Smart, Automated, and Reliable Security Service Platform for 6G

Deliverable D5.1
Library of Known PHYs Attacks and PLS Dataset

ROBUST-6G project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101139068.

Date of delivery:	25/12/2024	Version:	1.0
Project reference:	101139068	Call:	HORIZON-JU-SNS-2023
Start date of project:	01/01/2024	Duration:	30 months

Document properties:

Document Number:	D5.1
Document Title:	Library of Known PHY Attacks and PLS Dataset
Editor(s):	Cem Ayyıldız (GOHM), Fatih Emre YILDIZ (GOHM)
Authors:	Özgül Ayyıldız (GOHM), Veli Can Yıldırım (GOHM), Stefano Tomasin (UNIPD), Mattia Piana (UNIPD), Laura Luzzi (ENSEA), Yris Brice Wandji Piugie (ENSEA), Arsenia Chorti (ENSEA)
Contractual Date of Delivery:	31/12/2024
Dissemination level:	PU
Status:	Final
Version:	1.0
File Name:	ROBUST-6G D5.1_v1.0

Revision History

Revision	Date	Issued by	Description
0.1.0	26.03.2024	ROBUST-6G WP5	Initial document draft created.
0.1.1	08.07.2024	ROBUST-6G WP5	Headers added & Ch1 Introduction written
0.1.2	29.07.2024	ROBUST-6G WP5	Added threat and technology charts based on D2.1 analysis. Included description of internal dataset. Added Executive Summary and adjusted headers for consistency.
0.1.3	30.08.2024	ROBUST-6G WP5	Drafting the externals dataset descriptions
0.1.4	30.09.2024	ROBUST-6G WP5	Methodology reviewed and finalized.
0.2.0	19.10.2024	ROBUST-6G WP5	Reviewed the document and reorganized sections.
0.3.0	16.12.2024	ROBUST-6G WP3	Document reviewed.
1.0	25.12.2024	GOHM	Final version ready for submission.

Abstract

This deliverable, D5.1 - Library of Known PHY Attacks and PLS Datasets, analyzes physical layer security (PLS) challenges in 6G networks, building on insights from the D2.1 6G Threat Analysis Report. It maps relevant datasets to key threats—Spoofing, Tampering, Repudiation, Information Disclosure, and Denial of Service (DoS)—and emerging 6G technologies, such as D-MIMO, RIS, and sub-THz communications.

The report catalogs external datasets and introduces new datasets developed within the ROBUST-6G project, such as the RF Fingerprinting Migration Dataset. Gaps in dataset coverage are identified, especially for Repudiation and DoS threats. The methodology ensures comprehensive dataset validation, while the data management framework supports efficient access via repositories like Zenodo. This deliverable provides key findings and recommendations for future research to address gaps, helping researchers and industry professionals develop secure and resilient 6G networks.

Keywords

PLS, 6G, Physical Layer Attacks, Physical Layer Security, Known PHY Attacks, Wireless Physical Layer, PLS Dataset

Disclaimer

Funded by the European Union. The views and opinions expressed are however those of the author(s) only and do not necessarily reflect the views of ROBUST-6G Consortium nor those of the European Union or Horizon Europe SNS JU. Neither the European Union nor the granting authority can be held responsible for them.

Executive Summary

This deliverable, D5.1 - Library of Known PHY Attacks and PLS Datasets, provides a comprehensive resource for professionals and researchers focused on Physical Layer Security (PLS) in 6G networks. As 6G technologies, including Distributed Multiple-Input Multiple-Output (D-MIMO), Reconfigurable Intelligent Surfaces (RIS), and sub-terahertz (sub-THz) communication become more prevalent, new vulnerabilities arise at the physical layer (PHY). Safeguarding the confidentiality, integrity, and availability of 6G communications is key to ensuring secure and resilient networks capable of supporting critical applications. This document builds upon the insights of the 6G Threat Analysis Report (D2.1) by identifying key threats and mapping them onto relevant datasets. It also introduces new datasets developed within the ROBUST-6G project, focusing on areas such as RF fingerprinting and adversarial attack detection.

The document follows a structured methodology based on the STRIDE threat model—Spoofing, Tampering, Repudiation, Information Disclosure, and Denial of Service (DoS)—and the CIA framework, emphasizing Confidentiality, Integrity, and Availability. Each dataset is evaluated for its alignment with these security criteria and with technological enablers such as D-MIMO, RIS, IoT, and mmWave communications. The deliverable not only organizes and maps over 80 external datasets but also fills critical gaps with new contributions from the project. These datasets aim to address security vulnerabilities across different 6G technologies and application areas, creating a robust library for future research.

The findings reveal that while existing datasets provide strong coverage for Spoofing and Tampering threats, other areas remain underrepresented. In particular, Repudiation and Denial of Service (DoS) attacks lack sufficient datasets, limiting the development of comprehensive mitigation strategies. Similarly, key emerging technologies, such as sub-THz and D-MIMO, show significant gaps in dataset coverage, especially in areas related to information disclosure and jamming attacks. Although IoT and LPWAN technologies are well-represented in terms of Spoofing, additional focus is needed on DoS and Information Disclosure threats to ensure their secure deployment.

We emphasize that addressing these gaps is essential for the success of 6G networks, especially towards the support of ultra-reliable low-latency communications (URLLC) and massive IoT deployments. To bridge these gaps, new datasets targeting Repudiation and DoS threats across emerging technologies are recommended. Furthermore, expanding IoT and LPWAN datasets to cover Information Disclosure will help develop comprehensive security mechanisms for smart cities, healthcare, and industrial applications. Collecting multi-modal datasets—combining RF signals, network traffic, and sensor data—will also enable better threat modelling and detection.

Looking ahead, standardization of dataset formats and documentation will facilitate collaboration and interoperability within the research community. Additionally, we recommend collecting real-world datasets from urban, industrial, and vehicular environments to improve the relevance and reliability of PLS solutions. Leveraging AI/ML-powered techniques for dataset generation and analysis is identified as a key strategy to accelerate progress, enabling more sophisticated threat detection and response mechanisms.

This deliverable serves as a foundational resource for advancing physical layer security research in 6G networks. By addressing the identified vulnerabilities and proposing targeted solutions, the ROBUST-6G project aims to foster the development of secure, resilient, and future-proof communication systems. The insights and datasets that are provided in the present report will play a crucial role in supporting innovation and ensuring that 6G networks meet the demanding security requirements of tomorrow's digital ecosystems.

Table of Contents

List of Tables

List of Figures

Acronyms and abbreviations

Term	Description
6-DoF	Six Degrees of Freedom
6G	6th Generation
AGV	Automated Guided Vehicle
AI	Artificial Intelligence
AoA	Angle of Arrival
API	Application Programming Interface
B5G	Beyond 5G
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BS	Base Station
BVLOS	Beyond-line-of-sight
CBRS	Citizens Broadband Radio Service
CC BY 4.0	Creative Commons Attribution 4.0
CCNC	Consumer Communications & Networking Conference
CFR	Channel Frequency Response
CIA	Confidentiality, Integrity, Availability
CNN	Convolutional Neural Network
COTS	Commercial Off-The-Shelf
CRC	Cyclic Redundancy Check
CSI	Channel State Information
CSP	Cyclostationary Signal Processing
D-MIMO	Distributed Multiple Input Multiple Output
D(x.x)	Deliverable (x.x)
dB	Decibel
dBi	Decibel-isotropic
DDoS	Distributed Denial of Service
DOA	Direction of Arrival
DoS	Denial of Service
DSP	Digital Signal Processing
DSSS	Direct-Sequence Spread Spectrum
DtS	Direct-to-Satellite
FFT	Fast Fourier Transform
GB	Gigabyte

GHz	Gigahertz
GPS	Global Positioning System
GSM	Global System for Mobile Communications
HITL	Hardware-in-the-loop
I/Q	In-phase and Quadrature components
IoT	Internet of Things
ISAC	Integrated Sensing and Communication
LMMSE	Linear Minimum Mean Square Error
LoRa	Long Range (Low Power Wide Area Network)
LOS	Line of Sight
LPAN	Laplacian Pyramid Attention Network
LPAN-L	Lightweight Laplacian Pyramic Attention Network
LPWAN	Low-Power Wide-Area Network
LTE	Long-Term Evolution
MHz	Megahertz
MIMO	Multiple Input Multiple Output
MitM	Man-in-the-Middle
ML	Machine Learning
mMIMO	Massive MIMO
mmWave	Millimeter Wave
MS	Millisecond
MSPS	Mega Samples Per Second
NLOS	Non-Line-of-Sight
NoN	Network of Networks
O-RAN	Open - Radio Access Network
OBU	On-Board Unit
OFDM	Orthogonal Frequency-Division Multiplexing
OFDMA	Orthogonal Frequency-Division Multiple Access
ORACLE	Optimized Radio Classification Algorithm for Large-scale Environments
OTA	Over-the-air
PHY	Physical
PLS	Physical Layer Security
PSSCH	Physical Sidelink Shared Channel
PUF	Physically Unclonable Function
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service

QPSK	Quadrature Phase Shift Keying
RAN	Radio Access Network
RC	Remote Controller
RF	Radio Frequency
RFFI	Radio Frequency Fingerprinting Identification
RFID	Radio Frequency Identification
RIS	Reconfigurable Intelligent Surfaces
RRU	Remote Radio Unit
RSSI	Received Signal Strength Indicator
RSU	Roadside Unit
RX	Receiver
SDR	Software Defined Radio
SF	Spreading Factors
SINR	Signal-to-interference-plus-noise
SITL	Software-in-the-loop
SKG	Secret Key Generation
SN	Sequence Number
SNR	Signal-to-Noise Ratio
STRIDE	Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, Elevation of privilege
TB	Terabyte
TBPS	Terabit Per Second
THz	Terahertz
TOA	Time of Arrival
TX	Transmitter
UAV	Unmanned Aerial Vehicle
UCLA	University of California Los Angeles
UDP	User Datagram Protocol
UE	User Equipment
UJI	Universitat Jaume I
ULA	Uniform Linear Array
URA	Uniform Rectangular Array
URLLC	Ultra-Reliable Low Latency Communication
USRP	Universal Software Radio Peripheral
UT	User Terminal
UWB	Ultra-Wideband

V2I	Vehicle to Infrastructure
V2X	Vehicle-to-Everything
VANET	Vehicular Ad-hoc Networks
VHT	Very High Throughput
VLOS	Visual Line-Of-Sight
WBAN	Wireless Body Area Network
WP	Work Package
WPT	Wireless Power Transfer

1 Introduction

The introduction of 6G networks is poised to revolutionize communication technologies, offering ultra-high data rates, low latency, and enhanced network reliability. However, such advancements pose new security challenges at the physical layer (PHY), where information is transmitted over radio frequency (RF) propagation paths. Advanced technologies like Distributed MIMO (D-MIMO), Reconfigurable Intelligent Surfaces (RIS), and sub-terahertz (sub-THz) frequencies expand the attack surface, exposing the network to potential security risks. Unaddressed vulnerabilities could undermine the trust, integrity, and availability which are key to support critical 6G applications.

Physical Layer Security (PLS) is crucial to preserving the confidentiality, integrity, and availability of 6G communications. Effective mitigation of emerging threats such as Spoofing, Tampering, Repudiation, Information Disclosure, and Denial of Service (DoS) attacks requires access to comprehensive datasets for simulation and analysis. The core enablers of 6G —D-MIMO, ISAC, RIS, and pervasive Artificial Intelligence (AI) / Machine Learning (ML)— aim to optimize performance while promoting sustainability, trustworthiness, and digital inclusion [EAC+21]. Yet, these innovations also introduce complex challenges. AI and ML-powered mechanisms will be indispensable for managing network behavior, dynamically optimizing communication protocols, and addressing sophisticated security threats in real-time. AI-based security solutions [SPY+21] will play a key role in enhancing user experience and maintaining seamless network operation under threat conditions [CBK+22].

Securing 6G networks requires a deep understanding of physical-layer vulnerabilities and of strategic mitigation approaches grounded in prior research. This focus is crucial, as the physical layer forms the foundation of wireless communications and is often the first line of defense against potential attacks [MJC+21]. Leveraging AI/ML integration with existing knowledge will enable the development of targeted solutions to address these evolving threats [RFG+24]. A systematic analysis of vulnerabilities, combined with precise mitigation strategies, will bolster the resilience of 6G networks against physical-layer attacks.

This deliverable, D5.1 - Library of Known PHY Attacks and PLS Datasets, builds upon the insights from Deliverable D2.1 - 6G Threat Analysis Report, offering an in-depth mapping of datasets to the identified threat landscape. The goal is to assess the adequacy of existing datasets and pinpoint areas where additional data collection is necessary. In addition to external datasets, the deliverable introduces new datasets developed within the ROBUST-6G project, by focusing on critical areas such as RF fingerprinting and adversarial attack detection to bridge the identified gaps.

By integrating both external research and project-specific contributions, this deliverable serves as a comprehensive resource for professionals and researchers engaged in 6G physical-layer security. It provides a detailed overview of the security challenges, available datasets, and unresolved issues, fostering the development of secure and resilient 6G networks capable of supporting future digital ecosystems.

1.1 Scope and Objectives

The scope of this document focuses on the systematic collection, analysis, and presentation of datasets and research related to physical layer attacks in wireless communications. It encompasses both external datasets—comprising existing datasets that address various wireless technologies and threats—and internal datasets generated by the ROBUST-6G project and its stakeholders. These internal datasets build upon prior research and provide new insights to expand the understanding of Physical Layer Security (PLS).

A key component of this document is the 6G Physical Threat Analysis summary based on Deliverable D2.1, which outlines major threats to the physical layer in 6G networks. The document maps these identified threats to relevant external and internal datasets, ensuring alignment with the current threat landscape. The datasets are further categorized by their relevance to the threat matrix from D2.1 and classified according to the technologies they support, providing a structured framework for future research efforts.

The primary objective of this document is to act as a comprehensive resource for researchers and practitioners working on physical layer security. It aims to collect and organize datasets covering both established and emerging 6G technologies. Through systematic mapping to the D2.1 threat matrix, the document ensures that

all critical security challenges are addressed. Another key objective is to identify gaps in the current dataset landscape by comparing available datasets against the threat matrix. This analysis highlights under-researched areas requiring further attention and provides guidance for future research by pinpointing areas where data is lacking.

Additionally, the document introduces new datasets developed within the ROBUST-6G project to address specific gaps in the threat landscape, such as those related to RF fingerprinting and adversarial attack detection. These contributions are designed to enhance the resilience and security of future 6G networks, supporting ongoing research and innovation in PLS.

1.2 Structure of the Document

This deliverable, D5.1 - Library of Known PHY Attacks and PLS Datasets, is organized into five core sections, guiding the reader through the challenges, methodologies, and solutions for physical layer security (PLS) in 6G networks, as follows:

- **Section 1 – Introduction**: Provides the purpose, scope, and objectives of the document, along with an overview of the ROBUST-6G project.
- **Section 2 – Methodology**: Describes the approach for collecting and categorizing datasets. It builds on the D2.1 6G Threat Analysis Report, using the STRIDE framework to align datasets with identified threats and technological enablers. This section also covers validation processes to ensure comprehensive dataset mapping and introduces the data storage and access management framework.
- **Section 3 – Datasets**: Catalogs external datasets relevant to physical layer attacks and presents new datasets developed by the ROBUST-6G project, including the RF Fingerprinting Migration Dataset, which addresses gaps in IoT authentication research.
- **Section 4 – Assessment and Evaluation of Datasets for 6G Security**: Evaluates how well datasets align with the 6G threat landscape and technological enablers, identifying areas of strong coverage and gaps that require further research.
- **Section 5 – Conclusions and Future Work**: Summarizes the key findings and offers recommendations for future research, focusing on addressing dataset gaps, expanding IoT and LPWAN datasets, and fostering collaboration through standardized data formats and documentation.

2 Methodology

This section describes the methodology used to build a comprehensive library of known PHY attacks and generate new Physical Layer Security (PLS) datasets for 6G networks. Given the transformational potential of 6G technologies, physical layer security requires a systematic identification of threats and their mitigation strategies. To achieve this, we refer to insights from Deliverable D2.1 – 6G Threat Analysis Report [ROB24-D21], aligning our dataset framework with the STRIDE model and CIA principles to ensure security and resilience across communication systems.

The methodology involves several stages, beginning with a thorough analysis of the security challenges identified in the D2.1 threat analysis. This report serves as a foundation, outlining emerging technologies like Distributed MIMO (D-MIMO), Reconfigurable Intelligent Surfaces (RIS), mmWave and sub-THz communications, and IoT. Each of these enablers brings new opportunities, as well as novel attack surfaces, which demand detailed scrutiny [HAK22, CNA24, RH22].

Our dataset compilation and mapping approach draws heavily on existing frameworks for threat identification and mitigation. The STRIDE framework (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service) allows us to categorize threats systematically. Meanwhile, the CIA model ensures that our analysis considers Confidentiality, Integrity, and Availability as key dimensions of network security. By employing these complementary frameworks, we ensure a holistic view of the threat landscape.

To maintain practical relevance, we gather datasets from a variety of research repositories such as IEEE Xplore, Zenodo, and arXiv. Each dataset is evaluated for alignment with the identified threats and the technologies they impact. These datasets are not only classified by technological domain (e.g., IoT, D-MIMO)

but are also mapped onto relevant threat categories following the guidelines established in D2.1. This ensures that our framework reflects real-world security needs while remaining adaptable to future developments.

This section will first provide key insights from the D2.1 6G Threat Analysis Report, including an overview of the STRIDE and CIA models used to categorize threats. We will then explore technological enablers such as RIS, mmWave, and IoT, detailing both their benefits and vulnerabilities. Following this, we will describe the collection and categorization of relevant datasets, leading to the creation of a dataset mapping framework for physical layer threats. Finally, we will conduct a validation and relevance assessment to ensure that the mapped datasets are aligned with the objectives of the ROBUST-6G project.

This step-by-step approach ensures that we address all the critical security challenges while laying the foundation for future research. By the end of this section, readers will have a comprehensive understanding of how physical layer security strategies are formulated, how datasets are collected and mapped onto threats, and how technological innovations can impact security dynamics in the 6G ecosystem.

2.1 Key Insights from the D2.1 6G Threat Analysis Report

The D2.1 report provides a comprehensive analysis of the evolving security landscape for 6G networks, focusing on the challenges posed at the physical layer (PHY). As 6G aims to seamlessly integrate digital and physical worlds by leveraging advanced technologies like D-MIMO, RIS, mmWave, and IoT, it introduces novel attack surfaces. The D2.1 report emphasizes that physical layer security (PLS) will play a crucial role in safeguarding network performance, especially as 6G enables ultra-reliable low-latency communication (URLLC) for critical services.

This section summarizes the relevant insights, frameworks, and technological enablers identified in D2.1. These elements serve as the basis for threat identification and dataset mapping, which will be explored in subsequent sections.

2.1.1 Frameworks for Security Assessment: CIA and STRIDE

Security assessment in D2.1 relies on two key frameworks: CIA (Confidentiality, Integrity, Availability) and STRIDE (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service). These frameworks help structure the identification of vulnerabilities and ensure that the methodology addresses all dimensions of physical layer security.

- **CIA Model:** The CIA model is foundational for assessing network security. It evaluates whether communication systems can protect the confidentiality (unauthorized access prevention), integrity (unauthorized alteration prevention), and availability (ensuring reliable access) of transmitted data. For physical layer security, confidentiality is challenged by threats like eavesdropping over high-frequency channels, integrity by signal tampering, and availability by jamming attacks that disrupt communications.
- **STRIDE Model:** The STRIDE framework is used to systematically identify threats relevant to the physical layer. While it includes six categories—Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service (DoS), and Elevation of Privilege—the first five are most relevant to 6G PHY security. Elevation of Privilege typically applies to higher-layer attacks and is excluded from our analysis. The STRIDE framework provides a detailed view of how each threat manifests at the physical layer, guiding the dataset mapping process later in this chapter.

2.1.2 Technological Enablers: Security Benefits and Challenges

The D2.1 report identifies several key enablers that will define the 6G landscape, including D-MIMO, RIS, mmWave, and IoT. While these technologies provide opportunities to enhance performance and efficiency, they also introduce new vulnerabilities that need to be addressed.

D-MIMO: D-MIMO technology is poised to address several critical challenges in 6G networks by leveraging multi-user MIMO with coherent joint transmission and interference suppression. A key advantage of D-MIMO is its ability to enhance network capacity and coverage, particularly in dense urban and indoor environments. By distributing antenna arrays over a wide geographic area, D-MIMO systems can provide proximity gains, improving signal quality and reducing the effects of path loss. This leads to more reliable connections and higher spectral efficiency, even in challenging propagation conditions.

Another notable benefit of D-MIMO is its ability to enhance mobility support within the network. By enabling seamless multi-connectivity between distributed antenna nodes, D-MIMO reduces the frequency and complexity of handovers, ensuring a smoother experience for users on the move. Additionally, D-MIMO has the potential for energy-efficient deployments by distributing power demands across multiple nodes, optimizing resource allocation, and reducing the need for high-power, centralized base stations. This decentralized approach not only improves energy efficiency but also makes the system more flexible and scalable for future 6G applications [HYM+23].

However, D-MIMO also introduces several security vulnerabilities. The complexity of securing multi-connectivity and key agreement processes can lead to unauthorized access and data breaches (Information Disclosure). The denser network architecture and increased use of fronthaul links create more potential attack points, making the network more susceptible to disruption (Denial of Service). Ensuring MAC layer security becomes both critical and challenging, with the risk of data interception and alteration (Tampering). Additionally, the management of encryption keys for different signal paths in joint transmissions presents significant challenges, potentially leading to compromised communications (Spoofing and Information Disclosure) [JLS+24].

RIS: Reconfigurable Intelligent Surfaces (RIS) enhance wireless communication by dynamically manipulating electromagnetic waves to control how signals propagate in the environment [DRC+20]. This control allows RIS to improve key network performance metrics, such as capacity, coverage, positioning accuracy, security, and energy efficiency. By extending coverage to non-line-of-sight areas and facilitating smooth signal transitions between indoor and outdoor environments in both directions [HBK+24], RIS optimizes system performance in diverse scenarios. Additionally, by adjusting the phase, amplitude, and polarization of incident signals, RIS can steer, focus, or scatter wireless transmissions, reducing interference and enhancing signal strength in challenging conditions. As a result, RIS offers a versatile and efficient solution for various 6G applications.

Despite these advantages, RIS introduces several security threats. There's a risk of physical control of RIS devices being compromised, leading to unauthorized signal manipulation (Spoofing). The control channels for RIS, whether implicit or explicit, are vulnerable to tampering, which can cause operational disruptions (Tampering). In certain operational modes, the reliance on external entities for control increases the risk of information disclosure if control channels are intercepted (Information Disclosure). The complexity of managing RIS operations across multiple devices also presents challenges that could lead to service disruptions (Denial of Service) [SHC+24].

mmWaves and sub-THz: The integration of mmWaves (millimetre waves) and sub-terahertz (sub-THz) frequencies into 6G networks is poised to deliver unprecedented data rates, potentially exceeding 100 Gbps and even reaching terabits per second (Tbps). These frequencies, range from 30 GHz to 300 GHz [RXM+17] for mmWaves and from 100 GHz to 3 THz for sub-THz [RSC+19]. These high frequencies offer significant bandwidth, enabling ultra-high-speed communications. The narrow beamwidth and directional nature of these signals inherently enhance privacy and security by making eavesdropping more challenging.

However, these high frequencies face unique challenges. The exceptionally high propagation losses require the use of high-gain antennas [JKM23], necessitating precise real-time localization and tracking of devices [TKN+24]. While the directional nature of the signals enhances security, it also makes the system more vulnerable to beam misalignment and signal blockage. Additionally, the large available bandwidth makes these links potentially more susceptible to jamming attacks. Research is ongoing to address these vulnerabilities and develop countermeasures specific to mmWave and sub-THz communications.

Integrated Sensing and Communication (ISAC): ISAC merges sensing technologies like radar and LiDAR with communication, supporting applications like autonomous vehicles and smart cities. This convergence provides new opportunities but also creates risks related to data manipulation and repudiation. For instance, an attacker could tamper with sensor data to deny responsibility for an action. Datasets for ISAC include sensor fusion logs and validation datasets to ensure the integrity of transmitted data.

Internet of Things (IoT): IoT in 6G networks promises to connect an unprecedented number of devices, enabling smart environments, efficient resource management, and new services across various sectors including healthcare, agriculture, and transportation. Technologies like RFID, LoRa, and V2X sidelinks offer extensive coverage, low power consumption, and direct device-to-device communication capabilities, enhancing the flexibility and efficiency of IoT deployments.

However, IoT devices often have limited computational resources, making traditional security measures challenging to implement. This is particularly true for passive RFID tags, which are vulnerable to data tampering attacks. LoRa networks, while offering long-range communication, face challenges in ensuring end-to-end security across vast distances. V2X sidelinks, critical for applications like autonomous vehicles, face authentication and trust establishment challenges, especially in areas with limited cellular coverage. Furthermore, the limited security capabilities of IoT devices make them susceptible to exploitation, such as being compromised to act as bots in Distributed Denial of Service (DDoS) attacks, further increasing their threat to network operations. The sheer number and diversity of IoT devices also significantly increases the attack surface of the network, making comprehensive security management a complex task.

As emerging technologies like D-MIMO, RIS, mmWaves, ISAC, and IoT redefine the 6G landscape, they bring not only unprecedented opportunities but also unique security challenges, especially at the physical layer. The vulnerabilities inherent in these technologies—ranging from unauthorized access and data breaches in D-MIMO to signal manipulation risks in RIS, and jamming threats in mmWaves—highlight the critical need for robust physical layer security (PLS) measures. Addressing these challenges requires systematically analyzing potential attack vectors and vulnerabilities, which will guide the development of targeted PLS solutions and the compilation of a comprehensive PLS-related dataset.

The threat analysis in D2.1 serves as a crucial foundation for our work in compiling the library of known PHY attacks and generating new PLS datasets. It guides our research questions and experimental setups, ensuring that we address the most pressing security concerns in the physical layer of 6G networks. By aligning our efforts with the threats identified in D2.1, we can develop more targeted and effective security measures for the next generation of wireless communication systems.

2.1.3 Threat Matrix of the Physical Layer from D2.1

The D2.1 6G Threat Analysis Report identifies five primary threat categories that are particularly relevant to physical layer security in 6G networks. These categories form the foundation of our threat matrix and guide our approach to dataset mapping. Each category represents a distinct type of security challenge that 6G networks must address.

The threat categories are derived from the STRIDE model, a widely used threat modelling framework. STRIDE stands for Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, and Elevation of privilege. However, in the context of physical layer security for 6G networks, the Elevation of privilege category is less directly applicable, as it primarily relates to higher-layer security concerns. Therefore, our focus is on the first five categories, which are most relevant to physical layer threats.

