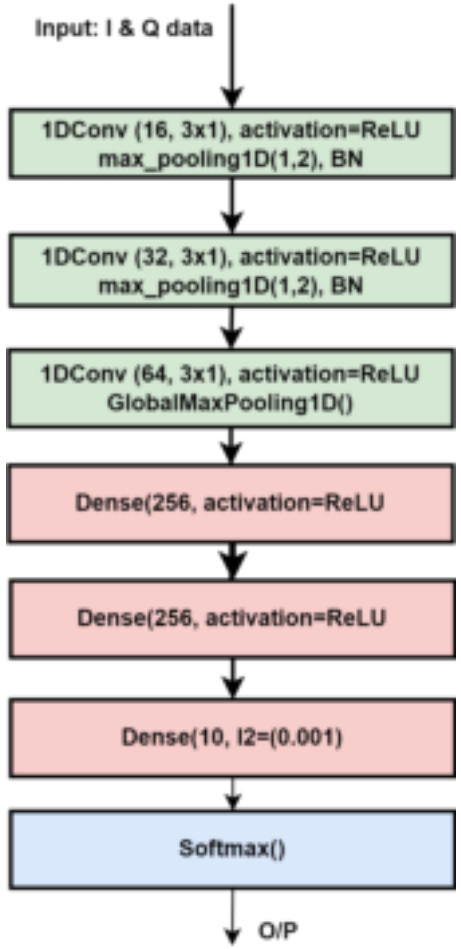


Fig 10: Model architecture (1D-CNN)

- **Lightweight 1D CNN**

- 3 Convolutional Layers
- Activation: ReLU
- Batch Normalization
- 1D Max_pooling(size=2)
- 3 Dense Layers (activation=ReLU)
- Loss function: Categorical Crossentropy
- Regularization: L2 (0.001)
- Optimizer: Adam (lr:0.001)
- Training/val ratio: 80:20
- Epochs:30
- Callbacks: Learning rate schedular
- Training stops if there is no further improvement in loss optimization
- Batch size=32
- Total trainable parameter: 58938
- Model size: 231.10 KB

Evaluation metrics:

$$Accuracy = \frac{(TP + TN)}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - score = 2 * \frac{(Precision * Recall)}{Precision + Recall}$$

TABLE I: Model Performance

1D-CNN	Accuracy (%) Vs Device Number			
Input Type	#3	#10	#20	#30
IQ	93.13	95.53	95.54	96.23
A/ϕ	90.47	90.60	84.12	91.98
FFT	35.25	88.65	86.60	84.97

