Experimental Approaches for Robust Authentication Based on Hardware Impairments

Valeria Loscrì, Ildi Alla and Selma Yahia

Inria Lille,
valeria.loscri@inria.fr, ildi.alla@inria.fr, selma.yahia@inria.fr

01

Context and Main Contributions

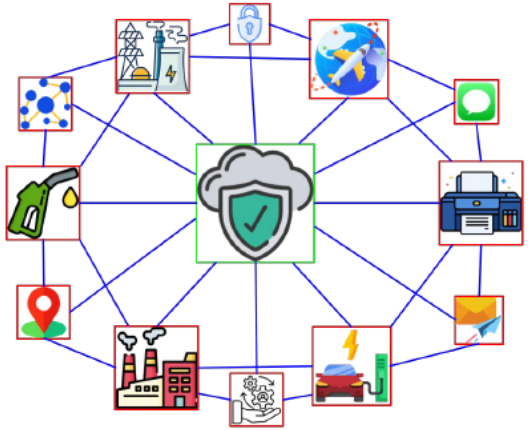
Why authentication

🌐Evolving Threat Landscape

📈IoT explosion with 30+ billion devices expected by 2025 across smart homes, healthcare, and industrial sectors.

🎯Unsecured devices become prime targets for sophisticated cyber attacks.

⚠️Increasing complexity of cyber threats targeting connected devices.



⚠️Security Risks

🔓Unauthorized access leading to critical data breaches and system compromises.

🕵️Device spoofing enabling dangerous Man-in-the-Middle attacks.

🤖Vulnerable devices being hijacked for large-scale DDoS botnet attacks.

✅Authentication Benefits

🔐Enhanced trust through robust device identity verification.

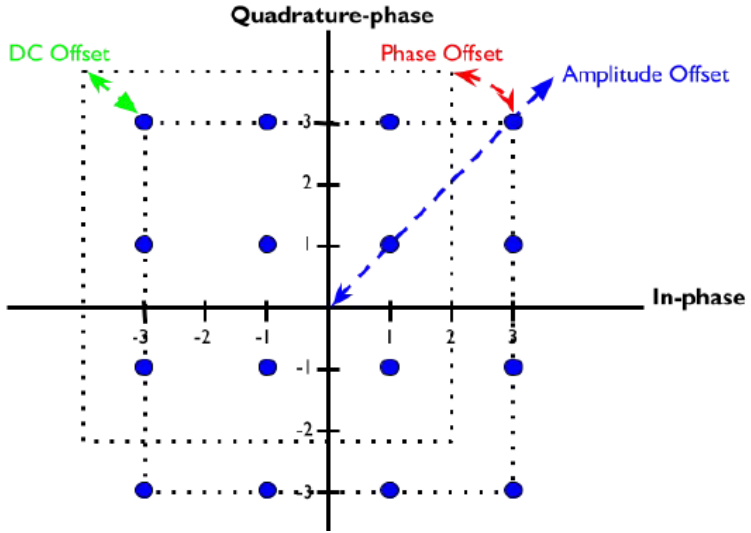
🛡️Comprehensive data protection against impersonation attempts.

💪Improved system resilience from sophisticated cyber threats.

Main Concept

📝Definition

PLA is a security approach that leverages the unique physical properties of wireless devices and communication channels to verify the authenticity of transmitted signals and communicating devices.



🔑Uses inherent hardware characteristics and signal properties as natural authentication factors.

📡Operates at the lowest layer of communication, where physical signals are transmitted.

🛡️Provides continuous and passive authentication without additional protocol overhead.

02

First Experiments

Carrier Frequency Offset (CFO)

Carrier Frequency Offset (CFO):

1- Frequency Spectrum Shift Analysis:

- Fourier Transform: Apply Fourier Transform to the received signal to observe the spectrum.
- Spectrum Observation: Look for a shift in the frequency spectrum, a hallmark of CFO.

2- Phase Rotation Analysis:

- Time-based Phase Change: Examine the phase change over time. CFO typically results in a linear phase rotation.
- Plot Phase vs. Time: Visualizing phase against time can clearly show this linear trend.