- **Spoofing** threatens the authenticity of network entities. It refers to attacks where an entity falsely claims to be another entity within the network. In 6G networks, this could manifest as impersonation attacks (e.g., false base station attacks), Man-in-the-Middle (MitM) attacks or replay attacks. The ultra-dense nature of 6G networks and the increased number of connected devices amplify the potential impact of spoofing attacks.
- **Tampering** compromises the integrity of data and systems. It involves unauthorized modification of data, signals, or system components. In the context of 6G physical layer security, this could include signal injection attacks, data manipulation in transit, or hardware tampering. The high data rates and low latency requirements of 6G networks make integrity protection particularly challenging and critical.
- **Repudiation** challenges the non-reputability of network transactions. It relates to the denial of participation in communications or transactions within the network. Examples include transaction denial, log alteration, and identity spoofing for repudiation purposes. In 6G networks, ensuring non-repudiation becomes more complex due to the network of networks (NoN) architecture and the increased number of interconnected devices and systems.
- **Information Disclosure** threatens the confidentiality of sensitive data. It refers to unauthorized access to or exposure of sensitive information. In 6G networks, this could occur through eavesdropping, side-channel attacks, or data interception in multi-connectivity scenarios. The use of higher frequency bands in 6G (mmWave, THz) introduces new challenges and opportunities for ensuring confidentiality.
- **Denial of Service** attacks compromise the availability of network resources. These attacks prevent legitimate users from accessing network resources or services. In 6G networks, this could include

jamming attacks, Distributed Denial of Service (DDoS) attacks, or resource exhaustion attacks. The ultra-reliable low latency communication (URLLC) requirements of 6G make availability a critical concern, as even short interruptions can have significant impacts on critical applications.

These threat categories are particularly relevant to 6G physical layer security due to the unique characteristics of 6G networks. The ultra-dense networks and massive connectivity increase the potential attack surface for various threats. The use of higher frequency bands (mmWave, THz) introduces new vulnerabilities to jamming and other denial of service attacks. The network of networks (NoN) architecture in 6G increases the complexity of ensuring end-to-end integrity and non-repudiation.

Our dataset mapping process aligns with these categories, ensuring comprehensive coverage of the 6G threat landscape and facilitating the development of robust physical layer security solutions. These are the threats mentioned on the D2.1 document related to Physical Layer Matrix.

Table 2-1: Summary of PHY Level Threat Matrix

Threat	Attack Surface	Description	Attacks
Spoofing	Authenticity	False claims of identity	Impersonation, MIM, Replay Attacks
Tampering	Integrity	Unauthorized modification	Signal Injection, Data manipulation
Repudiation	Non-reputability	Denial of participation	Transaction Denial, Log alteration
Information Disc.	Confidentiality	Unauthorized access to data	Eavesdropping, Side-channel attacks
DOS	Availability	Prevention access	Jamming, DDoS, Resource exhaustion

2.2 Compilation and Categorization of Relevant Datasets

The process of compiling and categorizing relevant datasets is a critical step toward building a robust library for Physical Layer Security (PLS) in 6G networks. The aim is to gather datasets that reflect real-world threats and align with the physical layer challenges outlined in D2.1 – 6G Threat Analysis Report.

This section describes the methods used to identify, assess, and organize datasets, ensuring comprehensive coverage of the STRIDE-based threat categories and relevant 6G technological domains. These datasets serve as the foundation for creating security strategies capable of mitigating threats like spoofing, tampering, information disclosure, and denial of service.

2.2.1 Identification of Online Sources

To ensure a wide-ranging dataset library, we utilized a variety of online platforms and academic repositories. Key sources include:

- **IEEE Xplore:** A major repository for peer-reviewed research articles and datasets related to wireless communications.
- **Zenodo:** An open-access repository hosting datasets across scientific fields, including IoT and networking technologies.
- **arXiv:** A preprint repository containing cutting-edge research in wireless networks and machine learning, providing early insights into 6G technologies.

Additionally, white papers, industry reports, and project-specific datasets (e.g., those developed within the ROBUST-6G project) have been included to expand the scope beyond academic literature.

2.2.2 Dual-Path Categorization

The dual-path categorization ensures that datasets comprehensively address the identified threats by following a two-step framework. This approach begins with the primary categorization based on threat types, ensuring alignment with the STRIDE model. The secondary categorization focuses on either technological enabler—representing cutting-edge 6G innovations—or technology domains, covering well-established communication technologies. This structure ensures flexibility, accommodating both emerging and traditional datasets.

2.2.2.1 Primary Categorization by Threat Type (STRIDE Model)

The primary categorization groups datasets based on the five primary threat categories outlined in the STRIDE framework. This ensures that datasets align with the security concerns critical to 6G networks.

2.2.2.2 Secondary Categorization: Technological Enablers vs. Technology Domains

Once datasets are categorized by threat type, they are further organized under technological enablers or technology domains, depending on their relevance.

Technological Enablers (6G-Specific Innovations)

These technological enablers represent the core innovations driving the evolution of 6G networks. While they introduce enhanced performance, efficiency, and new capabilities, they also bring unique security challenges that require specialized mitigation strategies. These technology enablers are as follows:

- Distributed Multiple-Input Multiple-Output (D-MIMO)
- Reconfigurable Intelligent Surfaces (RIS)
- High-Frequency Wireless Communication Technologies (mmWave and Sub-THz)
- Integrated Sensing and Communication (ISAC)
- Internet of Things (IoT)

Technology Domains (General Communication Technologies)

For datasets that do not align with 6G enablers, they are categorized into general communication domains based on use case and technical function.

- **Short-Range Communication:** This category includes technologies such as Wi-Fi, Bluetooth, and Zigbee [BWR+24], which are commonly used in personal area networks and smart devices. These technologies are essential for low-power embedded systems and are the subject of extensive research.
- **Low-Power Wide-Area Networks (LPWAN):** LPWAN technologies, including LoRa, LoRaWAN, and Sigfox [SBF+24], are designed for long-range communication with low power consumption, making them ideal for Internet of Things (IoT) deployments in smart cities and remote sensing applications.
- **Cellular & Mobile Communication:** This category encompasses technologies such as LTE, 5G NR [Sha22], and Cellular Vehicle-to-Everything (C-V2X) [CHS+20]. These technologies are critical for mobile communication, vehicular networks, and connected devices, playing a significant role in modern communication systems.
- **Vehicular Technologies:** Technologies such as Vehicular Ad-hoc Networks (VANETs), Unmanned Aerial Vehicles (UAVs or drones), and Vehicle-to-Everything (V2X) communication fall under this category. These are essential for autonomous and connected vehicle applications.
- **Software-Defined Radio (SDR) and Signal Processing:** SDR platforms and signal processing techniques are pivotal for experimenting with and implementing various communication protocols. They provide flexibility in studying and developing new communication methods.

Table 2-2: Grouped Technologies

Category	Technologies
Short-Range Communication	Wi-Fi, Bluetooth, Zigbee
LPWAN	LoRa, LoRaWAN, Sigfox
Cellular Technologies	LTE, 5G NR, Cellular (C-V2X)
Vehicular Technologies	VANETs, UAV (Drones), V2X
SDR	Signal Processing

2.3 Validation of Dataset Mapping and Relevance to 6G Threats

The Validation and Relevance Check phase ensures that the methodology used for mapping datasets to physical layer threats is comprehensive, consistent, and aligned with the objectives of improving security in the 6G landscape. This step builds confidence in the framework before proceeding with dataset collection and analysis

in subsequent sections. A detailed evaluation ensures that all critical threats and enabling technologies are adequately represented within the dataset mapping process.

2.3.1 Methodological Validation

Methodological validation ensures that the datasets align with essential threat categories and remain relevant to the technological innovations shaping 6G. We first validated the alignment of the mapping process with the five primary physical layer threats identified in the D2.1 Threat Analysis Report: Spoofing, Tampering, Repudiation, Information Disclosure, and Denial of Service (DoS). Ensuring comprehensive coverage of these threats guarantees that the datasets address the most critical security concerns within the 6G physical layer.

Additionally, we evaluated the applicability of the methodology to key technological enablers relevant to 6G networks, such as D-MIMO, RIS, mmWave, sub-THz frequencies, and IoT. This ensures that the datasets reflect the new attack surfaces and security challenges introduced by these innovations while also accounting for their performance-enhancing capabilities. Key parameters for each threat category were verified to ensure effective dataset mapping. For spoofing and repudiation, parameters like RF Fingerprinting and PLS-based Authentication were identified as essential to capture unauthorized access attempts and identity misrepresentation. For tampering and information disclosure, metrics such as Channel State Information (CSI) and Secret Key Generation (SKG) were deemed crucial to assess signal integrity and secure communication processes. This thorough verification ensures the datasets accurately capture both threat dynamics and security countermeasures.

Finally, a coverage assessment was conducted to identify any gaps in the threat categories or technologies covered by the datasets. This review ensures that the framework remains flexible and adaptable to emerging threats and technologies, helping us address new challenges in future work.

2.3.2 Process Validation

Process validation ensures the consistent and systematic application of the dataset mapping framework across all identified threats and technology domains. A consistency check was performed to verify that the mapping process was applied uniformly across all threats and technological enablers. This guarantees that datasets are categorized logically and without bias, providing a coherent framework for threat analysis.

Additionally, a relevance assessment was conducted to confirm alignment with the overall goals of the ROBUST-6G project. The framework was designed to support the development of effective Physical Layer Security (PLS) strategies by addressing the unique security challenges introduced by 6G technologies. Ensuring that the mapped datasets align with the project's objectives allows for targeted research and development efforts.

2.4 Data Storage and Access Management

This section outlines our approach to storing and managing the library of known PHY attacks and PLS datasets. It covers our choice of data repository, library structure, security measures, and access protocols. These strategies ensure the efficient organization, protection, and responsible use of both external and internally developed datasets, creating a valuable resource for 6G security research.

2.4.1 Repository Selection for the Library

For the datasets created within the ROBUST-6G project, we required a reliable and secure storage solution. After careful evaluation, we selected Zenodo as our primary repository, as it offers several advantages that align with our project needs:

- Long-term data retention (20 years), ensuring the longevity of our datasets;
- Support for large file uploads, accommodating diverse dataset sizes;
- Robust security measures, crucial for storing potentially sensitive data;
- Open-access capabilities, facilitating knowledge sharing within the research community.

The ROBUST-6G Zenodo community page, which hosts our library, can be accessed at: Zenodo - ROBUST-6G

2.4.2 Library Structure and Organization

Our library is structured to encompass internally developed datasets related to physical layer attacks and security measures. The data is organized into distinct categories for ease of navigation:

- PLS Countermeasures: Data related to physical layer security techniques and their effectiveness;
- Experimental Results: Findings from internal experiments conducted by project stakeholders;
- Simulation Data: Results from simulations of PHY attacks and defences.

Each dataset within the library is accompanied by comprehensive metadata, including:

- Detailed description of the security measure or experiment;
- Relevant 6G technologies (e.g., D-MIMO, RIS, mmWave);
- Data collection methodology;
- Potential use cases for researchers and developers.

3 Datasets

This section provides a comprehensive overview of datasets relevant to physical layer security (PLS) in 6G networks. It is divided into two sections: external datasets and the contributions made by the ROBUST-6G project. The external datasets cover a wide range of technologies and threats, providing critical resources for researchers and developers working on PLS. Each dataset description includes key features, threat coverage, and application areas to aid in selecting the most suitable data for research purposes.

The second section highlights new datasets developed within the ROBUST-6G project to fill gaps identified in existing datasets. These contributions address emerging threats in areas such as RF fingerprinting and adversarial attack detection, providing valuable additions to the dataset landscape.

3.1 External Datasets

This section documents publicly available external datasets that support research on physical layer security in wireless networks. These datasets encompass various technologies, including Wi-Fi, Bluetooth, LoRa, and cellular networks, and address different types of physical layer threats such as spoofing, tampering, jamming, and information disclosure.

Each dataset entry provides a brief description, highlights its primary features, and outlines the type of threat it addresses. In addition to traditional wireless communication datasets, this section includes specialized datasets focused on emerging technologies like Massive MIMO, millimeter-wave (mmWave) communication, and IoT networks. By cataloging these datasets, this section aims to serve as a resource for researchers, helping them identify the most appropriate datasets for their studies on physical layer security in 6G networks.

3.1.1 Physical-Layer Fingerprinting of LoRa devices using Supervised and Zero-Shot Learning

The Physical-Layer Fingerprinting of LoRa Devices Dataset [RML+17] contains raw signals (complex float I/Q (In-phase and Quadrature) samples) used in experiments for fingerprinting LoRa devices. The dataset includes four sets of data “lora1msps, lora2msps, lora5msps and lora10msps” corresponding to a sampling rate of 1, 2, 5 and 10 Mbit/s, respectively. These symbols are extracted from LoRa frames transmitted by 22 different devices and received by a USRP B210 (Universal Software Radio Peripheral). This dataset supports research on RF fingerprinting using supervised and zero-shot learning methods.

Dataset ID | Name	**DAT001 | Physical-Layer Fingerprinting of LoRa devices using Supervised and Zero-Shot Learning**
Key Features	• Contains I/Q samples from 22 different LoRa devices • Multiple data with varying sampling rates (1, 2, 5, and 10 Mega Samples Per Second (MSPS))
Quick Overview	• Supports RF fingerprinting using supervised and zero-shot learning • Facilitates identification of LoRa transmitters
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	IOT | LPWAN (LoRa)
Data Type	Complex Float I/Q Samples
Data Size	26.5 GB
Transmitter | Receiver	22 LoRa Devices | USRP B210
Owner | Access | License	U-Hasselt, COSIC KU Leuven | Public | MIT License

The dataset is accessible via: https://zenodo.org/records/583965

3.1.2 Long-Term Wi-Fi fingerprinting dataset and supporting material

The Long-Term Wi-Fi Fingerprinting Dataset [MRT+20] provides Wi-Fi measurements collected over 25 months at the library of Universitat Jaume I (UJI) in Spain. The dataset contains 103,584 Wi-Fi fingerprints collected across two floors of the building using Android smartphones. The data supports indoor positioning research and includes MATLAB® scripts for data filtering and analysis. Measurements were taken with both a Samsung Galaxy S3 and Galaxy A5 (2017) across multiple campaigns.

Dataset ID | Name	**DAT002 | Physical-Layer Fingerprinting of LoRa devices using Supervised and Zero-Shot Learning**
Key Features	• Wi-Fi RSS measurements collected across two floors over 25 months • 103,584 Wi-Fi fingerprints, organized into training and test datasets • Includes short- and long-term signal variations for indoor positioning research • Covers the impact of network changes on positioning accuracy
Quick Overview	• Enables the analysis of temporal variations in Wi-Fi RSS signals • Facilitates reproducibility and comparability in indoor positioning studies • Supports development and evaluation of positioning models that are resistant to signal changes • Useful for analysing access point dynamics and network configuration changes • Supports indoor positioning research using fingerprinting • Enables temporal signal variation studies
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	None | Short-Range Communication
Data Type	RSS values, timestamps, and device identifiers
Data Size	6.5 MB
Transmitter | Receiver	Wi-Fi Access Points (Total 448) | Samsung Galaxy S3 (used for 15 months), Samsung Galaxy A5 (2017) (used for later measurements, during the last 10 months)
Owner | Access | License	UJI, Tampere University of Technology | Public | CC BY 4.0

The dataset is accessible via: https://zenodo.org/records/3748719

3.1.3 Wi-Sig: RF Fingerprinting Dataset

Wi-Sig [HKC-22] (**Wi**-Fi **Sig**nal) is one of the best publicly available datasets for RFFI. It is made up of Wi-Fi captures taken over several days in the Orbit [RSO+05] testbed using various transmitters and receivers.

Over 10 million packets were received within a month. The packets were transmitted by 174 Wi-Fi transmitters and captured with 41 USRP Receivers. This database can be used for deployment-scale research about RF Fingerprinting.

The data is labelled and divided into prepackaged subsets for easier use. Furthermore, the collected raw data is published as “Raw Wi-Sig” dataset whose size is 1.4 TB.

Dataset ID | Name	**DAT003 | WiSig: Large-Scale Wi-Fi Signal Dataset**
Key Features	• Contains 10 million packets from Wi-Fi transmitters • Data from 174 Wi-Fi transmitters and 41 USRP receivers • Captured over four days, including raw and processed signals • Pre-packaged subsets: “ManySig”, “ManyTx-Rx”, “SingleDay”

Quick Overview	• Enables large-scale research into Wi-Fi RF fingerprinting • Allows analysis of channel and receiver variability on transmitter identification • Provides balanced and compact subsets for easier analysis • Includes pre-processing scripts and example use cases
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	None | Short Range Communication (Wi-Fi)
Data Type	Complex Float I/Q Samples
Data Size	76.9 GB (processed), 1.4 TB (raw)
Transmitter | Receiver	174 Wi-Fi devices (Atheros 5212, 9220, 9280, and 9580 chipsets) | 41 USRP devices (B210, X310, and N210)
Owner | Access | License	University of California, Los Angeles (UCLA) Cores | Public | CC BY 4.0

The dataset is accessible via: https://cores.ee.ucla.edu/downloads/datasets/wisig/

3.1.4 A Database for Radio Frequency Fingerprinting of Bluetooth Devices

The Database for Radio Frequency Fingerprinting of Bluetooth Devices [UDK+20] consists of Bluetooth signals samples, obtained at different sampling rates, collected from 27 different smartphones (6 manufacturers with various models). The authors provide a Matlab script to extract the I/Q samples from the raw sampled signals provided. The author's view of a possible use case of these data is RF fingerprinting.

Dataset ID | Name	**DAT 004 | A Database for Radio Frequency Fingerprinting of Bluetooth Devices**
Key Features	• Bluetooth signals from 27 smartphones, covering six manufacturers with multiple models • Multiple sampling rates. • Provided as voltage signals captured by high-sampling-rate oscilloscopes • Includes MATLAB scripts for time-series conversion and I/Q data generation
Quick Overview	• Supports RF fingerprinting for Bluetooth device identification • Supports analysis of I/Q components and time-series signal generation • Facilitates research on RF signal fingerprinting using real-world Bluetooth signals • Ideal for research on RF-based classification
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	None | Short-Range Communication (Bluetooth)
Data Type	Bluetooth RF Signals
Data Size	1.6 GB
Transmitter | Receiver	27 Different Smartphones| 1 Oscilloscope (TDS7404) 1 RF Frontend
Owner | Access | License	Atılım University, Norwegian University of S&T | Public | CC BY 4.0

The dataset is accessible via: https://zenodo.org/records/3876140

3.1.5 Comprehensive RF Dataset Collection: LoRa datasets

The RF dataset [EH21a] includes both time-domain I/Q samples and corresponding FFT samples, collected using an IoT testbed consisting of 25 identical Pycom IoT devices and a USRP B210 receiver. The receiver

operates at a center frequency of 915 MHz, with a sampling rate of 1 MS/s for recording the received signals. The data is stored in binary files, containing both time-domain I/Q and FFT samples, following the SigMF standard [HWO+18]. For each binary file, a plain-text JSON metafile is created to capture recording details such as sampling rate, date and time of recording, carrier frequency, and other relevant parameters.

More details and use cases of these LoRa datasets can be found in [EH21b].

Dataset ID | Name	**DAT 005 | Comprehensive RF Dataset Collection**
Key Features	• LoRa RF signals captured under varying indoor and outdoor scenarios • Data collected with multiple receivers and different LoRa configurations • I/Q and FFT representations provided for signal analysis • Includes metadata for all collected files
Quick Overview	• Enables exploration of RF fingerprinting accuracy under deployment changes • Supports analysis of hardware-induced signal impairments • Useful for studying sensitivity to environmental changes • Facilitates deep learning-based fingerprinting techniques evaluation
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	IOT | LPWAN (LoRa)
Data Type	IQ time-domain data, FFT frequency-domain data, and metadata
Data Size	1.2 TB
Transmitter | Receiver	25 Semtech SX1276 LoRa Tx | 1 - 2 USRP B210 Rx
Owner | Access | License	School of EECS, Oregon State University | Public | Citation Required

The dataset is accessible via:
https://research.engr.oregonstate.edu/hamdaoui/sites/research.engr.oregonstate.edu.hamdaoui/files/release_note_lora_datasets_final_oct2023_v2.pdf

3.1.6 Real-world Commercial Wi-Fi and Bluetooth Dataset for RF Fingerprinting

The Real-world Commercial Wi-Fi and Bluetooth Dataset for RF Fingerprinting [JKJ22] offers RF emissions collected from commercial off-the-shelf (COTS) Bluetooth and Wi-Fi transceivers under different conditions. The emissions were recorded using a National Instruments Ettus USRP X300 radio with a UBX160 daughterboard and VERT2450 antenna. The dataset focuses on identifying Wi-Fi-Bluetooth combo chipsets, helping validate models' generalization capability across two different setups (Day1 and Day2). The captured data follows the SigMF format and includes complex64 I/Q samples with JSON metadata files that document the collection environment.

Dataset ID | Name	**DAT 006 | Real-world Commercial Wi-Fi and Bluetooth Dataset for RF Fingerprinting**
Key Features	• RF fingerprinting data collected from Bluetooth and Wi-Fi devices • Two datasets per technology (Day1 and Day2) recorded under different conditions
Quick Overview	• Supports deep learning model evaluation and training • Helps study generalization capability in real-world scenarios
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	None | Short-Range Communication (Wi-Fi & Bluetooth)

Data Type	Complex64 I/Q Samples
Data Size	28 GB
Transmitter | Receiver	Bluetooth & Wi-Fi Devices | USRP X300 Receiver
Owner | Access | License	ANDRO Computational Solutions | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/real-world-commercial-wifi-and-bluetooth-dataset-rf-fingerprinting

3.1.7 ORACLE – RF Fingerprinting Dataset

ORACLE Dataset [SBA+19] (Optimized Radio Classification Algorithm for Large-scale Environments) is one of the most comprehensive publicly available datasets for RF fingerprinting. It comprises Wi-Fi transmissions captured over several weeks using a testbed of USRP X310 SDRs as transmitters and a USRP B210 as the receiver.

Over 20 million samples were collected for each radio within a month. The packets were transmitted by 16 bit-similar USRP X310 radios and captured by a single USRP B210 receiver. This dataset can be used for large-scale research on RF fingerprinting techniques.

The collected raw data is published as two separate datasets:

Dataset #1: Raw I/Q samples from over-the-air transmissions
Dataset #2: Demodulated I/Q symbols from over-the-cable transmissions with intentional I/Q imbalance

The data is labelled and divided into pre-packaged subsets for easier use. These subsets represent various experimental conditions, including different distances between transmitters and receivers.

The dataset structure follows industry-standard formats for signal metadata. Each recording includes both the raw data and associated metadata, providing detailed information about the capture conditions and device configurations.

Researchers can utilize this dataset to verify and expand upon the ORACLE methodology. It offers a valuable resource for investigating machine learning applications in wireless communications, with a focus on identifying distinct radio fingerprints among seemingly identical devices only using I/Q samples at the physical layer.

Dataset ID | Name	**DAT 007 | ORACLE RF Fingerprinting Dataset**
Key Features	Comprehensive Data Collection, Multiple Experiment, 20M Samples for each radio
Quick Overview	Ideal for radio classification using CNN
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	None | Short Range Communication (Wi-Fi)
Data Type	64-bit floating point & Metadata files
Data Size	30 GB
Transmitter | Receiver	16 USRP X310 | 1 USRP B210
Owner | Access | License	GENESYS Lab | Public | Citation Required

The dataset is accessible via: https://genesys-lab.org/oracle

3.1.8 LoRa_RFFI_Dataset

The LORA-RFFI Dataset [SZM+22] contains raw signals from 60 commercial off-the-shelf LoRa devices, captured using a USRP N210 software-defined radio (SDR) receiver. The dataset supports scalable and channel-robust radio frequency fingerprint identification (RFFI) using deep learning techniques. The signals are pre-processed into channel-independent spectrograms to mitigate wireless channel effects, ensuring robust classification and rogue device detection across diverse environments. This dataset enables the development and testing of scalable IoT device authentication systems based on hardware impairments without the need for cryptographic mechanisms.

Dataset ID | Name	**DAT 008 | LoRa_RFFI Dataset**
Key Features	• Raw I/Q signals collected from 60 commercial LoRa devices • Supports channel-independent spectrogram construction to mitigate channel effects • Pre-processed to enable deep metric learning-based fingerprint extraction • Facilitates scalable RFFI with enrolment and rogue device detection
Quick Overview	• Enables channel-robust RFFI for IoT device authentication • Useful for developing scalable authentication systems with minimal retraining • Supports testing under various channel conditions and antenna polarizations
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	IOT | LPWAN (LoRa)
Data Type	I/Q signals and channel-independent spectrograms
Data Size	15.55 GB
Transmitter | Receiver	60 LoRa Devices | USRP N210 SDR
Owner | Access | License	University of Liverpool | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/open-access/lorarffidataset

3.1.9 LORA_RFFI_DATASET_DIFFERENT_SPREADING_FACTORS

The LORA - RFFI Diff Spreading Factors dataset [SJM+23] contains radio frequency fingerprint identification (RFFI) data collected to classify devices through their physical-layer signal variations. This dataset was designed to investigate the impact of different LoRa spreading factors (SF) and low signal-to-noise ratio (SNR) conditions on device classification. It features data from 10 off-the-shelf LoRa devices collected using a USRP N210 software-defined radio (SDR) platform at different SF configurations (7, 8, and 9). The dataset provides 90,000 labelled packets, with a split between training and testing data across varying conditions.

Dataset ID | Name	**DAT 009 | LORA - RFFI Diff Spreading Factors Dataset**
Key Features	• RF data from 10 LoRa devices at SF 7, 8, and 9 • 90,000 labeled packets with different SF and SNR conditions
Quick Overview	• Evaluates device classification based on RF fingerprints • Explores performance under varying SF and SNR levels • Supports research on adaptive and robust RFFI models
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G-TE | Tech-Domain	IOT | LPWAN (Lora)
Data Type	RF signal recordings (preamble, device labels)

Data Size	6.10 GB
Transmitter | Receiver	10 LoRa Devices | USRP N210 SDR
Owner | Access | License	University of Liverpool | Public | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/lorarffidatasetdifferentspreadingfactors

3.1.10 Sigfox and LoRaWAN Datasets for Fingerprint Localization in Large Urban and Rural Areas

Sigfox and LoRaWAN Datasets for Fingerprint Localization in Large Urban and Rural Areas dataset [ABV+19] contains real-world data, where numerous devices with GPS receivers periodically obtain new location data, which is sent to a local data server via a Sigfox or LoRaWAN message. These LPWAN datasets provide a benchmark tool to evaluate fingerprint localization algorithms in large outdoor environments with various properties. Different dataset versions are available, and the data is organized in CSV and JSON format. JSON file format is:

- HDOP: Horizontal Dilution of Precision
- dev_addr: LoRaWAN device address
- dev_eui: LoRaWAN device EUI
- sf: Spreading factor
- channel: TX channel (EU region)
- payload: application payload
- adr: Adaptive Data Rate (1 = enabled, 0= disabled)
- counter: device uplink message counter
- latitude: Groundtruth TX location latitude
- longitude: Groundtruth TX location longitude
- airtime: signal airtime (seconds)
- gateways:
  - RSSI: Received Signal Strength
  - ESP: Estimated Signal Power
  - SNR: Signal-to-Noise Ratio
  - ts_type: Timestamp type. If this says "GPS_RADIO", a nanosecond precise timestamp is available
  - time: time of arrival at the gateway
  - id: gateway ID

Dataset ID | Name	**DAT 010 | Sigfox and LoRaWAN Datasets for Fingerprint Localization in Large Urban and Rural Areas**
Key Features	• LoRaWAN and Sigfox datasets collected in rural and urban areas • Contains GPS locations and gateway metadata
Quick Overview	• Supports localization with RSSI and TDoA techniques • Designed for research on outdoor fingerprint localization
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	IOT | LPWAN (Sigfox & LoRaWAN)
Data Type	Various
Data Size	0.178 GB
Transmitter | Receiver	LoRaWAN and Sigfox Devices | Multiple Gateways
Owner | Access | License	University of Antwerp IDLab - NV | Public | CC BY 4.0

The dataset is accessible via: https://zenodo.org/records/3904158#.X4_h7y8RpQI

3.1.11 ITU RayMobTime Datasets

The ITU RayMobTime dataset [KBG+18] provides multimodal data for simulating wireless communication environments with raytracing and mobility scenarios. It integrates simulation tools like Remcom's Wireless Insite, SUMO, and Blensor to generate realistic data over time, frequency, and space. The dataset supports research in wireless networks, enabling the study of mobility, signal propagation, and machine learning-based communication strategies. Multiple datasets are included, ranging across different scenarios, receivers, and channel conditions to facilitate advanced investigations.