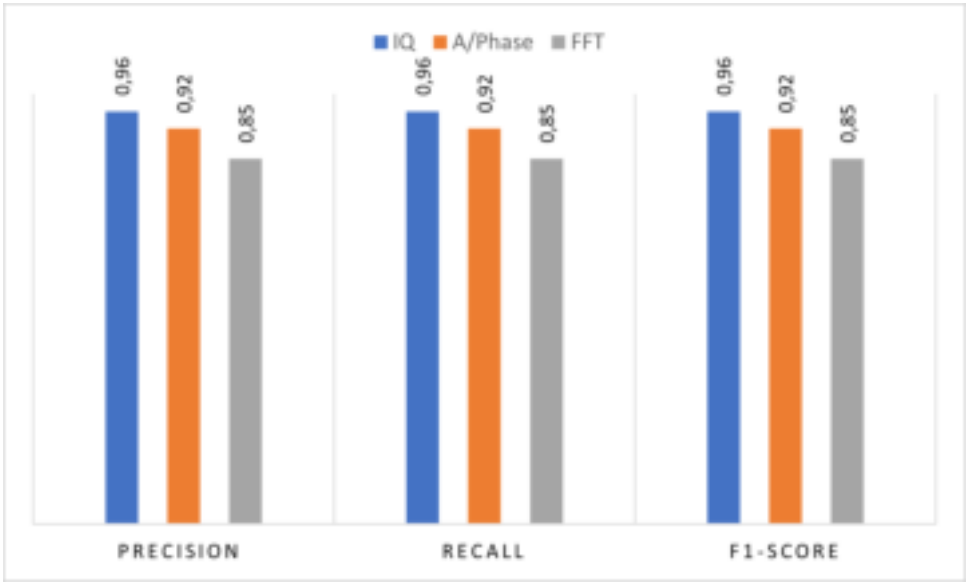
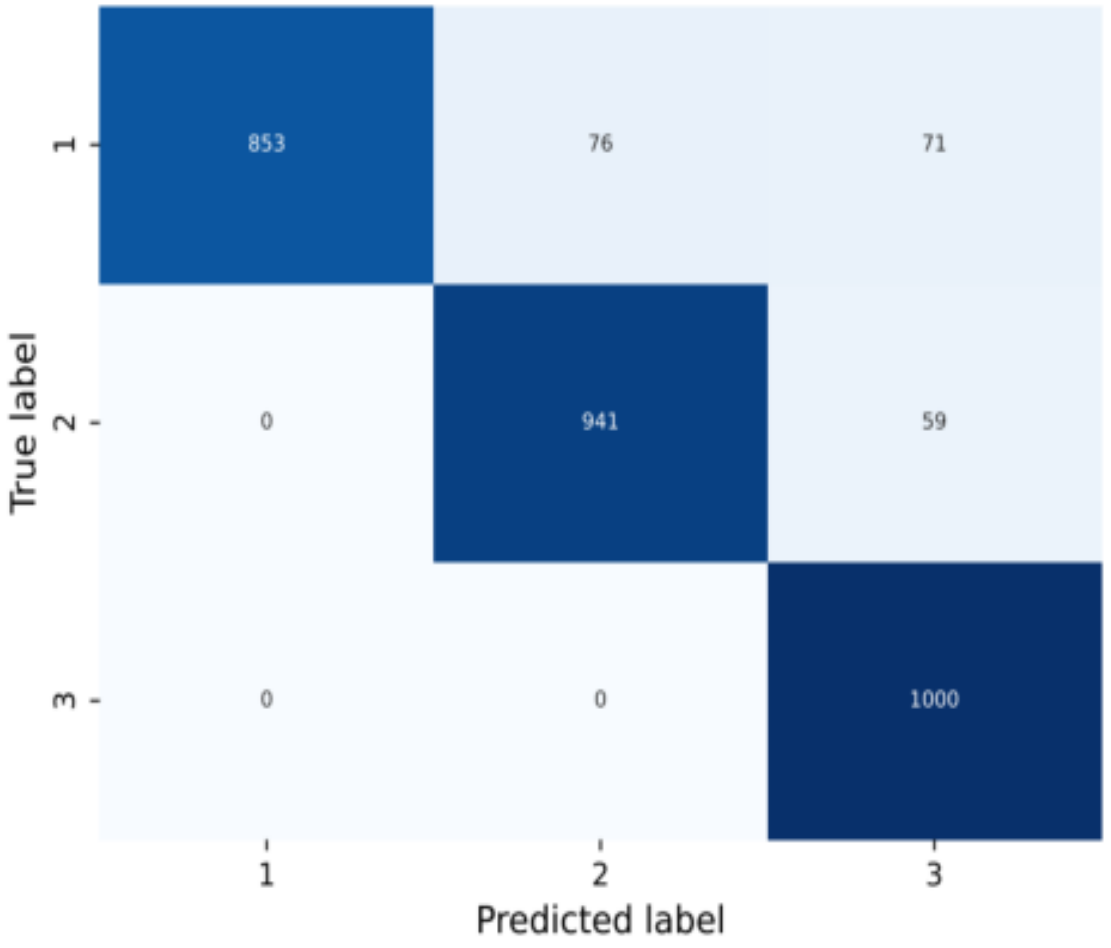