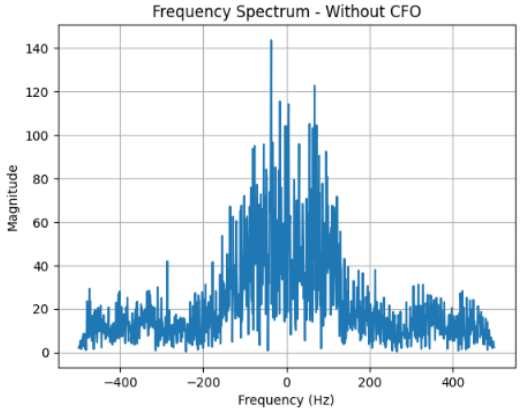
3- Constellation Diagram Distortion:

- Rotation Observation: CFO causes rotation of points in the constellation diagram.
- Time-based Analysis: Longer observation periods make this rotation more evident.

Frequency Spectrum Shift Analysis:

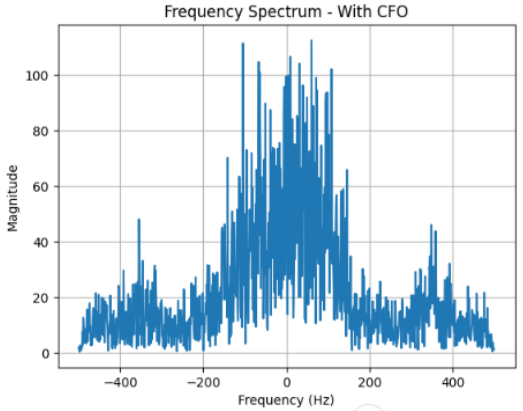
Without CFO:

- The spectrum is centered around 0Hz, as expected for a baseband signal
- Several peaks, multiple frequency components



With CFO:

- Similar pattern for the spectrum
- Not significant shift in the frequency components
- If there is a larger CFO, we expect a more important shift



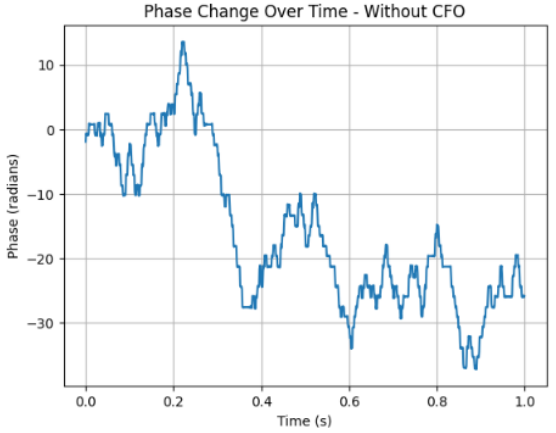
Phase Rotation Analysis

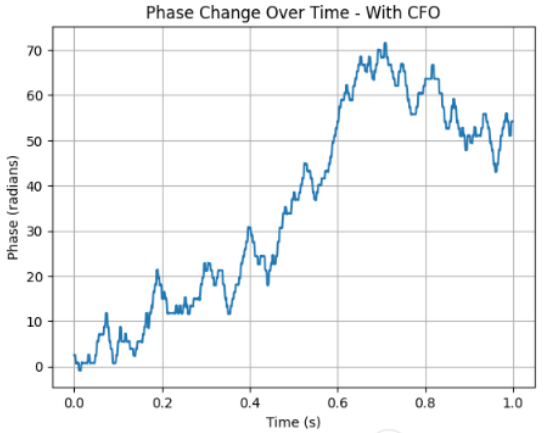
Without CFO:

- The phase changes are within a range of about 40 radians
- There is no clear patetrn of continuous linear increase or decrease (no significant frequency offset)
- Phase variations may be related to QPSK modulated signal, noise, etc.