Dataset ID | Name	**DAT 011 | ITU RayMobTime Datasets**
Key Features	• Contains ray-tracing data with multimodal inputs: LIDAR, camera images, and wireless signals • Includes multiple scenarios with 3D environments for frequency bands 2.8 GHz, 5 GHz, 28 GHz, and 60 GHz • Supports time evolution with varying intervals between scenes and episodes • Provides simulation data for both mobile and fixed receivers, facilitating vehicular and wireless network studies • Ready-to-use format with detailed technical documentation, including Python/Matlab code
Quick Overview	• Enables analysis of wireless communication channels in dynamic environments • Supports research on machine learning techniques for beam selection and communication strategies • Facilitates studies on vehicular networks with ray-tracing data integrated with SUMO traffic simulations • Allows evaluation of physical-layer performance over time and space
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G-TE | Tech-Domain	ISAC, mmWaves and sub-THz, IOT | Short-Range Communication
Data Type	Multimodal simulation data: wireless signals, LIDAR, camera images
Data Size	56.6GB
Transmitter | Receiver	Separate transmitters for 2.8 GHz, 5 GHz, 28 GHz, and 60 GHz bands | Fixed and mobile receivers (e.g., vehicles, RSUs)
Owner | Access | License	Klautau et al. | Public | Citation Required

The dataset is accessible via: https://www.lasse.ufpa.br/raymobtime

3.1.12 Deepsense-6G

The DeepSense 6G dataset [ACO+23] contains over one million real-world multi-modal sensing and communication data samples designed to advance deep learning research in wireless communication, sensing, and localization. This dataset offers coexisting data from several communication and sensing setups, including mmWave communication, GPS, LiDAR, Radar, and camera-based sensing, collected across more than 40 scenarios in diverse indoor and outdoor locations. The dataset captures environmental diversity across different times of the day, weather conditions, and deployment settings (such as vehicle-to-infrastructure (V2I), vehicle-to-vehicle (V2V), drone communication, and pedestrian movements)

Dataset ID | Name	**DAT 012 | Deepsense-6G**
Key Features	• Contains over 1 million multi-modal sensing and communication data samples • Data collected across more than 40 indoor and outdoor scenarios

	• Includes data from mmWave communication, GPS, LiDAR, Radar, and cameras • Scenarios cover different environments: parking lots, downtown streets, indoor spaces • Data collected at different times of the day and weather conditions
Quick Overview	• Supports applications in beam prediction, blockage prediction, user scheduling, resource management, and security enhancement • Facilitates benchmarking and comparison of machine learning models • Enables real-world development of multi-modal sensing-aided communication solutions • Includes task-specific datasets and benchmarking scripts
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	ISAC, mmWave/sub-THz, D-MIMO | Vehicular Technologies
Data Type	Multi-modal sensing and communication data
Data Size	Over 1 million samples (exact size not specified)
Transmitter | Receiver	mmWave Transmitters (SIVERS EVK06002) | mmWave Receivers (SIVERS EVK06002), Radar (TI AWR2243BOOST), GPS (SparkFun RTK2), Cameras (ZED2), LiDAR (SLAMTEC RPLiDAR)
Owner | Access | License	Arizona State University | Registration needed | CC BY 4.0

Dataset can be accessible via: https://www.deepsense6g.net

3.1.13 UCLA - RF Fingerprinting Dataset

The UCLA - RF Fingerprinting Dataset [HKC20] includes raw I/Q samples from WiFi modules to enable research on RF fingerprinting and transmitter identification. It has been captured using Atheros (5212, 9220, 9280) modules as transmitters and a USRP N210 as the receiver. The dataset contains more than 300,000 packets collected at a rate of 25 MSPS over a 1-second duration. Data was collected over multiple days from 163 nodes in the Orbit testbed, allowing analysis of the temporal variations in RF fingerprints. Each packet includes the first 256 I/Q samples from the preamble, ensuring data consistency across experiments.

Dataset ID | Name	**DAT 013 | UCLA - RF Fingerprinting Dataset**
Key Features	• Captures raw I/Q samples from Wi-Fi modules • Uses three types of Atheros modules (5212, 9220, and 9280) • Over 300,000 packets collected at a sampling rate of 25 Msps • Data captured over multiple days, ensuring temporal diversity • Packets extracted using energy detection, each containing 256 I/Q preamble samples
Quick Overview	• Enables evaluation of RF fingerprinting and transmitter authorization • Facilitates temporal analysis of fingerprints using multi-day data • Supports experimentation with supervised and open-set recognition techniques • Useful for investigating RF security challenges and anomalies
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	None | Short-Range Communication (Wi-Fi)
Data Type	Raw I/Q samples
Data Size	1.6 GB

Transmitter | Receiver	Atheros Wi-F modules (5212, 9220, 9280) | USRP N210
Owner | Access | License	UCLA Cores | Public | CC BY 4.0

The dataset is accessible via: https://cores.ee.ucla.edu/downloads/datasets/rf-fingerprinting-dataset/

3.1.14 The AWID2 Dataset

The AWID2 Dataset [KKS+16] consists of real traces of both normal and attack traffic in IEEE 802.11 wireless networks, focusing on intrusion detection. Captured within a simulated SOHO infrastructure, the dataset includes both full packet collections (F) and reduced collections (R). Each collection has separate training and testing sets. Scenarios cover 15 attack types along with regular activities such as Web browsing, VoIP, and file downloads. The goal is to enable machine learning-based development and evaluation of intrusion detection systems tailored to wireless networks.

Dataset ID | Name	**DAT 014 | The AWID2 Dataset**
Key Features	• Real traffic collected from SOHO infrastructure with WEP encryption • Includes normal and attack traffic • Separate training and testing set for two versions: full (F) and reduced (R) • 15 distinct attack scenarios (e.g., Man-in-the-Middle, Deauthentication, Honeypot) • Provided in CSV and pcap formats for machine learning and network analysis
Quick Overview	• Enables intrusion detection system development for 802.11 networks • Suitable for evaluating algorithms on both attack and benign wireless traffic • Facilitates research in detecting specific attack patterns and anomalies
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation, Denial of Service
6G-TE | Tech-Domain	None | Short-Range Communication (Wi-Fi)
Data Type	Wireless traffic in CSV and pcap formats
Data Size	Not specified
Transmitter | Receiver	Multiple SOHO clients (desktops, laptops, smartphones, tablets) | Netgear N150 WNR1000 v3 AP, Alpha AWUS036H monitoring card
Owner | Access | License	University of the Aegean | Available Upon Request | Citation Required

The dataset is accessible via: https://icsdweb.aegean.gr/awid/awid2

3.1.15 RF Jamming Dataset for Vehicular Wireless Networks

This dataset [KKA+21] comprises diverse scenarios of RF jamming attacks and interference in Vehicular Ad-hoc Networks (VANETs), along with corresponding ground truth labels. The dataset is designed to support the evaluation and development of detection algorithms for RF jamming attacks in VANETs.

In this dataset, there are data corresponding to three scenarios:

- **No attack:** there is no jammer, but there exists a source of unintentional interference. The goal is to assess the ability to distinguish an intentional jammer from unintentional interference.
- **Reactive jammer attack:** the jammer keeps transmitting jamming signals while moving toward the target, then backs off to a safe distance.

- **Constant jammer attack:** the jammer chases the target while transmitting jamming signal and, once it reaches the target, it transmits at full power.

This dataset includes the following features:

- Time: Timestamp of the measurement.
- SNR: Signal quality metric.
- Speed: Speed of the vehicles involved in the communication.
- RSSI: Signal strength measurement.
- PDR (Packet Delivery Ratio): Ratio of successfully delivered packets.
- Relative_Speed: Variations in estimated relative speed between the jammer and the receiver.
- Scenario: 1 (No Attack), 2 (Reactive Attack), 3 (Constant Attack)

This dataset can be used for the development of RF jamming detection techniques to enhance the robustness of VANETs against malicious interference.

Dataset ID | Name	**DAT 015 | RF Jamming Dataset Vehicular Wireless Networks**
Key Features	• CSV files containing measurements such as RSSI, SINR, PDR, speed, and timestamps • Includes two datasets with different maximum relative speed values: 25 m/s and 15 m/s • Ground truth labels for three scenarios: No Attack, Reactive Jammer, Constant Jammer • Data organized to support algorithm evaluation and feature analysis
Quick Overview	• Enables analysis of RF jamming effects in vehicular networks • Facilitates performance testing of detection algorithms • Useful for evaluating supervised models across multiple scenarios • Captures time-series data of vehicle movement and communication metrics
Threat Coverage	Denial of service
6G-TE | Tech-Domain	IOT | Vehicular Technologies
Data Type	CSV files containing signal and vehicle movement measurements
Data Size	0.00037 GB
Transmitter | Receiver	Vehicular transmitters using IEEE 802.11p | Vehicular receivers operating under urban mobility conditions
Owner | Access | License	Kosmanos et al. | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/rf-jamming-dataset-vehicular-wireless-networks

3.1.16 Indoor Skg Under an On-the-shoulder Eavesdropping Attack

The Indoor Skg Under an On-the-shoulder Eavesdropping Attack [MMC+23] dataset provides the received signal measurements of an eavesdropper called “Eve”. These measurements are done with 4 scenarios and 5 ios and 5 different locations of Eve. The 105 received signal measurements of Eve for each of the above combinations of scenarios and locations are further divided into batches of 20 .pkl files. Each file contains 5000 received signal measurements.

The dataset is obtained by configuring three universal software radio peripherals (USRPs), each with a single antenna, as two legitimate users, Alice and Bob, and an eavesdropper, Eve. We obtain the channel measurements for Alice, Bob, and Eve using USRP-2974 from National Instruments. Experiments were performed in 4 scenarios described below:

- **NLOS Dynamic:** Non-line-of-sight condition with dynamic channel measurements
- **LOS Dynamic:** Line-of-sight condition with dynamic channel measurements
- **NLOS Static:** Non-line-of-sight condition with static channel measurements

- **LOS Static:** Line-of-sight condition with static channel measurements

Static channel measurements are realized at nighttime and the channel remains static since there is no movement in the room. The LoS and NLoS scenarios were created in the absence or in the presence of absorbers between the antennas of Alice and Bob, respectively.

Supports research on physical layer security (PLS) in 6G, focusing on secret key generation (SKG) from wireless fading coefficients. Includes data from four indoor experiments.

Dataset ID | Name	**DAT 016 | Indoor Skg Under an On-the-shoulder Eavesdropping Attack**
Key Features	PLS | SKG | 6G
Quick Overview	Directly related to PLS and ideal to be used for the security challenge
Threat Coverage	Information Disclosure
6G -TE | Tech-Domain	None | SDR
Data Type	Signal Recordings & Syndromes during SKG
Data Size	18.81 GB
Transmitter | Receiver	USRP-2974 | USRP-2974
Owner | Access | License	Amitha Mayya & Miroslav Mitev & Arsenia Chorti & Gerhard Fettweis | Registration needed | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/dataset-paper-skg-security-challenge-indoor-skg-under-shoulder-eavesdropping-attack#files

3.1.17 Ultra-Wideband Channel State Information and Localization for Physical Layer Security

UWB Channel State Information Dataset [WKS21] is an essential resource for Physical Layer Security (PLS) research, offering real-world measurements of Ultra-Wideband (UWB) Channel State Information (CSI). This dataset was designed to support research on confidentiality, authentication, and key derivation, providing a comprehensive foundation for evaluating theoretical PhySec models in practical scenarios.

The dataset was collected using a highly detailed testbed consisting of consumer-grade UWB transceivers (DecaWave EVB1000) and a LEGO-based robot platform that autonomously moved within an indoor environment. Over the course of 112 hours, the setup recorded ≈1.7 million CSI samples under various scenarios, including symmetric, asymmetric, and varying-speed movements. Two eavesdroppers were integrated into the setup to enable the analysis of adversarial observations, making the dataset highly valuable for security applications.

With approximately 1.7 million samples, this dataset is particularly well-suited for machine learning and wireless security research. Each sample includes detailed CSI data, including both real and imaginary components, as well as corresponding RSSI values and location information, providing rich data for studying localization, interference, and security issues.

Dataset ID | Name	**DAT 017 | Ultra-Wideband Channel State Information Dataset**
Key Features	• Real-world UWB Channel State Information (CSI) measurements • Includes location data of terminals and eavesdroppers in each scenario • Over 112 hours of measurements, with more than 1.2 million samples • Incorporates two eavesdroppers for security analysis • Data provided as zipped NumPy arrays with custom headers

Quick Overview	• Supports analysis of channel characteristics for physical-layer security • Enables research on physical-layer authentication, key derivation, and wiretap codes • Simulates security-focused use cases involving eavesdroppers • Suitable for machine learning due to large sample size and varied scenarios
Threat Coverage	Tampering, Information Disclosure
6G -TE | Tech-Domain	None | SDR
Data Type	Channel State Information (CSI) data, Location data
Data Size	15.45 GB
Transmitter | Receiver	UWB transceivers with two eavesdroppers
Owner | Access | License	TU Dresden | Public | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/dataset-paper-skg-security-challenge-indoor-skg-under-shoulder-eavesdropping-attack

3.1.18 IRShield: A Countermeasure Against Adversarial Physical-Layer Wireless Sensing

The IRShield dataset [SMR+22] provides Wi-Fi Channel State Information (CSI) measurements along with experimental configurations used to evaluate IRShield, a countermeasure against adversarial physical-layer sensing attacks. The work focuses on addressing privacy violations that occur when attackers use Wi-Fi signals to infer human activities and movements through passive eavesdropping. IRShield leverages Intelligent Reflecting Surfaces (IRS) to introduce randomness into the wireless channel, making it difficult for adversaries to detect meaningful patterns related to human motion. IRS, which is closely related to Reconfigurable Intelligent Surfaces (RIS), plays a critical role in enhancing wireless privacy by dynamically altering the propagation environment [PZZ+22], making it a robust tool against adversarial sensing attacks.

The dataset captures measurements from multi-antenna IEEE 802.11n Wi-Fi routers acting as both transmitters and receivers, as well as channel variations induced by a 256-element IRS prototype. The IRS is deployed near the transmitters (anchors) to generate random signal reflections, obfuscating the wireless channel and reducing the adversary's capability to infer physical activities. The experiments were conducted in real-world environments, such as office spaces, to simulate motion detection scenarios by eavesdroppers placed in public areas.

IRShield provides an effective countermeasure by introducing random channel variations using reconfigurable intelligent surfaces (RIS), making it difficult for attackers to detect motion reliably.
The dataset is designed to facilitate further research into:

- Privacy-preserving wireless communication.
- Adversarial attacks on physical-layer sensing (e.g., motion detection, gesture recognition, and eavesdropping attacks).
- The effectiveness of IRS-based countermeasures.

Researchers can use this dataset to develop and test algorithms related to adversarial sensing, channel obfuscation, and wireless privacy enhancement techniques.

Dataset ID | Name	**DAT 018 | IRShield: A Countermeasure Against Adversarial Physical-Layer Wireless Sensing**
Key Features	• Contains Wi-Fi CSI data for physical-layer analysis. • Includes experiments on human motion detection with and without IRShield. • Python scripts provided for plotting and evaluating results. • Useful for analysing IRS-based wireless channel obfuscation.

Quick Overview	• Supports analysis of adversarial sensing and countermeasures on Wi-Fi signals. • Enables exploration of IRS-based privacy-preserving wireless communication. • Facilitates evaluation of motion detection attacks using CSI datasets.
Threat Coverage	Tampering, Information Disclosure
6G -TE | Tech-Domain	RIS | Short-Range Communication (Wi-Fi)
Data Type	Wi-Fi CSI data
Data Size	15.2 GB
Transmitter | Receiver	**IEEE 802.11n Wi-Fi devices** and **TP-Link N750 routers** acting as anchors. The receivers are **TP-Link N750** routers, functioning both as eavesdroppers and packet receivers.
Owner | Access | License	Max Planck Institute for Security and Privacy, Ruhr University Bochum, TH Köln | Public | CC BY 4.0

The dataset is accessible via: https://zenodo.org/records/6367411

3.1.19 BLE-WBAN: RF real-world dataset of BLE devices in human-centric healthcare environments

This dataset [KSK+23] contains raw RF data of Bluetooth Low Energy (BLE) signals focused on Wireless Body Area Network (WBAN). It consists of on-body and off-body recordings in an anechoic chamber.

Dataset features:

- It covers the entire bandwidth of the BLE technology. (recorded at 2.44GHz at 100MSps)
- The recording in an anechoic chamber reduces unwanted signals or interference.
- The on-body recordings were performed at 12 different locations including both left and right head, arm, wrist, chest, front and back waist.
- Off-body recording with the same devices at 7-different orientations

The author aims to study the physical layer characteristics of both on-body and off-body signals.

Dataset ID | Name	**DAT 019 | BLE-WBAN: RF real-world dataset of BLE devices in human-centric healthcare environments**
Key Features	• Contains RF signals recorded from 13 BLE devices (ESP32s) • Covers entire BLE spectrum with 100 Msps sampling rate at 2.44 GHz • Includes both on-body (12 anatomical locations) and off-body (7 orientations) recordings • Data recorded in an anechoic chamber to minimize interference
Quick Overview	• Enables analysis of BLE physical-layer signals for healthcare IoT environments • Facilitates the study of on-body and off-body signal variations • Suitable for machine learning tasks on RF signal data
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	IOT | Short Range Communications (Bluetooth Low Energy)
Data Type	Complex float I/Q samples
Data Size	28.5

Transmitter | Receiver	13 BLE devices (ESP32s) | USRP x310
Owner | Access | License	SeyedMohammad Kashani | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/ble-wban-rf-real-world-dataset-ble-devices-human-centric-healthcare-environments

3.1.20 IEEE 802.15.4 Backscatter Radio Frequency Fingerprinting

IEEE 802.15.4 Backscatter Radio Frequency Fingerprinting dataset [Yan24] focuses on IEEE 802.15.4 backscatter communication for Radio Frequency (RF) fingerprinting applications. It contains I/Q samples from transmitted frames generated by six carrier emitters, comprising two USRP B210 devices (denoted as c#) and four CC2538 chips (denoted as cc#), along with ten backscatter tags (labeled as tag#). The carrier emitters transmit an unmodulated signal, while the backscatter tags utilize Quadrature Phase Shift Keying (QPSK) modulation in the 2.4 GHz frequency band, adhering to IEEE 802.15.4 standards.

The dataset's importance lies in its ability to enhance security and improve device authentication in IoT networks by unique RF fingerprints emitted by devices. With a variety of emitter and tag combinations, this dataset provides a valuable resource for developing and testing RF fingerprinting techniques, essential for maintaining the integrity and authenticity of wireless communications in the expanding IoT landscape.

Dataset ID | Name	**DAT 020 | IEEE 802.15.4 Backscatter Radio Frequency Fingerprinting**
Key Features	• I/Q samples collected from six carrier emitters and ten tags • QPSK modulation, IEEE 802.15.4 protocol
Quick Overview	• Enables RF fingerprinting for IoT security 2.4 GHz backscatter signals for device authentication IEEE 802.15.4 backscatter communication dataset includes I/Q samples from six carrier emitters and ten backscatter tags. Supports RF fingerprinting for device authentication in IoT networks.
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	None | Short Range Communications (ZigBee - IEEE 802.15.4)
Data Type	I/Q samples
Data Size	70
Transmitter | Receiver	Six USRP and CC2538 Emitters | USRP B210 Receiver
Owner | Access | License	Uppsala University | Registration needed | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/ieee-802154-backscatter-radio-frequency-fingerprinting

3.1.21 LTE_RFF_IDENTIFICATION_DATASET

The LTE_RFF_IDENTIFICATION_DATASET [YL23] contains LTE uplink signals from ten different LTE devices collected using a USRP N210 across multiple locations. The original sampling rate was 25 MHz, with resampling to 30.72 MHz in MATLAB. The data is stored in MAT file format and includes signals processed using the WL method mentioned in the associated research paper. The dataset offers two collections: one in a line-of-sight scenario (DATASET1) and one in a non-line-of-sight scenario (DATASET2), enabling comprehensive research into LTE radio frequency fingerprint identification through deep learning techniques.

This dataset is particularly useful for researchers working on radio frequency fingerprinting using deep learning, providing high-quality LTE signal data for training and evaluation in both LOS and NLOS environments.

Dataset ID | Name	**DAT 021 | LTE RFF Identification Dataset**

Key Features	• Sampling rate: 25 MHz, resampled to 30.72 MHz • Collected from ten LTE devices using USRP N210 • Collected in both line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios
Quick Overview	• Contains LTE uplink signals for radio frequency fingerprinting • Enables analysis of signal variance across multiple locations and conditions • Suitable for developing deep learning models for RF fingerprinting • Useful for studying LTE device identification through uplink signal patterns
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	None | Cellular Technologies (LTE)
Data Type	Processed LTE signals (MAT files)
Data Size	11.5
Transmitter | Receiver	10 LTE devices (e.g., Mblu Note5, Huawei 7 Plus, Redmi 6a) | USRP N210
Owner | Access | License	Southeast University | Public | CC BY-NC-SA 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/lterffidentificationdataset

3.1.22 Securing the Physical Layer of IEEE 802.15.4 Through Cooperative Jamming

The Securing the Physical Layer of IEEE 802.15.4 Through Cooperative Jamming [PRT+23] dataset contains baseband signals recorded using software-defined radios (SDRs) to explore physical-layer security through cooperative jamming. These measurements simulate an IoT communication scenario using IEEE 802.15.4 transmissions with simultaneous known interference. The dataset captures various gain configurations and signal conditions, allowing researchers to study the effectiveness of known-interference cancellation algorithms. This dataset is structured to facilitate the analysis of signal processing and interference mitigation techniques in secure wireless communication environments.

Dataset ID | Name	**DAT 022 | Securing the Physical Layer of IEEE 802.15.4 Through Cooperative Jamming**
Key Features	• Contains baseband IQ samples of IEEE 802.15.4 signals and known interference • Multiple gain configurations from 0 to 89.5 dB recorded at 5 dB steps • 8 MHz bandwidth after digital down-conversion, aligned with 2.45 GHz center frequency • Includes signal-of-interest, interference, and noise floor measurements
Quick Overview	• Facilitates research on known-interference cancellation techniques • Enables evaluation of cooperative jamming for physical-layer security • Suitable for benchmarking algorithms in secure IoT communication • Includes measurement results for IEEE 802.15.4 O-QPSK modulation
Threat Coverage	Denial of service
6G -TE | Tech-Domain	None | SDR (IEEE 802.15.4)
Data Type	Baseband I/Q samples

Data Size	355
Transmitter | Receiver	USRP-2900 SDR | USRP-2900 SDR
Owner | Access | License	Tampere University | Public | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/securing-physical-layer-ieee-802154-through-cooperative-jamming

3.1.23 RF-Fingerprint-BT-IoT: Real-world Frequency Hopping Bluetooth dataset from IoT devices for RF fingerprinting

The RF-Fingerprint-BT-IoT: Real-world Frequency Hopping Bluetooth Dataset from IoT Devices for RF Fingerprinting [JJ22] contains RF signals collected from 10 commercial off-the-shelf (COTS) Bluetooth emitters, including 2 laptops and 8 commercial chips. These signals were captured using a National Instruments Ettus USRP X300 radio with a UBX160 daughterboard and VERT2450 antenna. The receiver was tuned to a 2 MHz bandwidth centered at 2.414 GHz. The dataset is split into two parts, Day1 and Day2, to allow for the evaluation of deep learning models under different conditions. The first part was collected under line-of-sight (LoS) conditions, and the second under non-line-of-sight (NLoS) with rich multipath propagation, providing a challenging testbed for RF fingerprinting.

Dataset ID | Name	**DAT 023 | RF-Fingerprint-BT-IoT: Real-world Frequency Hopping Bluetooth dataset from IoT devices for RF fingerprinting**
Key Features	• Contains Bluetooth RF signals from 10 COTS emitters, including laptops and commercial chips • Two datasets: Day1 (LoS, varying distances) and Day2 (NLoS, multipath propagation) • Each signal capture contains 40 million samples and is accompanied by metadata in JSON format • Follows SigMF specifications with extensions for additional metadata
Quick Overview	• Enables RF fingerprinting of Bluetooth emitters under real-world conditions • Provides a challenging testbed for evaluating generalization of machine learning models • Supports research on the impact of LoS and NLoS propagation on RF signals
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	None | Short-Range Communication (Bluetooth)
Data Type	Complex I/Q samples and metadata
Data Size	10.47
Transmitter | Receiver	10 COTS Bluetooth devices (2 laptops, 8 chips) | USRP X300 radio with UBX160 daughterboard and VERT2450 antenna
Owner | Access | License	Marconi-Rosenblatt AI/ML Innovation Lab | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/rf-fingerprint-bt-iot-real-world-frequency-hopping-bluetooth-dataset-iot-devices-rf

3.1.24 CSI Dataset towards 5G NR High-Precision Positioning

The CSI Dataset Towards 5G NR High-Precision Positioning [GWL+22] is designed to support research in integrated sensing and communication (ISAC) and high-precision 5G NR positioning. This fine-grained dataset strictly follows the 3GPP Release 18 standards, targeting indoor and outdoor urban environments. The

dataset mitigates the absence of commercial 5G ISAC base stations by providing synthetic data useful for machine learning research on positioning, channel estimation, and noise robustness.

Dataset ID | Name	DAT 024 | CSI Dataset Towards 5G NR High-Precision Positioning
Key Features	• Contains CSI data in 4D matrices representing channel features • Metadata includes 3D coordinates of User Equipment (UE) positions • Covers indoor office and outdoor urban canyon scenarios • Multiple SNR levels for robustness to noise
Quick Overview	• Supports high-precision positioning using 5G ISAC technology • Enables ML-based location estimation under varied channel conditions • Facilitates CSI-based positioning in multipath-rich environments • Includes Python data import examples for .mat files
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	ISAC | Cellular Technologies (5G NR)
Data Type	CSI data
Data Size	3.2
Transmitter | Receiver	gNBs operating at 3.5 GHz (FR1) and 40 GHz (FR2) | UEs equipped with 4x4 MIMO arrays
Owner | Access | License	Harbin Engineering University | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/open-access/csi-dataset-towards-5g-nr-high-precision-positioning

3.1.25 Toward receiver, modulation, carrier and symbol rate agnostic SEI

SEI Dataset [ZHA-23] is a groundbreaking dataset aimed at addressing the limitations of current Specific Emitter Identification (SEI) systems by incorporating variable signal parameters. SEI is an emerging technique in physical layer security (PLS) for identifying unique radio transmitters. This dataset uniquely accounts for variations in receiver, modulation, carrier frequency, and symbol rate, making it highly versatile for developing SEI systems that can generalize across different signal conditions.

The dataset is designed to overcome the common assumption that training and testing datasets must have consistent distributions. By introducing signal parameter variations, this dataset simulates real-world scenarios where signal conditions change, providing a challenging but realistic testbed for SEI research. This dataset is particularly useful for advancing SEI techniques under variable conditions, which is crucial for improving the robustness and generalization of SEI systems.