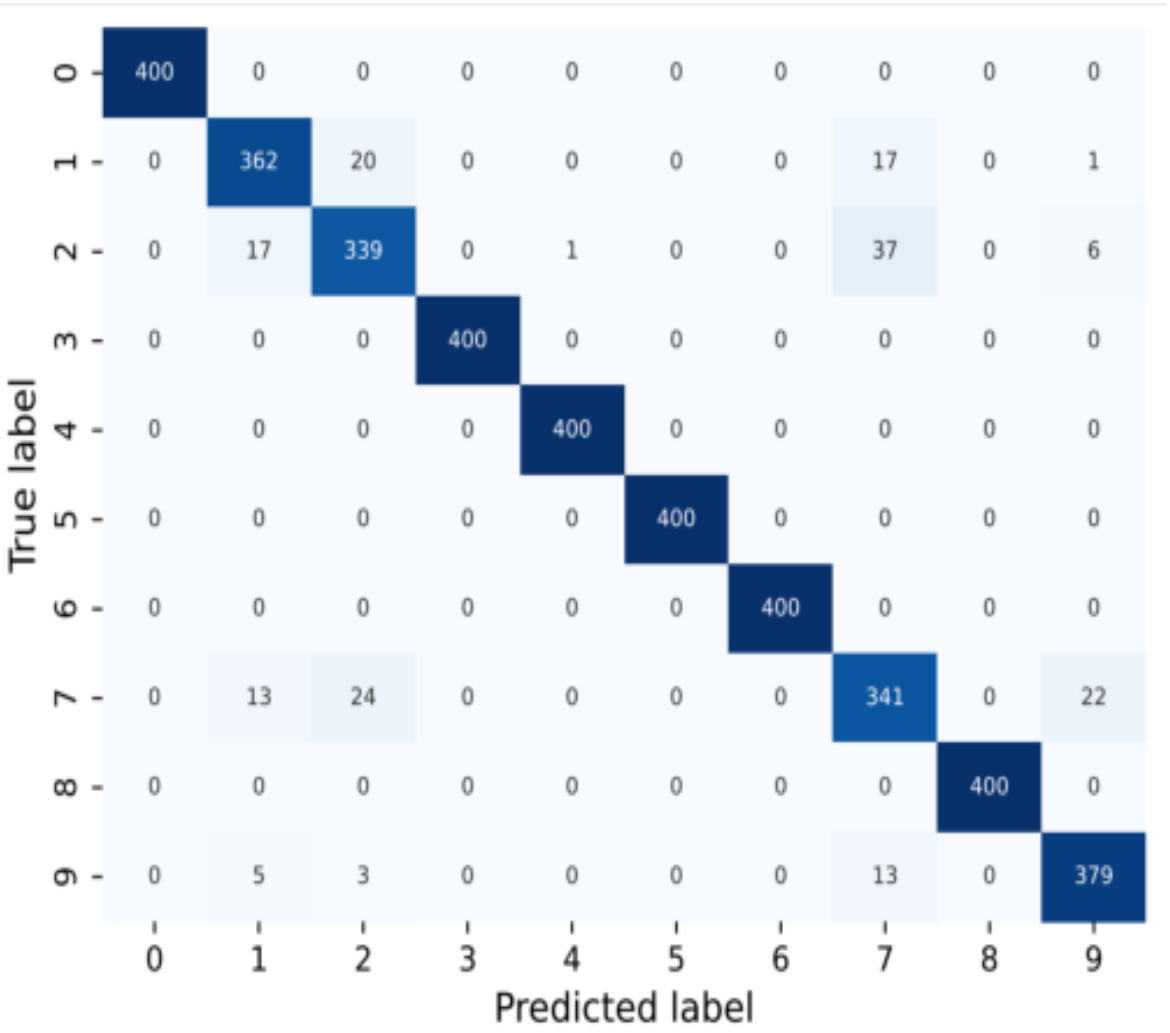