With CFO:

- The phase increases steadily over time (i.e., presence of CFO)
- The range of the phase change is larger (i.e., from 0 to over 60 radians in 1 second)
- Increasing phase, typical of frequency offset between Tx and Rx

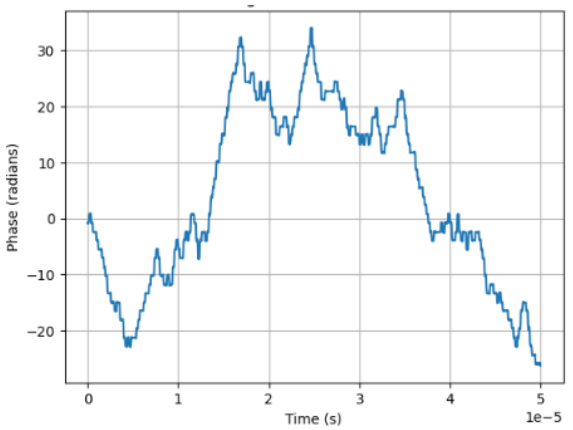




Phase Rotation Analysis – Small and Large CFP comparison

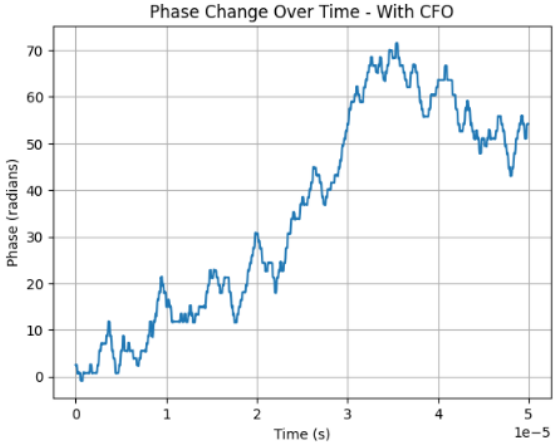
Small CFO:

- Phase changes with fluctuations, but in small ranges
- No steady increase or decrease
- Potential impact of noise or other imperfections



Large CFO:

- Clear upward trend, with significant offset
- Much larger phase change (over 60 radians)
- Continuous increase is typical of steady frequency difference between the Tx and the Rx oscillators



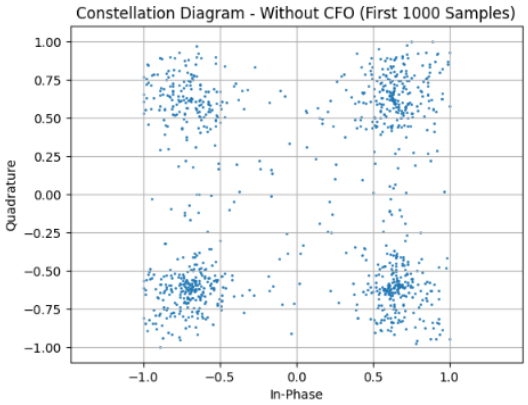
Constellation Diagram Distortion

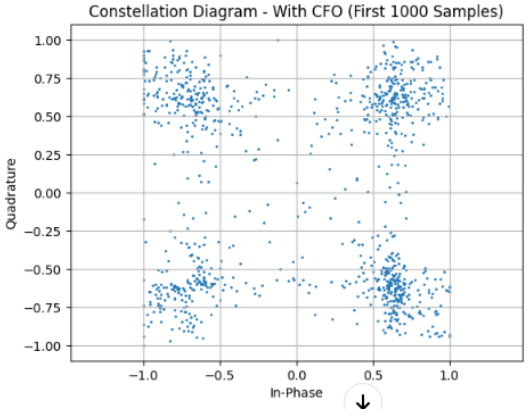
Without CFO:

- The points are spread around four clusters (the 4 states of QPSK modulation)
- Significant scatter in the points around the cluster (presence of noise)

With CFO:

- Scatter of noise is more pronounced
- Clusters are less distinct
- CFO causes the points to rotate around the constelaltion and spread out due to the pahse change due to the frequency offset





Direct Current Offset - DCO

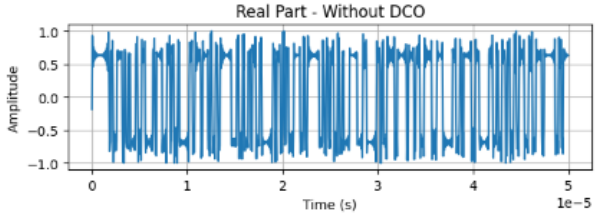
Without DCO:

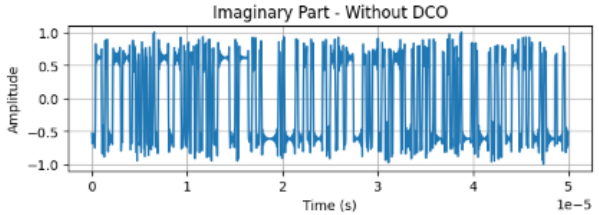
- Analysis in the time and frequency domain
- **Time Domain**

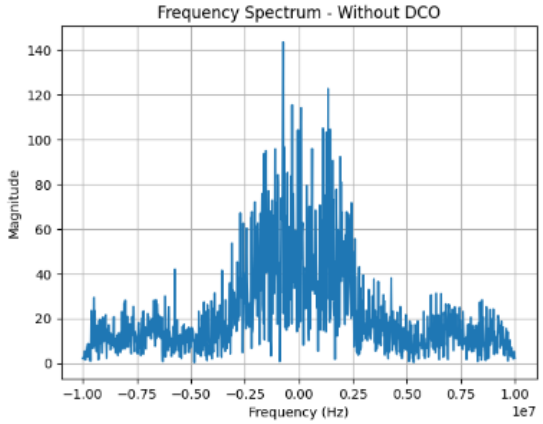
Real and imaginary part exhibit a centered oscillation around the zero amplitude

- **Frequency Domain**

Frequency spectrum simmetric around zero, no distinct peaks at zero.







Direct Current Offset - DCO

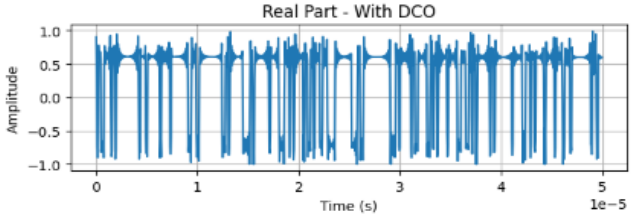
With DCO:

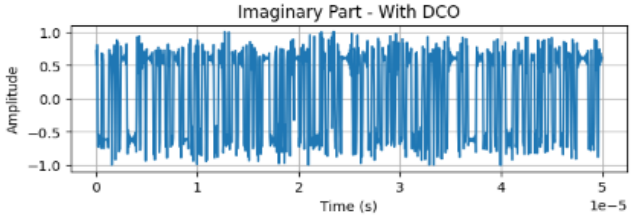
- Analysis in the time and frequency domain
- **Time Domain**

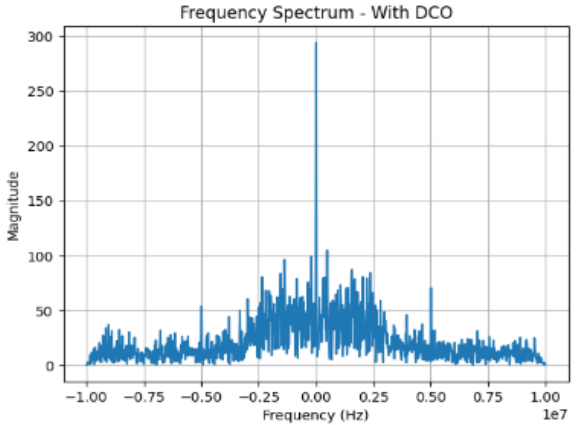
Slitght visibile shift in the baseline. Mean value different from zero confirms DCO presence

- **Frequency Domain**

Prominent peak at zero frequency, due to the constant component added to the signal







03

Some Takeaways

Observations & Questions

- One single impairment is not sufficient to characterize a device
- Combination of hardware impairments is more interesting
- Transient or steady state?
- How to process the signals?
- Statistic or ML approaches?