Dataset ID | Name	DAT 025 | Toward receiver, modulation, carrier and symbol rate agnostic SEI Dataset
Key Features	• Contains RF signals collected with varying modulation modes, carriers, and symbol rates • 5 emitters (USRP B210), 3 receivers (Tektronix RSA306B) • 4 modulation modes: BPSK, QPSK, 8PSK, and 16QAM • 3 carrier frequencies: 2.4 GHz, 2.2 GHz, 2.0 GHz • 2 symbol rates: 120 kHz, 100 kHz • Labelled dataset with emitter, receiver, carrier frequency, and symbol rate metadata
Quick Overview	• Supports research on SEI in varying conditions • Facilitates the study of signal distribution shifts caused by parameter changes

	• Useful for evaluating the robustness of SEI models to real-world variations • Enables development of receiver and modulation agnostic SEI solutions
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	None | SDR
Data Type	I/Q samples
Data Size	57.92 GB
Transmitter | Receiver	5 USRP B210 devices | 3 Tektronix RSA306B devices
Owner | Access | License	X. Zha | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/toward-receiver-modulation-carrier-and-symbol-rate-agnostic-sei-dataset

3.1.26 5G CFR/CSI dataset for wireless channel parameter estimation, array calibration, and indoor positioning

The 5G CFR/CSI dataset [PLQ+23] contains uplink channel frequency response (CFR) samples measured in an underground parking lot, supporting research in wireless channel parameter estimation, array calibration, and indoor positioning. The dataset captures CFR from 476 different locations, using a User Terminal (UT) equipped with a cylindrical antenna communicating with a 5G Remote Radio Unit (RRU). The data includes ground-truth locations of the UT, RRU Six Degrees of Freedom (6-DoF) pose, and obstacle locations, facilitating studies in Direction Of Arrival/Time Of Arrival (DOA)/(TOA) estimation, Line-of-Sight/Non-Line of Sight (LOS)/(NLOS) identification, and localization in multipath conditions.

Dataset ID | Name	**DAT 026 | 5G CFR/CSI dataset for wireless channel parameter estimation, array calibration, and indoor positioning**
Key Features	• Raw CFR data measured with UL-SRS symbols at multiple locations • Ground-truth positions of UT with centimeter-level accuracy • 6-DoF pose of the RRU, including pitch, roll, and yaw • Locations of obstacles (e.g., pillars) in the environment • System parameters such as carrier frequency and bandwidth
Quick Overview	• Enables research in wireless localization and positioning • Supports DOA and TOA parameter estimation • Provides LOS/NLOS identification data • Captures multipath effects with detailed CFR measurements
Threat Coverage	Tampering, Information Disclosure
6G -TE | Tech-Domain	None | Cellular Technologies (5G)
Data Type	Channel Frequency Response (CFR), assisted measurement data, including ground-truth locations, 6-DoF pose data, and obstacle locations
Data Size	1.13
Transmitter | Receiver	5G UT with omnidirectional antenna | 5G RRU
Owner | Access | License	Pervasive Communication Research Center | Public | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/5g-cfrcsi-dataset-wireless-channel-parameter-estimation-array-calibration-and-indoor

3.1.27 UAV Attack Dataset

The UAV Attack Dataset [WSM+20] consists of logs from both simulated and live UAV flights, capturing both benign and attack scenarios. The dataset aims to facilitate research on GPS spoofing and jamming attacks, as these are common threats against UAV systems, but are often challenging to conduct experimentally in real-world conditions. Logs are provided in both .ulog and .csv formats, covering multiple airframe types and both software-in-the-loop (SITL) and hardware-in-the-loop (HITL) setups. Live GPS attacks were conducted using the Keysight EXG N5172B signal generator and HackRF for GPS spoofing and jamming experiments, providing a comprehensive resource for studying UAV security threats.

Dataset ID | Name	**DAT 027 | UAV Attack Dataset**
Key Features	• Logs from both simulated and live UAV flights, including benign and attack scenarios • Simulated attacks include GPS spoofing using a Gazebo simulator and Ping DoS attacks using MAVLink pings • Live attacks include GPS spoofing with HackRF and GPS-SDR-SIM, and GPS jamming using white Gaussian noise broadcast • Data formats include .ulog files for raw logs and .csv files for easier analysis
Quick Overview	• Provides data for analysing the effects of GPS spoofing, GPS jamming, and DoS attacks on UAVs • Suitable for training and evaluating machine learning models for intrusion detection in UAV systems • Captures data for different UAV types including quadcopters and planes
Threat Coverage	Spoofing | Denial of Service
6G -TE | Tech-Domain	None | Vehicular Technologies
Data Type	Raw sensor data from UAVs in .ulog format, converted .csv files for structured analysis, including telemetry data such as GPS coordinates, sensor readings, flight control inputs, and system status information
Data Size	0.683 GB
Transmitter | Receiver	Great Scott Gadgets HackRF | Pixhawk GPS receiver, Pixhawk 4 flight controller
Owner | Access | License	Ontario Tech University, University of Tabuk | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/open-access/uav-attack-dataset

3.1.28 Dataset for Vehicle Indoor Positioning in Industrial Environments with Wi-Fi, inertial, and odometry data

The Dataset for Vehicle Indoor Positioning in Industrial Environments with Wi-Fi, Inertial, and Odometry Data [SPT+23] contains synchronized sensor data collected from a mobile unit resembling an industrial vehicle in an industrial environment. The dataset includes Wi-Fi signals from four interfaces, inertial measurements from two low-cost IMUs, and odometry data from an absolute encoder attached to a wheel. Ground truth positions were obtained using computer vision techniques with ArUco markers. This dataset facilitates the development and evaluation of indoor positioning and tracking systems in industrial settings.

Dataset ID | Name	**DAT 028 | Dataset for Vehicle Indoor Positioning in Industrial Environments with Wi-Fi, inertial, and odometry data**
Key Features	• Contains synchronized data from multiple sensors • Wi-Fi data collected from four interfaces

	• Inertial data from two low-cost IMUs (accelerometer, gyroscope, magnetometer) • Odometry data from an absolute encoder attached to a wheel • Ground truth positions obtained via computer vision using ArUco markers
Quick Overview	• Enables development of indoor positioning systems combining Wi-Fi and motion sensors • Facilitates research on sensor fusion techniques in industrial environments • Suitable for testing dead reckoning, Wi-Fi fingerprinting, and hybrid positioning methods • Provides realistic industrial environment data with heavy machinery and metallic structures
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	None | Short Range Communication (Wi-Fi)
Data Type	Wi-Fi RSSI values, inertial measurements, odometry data, ground truth positions
Data Size	0,008 GB
Transmitter | Receiver	Wi-Fi Access Points in industrial building | Mobile unit with Raspberry Pi, Wi-Fi interfaces, IMUs, and encoder
Owner | Access | License	ALGORITMI Research Center | Public | CC BY 4.0

The dataset is accessible via: https://zenodo.org/records/7826540

3.1.29 A Dataset of I/Q samples in Indoor Jamming Scenarios

I/Q Samples in Indoor Jamming Scenarios [ASO22] includes physical-layer radio information (I/Q samples) acquired from indoor communications affected by different types of jamming techniques. Specifically, it includes data acquired from 7 different Software Defined Radios (SDRs), i.e., the USRP Ettus Research X310, operating in an office environment while the transmitter and receiver communicate without Line of Sight, nLoS (Non-Line-of-Sight).

Each experiment is characterized by a transmitter, a receiver, and a jammer. While the hardware of the transmitter and the receiver are kept the same for all the experiments, the hardware of the jammer is changed adopting 5 different radios of the same model and brand.

The dataset includes different jamming types, e.g., no jamming (silent), tone (sinusoidal), and Gaussian noise. Moreover, the dataset includes different transmission distances and jamming power levels. In each experiment, a pre-determined sequence of bits ([0, 255]) has been modulated using the BPSK scheme and then stored, at the receiver, as a 2-column matrix of raw I/Q samples.

This dataset is highly valuable for physical layer security research as it provides extensive data on how various jamming techniques impact signal integrity and security. Researchers can use this dataset to develop and test robust physical layer security techniques such as anti-jamming algorithms and secure communication protocols. The diversity in jamming types and power levels allows for a thorough evaluation of security measures against different interference scenarios, helping to enhance resilience and reliability in real-world communication systems.

Dataset ID | Name	**DAT 029 | A Dataset of I/Q samples in Indoor Jamming Scenarios**
Key Features	• Contains raw I/Q samples captured from indoor wireless communications under jamming conditions • Includes jamming types: silent, tone (sinusoidal), and Gaussian noise • Data collected with 7 USRP X310 SDRs and VERT2450 omni-directional antennas

	• Captures experiments with varying jamming power, devices, and transmission distances • Experiments conducted in a dynamic office environment with non-line-of-sight (NLoS) conditions
Quick Overview	• Enables study of jamming effects on physical-layer communications • Supports the development of jamming detection and mitigation techniques • Useful for evaluating communication resilience under different jamming scenarios • Facilitates research on anti-jamming techniques and signal characterization
Threat Coverage	Denial of Service (DoS)
6G -TE | Tech-Domain	None | SDR (USRP X310)
Data Type	Complex I/Q samples captured under jamming conditions
Data Size	118.5 GB
Transmitter | Receiver	USRP Ettus Research X310 SDR | USRP Ettus Research X310 SDR
Owner | Access | License	Hamad Bin Khalifa University & Eindhoven University of Technology | Public | CC BY 4.0

The dataset is accessible via: https://zenodo.org/records/7119040

3.1.30 Mitigating RF Jamming Attacks at the Physical Layer with Machine Learning Dataset

The Mitigating RF Jamming Attacks at the Physical Layer with Machine Learning Dataset [JR22] contains data collected using software-defined radios (SDRs) to support research on RF jamming mitigation techniques. The dataset includes files for training and testing classifiers to detect various types of RF jamming attacks (constant, periodic, reactive) and corresponding mitigation strategies using machine learning. Organized into multiple directories, the dataset contains raw logs, classifier outputs, hyperparameter tuning studies, and performance results. It supports over-the-air (OTA) and ray-tracing emulated (RTE) testing scenarios for real-world evaluations with SDRs.

Dataset ID | Name	**DAT 030 | Mitigating RF Jamming Attacks at the Physical Layer with Machine Learning Dataset**
Key Features	• Contains SDR-collected logs for normal operation and under jamming conditions • Supports classification of constant, periodic, and reactive jammers • Includes hyperparameter tuning files for antenna state selection • Organized for OTA and RTE experiments with SDR hardware
Quick Overview	• Enables testing of RF jamming detection and mitigation strategies in realistic conditions • Facilitates machine learning-based classification of jammer types • Provides detailed logs and results for classifier performance analysis • Supports hardware-in-the-loop testing through ray-tracing emulation
Threat Coverage	Denial of Service
6G -TE | Tech-Domain	None | SDR (USRP)
Data Type	SDR logs, classifier outputs, hyperparameter tuning data
Data Size	0.172
Transmitter | Receiver	SDR-based jammers (constant, periodic, reactive) | USRP-based SDR system with reconfigurable antennas

Owner | Access | License	Drexel University | Public | CC BY 4.0

The dataset is accessible via: https://zenodo.org/records/6304194

3.1.31 Wi-Fi 2.4 GHz Jamming attack scenario P2 measurements using ADALM Pluto and Maia SDR

The Wi-Fi 2.4 GHz Jamming attack scenario P2 measurements using ADALM Pluto and Maia SDR [RGS24] comprises physical-layer data (I-Q samples) captured to study jamming and normal Wi-Fi operations. The measurements were collected using ADALM Pluto SDR with custom Maia-SDR firmware, in the WIRID-LAB at the Military University Nueva Granada, over a 250-square-meter area. The dataset is divided into two groups: "JAMMER" and "NORMAL," each containing 165 files, with data collected across 15 points on 11 Wi-Fi channels. Measurements were sampled at 15 Msps for one second per point and stored in SigMF format.

Dataset ID | Name	**DAT 031 | Wi-Fi 2.4 GHz Jamming attack scenario P2 measurements using ADALM Pluto and Maia SDR**
Key Features	• Physical-layer I-Q samples formatted in SigMF • Two measurement groups: "JAMMER" and "NORMAL" • 165 files per group, collected across 15 points and 11 Wi-Fi channels • Captured with ADALM Pluto SDR running Maia-SDR firmware • Sampling rate of 15 Msps per second per point
Quick Overview	• Enables analysis of jamming attacks in a controlled environment • Facilitates comparison between normal and jamming Wi-Fi traffic patterns • Useful for evaluating jamming mitigation techniques • Data captured under realistic deployment conditions at WIRID-LAB
Threat Coverage	Denial of Service
6G -TE | Tech-Domain	None | Short-Range Communication (Wi-Fi)
Data Type	I/Q samples
Data Size	13.5 GB
Transmitter | Receiver	ADALM Pluto SDR as a jammer running “Jammer.m” to transmit Legacy Short Training Field interference | ADALM Pluto SDR with Maia-SDR firmware
Owner | Access | License	Military University Nueva Granada, Institut National des Sciences Appliquée de Lyon | Public | CC BY 4.0

The dataset is accessible via: https://zenodo.org/records/10456777

3.1.32 Radio Frequency Fingerprinting Exploiting Power Amplifier Active Load-pulling

The Radio Frequency Fingerprinting Exploiting Power Amplifier Active Load-pulling [Liu24] dataset contains real-world radio frequency fingerprinting (RFF) data to enhance RF device classification by utilizing physical unclonable function (PUF) techniques. The dataset exemplifies how RF hardware impairments can be tuned, focusing on power amplifiers (PAs) in RF chains, through active load-pulling. The goal is to amplify small variations among transmitters with identical PAs, enhancing RFF classification accuracy, particularly in low to medium SNR conditions. The dataset provides I/Q samples from eight transmitters of the same PA model, collected in two experimental setups: cable-connected and over-the-air (OTA). Transmission was carried out using a USRP X310, while reception was handled by a USRP B210, with data saved in .bin format.

Dataset ID | Name	**DAT032 | Radio Frequency Fingerprinting Exploiting Power Amplifier Active Load-pulling**

Key Features	• Real-world I/Q samples from 8 transmitters with identical PA models • Two experimental setups: cable-connected and over-the-air (OTA) • Sampling rate: 2 MHz, Center frequency: 2.4 GHz • Data format: .bin, duration: 1 second per signal • Metadata includes transmitter tags, SNR levels, and link types (LOS/NLOS)
Quick Overview	• Supports research on tuning hardware impairments for enhanced RFF classification • Evaluates performance in varying SNR conditions (11-26 dB LOS, 12-15 dB NLOS) • Enables training and testing of machine learning models for RF classification • Facilitates CNN-based RFF feature extraction and classification
Threat Coverage	Spoofing
6G -TE | Tech-Domain	None | SDR
Data Type	I/Q samples with 16-QAM modulation
Data Size	1446
Transmitter | Receiver	USRP X310 with Mini-Circuits PGA-105+ PAs | USRP B210
Owner | Access | License	Yuepei Li | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/radio-frequency-fingerprinting-exploiting-power-amplifier-active-load-pulling

3.1.33 Bidirectional CSI Measurement for V2X Communications

The Bidirectional CSI Measurement for V2X Communications [FPL24] contains channel state information (CSI) measurements collected to test the reciprocity of V2X channels between a roadside unit (RSU) and an on-board unit (OBU) using PSSCH signals. The dataset was generated using two USRP X310 platforms with CBX daughter boards, configured to operate at 5.91 GHz and a sampling rate of 30.72 MSample/s. Measurements were taken in an outdoor environment at varying speeds, resulting in CSI data grouped according to different speed scenarios. The dataset aims to support research in V2X communications by providing detailed CSI and PSSCH signal data for analysis.

Dataset ID | Name	**DAT 033 | Bidirectional CSI Measurement for V2X Communications**
Key Features	• Contains CSI measurements from bidirectional V2X communications • Data collected using two USRP X310 SDR platforms with CBX daughter boards • Sampling rate of 30.72 MSample/s at a carrier frequency of 5.91 GHz • Includes CSI data captured at various speeds: 40 km/h, 30 km/h, 20 km/h, 10 km/h, 5 km/h, and stationary • Data organized into 23, 137, 45, 28, 15, and 96 CSI groups respectively
Quick Overview	• Enables testing of V2X channel reciprocity using PSSCH signals • Useful for evaluating CSI under varying speed conditions and environments • Supports analysis of signal performance in outdoor long-range V2X scenarios

	• Instructions provided for plotting CSI data using MATLAB (CSI_plot.m)
Threat Coverage	Tampering | Information Disclosure
6G -TE | Tech-Domain	None | Vehicular Technologies
Data Type	CSI measurements and PSSCH signal data
Data Size	1.22 GB
Transmitter | Receiver	USRP X310 (Alice - RSU) | USRP X310 (Bob - OBU)
Owner | Access | License	Purple Mountain Laboratory | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/bidirectional-csi-measurement-v2x-communications

3.1.34 Shake on it

The Shake on it dataset [AAW23] is designed to support research on generating secure shared keys between Wi-Fi devices using Channel State Information (CSI). The dataset focuses on randomness extraction through device movement. By shaking one device, variations in CSI values are collected to generate cryptographic keys securely. The dataset documents CSI readings from multiple experiments, ensuring reproducibility of key generation methodologies.

Dataset ID | Name	**DAT 034 | Shake on it**
Key Features	• CSI data collected from Wi-Fi devices at 2.4 GHz and 5 GHz bands • Data includes measurements during stationary and shaking conditions • Consists of CSI data from multiple sub-carriers with different bandwidths (20 MHz, 40 MHz) • Includes metadata to synchronize CSI values across experiments • Raw data size is approximately 60 MB
Quick Overview	• Allows study of secure key generation from device motion • Facilitates testing of cryptographic algorithms such as Cascade for key reconciliation • Demonstrates impact of sub-carrier selection and device shaking on key generation • Enables exploration of Secure Bit Generation Rate (SBGR) under varied conditions
Threat Coverage	Tampering, Information Disclosure
6G -TE | Tech-Domain	None | Short-Range Communication (Wi-Fi)
Data Type	CSI measurements from Wi-Fi sub-carriers
Data Size	0.06 GB
Transmitter | Receiver	Raspberry Pi 4 (modified Wi-Fi chip firmware) | Raspberry Pi 4 (with CSI-enabled firmware)
Owner | Access | License	Tel Aviv University | Public | MIT License

The dataset is accessible via: https://github.com/tomer-avrahami/shake-on-it

3.1.35 MalwSpecSys: A Dataset Containing Syscalls of an IoT Spectrum Sensor Affected by Heterogeneous Malware

MalwSpecSys [SHv+22] dataset accurately models the internal behaviour of an IoT spectrum sensor (belonging to the ElectroSense platform and consisting of a Raspberry Pi 3 with a software-defined radio kit) when it is functioning normally and under attack.

To accomplish this, the system calls of the IoT sensor are monitored under normal behaviour, gathered, cleaned, and stored in a centralized directory. Then, the device is infected with current malware affecting IoT devices, such as the Bashlite botnet, Thetick backdoor, Bdvl rootkit, and a ransomware proof of concept.

The monitoring process is repeated for each malware, and infections are sequential, meaning that the device is not infected with more than one malware at a time.

The MalwSpecSys dataset is crucial for studying the impact of malware on the operation of IoT spectrum sensors and their security. It provides insights into how malware affects system behaviour, which is essential for developing physical layer security mechanisms tailored to IoT devices. Researchers can use this dataset to investigate vulnerabilities at the physical layer caused by malware and to design security solutions that enhance the robustness of IoT spectrum sensors against malicious attacks. It also aids in developing detection mechanisms that can identify abnormal behaviours indicative of security breaches.

Dataset ID | Name	**DAT 035 | MalwSpecSys: A Dataset Containing Syscalls of an IoT Spectrum Sensor Affected by Heterogeneous Malware**
Key Features	• System call logs of an IoT spectrum sensor captured during normal operation and under malware attacks • Contains data infected sequentially with Bashlite botnet, Thetick backdoor, Bdvl rootkit, and ransomware PoC • Logs stored with precise timestamps to reflect infection events and attack phases • Structured in folders based on behaviour phases and infection events • Malware types: Bashlite botnet, Thetick backdoor, Bdvl rootkit, Ransomware PoC • Logs from both normal and malware-infected states
Quick Overview	• Facilitates analysis of malware behaviour on IoT devices through system call monitoring • Enables studying sequential infections from multiple malware strains on a spectrum sensor • Useful for developing malware detection models using system-level data • Provides insights into individual attack phases and infection timelines
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	IOT | SDR
Data Type	System Call Logs (*.log)
Data Size	73.71 GB
Transmitter | Receiver	N/A (IoT device operations)
Owner | Access | License	University of Zurich | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/malwspecsys-dataset-containing-syscalls-iot-spectrum-sensor-affected-heterogeneous-malware

3.1.36 Radio Frequency Fingerprint LoRa Dataset with Multiple Receivers

The Radio Frequency Fingerprint LoRa Dataset with Multiple Receivers [SZM+24] contains experimental LoRaWAN signals collected using 10 commercial LoRa devices and 20 software-defined radio (SDR) receivers. It is designed to study radio frequency fingerprint identification (RFFI) systems, focusing on receiver-agnostic models and collaborative inference. This dataset includes multiple subsets representing different training and testing conditions, enabling robust testing across various scenarios. The data is provided in HDF5 format, allowing direct access and analysis for researchers. The goal of the dataset is to improve device authentication using non-cryptographic RF fingerprinting techniques while mitigating the impact of receiver hardware impairments.

Dataset ID | Name	DAT 036 | Radio Frequency Fingerprint LoRa Dataset with Multiple Receivers
Key Features	• Contains signals from 10 LoRa devices and 20 SDR receivers • Multiple subsets to facilitate receiver-agnostic and collaborative analysis • Metadata includes device identifiers and receiver configurations • Designed to study receiver drift and collaborative RFFI systems • Provided in HDF5 format for seamless analysis
Quick Overview	• Enables evaluation of RFFI models in multi-receiver settings • Facilitates collaborative and fine-tuning strategies for receiver-agnostic fingerprinting • Useful for studying receiver drift and its impact on device identification • Provides real-world experimental signals captured from commercial off-the-shelf devices and SDR platforms
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	IOT | LPWAN (LoRa)
Data Type	I/Q Samples
Data Size	27.67 GB
Transmitter | Receiver	10 commercial LoRa devices | 20 SDR platforms including RTL-SDRs and USRP devices
Owner | Access | License	Southeast University & University of Liverpool | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/radio-frequency-fingerprint-lora-dataset-multiple-receivers

3.1.37 Drone Remote Controller RF Signal Dataset

The Drone Remote Controller RF Signal Dataset [EEA+20] contains RF signals captured from 17 remote controllers (RCs) used to control drones, spanning 8 manufacturers. The RF signals, transmitted in the 2.4 GHz band, were collected using a passive RF surveillance system comprising a high-frequency oscilloscope, a 24 Decibel-Isotropic (dBi) parabolic antenna, and a low-noise amplifier. Each signal consists of 5 million samples, covering a duration of 0.25 millisecond (ms). The dataset aims to support research in the detection and classification of drones using RF signals. It provides ~1000 RF signals for each RC, stored in ".mat" format. MATLAB scripts included in the dataset allow users to analyse, visualize, and extract features from the data.

Dataset ID | Name	DAT 037 | Drone Remote Controller RF Signal Dataset
Key Features	• Contains RF signals from 17 drone RCs, spanning 8 manufacturers • ~1000 signals per RC, with each signal containing 5 million samples • Each signal spans 0.25 ms and operates in the 2.4 GHz band • Signals are stored in ".mat" format, including metadata such as make, model, and timestamp • MATLAB scripts provided for visualization and feature extraction
Quick Overview	• Enables the study of RF signals for drone detection and classification • Facilitates research on transient signal analysis and machine learning for RF security • Useful for developing models to detect malicious drones based on controller signals

	• Includes tools for creating databases and visualizing RF signals
Threat Coverage	Spoofing
6G -TE | Tech-Domain	IOT | Vehicular Technologies
Data Type	RF signals stored in .mat format
Data Size	124
Transmitter | Receiver	Drone Remote Controllers (DJI, Spektrum, Futaba, Graupner, HobbyKing, FlySky, Turnigy, Jeti Duplex) | Keysight MSOS604A Oscilloscope, 24 dBi parabolic antenna, low-noise amplifier
Owner | Access | License	North Carolina State University | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/open-access/drone-remote-controller-rf-signal-dataset

3.1.38 Cardinal RF (CardRF): An Outdoor UAV/UAS/Drone RF Signals with Bluetooth and Wi-Fi Signals Dataset

Cardinal RF (CardRF): An Outdoor UAV/UAS/Drone RF Signals with Bluetooth and Wi-Fi Signals Dataset [MEL+22] contains RF signals collected from a variety of UAV controllers, drones, Bluetooth devices, and Wi-Fi devices for the purpose of developing RF-based detection and identification systems. The dataset captures signals both at visual line-of-sight (VLOS) and beyond-line-of-sight (BVLOS), providing diverse conditions for analysis. It includes raw signals recorded in .mat format, each containing 5 million sampling points over a 0.25ms duration. Processed data with 1024-point signals is available in supplementary files for direct analysis. MATLAB code for signal plotting and manipulation is also included. The dataset totals over 65 GB and supports research in RF fingerprinting, UAV detection, and wireless communication.

Dataset ID | Name	**DAT 038 | Cardinal RF (CardRF): An Outdoor UAV/UAS/Drone RF Signals with Bluetooth and Wi-Fi Signals Dataset**
Key Features	• Contains raw RF signals from UAVs, their controllers, Bluetooth, and Wi-Fi devices • Includes both VLOS (line-of-sight) and BVLOS (beyond-line-of-sight) signals • 5 million sampling points per signal spanning 0.25ms • Processed signals are resampled to 1024 points for analysis • Provided in ".mat" format with MATLAB scripts for visualization and analysis
Quick Overview	• Facilitates research on UAV detection, RF fingerprinting, and wireless device identification • Useful for studying the effects of channel conditions (LOS vs. NLOS) on RF signatures • Supports machine learning-based signal classification using processed RF data • Enables RF-based UAV monitoring under real-world outdoor conditions
Threat Coverage	Spoofing
6G -TE | Tech-Domain	IOT | Short-Range Communication (Bluetooth, Wi-Fi)
Data Type	Raw RF signals in ".mat" format, supplementary processed signals
Data Size	66.26
Transmitter | Receiver	DJI UAVs, Apple Bluetooth devices, Cisco and TP-Link Wi-Fi devices | RF Signal Sensing and Capturing System (RFSSCS)

Owner | Access | License	University of Louisville, North Carolina State University | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/cardinal-rf-cardrf-outdoor-uavuasdrone-rf-signals-bluetooth-and-wifi-signals-dataset

3.1.39 LoRa sensor data sets for RF finger printing via Self-Organizing Feature Maps

The LoRa sensor data sets for RF fingerprinting via Self-Organizing Feature Maps [NDB22] contain two datasets designed for RF fingerprinting analysis using self-organizing feature maps (SOFMs). The first dataset includes raw LoRa I/Q signals collected from sensors in a controlled testbed. The second dataset consists of PNG images representing the output of SOFM-based processing applied to the raw data, facilitating dimensionality reduction and improved machine learning performance. The datasets are provided in MATLAB and image formats, enabling further research on IoT security and device authentication through RF fingerprinting techniques.