Fig 11: Precision, Recall and F1-score (IQ, FFT, A/Phase)

RESULTS

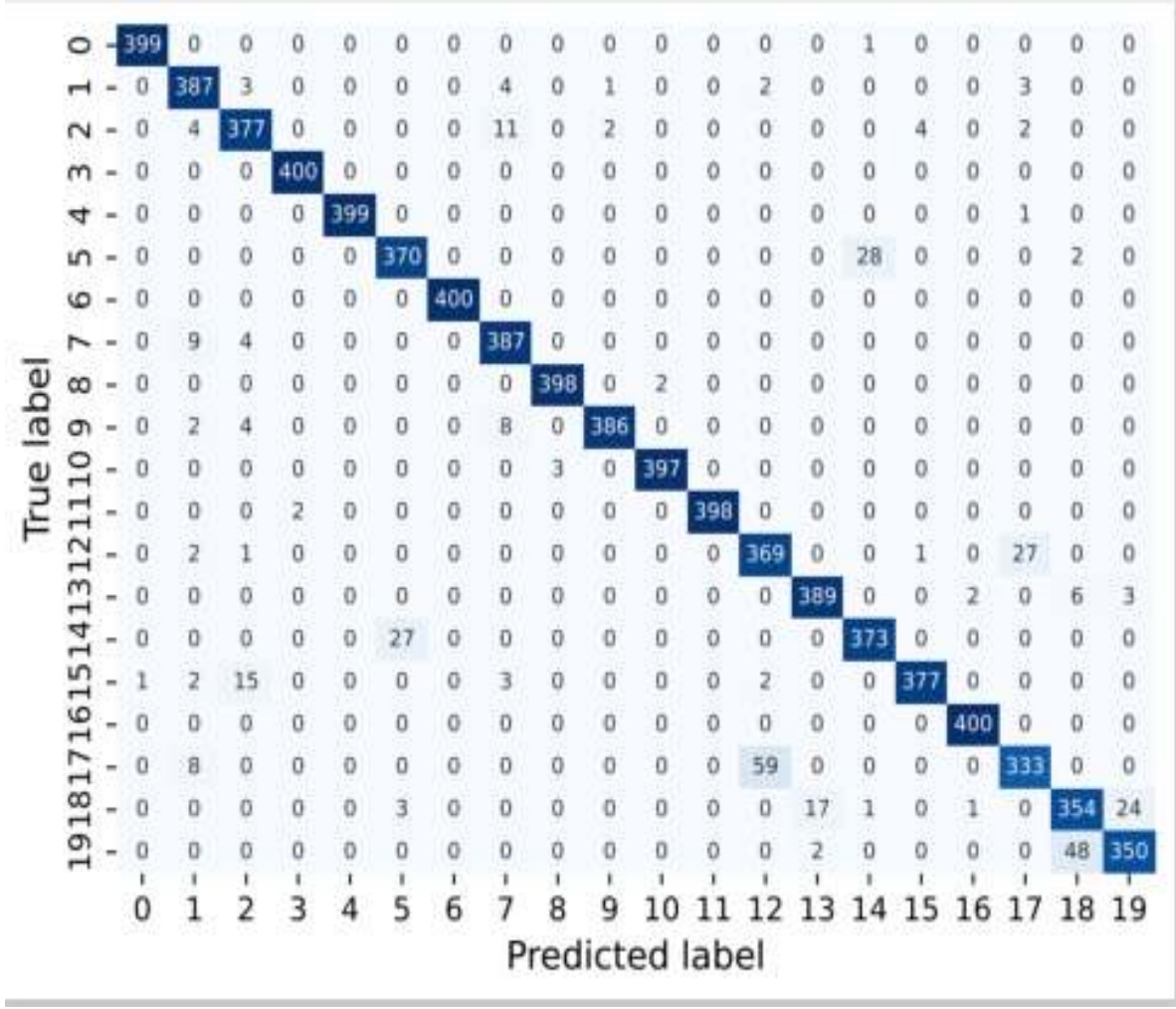


Confusion matrix using our collected data (3 devices only)