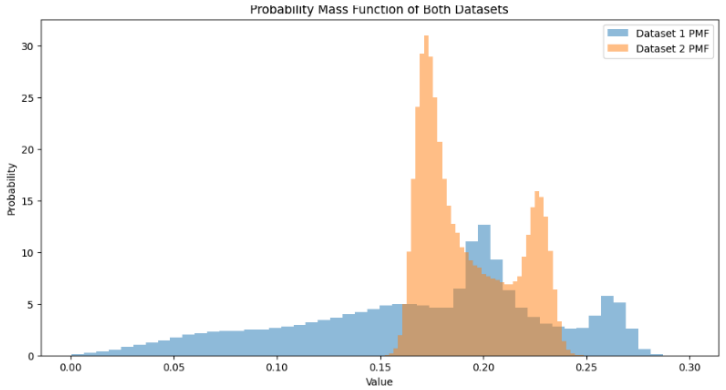
04

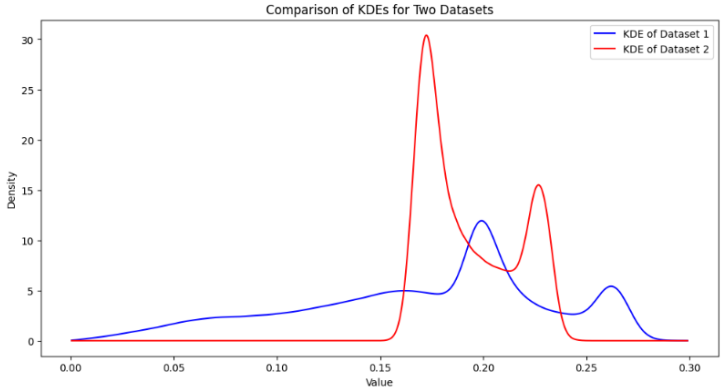
Features selections from KDE

Features Selection from KDE

Dataset1: small imperfection values
Dataset2: high imperfection values

Statistical approaches and data-driven methods have been considered to evaluate and quantify the difference between distributions





Feature Selection from KDE

Statistical Distance Measures:

- **Kullback-Leiber Divergence** (KL Divergence): meausres the difference between two probability distributions
- **Jensen-Shannon Divergence:** a symmetric and finite measure based on KL divergence

Numerical Metrics:

- **Mean Squared Error (MSE):** MSE can be computed on any two sets of KDE values
- **Root Mean Squared Deviation (RMSD):** average squared difference between the estimated and true values

Overlap Coefficient:

- How much one distribution overlaps with another

Features Extraction:

- **Peak Analysis:** features extraction of height, width, area, etc.
- **Moments:** skewness, kurtosis, variance, average, calculation for each distribution