Dataset ID | Name	**DAT 039 | LoRa sensor data sets for RF finger printing via Self-Organizing Feature Maps**
Key Features	• Contains raw LoRa I/Q samples in MATLAB format • Includes SOFM-processed output as PNG images • Provided in two datasets: raw I/Q data and SOFM-processed output • Metadata includes RF parameters for each LoRa device and sensor
Quick Overview	• Enables RF fingerprinting analysis using self-organizing feature maps • Facilitates dimensionality reduction for efficient machine learning models • Suitable for research on IoT device authentication and RF security • Allows comparison of raw and SOFM-processed data for RF pattern analysis
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	IOT | LPWAN (LoRa)
Data Type	LoRa I/Q samples in MATLAB format, SOFM images in PNG format
Data Size	0.065 GB
Transmitter | Receiver	LoRa sensors used in the controlled testbed | Unspecified, as data includes processed results
Owner | Access | License	University of Bristol | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/lora-sensor-data-sets-rf-finger-printing-self-organizing-feature-maps

3.1.40 Dataset for Authentication and Authorization using Physical Layer Properties in Indoor Environment

The Dataset for Authentication and Authorization Using Physical Layer Properties in Indoor Environment [ATL+24] provides raw and analysed data focusing on indoor wireless communication between IoT devices using ZigBee Zolertia Z1 nodes. The dataset includes essential physical-layer metrics such as Received Signal Strength Indicator (RSSI), Link Quality Indicator (LQI), internal temperature, battery level, and antenna orientation. The data is collected from three client nodes at varying distances (1m, 2m, and 3m) from a gateway, exploring multiple antenna orientations (0°, 90°, and 180°) to analyse environmental impacts on the communication link. The dataset is structured in reusable formats to aid machine learning research for secure authentication and authorization in IoT networks.

Dataset ID | Name	DAT 040 | Dataset for Authentication and Authorization using Physical Layer Properties in Indoor Environment
Key Features	• Contains RSSI, LQI, device temperature, and battery levels from ZigBee Zolertia Z1 nodes • Three client nodes positioned at 1m, 2m, and 3m distances from the gateway • Multiple antenna orientations: 0°, 90°, and 180° • Data is formatted for direct import and analysis in machine learning applications • Provides 347,200 instances of communication events collected over 24-hour periods
Quick Overview	• Facilitates research on IoT authentication using physical-layer properties • Useful for examining environmental impacts on signal strength and quality • Enables evaluation of secure communication models for ZigBee networks • Supports exploration of device-to-device (D2D) authentication within indoor IoT ecosystems
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	None | Short-Range Communication (ZigBee)
Data Type	Raw and analysed data (tables, graphs, and figures)
Data Size	0.166
Transmitter | Receiver	ZigBee Zolertia Z1 nodes | ZigBee Zolertia Z1 gateway
Owner | Access | License	K. I. Ahmed et al. | Public | CC BY 4.0

The dataset is accessible via: https://zenodo.org/records/10706416

3.1.41 A dataset for RSSI based outdoor localization using LoRaWAN in a harbor as a harsh and industrial environment

The A dataset for RSSI-based outdoor localization using LoRaWAN in a harbour as a harsh and industrial environment [MKL23] contains RSSI and SNR measurements recorded from a LoRaWAN deployment in a harbor to assess the impact of a dynamic industrial environment on wireless localization. The data includes measurements from three gateways and one mobile end node equipped with GPS, along with two stationary nodes used for environmental monitoring. The dataset is provided in MongoDB format and enables researchers to evaluate the performance of RSSI-based localization techniques, particularly in harsh industrial conditions affected by humidity, temperature changes, and high metal infrastructure.

Dataset ID | Name	DAT 041 | A dataset for RSSI based outdoor localization using LoRaWAN in a harbor as a harsh and industrial environment
Key Features	• RSSI and SNR measurements from three gateways • Measurements include timestamp, device ID, latitude, longitude, and RSSI values • Data collected from mobile and stationary nodes to explore dynamic and static conditions • Provided in a structured format for direct import and analysis
Quick Overview	• Enables the evaluation of RSSI-based localization under real-world industrial conditions • Facilitates comparison of signal behaviour between dynamic and static environments

	• Supports development of location estimation models in harsh industrial environments
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	IOT | LPWAN (Lora WAN)
Data Type	Spreadsheet file consists of the Timestamp, Latitude-longitude, Device ID, RSSI and SNR measurements
Data Size	0.000114
Transmitter | Receiver	RAK811 LoRa Tracker Boards (SX1276 transceivers) | UG87 LoRaWAN Gateways (SX1301 transceivers)
Owner | Access | License	Polytechnic Institute of Viana do Castelo, Persian Gulf University | Public | CC BY 4.0

The dataset is accessible via: https://zenodo.org/records/10142174

3.1.42 Waldo Spectrogram Dataset for Signal Detection and Localization in the Citizen Broadband Radio Service (CBRS) Band

The Waldo Radar Detection Dataset [SCR+22] contains approximately 1,300 spectrograms generated in MATLAB, focusing on signal detection and localization within the CBRS band. These spectrograms feature a mix of 5G, LTE, DSSS, and radar signals, recorded with a sampling rate of 100 MHz over 20 ms. The dataset is provided in a structured format, with each spectrogram accompanied by an XML annotation file indicating signal labels and their pixel bounding boxes. The dataset aims to support research in signal detection and interference analysis, simulating real-world scenarios where signals overlap in frequency, complying with FCC regulations.

Dataset ID | Name	**DAT 042 | Waldo Spectrogram Dataset for Signal Detection and Localization in the Citizen Broadband Radio Service (CBRS) Band**
Key Features	• 1,300 spectrogram images generated from CBRS band signals in MATLAB • Includes signals: 5G, LTE, DSSS, and radar • Annotation files in XML format for bounding box detection • Sampled with 100 MHz sampling rate over a 20 ms period • Compliant with FCC power regulations for CBRS spectrum
Quick Overview	• Enables analysis of overlapping signals within the CBRS band • Facilitates training and testing deep learning models for signal detection (e.g., YOLOv3) • Useful for research on radar signal detection within noisy environments • Helps explore interference patterns between 5G, LTE, and radar signals
Threat Coverage	Spoofing, Tampering
6G -TE | Tech-Domain	None | SDR, Cellular Technologies (LTE, 5G)
Data Type	Spectrogram images with XML annotations
Data Size	0.25
Transmitter | Receiver	Simulated transmitters: 5G, LTE, DSSS, and naval radar signals | Simulated MATLAB environment
Owner | Access | License	GENESYS Lab | Public | Citation Required

The dataset is accessible via: https://genesys-lab.org/waldo

3.1.43 SenseORAN - Spectrogram Dataset for O-RAN based Radar Detection in the CBRS band

SenseORAN Dataset [RUD+23] provides spectrogram data for radar detection in the Citizens Broadband Radio Service (CBRS) band using O-RAN-based cellular networks. This dataset supports research in Open RAN (O-RAN) technology, specifically in detecting radar pulses that overlap with cellular signals, which is a critical challenge for spectrum sharing. The dataset was created as part of the study "SenseORAN: O-RAN based Radar Detection in the CBRS Band," which demonstrated the capability of machine learning-based xApps to detect radar signals in the CBRS band while reducing interference with LTE operations.

The dataset consists of spectrogram images, generated from overlapping radar and LTE signals, captured under varying noise levels and traffic conditions. It is intended for training and validating machine learning models, specifically for radar detection within O-RAN (Open Radio Access Network) architectures. The dataset includes images and labels, with the images detailing signal-to-noise ratio, signal-to-interference-plus-noise ratio, and UDP (User Datagram Protocol) values. It provides a critical resource for research aimed at enhancing spectrum sensing and management in shared bands.

Dataset ID | Name	**DAT 043 | SenseORAN - Spectrogram Dataset for O-RAN Based Radar Detection in the CBRS Band**
Key Features	• Spectrograms representing over-the-air radar and cellular signals • Organized into two folders: images (spectrograms) and labels (metadata) • Includes 3 split files: training, validation, and test directories • Captures different SNR, SINR, and UDP traffic levels in the CBRS band • Labels radar pulses with thin horizontal lines within spectrogram images Spectrogram images of overlapping radar and cellular signals • Includes different noise levels and traffic conditions for robust testing
Quick Overview	• Ideal for developing machine learning models for radar detection in shared spectrum environments • Supports O-RAN architecture research and radar detection in the CBRS band • Over-the-air data collection with USRP X310 SDRs and srsRAN-based LTE network setup
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	None | Cellular Technologies (O-RAN, LTE)
Data Type	I/Q samples
Data Size	9.12
Transmitter | Receiver	3 Ettus Research USRP X310 SDRs (1 for radar, 2 for 4G LTE)
Owner | Access | License	GENESYS Lab | Public | Citation Required

The dataset is accessible via: https://genesys-lab.org/senseoran

3.1.44 COPILOT - Dataset for leveraging Co-Operative Perception using LiDAR for Handoffs between Road Side Units

The COPILOT Dataset for Leveraging Co-Operative Perception Using LiDAR for Handoffs Between Road Side Units [PRC+24] contains GPS, LiDAR, and wireless connectivity data collected from a V2X environment. The dataset replicates real-world vehicular network scenarios and is categorized into four types: Line of Sight (LOS), Non-Line of Sight (NLOS) with pedestrian blockages, NLOS with vehicular obstructions,

and NLOS with both vehicles and a pedestrian in motion around the RSU. Each scenario contains multiple episodes, capturing dynamic handoff conditions between RSUs. The dataset provides a valuable resource for developing machine learning-based strategies for RSU selection and proactive handoffs in mmWave networks.

Dataset ID | Name	**DAT 044 | COPILOT - Dataset for leveraging Co-Operative Perception using LiDAR for Handoffs between Roadside Units**
Key Features	• Captures GPS, LiDAR, and wireless connectivity data • Covers 4 categories: LOS, static pedestrian, static vehicle, and dynamic blockages • 10 episodes per scenario, with varying sample counts • Includes real-world measurements from mmWave radios and RSUs • Provided in sync with timestamps, GPS data, and point cloud information
Quick Overview	• Enables V2X researchers to explore RSU handoffs under various real-world conditions • Supports analysis of mmWave communication performance with changing LOS and NLOS scenarios • Facilitates machine learning models for cooperative perception and proactive RSU handoffs • Suitable for studying throughput and latency in mmWave networks
Threat Coverage	Retrieving data. Wait a few seconds and try to cut or copy again.
6G -TE | Tech-Domain	IOT, ISAC | mmWaves and sub-THz
Data Type	GPS, LiDAR point cloud data, wireless connectivity metrics
Data Size	35 GB
Transmitter | Receiver	Autonomous vehicle with GPS and Ouster OS1-64 LiDAR | Talon AD7200 60GHz routers, serving as RSUs
Owner | Access | License	GENESYS Lab | Public | Citation needed

The dataset is accessible via: https://genesys-lab.org/copilot

3.1.45 AirID RF Fingerprinting Dataset

The AirID RF Fingerprinting Dataset [MSS+20] contains recordings of raw IQ samples collected from over-the-air transmissions of UAVs. This dataset was created to evaluate the performance of UAV identification using custom RF fingerprints injected into the transmitted IQ symbols. The dataset includes transmissions from 4 Ettus B200mini SDR radios, each installed on DJI M100 UAVs, with unique IQ imbalances for identification. Metadata and data files are provided, following the SigMF format, allowing researchers to reproduce the original work or to further explore machine learning problems in wireless communications.

Dataset ID | Name	**DAT 045 | AirID RF Fingerprinting Dataset**
Key Features	• Raw IQ samples from UAV transmissions with custom IQ imbalances • 4 UAV transmitters (Ettus B200mini radios) mounted on DJI M100 UAVs • Metadata and dataset files in SigMF format • UAV-to-UAV and ground-to-UAV transmissions with varied hovering conditions • Includes controlled interference to test identification performance
Quick Overview	• Enables development of RF fingerprinting methods for UAV identification

	• Allows assessment of deep learning models for wireless communication problems • Facilitates experiments with real-world conditions such as interference and dynamic motion • Useful for evaluating RF signatures in changing environments
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	IOT | SDR, Vehicular Technologies
Data Type	Complex IQ samples from SDR transmissions
Data Size	4.5
Transmitter | Receiver	4 DJI M100 UAVs with Ettus B200mini radios | 5 UAV and ground receivers with Ettus B200mini radios
Owner | Access | License	GENESYS Lab | Public | Citation needed

The dataset is accessible via: https://genesys-lab.org/airid

3.1.46 POWDER RF Fingerprinting Dataset

The POWDER RF Fingerprinting Dataset [RJS+20] contains raw I/Q samples collected from over-the-air transmissions on the POWDER platform in Salt Lake City, Utah. This dataset includes signals from four base stations (MEB, Browning, Beavioral, Honors) transmitted to a fixed endpoint (Humanities). The base stations use USRP X310 radios to emit standards-compliant frames, including IEEE 802.11a (Wi-Fi), LTE, and 5G-NR. Data was collected on two independent days, with five sets of I/Q samples for each base station per waveform. Sampling rates were set to 5 MS/s for Wi-Fi and 7.69 MS/s for LTE and 5G. The dataset is stored in the SigMF format, with each recording consisting of a metadata file and a binary file of I/Q samples. This data supports research in RF fingerprinting and physical-layer authentication for Open RAN environments, enabling the identification of emitters based on their unique signal characteristics.

Dataset ID | Name	**DAT 046 | POWDER RF Fingerprinting Dataset**
Key Features	• Contains raw I/Q samples from over-the-air transmissions on the POWDER platform • Includes data from 4 USRP X310 base stations transmitting Wi-Fi, LTE, and 5G signals • Data was collected on two separate days, with 5 sets of samples per day • Stored in SigMF format with binary sample files and corresponding metadata • Captures signals sampled at 5 MS/s (Wi-Fi) and 7.69 MS/s (LTE, 5G) at 2.685 GHz
Quick Overview	• Enables RF fingerprinting analysis of multi-protocol wireless signals • Facilitates research on emitter detection and physical-layer authentication in Open RAN networks • Supports studies on the impact of environmental changes and channel variations on classification performance • Useful for evaluating machine learning models, including those with triplet-loss functions
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	None | Cellular Technologies (5G, LTE), Short Range Communication (Wi-Fi)
Data Type	Complex I/Q samples
Data Size	4 GB (compressed)

Transmitter | Receiver	USRP X310 | USRP B210
Owner | Access | License	GENESYS Lab | Public | Citation needed

The dataset is accessible via: https://genesys-lab.org/powder

3.1.47 Data Augmentation RF Fingerprinting Dataset

The Data Augmentation RF Fingerprinting Dataset [SSD+20] consists of MATLAB-simulated datasets developed to improve RF fingerprinting by providing channel-resilient training data. The dataset comprises three components: TxData, Day1, and Day2. TxData contains transmitter-side data without channel or noise effects, while Day1 and Day2 provide receiver-side raw IQ samples subjected to different channel models and SNR variations. The dataset is released in SigMF format, where each transmission is represented as binary interleaved I/Q samples (.bin) with accompanying metadata (.json). This dataset supports deep learning research for channel-invariant RF fingerprinting.

Dataset ID | Name	**DAT 047 | Data Augmentation RF Fingerprinting Dataset**
Key Features	• MATLAB-simulated RF fingerprinting datasets • TxData: 10 radios transmitting 2042 packets each, recorded at the transmitter side (2 GB) • Day1: 10 radios, 16 SNR levels (-10 dB to 20 dB), 2042 packets per radio, recorded at the receiver side (45 GB) • Day2: Same structure as Day1 but with new channel seeds (45 GB) • Datasets are provided in SigMF format, requiring binary-to-float64 conversion
Quick Overview	• Enables research on channel-invariant RF fingerprinting using deep learning • Useful for training DNNs to handle channel and noise variations • Supports analysis of raw I/Q samples for RF fingerprinting tasks • Accompanies a paper demonstrating up to 75% accuracy improvement with data augmentation
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	None | Short-Range Communication (Wi-Fi 802.11a)
Data Type	Interleaved binary I/Q samples (float64)
Data Size	92 GB (2 GB + 45 GB + 45 GB)
Transmitter | Receiver	MATLAB virtual radios | MATLAB WLAN Toolbox with TGn channel model
Owner | Access | License	GENESYS Lab | Public | Citation needed

The dataset is accessible via: https://genesys-lab.org/dataaugmentation

3.1.48 Fast mmWave Beamforming Dataset with Camera Images

The Fast mmWave Beamforming Dataset with Camera Images [SBG+20] contains two datasets collected to facilitate research on improving mmWave beamforming using machine learning and visual data. The datasets include images from a testbed where transmitter and receiver devices are monitored with GoPro Hero 4 cameras. Two different obstacles, a wooden board and a cardboard box, were placed to block the line-of-sight (LOS) path between devices, with measurements taken for each configuration. Each dataset contains the associated best beam pair configuration, raw SNR measurements, and MATLAB code to process the optimal beam configuration. The datasets are provided in HDF5 format, organized into groups based on different views and antenna setups.

Dataset ID | Name	**DAT 048 | Fast mmWave Beamforming Dataset with Camera Images**

Key Features	• Includes two datasets with camera images and associated best beam pair configurations • Raw SNR measurements recorded with 169 different beam pair configurations • Covers scenarios with two obstacles: wooden board and cardboard box • Provided in HDF5 format with metadata for transmitter and receiver locations • Contains MATLAB code to process measurements and identify the best beam pair
Quick Overview	• Facilitates research on visual information for guiding mmWave beamforming • Enables evaluation of ML models to reduce beam alignment time • Provides real-world data from indoor testbed experiments • Useful for studying beamforming with varying obstacles and lighting conditions
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	mmWave/sub-THz, ISAC | SDR
Data Type	Camera images, SNR measurements, MATLAB code
Data Size	6.4 GB
Transmitter | Receiver	NI mmWave Transceiver System with SiBeam RF heads
Owner | Access | License	GENESYS Lab | Public | Citation Required

The dataset is accessible via: https://genesys-lab.org/mmWave-beamforming

3.1.49 Hovering UAVs RF Fingerprinting Datasets

The Hovering UAVs RF Fingerprinting Datasets [SRS+20] contain non-standard waveforms collected from 7 identical DJI M100 UAVs flying inside an RF anechoic chamber. The dataset aims to facilitate RF fingerprinting of identical UAVs using raw I/Q samples. UAVs were flown at distances of 6, 9, 12, and 15 feet from the receiver, with transmissions captured using an Ettus USRP X310 equipped with a UBX 160 daughterboard. At each distance, 4 bursts of data were collected, each burst containing ~140 interleaved periods of data and noise, resulting in a total of over 13,000 transmissions. The dataset is released in SigMF format, with the raw signals stored in binary (.bin) files alongside metadata in JSON files.

Dataset ID | Name	**DAT 049 | Hovering UAVs RF Fingerprinting Datasets**
Key Features	• Contains non-standard waveforms from 7 identical DJI M100 UAVs • I/Q samples captured across four distances: 6, 9, 12, and 15 feet • Over 13,000 transmissions stored in SigMF format with accompanying JSON metadata • Metadata includes transmission distance, protocol, environment, and device details • Binary format (.bin) with interleaved I/Q values in float16 data type
Quick Overview	• Enables RF fingerprinting of identical UAVs using non-standard waveforms • Facilitates analysis of channel variations and signal imperfections caused by hovering • Supports deep learning models for UAV classification and identification • Useful for studying the impact of UAV movement on signal stability

Threat Coverage	• Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	None | SDR (USRP X310)
Data Type	I/Q samples with metadata in SigMF format
Data Size	4.4
Transmitter | Receiver	7 identical DJI M100 UAVs | Ettus USRP X310 with UBX 160 daughterboard
Owner | Access | License	GENESYS Lab | Public | Citation needed

The dataset is accessible via: https://genesys-lab.org/hovering-uavs

3.1.50 Channel Estimation in Beyond-5G Massive MIMO Datasets

The Channel Estimation in Beyond-5G (B5G) Massive MIMO Datasets [BSB+21] contain simulated data for massive MIMO (mMIMO) systems, focusing on deep learning-based channel estimation methods. The dataset simulates a downlink transmission scenario involving a base station (BS) equipped with 32 Uniform Rectangular Array (URA) antennas and user equipment (UE) with 4 Uniform Linear Array (ULA) antennas. The system operates at 28 GHz with 100 MHz bandwidth, generating 9,000 training samples and 500 testing samples across various SNR levels. Each transmission includes channel sounding preambles and data transfer phases, enabling detailed study of CSI estimation for both training and testing purposes. MATLAB's Communication and Phased Array toolboxes were used to generate the dataset.

Dataset ID | Name	**DAT 050 | Channel Estimation in Beyond-5G Massive MIMO Datasets**
Key Features	• Simulates downlink transmissions with a 32x4 mMIMO configuration • Operates at 28 GHz with 100 MHz bandwidth and 234 sub-carriers • Contains 9,000 training samples and 500 testing samples with various SNR levels • Includes channel sounding frames and CSI data across multiple scatterers • Generated using MATLAB's Communication and Phased Array toolboxes
Quick Overview	• Supports evaluation of mMIMO channel estimation algorithms under varying SNR levels • Enables study of CSI estimation methods, such as deep learning and Linear Minimum Mean Square Error (LMMSE) • Suitable for testing end-to-end performance improvements in low SNR conditions • Facilitates development of edge AI-based networking optimization for B5G
Threat Coverage	Tampering, Information Disclosure
6G -TE | Tech-Domain	None | Celullar Technologies (5G)
Data Type	Simulated OFDM data (MATLAB-generated)
Data Size	19 GB (compressed)
Transmitter | Receiver	Base Station with 32 URA antennas | User Equipment with 4 ULA antennas
Owner | Access | License	GENESYS-LAB | Public | Citation needed

The dataset is accessible via: https://genesys-lab.org/CS-5g-beyond

3.1.51 FLASH

The FLASH dataset [SGR+22] contains multimodal data and RF ground truth collected from an autonomous vehicle equipped with sensors. This dataset supports the validation of federated learning architectures for high band mmWave sector selection. It includes recordings from four sensor modalities: mmWave radio, GPS, camera, and LiDAR. The data captures both line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios with variations, such as obstacles (pedestrians, static cars, and moving cars) between the transmitter (Tx) and receiver (Rx). This setup ensures synchronized collection across sensors with the Robot Operating System (ROS) managing sensor data in real time.

Dataset ID | Name	**DAT 051 | FLASH**
Key Features	• Multimodal sensor data: mmWave radio, GPS, camera, and LiDAR • Synchronized RF ground truth, including sector IDs and RSSI values • Captures LOS and NLOS scenarios across multiple categories • Includes metadata such as speed, lane, and obstacle type • Provided in ROS bag files for seamless import and processing
Quick Overview	• Enables testing of machine learning models for mmWave sector selection • Facilitates research on federated learning with real-world sensor data • Supports analysis of the impact of environmental changes (e.g., moving obstacles) on RF performance • Suitable for evaluating multimodal learning models combining sensor and RF data
Threat Coverage	Spoofing, Tampering
6G -TE | Tech-Domain	ISAC, mmWaves and sub-THz | Vehicular Technologies, Short Range Communication (IEEE 802.11ad)
Data Type	Multimodal sensor data with RF ground truth
Data Size	20
Transmitter | Receiver	TP-Link Talon AD7200 (Qualcomm QCA9500)
Owner | Access | License	GENESYS-LAB | Public | Citation needed

The dataset is accessible via: https://genesys-lab.org/multimodal-fusion-nextg-v2x-communications#flash

3.1.52 ICARUS: Detecting Anomalous RF Signals Dataset

The ICARUS: Detecting Anomalous RF Signals Dataset [RCS+23] consists of synthetically generated and over-the-air (OTA) captured in-phase and quadrature samples, along with cyclostationary signal processing (CSP) features. The dataset is organized into three modules: (i) synthetic datasets of LTE and DSSS signals generated using MATLAB, (ii) indoor OTA datasets from the NSF POWDER testbed, and (iii) OTA datasets with commercial cellular LTE signals combined with DSSS signals. It enables anomaly detection in wireless signals by analysing DSP and machine-learning-based features. The dataset offers metadata for signal classification tasks and is prepared for import and use in research and experimentation on RF signal detection and analysis.

Dataset ID | Name	**DAT 052 | ICARUS: Detecting Anomalous RF Signals Dataset**
Key Features	• IQ samples and CSP features for LTE and DSSS signals • Three modules: synthetic, indoor OTA, and OTA cellular datasets • Collected using USRP X310 and B210 at multiple frequencies and sampling rates • Contains metadata about the collection setup and signal configurations

Quick Overview	• Enables detection and modulation classification of DSSS anomalies within LTE frames • Supports both synthetic and real-world OTA signals for comprehensive analysis • Useful for evaluating signal processing and machine learning models on wireless signals • Suitable for testing under different SNR and interference conditions
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	None | Cellular Technologies (LTE) SDR (DSSS, cyclostationary, Signal Processing)
Data Type	I/Q Samples, CSP Features
Data Size	17.5
Transmitter | Receiver	USRP X310 | USRP X310 & USRP B210
Owner | Access | License	GENESYS Lab | Public | Citation Required

The dataset is accessible via: https://genesys-lab.org/ICARUS

3.1.53 CBRS: Real-world Radar and LTE Signals Dataset Collected Over-the-air in Shared CBRS Band

The Real-world Radar and LTE Signals Dataset Collected Over-the-air in Shared CBRS Band [TCG+23] contains 5,640 frames of IQ samples, each 40ms long, collected at 30.72 MHz sampling rate. The data captures overlapping and non-overlapping transmissions of LTE and radar signals in the shared CBRS band (3.55-3.7 GHz) under various SINR conditions (15–35 dB). The dataset was generated using an experimental setup with software-defined radios (SDRs) in an RF anechoic chamber, following a controlled API-based configuration. It is provided in SigMF format with metadata files that document essential collection parameters such as sampling rates and center frequencies. This dataset enables researchers to study and develop ML-based solutions for interference detection and signal localization in shared spectrum environments.

Dataset ID | Name	**DAT 053 | CBRS**
Key Features	• Contains 5,640 IQ sample frames of 40ms duration each • Collected at a 30.72 MHz sampling rate using SDRs in an RF anechoic chamber • Includes overlapping and non-overlapping LTE and radar transmissions in the CBRS band • Metadata stored in SigMF format documenting parameters like SINR, sampling rates, and frequencies
Quick Overview	• Facilitates research on ML-based interference detection and localization in shared spectrum environments • Enables analysis of LTE and radar coexistence in real-world OTA settings • Supports high-SINR and low-SINR scenario evaluations for ML model training • Demonstrates API-based automated dataset collection for spectrum research
Threat Coverage	Spoofing, Tampering
6G -TE | Tech-Domain	None | Cellular Technologies (LTE)
Data Type	IQ samples with SigMF metadata files
Data Size	55.5
Transmitter | Receiver	Ettus X310 SDR | Ettus X310 SDR

Owner | Access | License	GENESYS Lab | Public | Citation Required

The dataset is accessible via: https://genesys-lab.org/CBRS

3.1.54 Berlin V2X

The Berlin V2X dataset [HGP+23] offers high-resolution GPS-located wireless measurements across urban environments in Berlin, Germany, for both cellular and sidelink radio access technologies. The data was collected over three days using up to four cars in different driving modes, including platoon and dispersed configurations. The dataset provides insights into physical layer parameters, radio resource management, wireless Quality of Service (QoS), and GPS positioning. Additionally, it includes side information such as traffic and weather conditions. The dataset is designed to support machine learning (ML) research in tasks like QoS prediction, transfer learning, and proactive radio resource management.