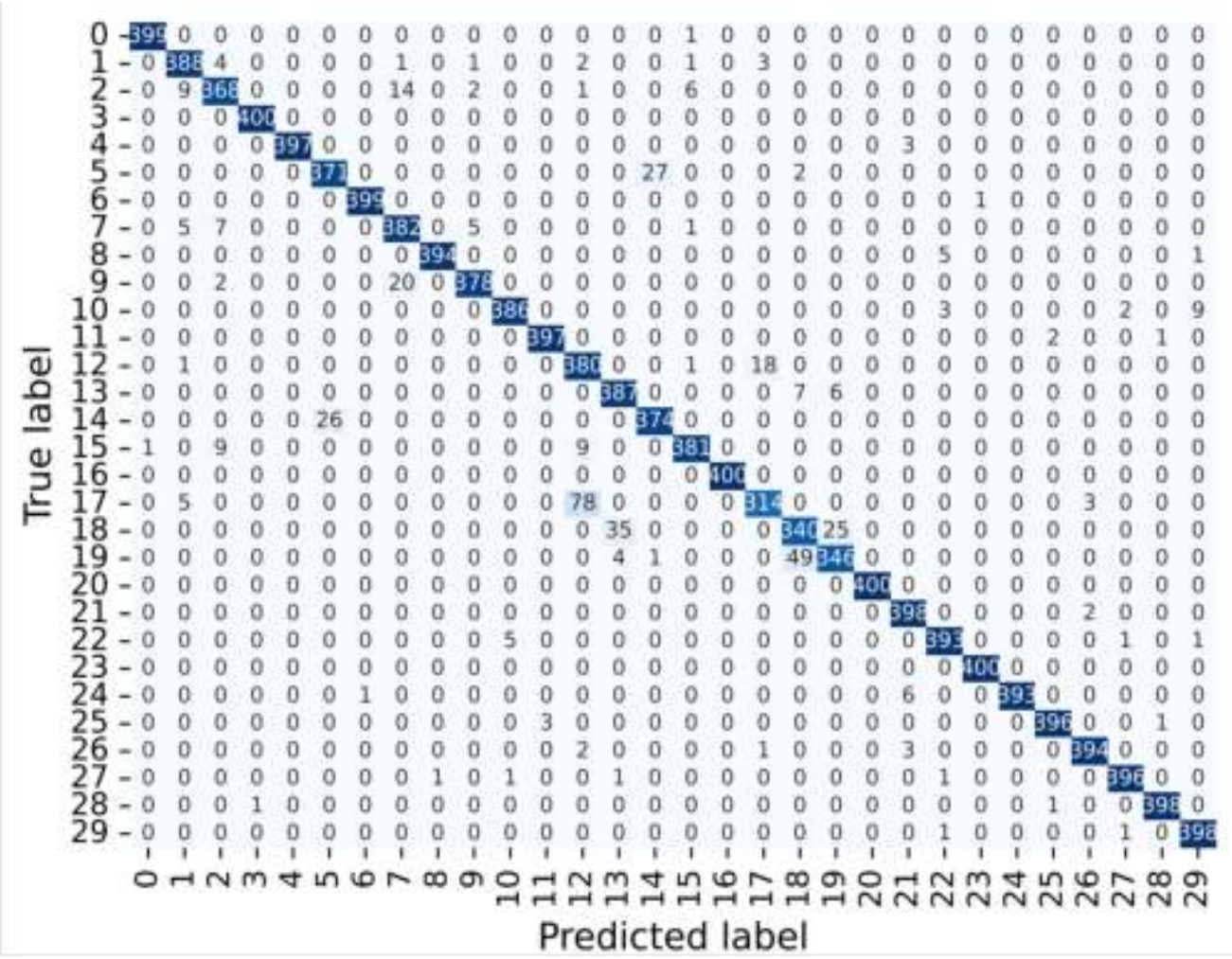


Confusion matrix using publicly available data (10 devices only)

RESULTS



Confusion matrix using publicly available data (20 devices only)



Confusion matrix using publicly available data (30 devices only)

6 Conclusion and Future Work

Conclusion: Challenges and Future Work

Despite end-end security LoRa device are still vulnerable to cyber attacks.
DL based LoRa device identification can be an effective system to identify malicious node without imposing any resource constraints on end device.

Challenges:

- Robustness of DL model
- Large scale publicly available dataset
- Scalability
- Features stability and model portability
- Open-set device identification

Future Work

- Currently: using FFT and Spectrogram representation of the IQ signals
- Transfer learning
- Solution for open-set device identification
- **Future:**
- Create a large-scale test-bed to collect real LoRa RF signal in different environments
- Propose Two factor authentication for LoRa device by complementing existing technique with DL based identification

Bibliography

[1] mproving Quality-of-Service in LoRa Low-PowerWide-Area Networks through Optimized RadioResource Managementhttps://www.statista.com/statistics/1101442/iot-number-of-connected-devices-worldwide/ accessed: 19/5/2023
[2] A. Elmaghbub and B. Hamdaoui, "LoRa Device Fingerprinting in the Wild: Disclosing RF Data-Driven Fingerprint Sensitivity to Deployment Variability," *IEEE Access*, vol. 9, pp. 142893–142909, 2021, doi: 10.1109/ACCESS.2021.3121606.
[3] S. U. Rehman, K. W. Sowerby, S. Alam, and I. Ardekani, "Radio frequency fingerprinting and its challenges," in 2014 IEEE Conference on Communications and Network Security, pp. 496–497, 2014
[4] G. Shen, J. Zhang, A. Marshall, L. Peng, and X. Wang, "Radio Frequency Fingerprint Identification for LoRa Using Deep Learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2604–2616, 2021, doi: 10.1109/JSAC.2021.3087250.
[5] P. Robyns, E. Marin, W. Lamotte, P. Quax, D. Singelee, and B. Preneel, "Physical-layer fingerprinting of LoRa devices using supervised and zero-shot learning," *Proc. 10th ACM Conf. Secur. Priv. Wirel. Mob. Networks, WiSec 2017*, pp. 58–63, 2017, doi: 10.1145/3098243.3098267.
[6] G. Shen, J. Zhang, A. Marshall, and J. R. Cavallaro, "Towards Scalable and Channel-Robust Radio Frequency Fingerprint Identification for LoRa," *IEEE Trans. Inf. Forensics Secur.*, vol. 17, pp. 774–787, 2022, doi: 10.1109/TIFS.2022.3152404.
[7] A. Elmaghbub, B. Hamdaoui, and A. Natarajan, "WideScan: Exploiting Out-of-Band Distortion for Device Classification Using Deep Learning," *2020 IEEE Glob. Commun. Conf. GLOBECOM 2020 - Proc.*, vol. 2020-Janua, 2020, doi: 10.1109/GLOBECOM42002.2020.9348138.

Thank you for your attention ☺

Questions Please …!