05

Conclusions

Conclusions

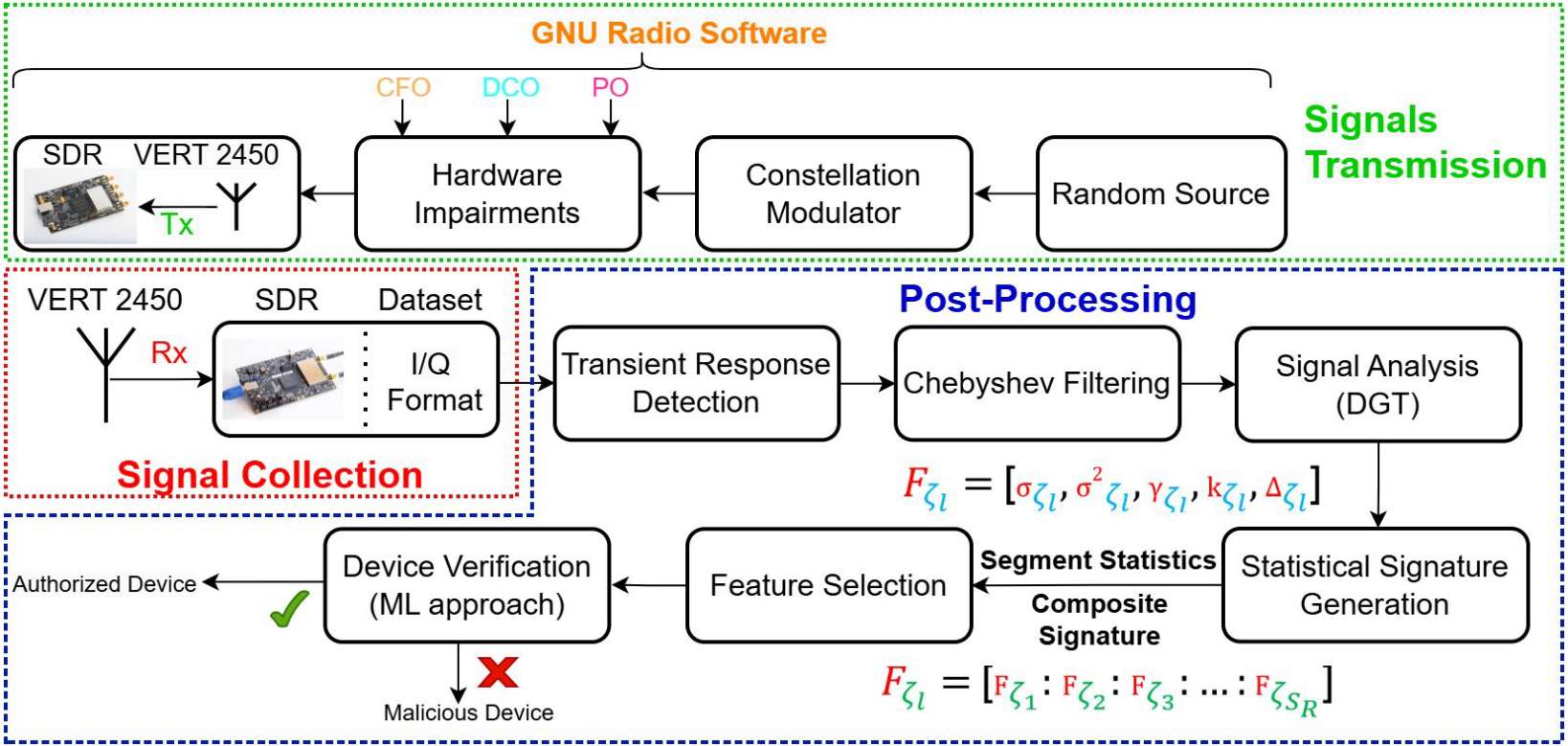
- Evidence that device impairments permit to distinguish one device from another «identical» device

- Anyway, one single imperfection is not sufficient to uniquely characterize a device

- Statistical approaches based on threshold were not effective to capture the differend induced by hardware impariments, even when combination of hardware impairments is considered

- Even though the KDE was not effective, some features extraction considered, in terms of kurtosis, skewness, etc. have been integrated in the ML based authentification approach

Conclusions – The Final Framework



Meta Discussion

- **Did you use any experimentation artifacts from the community?**

No, we did not find some open access datasets encompassing the combination of hardware impairments we needed.

Moreover, several similar appraoches in literature are based on theoretical/simulation validation.

- **Did you attempt to replicate or reproduce earlier research?**

No, for the same reasons as before, since we did not find the dataset and code availavle for what we aimed to do.

- **What insights can be gained from your methodology?**

We did an extended campaign of mearuements in order to collect data, combine them and search for the specifc features we needed. The main point is that to be valid signature need to be robust in different conditions and different scenarios. By making the measurements, we exactly searched for such a type of characteristics to motivate the netx steps of our work.

Meta Discussion

- **What challenges did you face, and how did you overcome them?**

At the beginning it was not very clear which types of data were the most inetrestng and representative, how to procees them and how to deal with the measurements. We did several measurements with data collected that we realized later were not very useful. After some initial difficults in thi sense, we re-organized the experiments, and we also applied data augmentiation techniques to characterize various scenarios. Moreover, at the beginnign we implemented statistical approaches, that are not idoneous for such a type of problems, since it was not possible to generalise the results obtained based on these approaches. In particular, we considered threshold-based statistical approaches, that were not suitable for authentication.

Meta Discussion

- **Did you produce any intermediate or unsuccessful results?**

Yes, as mentioned before, the first measurements were completely useful and the first appraoches we realized, based on statistical methods, by considering threshold appraoches, were not suitable for our scope.

- **Have you shared your experimental artifacts with the community?**

Yes, we have shared the experimental artifacts, and both the code and data are accessible and reproducible

Thank you