Dataset ID | Name	**DAT054 | Berlin V2X**
Key Features	• GPS-located wireless measurements collected in urban environments • Covers cellular (LTE) and sidelink technologies with data from multiple operators • Includes physical layer parameters, radio resource management, and QoS metrics • Provides GPS positioning data along with traffic and weather side information • Data labelled and pre-filtered for easy analysis with ML tools
Quick Overview	• Facilitates QoS prediction and link selection research • Enables transfer learning across multiple radio access technologies • Supports proactive radio resource management studies • Includes cellular (LTE) and sidelink (5.9 GHz) data from two mobile network operators
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	IOT, ISAC | Cellular Technologies (LTE), Vehicular Technologies
Data Type	QoS, Cellular RRM, Signal Quality, GPS Data
Data Size	21.10
Transmitter | Receiver	Multiple Vehicles with GPS | Multiple Vehicles with GPS
Owner | Access | License	Hernangómez et al. | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/open-access/berlin-v2x

3.1.55 AI4Mobile Industrial Wireless Datasets: iV2V and iV2I+

The AI4Mobile Industrial Wireless Datasets: iV2V and iV2I+ [HPW+22] contain wireless measurements from two industrial testbeds. The iV2V dataset features 10 hours of sidelink communication between three Automated Guided Vehicles (AGVs). The iV2I+ dataset spans 16 hours of data collection at an industrial site, where an autonomous cleaning robot interacts with a private cellular network. These datasets include physical-layer parameters (e.g., signal strength, quality), Quality of Service (QoS) metrics (e.g., throughput, delay), and positioning data, facilitating applications such as fingerprinting, line-of-sight detection, and link selection.

Dataset ID | Name	**DAT 055 | AI4Mobile Industrial Wireless Datasets: iV2V and iV2I+**
Key Features	• Includes wireless measurements from two industrial testbeds (iV2V and iV2I+) • iV2V covers 10 hours of sidelink communication among 3 AGVs

	• iV2I+ covers 16 hours of communication between an autonomous robot and a private cellular network • Data includes physical layer parameters, QoS metrics, and positioning information • Parquet files are provided for easy import and analysis
Quick Overview	• Supports tasks such as fingerprinting, line-of-sight detection, QoS prediction, and link selection • Useful for machine learning research in industrial wireless communication • Pre-filtered and labeled data ensures fast onboarding
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	IOT, ISAC | Cellular Technologies, Vehicular Technologies
Data Type	Physical-layer parameters, QoS metrics, and sensor data
Data Size	291
Transmitter | Receiver	AGVs and autonomous cleaning robot | Private cellular base stations and AGV sensors
Owner | Access | License	Hernangómez et al. | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/open-access/ai4mobile-industrial-wireless-datasets-iv2v-and-iv2i

3.1.56 Bistatic MIMO Radar Sensing

The Bistatic MIMO Radar Sensing [DGW+23] dataset contains Matlab code and synthetic aperture measurement data used to investigate wireless power transfer (WPT) through radar imaging. The measurements include scattering parameter S21S21 data collected over 1,000 frequency steps (3–10 GHz) using a synthetic aperture measurement testbed. This dataset consists of channel matrices, antenna configurations, and characterization data that facilitate the study of radar sensing using ultra-wideband MIMO systems. The used antennas include a 51-element uniform linear array (ULA) and a 13×13 uniform rectangular array. These data support research on detecting reflective surfaces and predicting channel state information for efficient power transfer.

Dataset ID | Name	**DAT 056 | Bistatic MIMO Radar Sensing**
Key Features	• Measurement data collected at 1000 frequency steps across 3-10 GHz • Includes channel matrices, antenna configurations, and gain data • Provides antenna positions for ULA and URA configurations • Contains XETS antenna characterization measurements in an anechoic chamber
Quick Overview	• Enables study of radar sensing through synthetic aperture measurements • Facilitates wireless power transfer (WPT) experiments using UWB MIMO systems • Supports development of beamforming techniques using channel state information (CSI) • Useful for examining reflective surface detection using MIMO radar technology
Threat Coverage	Spoofing, Tampering, Information Disclosure, Repudiation
6G -TE | Tech-Domain	ISAC | SDR
Data Type	MIMO Radar and CSI data

Data Size	0.167
Transmitter | Receiver	Synthetic ULA with 51 antennas | Synthetic URA with 13×13 antennas
Owner | Access | License	Graz University of Technology | Public | CC BY 4.0

The dataset is accessible via: https://gitlab.com/baenshy/bistatic-mimo-radar-sensing

3.1.57 IEEE 802.11p Wireless Congestion and Jamming Experiments

The IEEE 802.11p Wireless Congestion and Jamming Experiments [Kih22] provides complex-baseband samples simulating vehicle-to-vehicle (V2V) communications under various jamming scenarios. Up to 600 vehicles were simulated using commercial IEEE 802.11p radios, broadcasting Basic Safety Messages (BSMs) at a 10 Hz rate per vehicle. The dataset is structured to support research on jamming detection and mitigation using machine learning for vehicular networks. Data files contain float32 complex-baseband samples stored in IQIQIQ order, captured using GNURadio and saved as .dat files. The dataset offers multiple jamming scenarios, with the file names indicating the number of simulated vehicles (e.g., 2radio.dat and 600radio.dat). Each subset uses a 1 MHz sampling rate, meaning 1 million samples correspond to 20 BSMs for two vehicles or 6,000 BSMs for 600 vehicles.

Dataset ID | Name	**DAT 057 | IEEE 802.11p Wireless Congestion and Jamming Experiments**
Key Features	• Contains complex-baseband samples of IEEE 802.11p radios in V2V scenarios • Simulation-based dataset with up to 600 vehicles transmitting at 10 Hz per vehicle • Data captured in .dat files using float32 IQIQIQ order • Includes example code and tutorials on CodeOcean for signal analysis
Quick Overview	• Enables research on jamming detection and mitigation strategies for V2V networks • Provides multiple jamming scenarios, with files named to indicate the number of vehicles • Supports physical layer-based machine learning experiments • Suitable for vehicular network research using IEEE 802.11p technology
Threat Coverage	Denial of Service
6G -TE | Tech-Domain	IOT | Vehicular Technologies (IEEE 802.11p)
Data Type	Complex-baseband samples (IQ data)
Data Size	0.175
Transmitter | Receiver	Simulation-based IEEE 802.11p V2V radios | Simulation data captured using GNU Radio (File Sink object)
Owner | Access | License	Billy Kihei | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/ieee-80211p-wireless-congestion-and-jamming-experiments#files

3.1.58 e-FLASH

The e-FLASH dataset [Gua23] provides multimodal data collected for enhancing situational awareness in millimetre-wave (mmWave) multiple-antenna systems, specifically aiding in beam selection for vehicle-to-everything (V2X) communications. This dataset, which extends the FLASH dataset by incorporating additional sensor modalities, consists of 23 GB of real-world data organized by category, scenario, and episode. The dataset features time-synchronized recordings from various sensors, including mmWave radios, GPS, camera systems, and LiDAR, alongside RF ground truth data for selected beams in the mmWave band. This

structured approach aims to facilitate machine learning (ML) applications in wireless communication required for autonomous driving.

The e-FLASH dataset differs from the original FLASH dataset primarily in the inclusion of additional sensor modalities and enhancements in data organization. While the FLASH dataset utilizes mmWave radios, GPS, a single GoPro camera, and Velodyne LiDAR for collecting multimodal data, the e-FLASH dataset expands on this by integrating both a GoPro Hero4 and a Hero9 camera, thereby providing front- and side-facing perspectives. Additionally, the e-FLASH dataset incorporates Ouster LiDAR, which offers a higher resolution with 64 channels compared to the Velodyne's 16 channels. This increased sensor diversity enables more comprehensive data collection, contributing to improved situational awareness and decision-making capabilities in V2X communications. Moreover, the e-FLASH dataset is organized hierarchically, allowing for a more structured approach in managing and analysing the multimodal data collected, which may facilitate machine learning applications more effectively than its predecessor.

Dataset ID | Name	**DAT058 | e-FLASH Dataset**
Key Features	• Multimodal data for mmWave beam selection • Organized by category, scenario, and episode in a hierarchical format • Includes sensor modalities: mmWave radio, GPS, multiple cameras, Velodyne and Ouster LiDAR • Time-synchronized and heterogeneous data types paired with RF ground truth data
Quick Overview	• Supports machine learning in wireless communication for autonomous driving • Aids in decision making for beam selection by including non-RF modalities • Facilitates research in ML-based PHY-layer optimization areas, such as beamforming and localization
Threat Coverage	Spoofing, Information Disclosure, Tampering, Repudiation
6G -TE | Tech-Domain	IOT, mmWaves and sub-THz, ISAC | Vehicular Technologies (V2X)
Data Type	Multimodal data: LiDAR, GPS, camera images, and mmWave RF data
Data Size	23
Transmitter | Receiver	TP-Link Talon AD7200 tri-band routers with Qualcomm QCA9500 chips | Various sensors: GPS, Hero 4 and Hero 9 cameras, Velodyne LiDAR, Ouster LiDAR
Owner | Access | License	GENESYS LAB | Public | CC BY

The dataset is accessible via: https://genesys-lab.org/multimodal-fusion-nextg-v2x-communications#eflash

3.1.59 Dataset: IQ samples of LTE, 5G NR, Wi-Fi, ITS-G5, and C-V2X PC5

The IQ samples of LTE, 5G NR, Wi-Fi, ITS-G5, and C-V2X PC5 [GS23] contains raw IQ samples captured from various wireless technologies, including LTE, 5G NR, Wi-Fi, ITS-G5, and C-V2X PC5, as well as Noise. The data is organized into six sub-datasets based on sampling rates of 1, 5, 10, 15, 20, and 25 Msps, with each sub-dataset comprising 7,500 examples per technology. Each example consists of IQ samples corresponding to a specific sampling rate and technology, allowing detailed analysis and classification across different wireless standards. This dataset was developed to support research in wireless signal classification, interference analysis, and the evaluation of machine learning models across multiple communication standards and varying sampling rates.

Dataset ID | Name	**DAT 059 | Dataset: IQ samples of LTE, 5G NR, Wi-Fi, ITS-G5, and C-V2X PC5**

Key Features	• Contains raw IQ samples from LTE, 5G NR, Wi-Fi, ITS-G5, C-V2X PC5, and Noise • Six sub-datasets based on sampling rates: 1, 5, 10, 15, 20, and 25 Msps • 7,500 examples per technology within each sampling rate • Varying modulation and traffic load configurations
Quick Overview	• Enables comprehensive analysis of multiple wireless technologies across different sampling rates • Facilitates development and evaluation of classification models for diverse signal types • Supports research on the impact of sampling rates and hardware configurations on signal characteristics • Provides a robust dataset for studying interference and noise across various communication standards
Threat Coverage	Tampering
6G -TE | Tech-Domain	None | Short-Range Communication (Wi-Fi), Cellular Technologies (LTE, 5G NR), Vehicular Technologies (ITS-G5, C-V2X PC5)
Data Type	Signal recordings (IQ samples) from LTE, 5G NR, Wi-Fi, ITS-G5, C-V2X PC5, and Noise
Data Size	1.2
Transmitter | Receiver	USRP X310, RSU4 | USRP X310, USRP N310
Owner | Access | License	IDLab Ghent University-IMEC | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/dataset-iq-samples-lte-5g-nr-wi-fi-its-g5-and-c-v2x-pc5

3.1.60 Statistical Characterization of 28GHz V2X Channels via Autonomous Beam-Steered Measurements

The Statistical Characterization of 28GHz V2X Channels via Autonomous Beam-Steered Measurements [KZA+24] contains raw complex I/Q samples from 28GHz Vehicle-to-Everything (V2X) transmissions collected to analyse channel characteristics. The data is divided into multiple sub-datasets based on different vehicular routes, including urban, suburban, and foliage environments. Each entry includes geo-positioning logs, alignment specifics, and signal propagation measurements captured using a fully autonomous robotic beam-steering platform equipped with a custom broadband sliding correlator channel sounder. The dataset is provided in JSON format (compressed as tar.gz), facilitating comprehensive analysis and modelling of millimetre-wave (mmWave) V2X channels.

Dataset ID | Name	**DAT 060 | Statistical Characterization of 28GHz V2X Channels via Autonomous Beam-Steered Measurements**
Key Features	• Raw complex I/Q samples from 28GHz V2X signals • Multiple sub-datasets based on vehicular routes: Urban Campus Route I, II, III, Foliage Route, Garage Route, Stadium Route, Suburban Neighbourhood Route • Geo-positioning logs, alignment specifics, and signal propagation measurements • Data captured using autonomous robotic beam-steering platform and USRP B210 receiver Provided in JSON format
Quick Overview	• Facilitates analysis of 28GHz V2X channel characteristics across diverse environments

	• Enables evaluation of channel models and propagation standards specific to V2X communications • Supports studies on spatial consistency, multipath clustering, shadowing, and fading properties • Suitable for developing and validating mmWave V2X communication strategies
Threat Coverage	Spoofing, Information Disclosure, Tampering, Repudiation
6G -TE | Tech-Domain	IOT, mmWaves and sub-THz | Cellular Technologies (mmWAve, V2X, 5G, ITS), Vehicular Technologies (V2X, ITS)
Data Type	JSON files containing complex I/Q samples and metadata
Data Size	228
Transmitter | Receiver	Custom Broadband Sliding Correlator Channel Sounder | USRP B210
Owner | Access | License	Arizona State University | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/statistical-characterization-28ghz-v2x-channels-autonomous-beam-steered-measurements

3.1.61 Cooperative Localization using CARLA-SUMO-Artery simulators

The Cooperative Localization using CARLA-SUMO-Artery simulators [AP23] contains data for assessing cooperative localization algorithms using realistic driving patterns from multiple vehicles in the CARLA simulator, alongside V2V communication and network quality conditions. The dataset includes trajectories of 60 vehicles over 3,000-time instances in a Manhattan-like town layout, with detailed driving parameters and communication metadata. The data is provided in a format compatible with CARLA, SUMO, and Artery simulators, enabling comprehensive analysis and simulation of vehicular communication scenarios.

Dataset ID | Name	**DAT 061 | Cooperative Localization using CARLA-SUMO-Artery simulators**
Key Features	• Contains trajectories of 60 vehicles over 3000-time instances • Includes driving parameters such as 3D position, rotation, velocity • Utilizes CARLA simulator's Town01 for a Manhattan grid-like topology • Provides V2V communication data with transmitting power and middleware update intervals • Includes local dynamic maps with transmitted messages from neighbouring vehicles
Quick Overview	• Enables assessment of cooperative localization algorithms using realistic vehicle movement patterns • Facilitates analysis of V2V communication and network quality conditions • Supports simulation studies involving CARLA, SUMO, and Artery simulators • Suitable for evaluating the impact of driving patterns and communication parameters on localization accuracy
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	IOT | Vehicular Technologies (V2V Communication)
Data Type	Trajectory and communication data from simulated vehicular networks
Data Size	0.215
Transmitter | Receiver	Simulated vehicles in CARLA with V2V communication | Simulators CARLA, SUMO, Artery

Owner | Access | License	Piperigkos & Anagnostopoulos | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/cooperative-localization-using-carla-sumo-artery-simulators

3.1.62 RIS_CE

The RIS_CE dataset [Xia23] contains synthetic data for channel estimation in reconfigurable intelligent surface (RIS)-aided multi-user mmWave massive MIMO systems. It is designed to evaluate the performance of a multi-scale attention-based channel estimation framework. The dataset models RIS with 256 elements (16×16), a base station (BS) with 64 antennas (8×8), and 6 users transmitting pilots with a length of 32 symbols at a 28 GHz frequency. The dataset includes various channel environments, considering hardware imperfections and time-varying channel characteristics, providing high-dimensional cascaded channel data for testing advanced models like the Laplacian pyramid attention network (LPAN) and its lightweight variant LPAN-L.

Dataset ID | Name	**DAT 062 | RIS_CE**
Key Features	• Synthetic data for RIS-aided multi-user mmWave MIMO channel estimation • Models a BS with 64 antennas, RIS with 256 elements, and 6 users • Contains channel data under different scenarios considering hardware imperfections and time-variance • Designed for evaluating LPAN and LPAN-L models with limited pilot overhead • Includes training, validation, and test datasets in RAR format
Quick Overview	• Supports evaluation of multi-scale attention-based channel estimation algorithms • Useful for studying spatial correlations of cascaded channels in RIS environments • Enables testing of lightweight LPAN-L models with lower computational complexity
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	RIS | SDR
Data Type	Synthetic channel estimation data
Data Size	5
Transmitter | Receiver	6 users transmitting pilots with 32 symbols each | BS with 64 antennas and RIS with 256 elements
Owner | Access | License	Jian Xiao | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/risce

3.1.63 Reconfigurable Intelligent Surface (RIS) benchmarking results and simulation code

The Reconfigurable Intelligent Surface (RIS) Benchmarking Results and Simulation Code [AR21] provides data and code for benchmarking RIS configurations, facilitating research on trade-offs in unit-cell and surface-level design. The dataset includes MATLAB files, figures, and sample unit-cell data needed to generate and evaluate RIS configurations. Key components include benchmark pattern definitions, example unit-cell data, and scripts for merging benchmark plots.

Dataset ID | Name	**DAT 063 | Reconfigurable Intelligent Surface (RIS) benchmarking results and simulation code**

Key Features	• Contains MATLAB scripts (.m), data files (.mat), figures (.png), and documentation (.txt) • Includes RIS configuration generation and benchmarking scripts • Contains predefined benchmark patterns and sample unit-cell data • Plots generated from benchmark tests provided as supplementary material • Provided in a compressed format (.zip) for easy access and integration
Quick Overview	• Enables benchmarking of RIS configurations and unit-cell designs • Facilitates the generation of custom RIS configurations for arbitrary sizes and grouped control • Supports evaluation of surface- and unit-cell-level designs under quantifiable benchmarks • Useful for generating 3D field plots and stitching multiple benchmark results into a single figure
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	RIS | SDR
Data Type	MATLAB scripts, configuration files, figures, and benchmark data
Data Size	1.32
Transmitter | Receiver	Not applicable (simulation-based)
Owner | Access | License	Ammar Rafique | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/reconfigurable-intelligent-surfaceris-benchmarking-results-and-simulation-code

3.1.64 A comprehensive dataset of RIS-based channel measurements in the 5GHz band

The A Comprehensive Dataset of RIS-Based Channel Measurements in the 5GHz Band [THW+23] contains S21 channel measurements collected using a reconfigurable intelligent surface (RIS) prototype system operating in the 5 GHz band. The dataset covers various configurations of antennas and RIS placements, including specular and non-specular reflection angles, as well as measurements with the RIS mounted on a rotating stage. Measurements were taken in a fully anechoic chamber to minimize multipath effects and optimize measurement accuracy. The dataset, provided in MATLAB .mat format, can support research in machine learning, wireless communications, and performance optimization for RIS-assisted communication links.

Dataset ID | Name	**DAT 064 | A Comprehensive dataset of RIS-based channel measurements in the 5GHz band**
Key Features	• S21 measurements collected from RIS-based systems • Includes measurements with various angles and distances between antennas and RIS • Covers specular and non-specular reflection scenarios • Data taken in a fully anechoic chamber to reduce environmental interference • Provided in MATLAB .mat format
Quick Overview	• Enables analysis of RIS-assisted wireless communication performance • Facilitates machine learning research using real-world channel measurements • Supports comparison of iterative algorithms and model-based approaches for RIS optimization

	• Useful for evaluating received power optimization and attenuation minimization strategies
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	RIS | SDR
Data Type	Channel measurement data (S21)
Data Size	1.21
Transmitter | Receiver	Horn antennas (AInfo LB-187-15) | PicoVNA 106 (Vector Network Analyzer)
Owner | Access | License	Ruhr University Bochum, TH Cologne - University of Applied Sciences, Cologne | IEEE Membership | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/comprehensive-dataset-ris-based-channel-measurements-5ghz-band

3.1.65 Dataset for Channel Estimation in RIS-assisted Satellite IoT Communications

The Dataset for Channel Estimation in RIS-assisted Satellite IoT Communications [TKK+23] contains data supporting the evaluation of Graph Attention Network (GAT)-based channel estimation in Direct-to-Satellite (DtS) IoT networks with reconfigurable intelligent surfaces (RIS). The dataset provides training and testing data to explore RIS configurations aimed at reducing path loss and computational overhead. It enables researchers to compare the performance of the GAT model with conventional deep learning techniques under dynamic conditions, ensuring robustness with lower complexity. The dataset is available in .csv and .mat formats, facilitating easy integration with machine learning models and numerical simulations.

Dataset ID | Name	**DAT 065 | Dataset for Channel Estimation in RIS-assisted Satellite IoT Communications**
Key Features	• Training and testing data for channel estimation in RIS-based satellite IoT networks • Provided in .csv and .mat formats • Supports comparisons between GAT and conventional deep learning methods • Contains multiple RIS configurations for robust performance evaluation
Quick Overview	• Facilitates research on energy-efficient IoT communication using RISs • Allows testing of GAT-based models for channel estimation • Useful for studying the impact of RIS configurations on channel conditions
Threat Coverage	Spoofing Tampering Information Disclosure
6G -TE | Tech-Domain	RIS, IOT | Cellular Technologies
Data Type	Tabular and structured data in .csv and .mat formats
Data Size	0.45
Transmitter | Receiver	Satellite Transceiver & RIS enables Satellite Systems | IoT Nodes with Limited Computational Capacity
Owner | Access | License	Tekbiyik et al. | Public | CC BY 4.0

The dataset is accessible via: https://ieee-dataport.org/documents/dataset-channel-estimation-ris-assisted-satellite-iot-communications

3.1.66 Bluetooth Wearable Device Dataset

The Wearable Device Bluetooth/BLE Physical Layer Dataset [RTD+24] is a valuable dataset containing Bluetooth/BLE radio recordings from 32 different wearable devices. Captured in an RF isolated environment using software-defined radio (SDR), it offers comprehensive insight into Bluetooth communication across a wide variety of devices. The dataset includes raw complex I/Q data files with detailed YAML metadata for each recording, describing device details, session parameters, and event timelines. Channelized versions of the recordings are provided in both 25 MHz and 5 MHz segments, and demodulated data files are saved in JSON format, capturing decoded Bluetooth packets for detailed analysis.

Dataset ID | Name	**DAT 066 | Bluetooth Wearable Device Dataset**
Key Features	• Raw complex I/Q samples of Bluetooth/BLE signals • Channelized data in 25 MHz and 5 MHz bandwidth segments • Metadata in YAML format detailing device parameters, timestamps, and setup conditions • JSON files containing demodulated Bluetooth data per recording
Quick Overview	• Provides insights into Bluetooth communication phases in isolated RF conditions • Supports physical layer analysis and device fingerprinting based on RF signatures • Useful for studying Bluetooth packet structure, event timing, and protocol behaviour • Suitable for wireless security analysis on a range of Bluetooth/BLE-enabled wearable devices
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	IOT | Short-Range Communication (Bluetooth/BLE)
Data Type	Complex I/Q samples and JSON metadata
Data Size	4600
Transmitter | Receiver	Wearable Bluetooth/BLE devices (various) | Ettus Research USRP X310 with CBX-120 daughterboard
Owner | Access | License	University of Latvia | Public | CC BY 4.0

The dataset is accessible via: https://github.com/edi-riga/Wearable_device_dataset/tree/main

3.1.67 DICHASUS Massive MIMO CSI Dataset Collection

The DICHASUS (DIstributed CHAnnel Sounder by University of Stuttgart) Massive MIMO CSI Dataset Collection [EGD+21] comprises various channel state information (CSI) datasets tailored for machine learning applications. These datasets cover numerous scenarios, including line-of-sight (LoS), non-line-of-sight (NLoS), indoor, outdoor, collocated, and distributed antenna configurations, complete with "ground truth" positioning labels. The collection includes multiple datasets captured in diverse environments, such as industrial areas, hallways, and outdoor campuses, using configurations of up to 64 antennas.

DICHASUS leverages an over-the-air (OTA) synchronization protocol to maintain precise alignment across receivers in terms of frequency, time, and phase, ensuring phase-coherent channel measurements even when receivers are spread over extensive areas. This robust synchronization is crucial for generating repeatable and environment-specific results, as the system achieves highly accurate clock synchronization across distributed receivers. Moreover, "ground truth" position labels for each dataset are meticulously recorded using advanced tools like differential GNSS, tachymeters, and LiDAR sensors, achieving centimetre-level or millimetre-level accuracy.

Dataset ID | Name	**DAT 067 | DICHASUS Massive MIMO CSI Dataset Collection**

Key Features	• Diverse CSI measurements for LoS, NLoS, indoor, outdoor, colocated, and distributed configurations • OTA synchronization protocol ensures precise phase coherence across receivers • High positional accuracy in "ground truth" labels using differential GNSS, tachymeters, and LiDAR • Variety of environments and setups, from industrial to campus outdoor environments
Quick Overview	• Provides CSI data for machine learning applications across diverse physical settings and antenna configurations • Suitable for machine learning model training on channel state data, with high synchronization and position accuracy • Enables repeatable and stable measurements across different times and environmental conditions • Supports research on distributed antenna setups and phase-coherent signal analysis
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	D-MIMO | Short Range Communication (mmWave, M-MIMO)
Data Type	Mean SNR by position, SNR by antenna, and channel coefficients
Data Size	Multiple dataset sizes, from 5.4 to 224 GB, totalling over 700 GB
Transmitter | Receiver	Multiple configurations (varies per dataset) | Distributed and collocated arrays, synchronized using OTA protocol
Owner | Access | License	University of Stuttgart | Public | Any use requires citation to author's dataset identified by DOI

The dataset is accessible via: https://dichasus.inue.uni-stuttgart.de/

3.1.68 DeepMIMO: A Generic Deep Learning Dataset for Millimeter Wave and Massive MIMO Systems

The DeepMIMO dataset [Alk19] is a comprehensive and customizable dataset designed to support machine learning research in millimetre-wave (mmWave) and massive MIMO systems. It encompasses various ray-tracing scenarios, generated using the REMCOM Wireless InSite software [R], enabling researchers to simulate realistic channels based on specific environmental geometry, materials, and transmitter/receiver locations. The dataset is highly flexible, allowing customization of critical parameters such as the number of antennas, OFDM subcarriers, system bandwidth, and more, to suit diverse machine learning applications. Each dataset within DeepMIMO is defined by a selected ray-tracing scenario and parameter set, allowing for precise replication and benchmark comparisons across studies.

Dataset ID | Name	**DAT 068 | DeepMIMO: A Generic Deep Learning Dataset for Millimeter Wave and Massive MIMO Systems**
Key Features	• Includes multiple datasets based on parameterized ray-tracing scenarios • Customizable parameters: antenna configuration, OFDM subcarriers, number of channel paths • Channels modelled with environment-dependent data capturing geometry and material effects • Generated with REMCOM Wireless InSite for high fidelity in simulating real-world channel characteristics • Provides channel matrices, angles of arrival/departure, path power, delays, and user location data

Quick Overview	• Supports applications in mmWave/massive MIMO beam prediction, channel estimation, and user positioning • Tailored datasets for flexible pre- and post-processing in wireless machine learning research • Enables comparative benchmarking and reproducibility of results across machine learning models • Facilitates large-scale dataset generation for training deep learning models
Threat Coverage	Spoofing, Information Disclosure, Tampering
6G -TE | Tech-Domain	RIS, mmWaves and sub-THZ, ISAC | Cellular Technologies
Data Type	Ray-traced channel data with adjustable parameters
Data Size	Variable, based on selected scenario and parameters
Transmitter | Receiver	Simulated base stations in parameterized configurations | Simulated users in specified locations within ray-tracing scenarios
Owner | Access | License	Arizona State University | Public | Citation needed

The dataset is accessible via: https://www.deepmimo.net/versions/v2-matlab/

3.1.69 RIS-Power-Measurements-Dataset

The RIS-Power-Measurements-Dataset [RMG+22] contains detailed power measurements taken in the anechoic chamber at TU Darmstadt to evaluate the performance of Reconfigurable Intelligent Surfaces (RIS) based on RF switches. The measurements were conducted with the transmitter (TX) positioned 1.1 meters from the RIS at a 33° elevation, while the receiver (RX) was placed 6.3 meters from the RIS at a -3° elevation. Each entry in the dataset provides power readings under various angles, with a horn antenna gain of 13.5 dBi, using OFDM QPSK-modulated symbols with 5 MHz bandwidth. The CSV file contains azimuth (θn) and elevation (Φn) angles for the RIS beam location, as well as the RIS_beamforming file, which includes data on RIS beamforming capabilities as antenna elements are activated. This dataset supports research into RIS configurations and their effects on signal power and beamforming.

Dataset ID | Name	**DAT 069 | RIS-Power-Measurements-Dataset**
Key Features	• Includes power measurements from RIS at specified TX-RX placements • Covers beamforming data with varied RIS configurations • Measurements in CSV format: RIS main beam (azimuth θn, elevation Φn), radiation pattern (θr) • Suitable for evaluating RF switch-based RIS performance in controlled conditions
Quick Overview	• Provides data on RIS-based beamforming and power measurements • Useful for research on signal control and beamforming in RIS setups • Supports exploration of how antenna element activation affects RIS performance
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	RIS | Cellular Technologies (LTE)
Data Type	Power measurements, beamforming configurations
Data Size	0.005
Transmitter | Receiver	SDR device with horn antenna, positioned 1.1m from RIS, transmitting OFDM QPSK-modulated symbols with 5 MHz bandwidth and TX power of -30 dBm per subcarrier | SDR device with horn antenna,

	positioned 6.3m from RIS, sampling Reference Signal Received Power (RSRP)
Owner | Access | License	TU Darmstadt | Public | Citation Needed

The dataset is accessible via: https://github.com/marcantonio14/RIS-Power-Measurements-Dataset

3.1.70 MaMIMO-UAV 3D Channel State Information Dataset

The MaMIMO-UAV 3D Channel State Information Dataset [CTC+23] contains channel state information (CSI) measurements from a massive MIMO system with an 8x8 rectangular patch antenna array pointing skyward, located on the campus of KU Leuven. This dataset includes CSI estimates based on pilot sequences transmitted by a drone flying various trajectories, with corresponding GPS and flight data, allowing 3D spatial analysis. The system operates with OFDM signals comprising 100 subcarriers, each with 180 kHz spacing, covering an 18 MHz total bandwidth. Data files are provided in binary and CSV formats, with supporting scripts for reading and integrating the CSI data and flight trajectory data.

Dataset ID | Name	**DAT 070 | MaMIMO-UAV 3D Channel State Information Dataset**
Key Features	• Contains CSI estimates between a base station and UAV, captured via an 8x8 antenna array • Includes 100 subcarriers per OFDM signal at 180 kHz subcarrier spacing (18 MHz total bandwidth) • Binary files include timestamp, antenna count, layer count, symbol count, subsample rate, frequency, and gain information • GPS and drone metrics (e.g., speed, height) available in CSV files • Two Python scripts for parsing and integrating channel and trajectory data
Quick Overview	• Enables 3D spatial analysis of MaMIMO-UAV channel states • Facilitates study of non-stationary UAV-to-MIMO base station channels • Supports evaluation of UAV trajectory impact on channel behaviour • Useful for testing channel estimation models and 3D channel modelling
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	None | Cellular Technologies (LTE-based M-MIMO), Vehicular Technologies
Data Type	Channel state information (binary), GPS and drone metrics (CSV)
Data Size	7.2
Transmitter | Receiver	UAV transmitting pilot sequences | 8x8 patch antenna array on MaMIMO base station
Owner | Access | License	KU Leuven | Public | CC-BY-NC-4.0

The dataset is accessible via: https://rdr.kuleuven.be/dataset.xhtml?persistentId=doi:10.48804/0IMQDF

3.1.71 Ultra Dense Indoor MaMIMO CSI Dataset

The Ultra Dense Indoor MaMIMO CSI Dataset [DP21] contains thousands of Channel State Information (CSI) samples collected using the KU Leuven Massive MIMO testbed with 64 antennas, focusing on four distinct antenna array topologies: URA (Uniform Rectangular Array) LoS (Line-of-Sight), URA NLoS (Non-Line-of-Sight), ULA (Uniform Linear Array) LoS, and DIS (Distributed) LoS. Each sample is spatially labeled with sub-millimetre accuracy, covering a 9 m² area in 5 mm increments. A total of 252,004 samples are provided per topology, resulting in one of the largest open datasets available for measured MaMIMO CSI data. This dataset facilitates extensive analysis, including visualizing precoders and validating positioning algorithms.

Dataset ID | Name	DAT 071 | Ultra Dense Indoor MaMIMO CSI Dataset
Key Features	• Thousands of Channel State Information (CSI) samples from a 64-antenna Massive MIMO (MaMIMO) testbed • Includes four antenna array topologies: URA (Uniform Rectangular Array) LoS, URA NLoS, ULA (Uniform Linear Array) LoS, and DIS (Distributed) LoS • Positioning accuracy less than 1 mm using CNC (Computer Numerical Control) tables • Each .npy file contains a 64x100 complex CSI matrix, recorded across a 9 m² grid with 5 mm spacing • Provides spatial labels for each sample, with measurements at 2.61 GHz over 20 MHz
Quick Overview	• Allows in-depth study of MaMIMO (Massive MIMO) scenarios with diverse antenna topologies • Facilitates visualization of spatially labelled CSI samples, with high positional accuracy • Supports validation of MaMIMO precoder algorithms and positioning studies • Enables analysis of user influence on CSI in real-time, with time series data from a moving individual and static references
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	D-MIMO | SDR
Data Type	Complex CSI data
Data Size	17.87 GB
Transmitter | Receiver	64 mMIMO antenna | 4 single-antenna receivers
Owner | Access | License	KU Leuven ESAT Networked Systems | Public | CC-BY

The dataset is accessible via: https://ieee-dataport.org/open-access/ultra-dense-indoor-mamimo-csi-dataset

3.1.72 SoftNull

The SoftNull dataset [ESZ+16] was created to support the development of digital beamforming-based self-interference reduction for massive MIMO systems, enabling full-duplex behaviour without the use of analog cancellers. Data was collected in three distinct environments—Anechoic Chamber, indoor (rich scattering), and outdoor—at NASA facilities. Each environment setup involves a system with 80 base station antennas (a 72-element planar array and an 8-element circular array) and four single-antenna users, totalling 84 antennas across five nodes. Channel responses were measured by having each antenna transmit pilots sequentially while others received, yielding an 84 x 84 matrix per subcarrier per packet. Each .mat file includes data for 25 packets, structured with experiment settings and channel response matrices.

This dataset contributes to the RENEW (Reconfigurable Eco-system for Next-generation End-to-end Wireless) project [DBZ+18], which aims to establish the world's first fully programmable and open-source massive MIMO platform. RENEW enables programmable, large-scale wireless experimentation and data collection for innovative wireless applications and protocols, supporting advancements from the PHY layer to network stacks.

Dataset ID | Name	DAT 072 | SoftNull
Key Features	• Channel response matrices captured from 84 antennas in 3 environments • Includes data from the Anechoic Chamber, indoor (rich scattering), and outdoor setups • 84x84 channel response matrix per subcarrier and packet, representing comprehensive antenna interactions

	• Each .mat file provides data for 25 packets, structured by experiment settings and packet-specific channel responses
Quick Overview	• Supports the study of self-interference mitigation in massive MIMO with full-duplex capabilities • Allows assessment of the SoftNull technique across varied environments • Provides data for developing beamforming techniques that leverage large antenna arrays
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	None | SDR
Data Type	Channel response matrices (.mat files)
Data Size	37.5
Transmitter | Receiver	84-antenna setup: 72 planar antennas, 8 circular antennas, 4 single-antenna users | 84-antenna setup (rotating transmit/receive antennas for response capture)
Owner | Access | License	RENEW | Public | Citation Required

The dataset is accessible via: https://renew.rice.edu/dataset-softnull.html

3.1.73 Argos Channel Survey

The Argos Channel Survey [SDG+16] dataset provides a comprehensive collection of many-antenna MU-MIMO channel measurements. This dataset includes high time-frequency resolution channel traces across UHF, 2.4 GHz, and 5 GHz bands, captured in diverse environments with configurations supporting up to 104 base-station antennas and up to 8 simultaneous users. Data was collected both indoors and outdoors on the Rice University campus, with additional mobile measurements taken using a CineMoco track system. The dataset is available in MongoDB format, accompanied by metadata including device configuration, environmental context, and mobility patterns.

Dataset ID | Name	**DAT073 | Argos Channel Survey Dataset**
Key Features	• High-resolution complex channel measurements for MU-MIMO analysis • Covers UHF, 2.4 GHz, and 5 GHz frequency bands • Data from up to 104 base-station antennas and 8 simultaneous users • Measurements across indoor, outdoor, and mobile environments on Rice University campus • Supports diverse mobility profiles with CineMoco tracking data
Quick Overview	• Enables in-depth MU-MIMO channel analysis in diverse, realistic conditions • Suitable for evaluating channel models and antenna array configurations • Supports mobility-based analysis with synchronized tracking • Allows examination of frequency and spatial diversity across multiple frequency bands
Threat Coverage	Tampering, Information Disclosure
6G -TE | Tech-Domain	None | Cellular Technologies, Short-Range Communication
Data Type	Complex channel state information (CSI) measurements
Data Size	257 GB
Transmitter | Receiver	Multiple mobile and stationary MU-MIMO transmitters | Up to 104-antenna array base-station

Owner | Access | License	RENEW | Public | Citation Required

The dataset is accessible via: https://renew.rice.edu/dataset-argos.html

3.1.74 FDD Massive MIMO

The FDD Massive MIMO Dataset [ZZS18, DS21] contains channel state information (CSI) data collected using a 64-antenna base station operating in the 2.4 GHz ISM band, with channels separated by 72 MHz. Data was recorded in both indoor and outdoor environments on the Rice University campus, covering a range of scenarios: 8 indoor line-of-sight (LOS), 16 indoor non-line-of-sight (NLOS), 4 outdoor LOS, and 21 outdoor NLOS mobile node locations. Each location captures up to 5,000 frames outdoors and 250 frames indoors, enabling thorough analysis of FDD Massive MIMO channel characteristics under varied conditions. Data is structured in several dimensions and includes LOS/NLOS conditions, multiple subcarrier and antenna measurements, and user index configurations.

Dataset ID | Name	**DAT 074 | FDD Massive MIMO**
Key Features	• Channel measurements from 64 antennas across two ISM channels (2.4 GHz) • Includes indoor/outdoor data with LOS and NLOS conditions • Contains up to 5000 frames for outdoor locations and 250 frames for indoor locations • Organized by location (total of 49 distinct mobile node locations) and environment type • Detailed per-frame metadata: subcarrier number, antenna index, user index, and calibration configurations
Quick Overview	• Enables analysis of channel characteristics across multiple LOS/NLOS scenarios • Provides high-resolution CSI data for use in FDD Massive MIMO simulations and experiments • Useful for developing and validating angle-correlation models for user interference scenarios • Offers extensive raw and processed data formats for flexible experimentation
Threat Coverage	Spoofing, Information Disclosure, Tampering
6G -TE | Tech-Domain	None | SDR
Data Type	Channel state information (CSI) data
Data Size	75
Transmitter | Receiver	64-antenna base station operating at 2.4 GHz | Mobile nodes positioned across LOS/NLOS environments
Owner | Access | License	RENEW | Public | Citation Required

The dataset is accessible via: https://renew.rice.edu/dataset-fdd.html

3.1.75 Multi-User MIMO Dataset

The Multi-User MIMO Dataset [DS21] contains measurements collected using the Argos V2 platform [SYZ13] at Rice University. This dataset includes mobile channels recorded to a 64-antenna massive MIMO base station in outdoor environments. Measurements were taken across 4 Line-of-Sight (LoS) and 5 Non-Line-of-Sight (NLoS) clusters in near-stable conditions, covering over 225 unique user locations. For each location, the dataset includes channels measured across 52 data sub-carriers and up to 300 frames in the 2.4-GHz ISM band. The dataset comprises two main files: one HDF5 file with the primary data and a ZIP file containing processing scripts.

Dataset ID | Name	DAT 075 | Multi-User MIMO Dataset
Key Features	• Collected with a 64-antenna massive MIMO base station • Mobile user channels measured in 4 LoS and 5 NLoS clusters • Includes data from over 225 unique locations • Contains measurements for 52 data sub-carriers and up to 300 frames per location • Frequency band: 2.4 GHz ISM
Quick Overview	• Enables analysis of multi-user MIMO channels under LoS and NLoS conditions • Suitable for testing massive MIMO channel models and validating user correlation effects • Useful for training and validating machine learning models in wireless communications
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	None | SDR
Data Type	HDF5 files (channel data), ZIP (processing scripts)
Data Size	18 GB
Transmitter | Receiver	Multi-antenna base station (64 antennas) | Mobile users in clustered environments
Owner | Access | License	RENEW | Public | Citation Required

The dataset is accessible via: https://renew.rice.edu/dataset-iuc.html

3.1.76 Full-Duplex Massive MIMO

The Full-Duplex Massive MIMO [DBZ+18] dataset was collected in an anechoic chamber setup on the POWDER platform, focusing on self-interference measurement across all radios on a massive MIMO base station with 48 antennas. In full-duplex systems, self-interference—where outgoing transmissions interfere with incoming signals—is a significant challenge. To capture these effects, the RENEW team at the University of Utah used their Sounder tool, by continuously transmitting pilot signals from each antenna in a round-robin manner, while the other antennas were used in listening mode. This dataset provides valuable insights into self-interference characteristics, with data files organized to include raw I/Q samples and relevant metadata.

Dataset ID | Name	DAT 076 | Full-Duplex Massive MIMO
Key Features	• Collected in an anechoic chamber on the POWDER platform • Self-interference measurement across a 48-antenna massive MIMO base station • Raw I/Q samples and metadata from pilots sent by each antenna • Organized into multiple data files, each covering various antenna chains and configurations
Quick Overview	• Enables analysis of self-interference in a full-duplex massive MIMO environment • Provides baseline data for evaluating interference mitigation techniques • Useful for studying the impact of antenna configuration on interference and performance
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	None | Cellular (Massive MIMO - 5G)
Data Type	Raw I/Q samples with metadata

Data Size	46
Transmitter | Receiver	48-antenna massive MIMO base station | Same 48-antenna setup (full-duplex mode)
Owner | Access | License	RENEW | Public | Citation Required

The dataset is accessible via: https://renew.rice.edu/dataset-fullduplex.html

3.1.77 Angle-of-Arrival for Massive MIMO

The Angle-of-Arrival for Massive MIMO [Mir22] dataset was collected by Joshua Miraglia at the University of Utah within the anechoic chamber of the POWDER platform. This dataset supports research on signal classification in a massive MIMO system, capturing signals transmitted by two UEs (user equipment) to a 48-antenna base station across an 80-point grid representing different angles (8 azimuth and 10 elevation positions). The primary purpose is to enable deep learning model training for angle-of-arrival (AoA) estimation. The data includes parameters like modulation order, signal bandwidth, center frequency, and SNR proxy values, provided as HDF5 files for each location.

Dataset ID | Name	**DAT 077 | Angle-of-Arrival for Massive MIMO Dataset**
Key Features	• Contains raw I/Q samples from MIMO transmissions • 80-point grid with 8 azimuth and 10 elevation angles • Data collected from a 48-antenna MIMO base station with 2 UEs • Includes metadata: modulation order, signal bandwidth, center frequency, TX gain (SNR proxy), and actual coordinates • Provided in HDF5 format for structured data storage
Quick Overview	• Facilitates analysis and training for signal classification models in massive MIMO systems • Captures realistic AoA scenarios with grid-based signal variation • Supports deep learning applications for MIMO signal analysis • Useful for studying effects of different signal parameters and client locations on received signal characteristics
Threat Coverage	Spoofing, Information Disclosure, Tampering, Repudiation
6G -TE | Tech-Domain	Positioning and Sensing
Data Type	I/Q samples and Power Spectral Densities (PSDs)
Data Size	4.3 TB
Transmitter | Receiver	48 Antenna Base Station | 2 Iris SDR Devices
Owner | Access | License	RENEW | Public | Citation Required

The dataset is accessible via: https://renew.rice.edu/dataset-aoa.html

3.1.78 Coherent vs. Non-Coherent MU-MIMO with Uplink Data

The Coherent vs. Non-Coherent MU-MIMO with Uplink Data [DZS22] dataset provides experimental data for analysing distributed MIMO (D-MIMO) systems under both synchronized and non-synchronized configurations. Collected using the RENEW platform, the dataset includes uplink pilot and OFDM data symbols generated from random bits. Data was gathered from a 64-antenna RENEW massive MIMO base station in an indoor lab environment with up to 8 user clients (4 dual-polarized antennas).

This dataset emulates both centralized and distributed MIMO scenarios: in the coherent mode, the antennas are fully synchronized in time and frequency, representing a typical centralized massive MIMO system. In the non-coherent mode, each Iris radio antenna in the base station operates with its own local oscillator (LO), introducing residual frequency and time offsets that simulate a D-MIMO environment. These offsets mimic real-world D-MIMO conditions, where synchronization occurs over-the-air rather than through a central clock.

The dataset covers multiple configurations (64x1, 64x2, 64x4, and 64x8) and modulation schemes (QPSK for all datasets, with some 16-QAM samples available) across two channel conditions: Line-of-Sight (LoS) and Non-Line-of-Sight (NLoS). This setup allows for detailed experimentation with D-MIMO-specific challenges, such as mitigating carrier frequency offset (CFO) and handling independent antenna synchronization.

Dataset ID | Name	**DAT 078 | Coherent vs. Non-Coherent MU-MIMO with Uplink Data**
Key Features	• Uplink pilot and OFDM data symbols generated from random bits • 64-antenna RENEW massive MIMO base station with up to 8 user clients • Two channel conditions: Line-of-sight (LoS) and Non-line-of-sight (NLoS) • Covers configurations 64x1, 64x2, 64x4, and 64x8 • QPSK modulation for all datasets; 16-QAM available for some datasets • HDF5 format with data processing instructions available on RENEW Wiki
Quick Overview	• Enables study of distributed beamforming in time and frequency offset conditions • Supports testing MU-MIMO detectors under coherent and non-coherent scenarios • Suitable for analysis of residual frequency and time offsets in distributed arrays • Provides realistic indoor lab data with controlled LoS and NLoS conditions
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	D-MIMO | SDR
Data Type	Pilot and data OFDM modulated signals
Data Size	175 GB
Transmitter | Receiver	Up to 8 user clients (4 dual-polarized antennas) | 64-antenna RENEW massive MIMO base station
Owner | Access | License	RENEW | Public | Citation Required

The dataset is accessible via: https://renew.rice.edu/dataset-cfo.html

3.1.79 Experimental Evaluation of AoA Estimation for UAV to Massive MIMO

The Experimental Evaluation of AoA Estimation for UAV to Massive MIMO [Ric22] dataset contains measurements collected to study the AoA estimation between a UAV and a Massive MIMO system. Collected by Tarence Rice from Rice University at the university's football stadium, this dataset includes channel sounding measurements from uplink packets sent from a drone to a Massive MIMO array. The drone, controlled via the ASTRO autonomous framework with an integrated RENEW SDR, hovered at five distinct GPS-mapped locations at a 20-meter altitude. Three antenna positions on the base station were missing data due to equipment limitations; these were zeroed out during processing to maintain array structure, introducing only minimal error (1-2 degrees).

Dataset ID | Name	**DAT 079 | Experimental Evaluation of AoA Estimation for UAV to Massive MIMO**
Key Features	• Collected measurements of UAV-to-Massive MIMO channels • Signal data was recorded for 5 drone locations with GPS coordinates and azimuth angles relative to the base station

	• RENEW-Argos V3 Massive MIMO setup used, with 3.6 GHz signal frequency, 5 MHz sample rate, QPSK modulation • Includes metadata on array geometry, antenna spacing, and schedules for base station, reference, and drone nodes • Data in 2240x74x4x1x2001 format, with 1120 samples at 45° and -45° polarity
Quick Overview	• Enables analysis of AoA estimation in drone-to-MIMO scenarios • Facilitates testing of Massive MIMO channel modeling and spatial analysis using real-world UAV data • Useful for evaluating the impact of missing antennas on array structure and AoA accuracy • Matlab processing files included for dataset handling
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	IOT | Vehicular Technologies (UAV)
Data Type	Complex channel I/Q samples from UAV-to-MIMO
Data Size	13
Transmitter | Receiver	Drone with ASTRO framework and RENEW SDR | Argos V3 Massive MIMO base station
Owner | Access | License	RENEW | Public | Citation Required

The dataset is accessible via: https://renew.rice.edu/dataset-drone.html

3.1.80 LensFD

The LensFD [CBV+23] dataset was collected as part of a project comparing performance across four lens configurations at a massive MIMO base station. The configurations tested include: no lens (baseline), small lens array, medium lens array, and large lens array. Each configuration was applied to an 80-antenna Uniform Planar Array (UPA) at a CBRS-band (3.5 GHz) base station consisting of 40 radios, with each radio supporting two independent antenna ports due to slant-linear polarization. This dataset includes both raw IQ samples and processed Channel State Information (CSI) and captures data from both operational radios and those experiencing hardware failure, across various indoor and outdoor environments.

Dataset ID | Name	**DAT 080 | LensFD Dataset**
Key Features	• Four lens configurations applied to massive MIMO at 3.5 GHz • Includes both raw IQ samples and processed CSI data • Data collected from self-interference and array-to-client channels across multiple environments • Captures both indoor (Duncan Hall) and outdoor (Rice University Stadium) data
Quick Overview	• Enables study of lens array impact on massive MIMO performance • Allows direct comparison of four lens configurations at the MIMO base station • Facilitates evaluation of self-interference channels and static/mobile client channels • Useful for understanding full-duplex, sub-6 GHz massive MIMO performance
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	None | SDR
Data Type	Raw IQ samples and Channel State Information (CSI)
Data Size	18 GB

Transmitter | Receiver	80-antenna UPA at 3.5 GHz | 40 radio chains with dual-polarized antenna ports
Owner | Access | License	RENEW | Public | Citation Required

The dataset is accessible via: https://renew.rice.edu/dataset-lens.html

3.1.81 Indoor Mobility Channel Measurement for Massive MIMO

The Indoor Mobility Channel Measurement for Massive MIMO [ASD+23] dataset consists of massive MIMO channel measurements conducted in an indoor setting on the Rice University campus. The experiment used a 64-antenna RENEW software-defined massive MIMO base station and 7 software-defined clients in a large open area. Six clients were arranged in a fixed circle, 15m from the base station, while the seventh client was mobile, mounted on a robot moving at varying speeds (0.5m/s, 1m/s, and 2m/s). Both Line-of-Sight (LoS) and Non-Line-of-Sight (NLoS) measurements were collected, with uplink pilots based on the IEEE 802.11 LTS OFDM signal across 52 non-zero subcarriers. The dataset, collected on November 20, 2022, is divided into six files with different configurations and movement speeds, totalling 33.5 GB.

Dataset ID | Name	**DAT 081 | Indoor Mobility Channel Measurement for Massive MIMO**
Key Features	• Massive MIMO channel measurements collected with 64-antenna base station and 7 clients • 6 fixed clients arranged in a circle 15m from base station, 1 mobile client moving at various speeds • Measurements include both LoS and NLoS conditions • Uplink pilots based on 802.11 LTS OFDM signal with 52 non-zero subcarriers • Data provided in HDF5 format with visualization tools available in RENEWLab
Quick Overview	• Suitable for deep reinforcement learning model training for massive MIMO networks • Enables analysis of MIMO channel characteristics under various mobility and LoS/NLoS scenarios • Useful for developing and testing resource scheduling algorithms in massive MIMO
Threat Coverage	Spoofing, Tampering, Information Disclosure
6G -TE | Tech-Domain	None | Short Range Communication (IEEE 802.11)
Data Type	Channel measurement data in HDF5 format
Data Size	33.5 GB
Transmitter | Receiver	RENEW 64-antenna massive MIMO base station | 7 software-defined clients (6 fixed, 1 mobile)
Owner | Access | License	RENEW | Public | Citation Required

The dataset is accessible via: https://renew.rice.edu/dataset-indoor-channel.html

3.1.82 M3A

The M3A (Multipath Multicarrier Misinformation to Adversaries) [LDM+23] dataset provides physical-layer data from a multi-antenna, multicarrier OFDM/OFDMA transmission system. The setup allows a sender (Alice) to deliver data to legitimate users (Bob) while simultaneously sending misinformation to eavesdroppers (Eve), resulting in degraded symbol detection at Eve while maintaining data integrity at Bob. Data was collected in a multipath-rich lab environment, with various obstacles influencing signal quality across 20 locations. This dataset includes detailed PHY-layer metrics such as SNR, EVM, and BER for both Bob and

Eve across multiple beamforming strategies (M3A and baseline), modulation parameters, and experimental topology. The dataset is provided in HDF5 format with MATLAB analysis scripts available.

Dataset ID | Name	DAT 082 | M3A
Key Features	• Multi-antenna OFDM/OFDMA transmission system with adaptive misinformation technique • Physical layer metrics: SNR, EVM, and BER across 20 locations • Data collected for M3A and beamforming comparison over Sub-6GHz frequency • Rich multipath lab environment with multiple obstacles • HDF5 format with analysis scripts for MATLAB
Quick Overview	• Enables analysis of physical-layer security techniques and multipath effects • Facilitates comparison between M3A and baseline beamforming in terms of BER and other PHY metrics • Useful for evaluating eavesdropper resilience and legitimate receiver reliability • Provides in-depth data on system behaviour across different distances and modulation settings
Threat Coverage	Tampering, Information Disclosure
6G -TE | Tech-Domain	None | SDR
Data Type	Physical-layer performance metrics in HDF5 format
Data Size	1.2
Transmitter | Receiver	Base Station (Multi-antenna OFDM/OFDMA transmitter) | Iris SDRs configured as Bob (legitimate receiver) and Eve (eavesdropper)
Owner | Access | License	RENEW | Public | Citation Required

The dataset is accessible via: https://renew.rice.edu/dataset-m3a.html

3.1.83 Distributed Multi-user MIMO Datasets

The Distributed Multi-User MIMO Datasets [DBZ+18] were collected in a controlled indoor environment on the Rice University campus to explore the performance impact of base station (BS) array placement in co-located and distributed configurations in a massive MIMO setup. The experiment utilized the RENEW massive MIMO platform with four Iris SDR chains, totalling 64 dual-polarized antennas. Each chain functioned as an access point (AP), connected through fibre links to a central hub and operated under two modes, HUB-mode (for synchronized clocking) and LO-mode (independent clocks). Six standalone Iris SDRs, spaced on a 14-meter diameter circle, were used as user devices in both line-of-sight (LoS) and non-line-of-sight (NLoS) conditions. This dataset provides extensive uplink pilot and data transmission logs, supporting research on spatial diversity and phase coherence in multi-user massive MIMO setups.

Dataset ID | Name	DAT 083 | Distributed Multi-User MIMO Dataset
Key Features	• Contains 257 GB of multi-user MIMO channel measurements • Data captured in both co-located and distributed AP setups • Two clocking configurations: HUB-mode and LO-mode • Six Iris SDRs function as users, transmitting orthogonal pilot and data symbols • Co-located and distributed configurations with variations in LoS and NLoS channels
Quick Overview	• Enables analysis of massive MIMO configurations and channel coherence

	• Supports performance studies on user-device signal reception and phase synchronization • Useful for beamforming studies, spatial diversity analysis, and multi-user interference evaluation
Threat Coverage	Spoofing, Tampering, Repudiation, Information Disclosure
6G -TE | Tech-Domain	D-MIMO | SDR
Data Type	Multi-user MIMO channel measurement data
Data Size	257 GB
Transmitter | Receiver	Six Iris SDRs as users | RENEW base station with four Iris SDR chains (64 dual-polarized antennas)
Owner | Access | License	RENEW | Public | Citation Required

The dataset is accessible via: https://renew.rice.edu/dataset-distributed.html

3.2 Contribution of ROBUST-6G to PLS Datasets

This section focuses on the datasets developed as part of the ROBUST-6G project, addressing the gaps identified in existing datasets for physical layer security (PLS). The ROBUST-6G project aims to enhance the resilience of 6G networks by contributing specialized datasets that cover underrepresented threat categories, such as RF fingerprinting, adversarial attacks, and authentication challenges in IoT environments.

The datasets introduced here are designed to align with the project's threat matrix, addressing specific vulnerabilities in emerging technologies like Distributed MIMO (D-MIMO), Reconfigurable Intelligent Surfaces (RIS), and mmWave communication. Each dataset description outlines the purpose behind its creation, the collection methodology, and its relevance to threat mitigation. These contributions play a crucial role in advancing PLS research by filling critical gaps and setting benchmarks for future studies. Through this section, readers will understand how the ROBUST-6G datasets provide real-world, actionable data for the development of innovative security solutions and robust threat mitigation strategies for next-generation 6G networks.

3.2.1 RF Fingerprinting Migration Dataset

The RF Fingerprinting Migration Dataset is a newly generated dataset designed to address the critical issue of authentication in Internet of Things (IoT) devices. In the context of IoT, ensuring the authenticity of devices is paramount, especially when considering the low power nature of IoT sensors. A significant problem arises when these devices are required to change their gateway, leading to a necessary adaptation of their RF fingerprint. During this adaptation phase, IoT sensors typically need to send many packets to retrain the machine learning (ML) model responsible for their identification, which in turn consumes a considerable amount of battery life [ACK+21].

The main purpose of the RF Fingerprinting Migration Dataset is to facilitate the development of more efficient methods for adapting RF fingerprints to new gateways with minimal data transmission. By doing so, it aims to significantly reduce the battery consumption of IoT sensors during this migration phase. This dataset specifically targets the threat of authentication attacks in IoT networks, where attackers might exploit the migration phase to impersonate or tamper with devices.

The dataset can be accessed via https://zenodo.org/records/14436621 and will be available in 2026.

Table 3-1: RF Fingerprinting Migration Dataset Overview

Category	Details
Purpose	Adapt RF fingerprints to new gateways with minimal data transmission
Threat Addressed	Authentication attacks during IoT device migration
Devices used	3 SDRs (Two Fairwaves XTRX, one B200 Mini), 30 IoT sensors

RF Parameters	2-GFSK, 50k Baud Rate, 866 MHz, 100 ms, 25 kHz deviation
Packet Structure	16 bytes (preamble, syncword, length, SN, data, CRC)
Dataset Size	810,000 packets, ~6.18 GB

Data Collection of this dataset consists of 3 main parts. Receivers, Transmitters and the Environment.

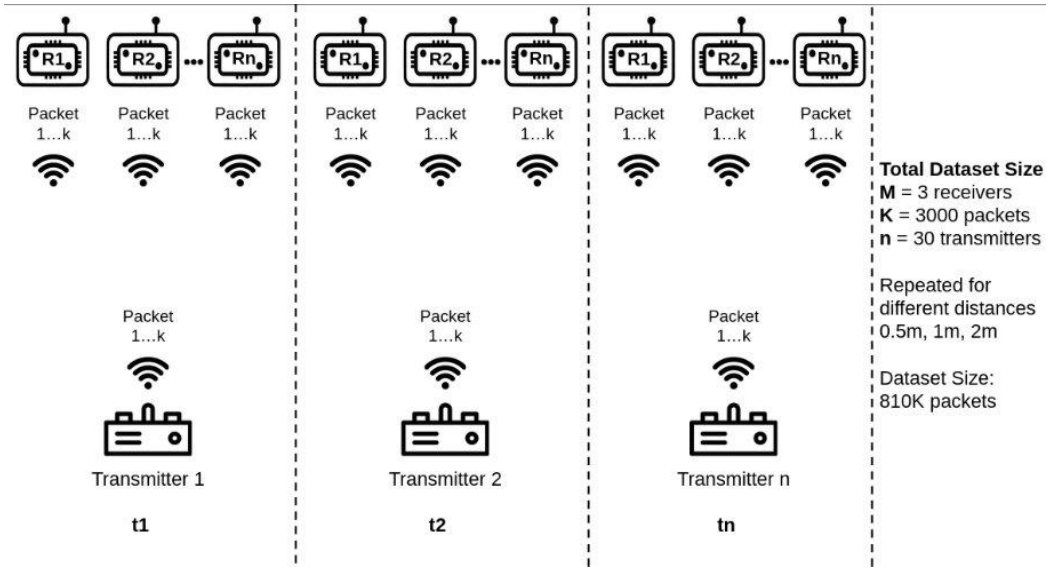

Figure 3-1: Schematic Diagram of the RF Fingerprinting Migration Dataset

Receiver: Utilizing 3 Software Defined Radios (Two Fairwaves XTRX (identical) and a single B200 Mini) placed close to each other. Each SDR is working with 20-30 seconds of intervals for at least 7-8 minutes, collecting the raw I/Q signals in each interval and identifying possible packet transmissions. Each detected packet transmission is demodulated and decoded. If the packet is demodulated and decoded successfully, it is saved according to the naming convention of this dataset. If the detected packet is not decoded correctly, then it is deleted. After each transmitter has sent 4,000 packets and the receivers have collected those packets, the packets are structured according to their Packet Numbers (Sequence Numbers). If a packet number is not captured by one of the 3 receivers, then it is not saved.

All receivers used a sampling rate of 400 KHz, while collecting the packets.

Transmitter: The transmitter side has 30 sensors developed in-house.

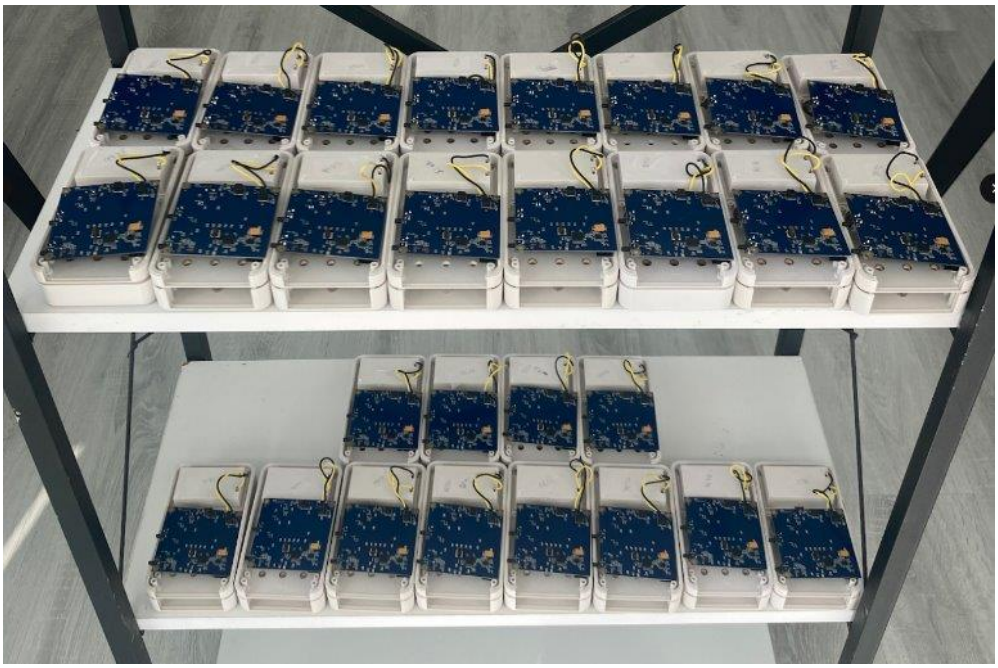

Figure 3-2: TI13XX-based designed IoT sensors used in RF Fingerprinting Migration Dataset

Table 3-2: Packet Structure of the Dataset

Packet Feature	Data	Length (Bytes)
Preamble	55 55 55 55	4
Sync word	93 0b 51 de	4
Length	05	1
Sequence Number	Various (0-4000)	2
Data	aa bb cc	3
CRC	Various	2

Environment: The lab used to generate the RF Fingerprinting Migration Dataset has 3 receivers placed next to each other. A frequency of 866 MHz is used to generate this dataset. After a long period of inspection, the frequency band was clear and empty for us to develop this dataset.

The SDR's and transmitters had 3 different distances between each other. These are 0.5 metres, 1 metre and 1.5 metres.

Figure 3-3: Data Collection Setup & Transmitter and Receiver while on operate at 1 meter

4 Assessment and Evaluation of Datasets for 6G Security

The development of secure 6G networks relies on a solid research foundation supported by diverse datasets. These datasets play a pivotal role in identifying vulnerabilities, testing mitigation strategies, and fostering innovative solutions for Physical Layer Security (PLS). This section presents a comprehensive evaluation of datasets supporting PLS research, focusing on their relevance to 6G security requirements.

The datasets analysed encompass technologies such as Wi-Fi, Bluetooth, LoRa, and millimetre waves, along with key threats like spoofing, tampering, repudiation, information disclosure and denial of service (DoS). Mapping these datasets onto technological enablers and threat categories ensures that researchers can identify strengths and gaps, by promoting the development of robust PLS solutions tailored for 6G networks. Additionally, this section highlights critical gaps, such as insufficient representation of repudiation threats and LPWAN technologies, providing direction for future research and dataset development efforts.

4.1 Overview of Dataset Collection

The collection included extensive searches across online databases, academic repositories, and scientific publications. Key criteria for selection included relevance to physical layer security, alignment with use cases from the 6G Threat Analysis Report (D2.1), dataset size, and the availability of an active community supporting the dataset. In total, the collection effort yielded 80 external datasets and 1 internal dataset. The collection effort yielded both internal datasets developed within the project and external datasets sourced from repositories such as IEEE Xplore, Zenodo, and arXiv. These datasets cover a wide array of communication

technologies, ranging from conventional technologies (Wi-Fi, Bluetooth) to emerging 6G enablers like Reconfigurable Intelligent Surfaces (RIS) and sub-THz frequencies.

This section explores the technologies and threat categories represented in the datasets, categorized based on the dual-path approach described in Section 2. The datasets span a wide range of technologies relevant to 6G, categorized into technological enablers and general communication domains:

6G Technological Enablers

- Distributed Multiple-Input Multiple-Output (D-MIMO)
- Reconfigurable Intelligent Surfaces (RIS)
- Millimetre waves (mmWaves) and Sub-THz frequencies
- Integrated Sensing and Communication (ISAC)
- Internet of Things (IoT)

General Communication Domains

- Short-Range Communication: Wi-Fi, Bluetooth, Zigbee
- Low-Power Wide-Area Networks (LPWAN): LoRa, LoRaWAN, Sigfox
- Cellular & Mobile Communication: LTE, 5G NR, C-V2X
- Vehicular Technologies: VANETs, UAVs (drones), V2X
- Software-Defined Radio (SDR) and Signal Processing

The datasets are mapped onto **threat categories** based on the STRIDE framework, focusing on five primary physical layer threats:

- **Spoofing**: Impersonation or unauthorized access.
- **Tampering**: Manipulation of signals or data.
- **Repudiation**: Lack of mechanisms to verify the source of actions.
- **Information Disclosure**: Exposure of sensitive data.
- **Denial of Service (DoS)**: Disruption or exhaustion of resources.

4.2 Mapping Results: Threats and Technologies

This section details the process of mapping the external datasets collected against the threat matrix outlined in Deliverable D2.1 - 6G Threat Analysis Report. The objective of this mapping is to evaluate how well these datasets align with the five primary physical layer threats: Spoofing, Tampering, Repudiation, Information Disclosure, and Denial of Service (DoS). By conducting this detailed mapping, we aim to establish a comprehensive understanding of how these datasets support research in Physical Layer Security (PLS) for 6G networks.

The mapping also serves to align datasets with relevant 6G technical enablers—such as Distributed Multiple-Input Multiple-Output (D-MIMO), Reconfigurable Intelligent Surfaces (RIS), mmWave frequencies, sub-THz frequencies, and IoT technologies—offering insights into the interplay between technology and threats. This multi-dimensional classification approach ensures that researchers can identify well-covered areas and recognize where additional datasets are required for comprehensive security research.

4.2.1 Mapping Results by Threat Category

The datasets were systematically examined for their relevance to the five primary physical layer threats identified in D2.1. Each dataset was assessed and categorized to understand its contribution to specific threat scenarios. The following summarizes the coverage of datasets across the threat categories:

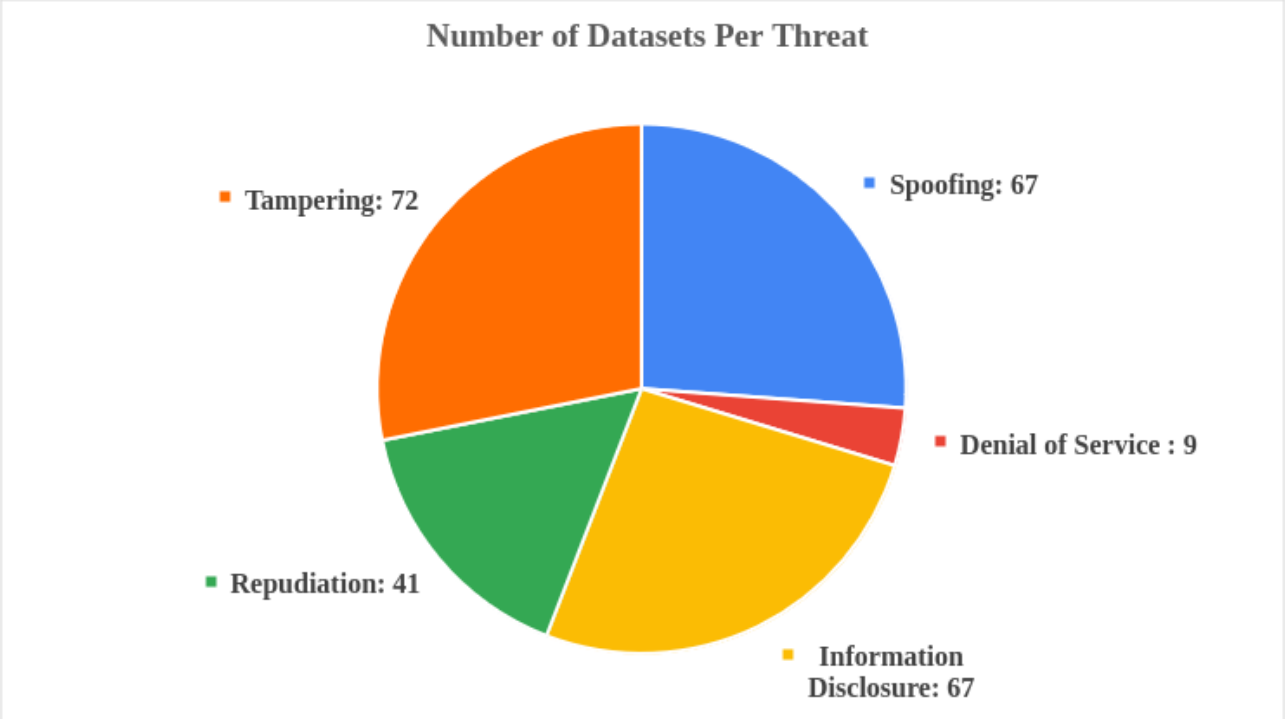

Figure 4-1: Distribution of Datasets by Threat Category

- **Spoofing:** Spoofing threats are represented by 67 datasets, showing significant coverage across multiple areas. This spread indicates a broad-based concern across different communication types, although some variance exists. The representation of spoofing in certain areas is comparatively low, suggesting these areas may require additional investigation to ensure comprehensive protection across the full spectrum of applications.
- **Tampering**: The datasets reveal that tampering threats are prominent, with 72 datasets addressing this category, making it the most documented threat. This high representation indicates substantial attention to tampering issues across various areas, suggesting these areas are either particularly susceptible to this threat or prioritized in research. However, there is a notable presence across multiple types, indicating a multi-domain concern for tampering risks. Future research might benefit from increased focus on underrepresented areas where tampering-related data seems sparse, suggesting a potential gap in threat documentation.
- **Repudiation**: With 41 datasets, repudiation threats receive moderate attention compared to other categories. The distribution of datasets suggests that this threat is more thoroughly investigated in certain domains, possibly due to the need for reliable authentication mechanisms. However, the limited data on repudiation within some other areas highlights where more investigation might be necessary to bolster accountability and non-repudiation measures across all communication types. Addressing these gaps could enhance resilience against repudiation across a broader range of platforms.
- **Information Disclosure**: Information disclosure threats, also represented by 67 datasets, highlight an area of extensive research, particularly in emerging areas. The high dataset count signals a strong focus on data confidentiality. Information disclosure is similarly well-documented, while other areas appear less addressed, pointing to potential vulnerabilities or research gaps. Expanding research in these underrepresented areas could provide a more balanced security posture across diverse systems.
- **Denial of Service (DoS)**: Denial of Service, represented by only 9 datasets, is the least documented threat in this analysis, indicating a potential under-prioritization in research relative to other threat categories. Certain areas receive the majority of the limited DoS dataset focus, reflecting particular concern for high-demand domains. Some areas show minimal representation, while others lack substantial DoS dataset coverage altogether. However, if a dataset includes relevant physical layer features—such as RSSI, SNR, packet loss, traffic volumes, or bit error rates—it can still be leveraged to study DoS impacts even without explicit DoS labelling. For example, datasets on communication or signal data that capture interference, jamming, or abnormal traffic conditions can provide valuable insights for modelling and analysing DoS scenarios. Nevertheless, given the potential disruptions DoS attacks can cause, especially in critical infrastructure and real-time communication, there remains a

pressing need for further dedicated research across all areas to ensure comprehensive resilience against service interruptions.

4.2.2 Mapping Results by Technology Domain

The datasets were systematically categorized to understand their relevance across six primary technology domains. Each dataset reflects the focus and research activity on these communication technologies, highlighting the emphasis on both emerging and established domains. Below is an analysis of the distribution of datasets across each technology domain.

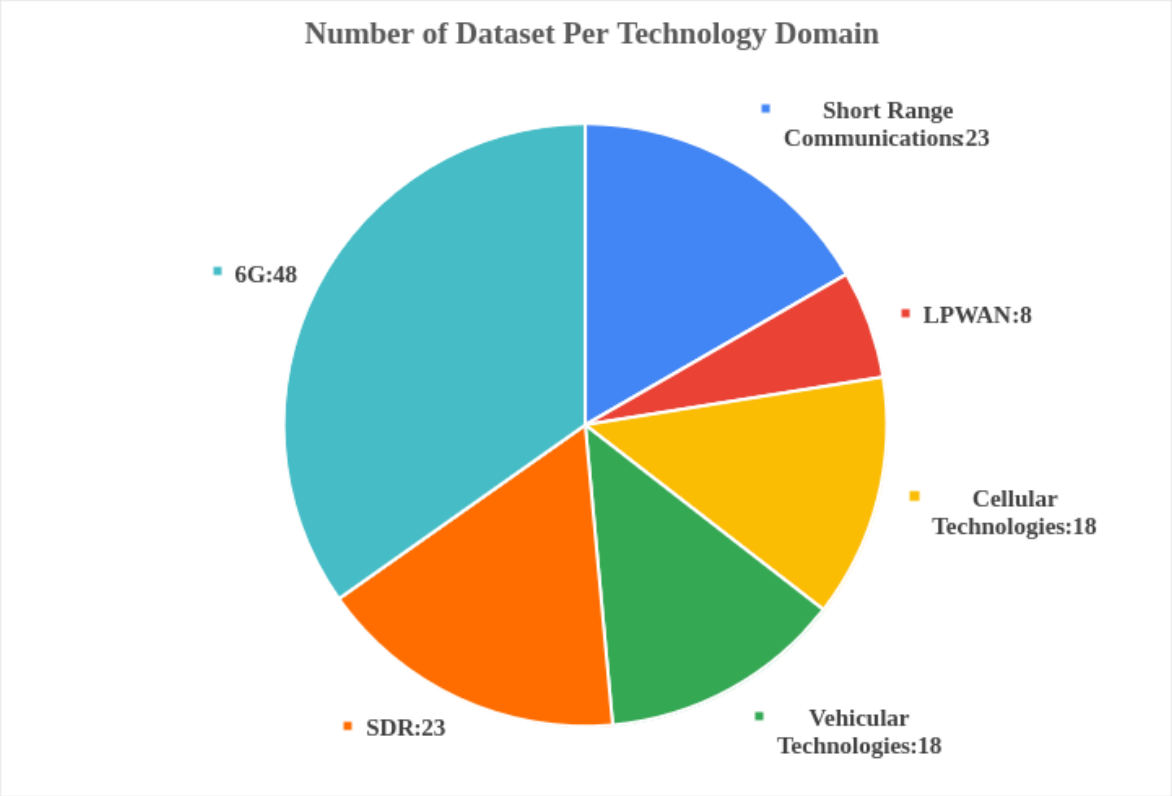

Figure 4-2: Distribution of Datasets by Threat Category

- **6G Technology:** With 48 datasets, 6G technology received the highest coverage. This includes datasets that relate to both 6G as a whole and specific 6G technical enablers, which will be discussed in more detail later in the document. The high number of datasets shows a strong interest in researching and developing next-generation communication systems. This focus on 6G suggests that it represents a key area for future innovations.
- **Short-Range Communications:** Short-range technologies, like Wi-Fi, Bluetooth, and Zigbee, are represented by 23 datasets. These technologies are essential for personal devices and systems that require low power and efficient data transfer. The existing datasets highlight important areas in short-range communications, but additional research could benefit new applications, especially on the Internet of Things (IoT), where these technologies play a big role.
- **Low-Power Wide-Area Networks (LPWAN):** With only 8 datasets, LPWAN is the least covered domain. LPWAN technologies, used in IoT and remote sensing for long-range, low-power communications, appear to have research gaps, especially regarding security and performance. More datasets focused on LPWAN could help address potential vulnerabilities and improve the reliability of these networks.
- **Cellular & Mobile Technologies:** Cellular technologies, such as LTE, 5G NR, and C-V2X, have 18 datasets. This moderate representation reflects the importance of cellular networks in mobile communication for various applications, from smartphones to smart city infrastructure. While current datasets cover key areas, more research could help tackle new challenges in mobile resilience as mobile and IoT use grows.

- **Vehicular Technologies:** Also with 18 datasets, vehicular technologies focus on communication systems for connected and autonomous vehicles. These datasets address critical needs in vehicular networks, which are important for safety and efficiency in smart and autonomous transportation. However, as these systems grow more complex, further datasets could help strengthen security and performance in vehicular communication.
- **Software-Defined Radio (SDR) and Signal Processing:** SDR and signal processing technologies are represented by 23 datasets. These flexible platforms allow researchers to experiment with different communication protocols, making them important for developing resilient communication systems. The dataset count shows a solid interest in SDR, but expanding research here could lead to stronger, more adaptable communication methods across various applications.

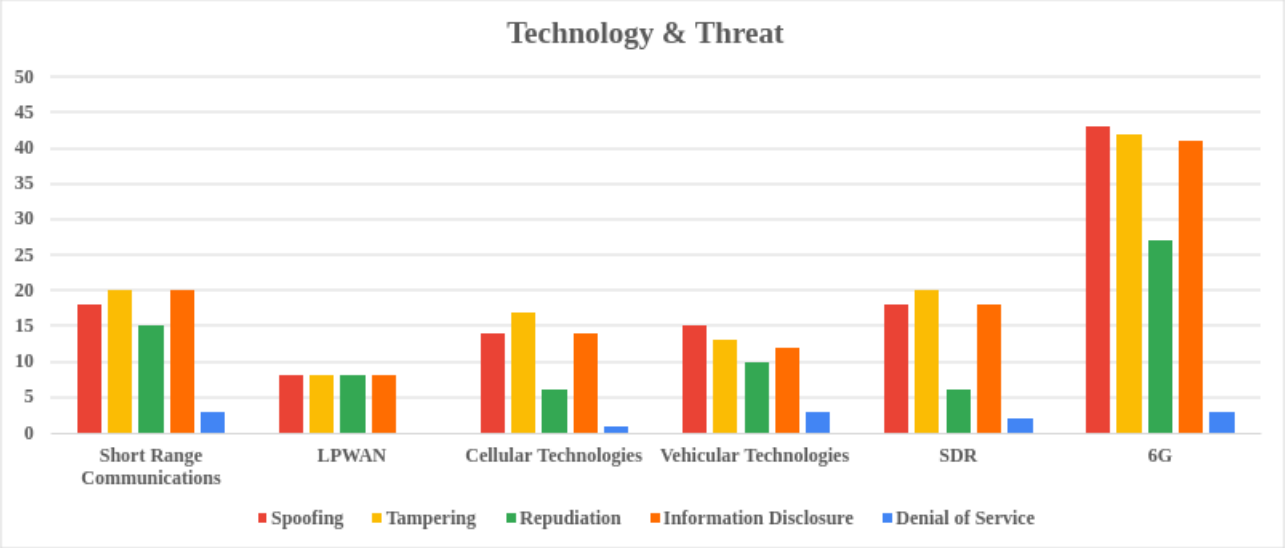

Figure 4-3: Threat Distribution Across Technologies

As seen in Figure 4-3, the results indicate that the number of mapped datasets exceeds the total count of unique datasets. This is because individual datasets can be associated with multiple threat categories and span various technology domains. In many cases, a single dataset addresses more than one type of threat, reflecting the multifaceted nature of security challenges in modern communication systems. Additionally, datasets can overlap across technological classifications, as certain data sources or studies may be applicable to more than one technological domain or advancement area. This overlap highlights the interconnected nature of emerging technologies and the shared security concerns that span across them. A detailed explanation is provided below.

- **Short-Range Communications**: Short-Range Communications show a strong focus on tampering, with 20 datasets, followed closely by spoofing with 18 datasets and information disclosure with 15 datasets. Repudiation has moderate representation at 10 datasets, while denial of service is the least represented, with only 3 datasets. This suggests that while there is substantial attention to tampering, spoofing, and information disclosure, denial of service may require further research to ensure robust security.
- **LPWAN (e.g., LoRa, Sigfox)**: LPWAN exhibits a lower overall dataset count across all threats, with each threat category—spoofing, tampering, repudiation, and information disclosure—represented by 8 datasets. Denial of service, however, has no representation, which highlights a significant gap in research on service availability for LPWAN, indicating a need for further investigation in this area.
- **Cellular & Mobile Networks (5G, LTE, C-V2X)**: In Cellular & Mobile Networks, tampering leads to 17 datasets, indicating a primary focus on integrity threats. Information disclosure follows with 14 datasets, while spoofing is represented by 14 datasets as well. Repudiation has a lower representation with 6 datasets, and denial of service is minimally covered with just 1 dataset. This distribution suggests that while tampering and data confidentiality are prioritized, denial of service remains an area with limited focus.
- **Vehicular Networks and UAVs**: Vehicular Networks have a balanced distribution, with tampering being the most represented threat at 15 datasets, followed closely by spoofing and information disclosure, each with 13 datasets. Repudiation has a moderate level of focus with 10 datasets, whereas denial of service is underrepresented, with only 3 datasets. This implies a need for additional research on denial of service to improve service resilience in vehicular networks.

- **Software-Defined Radio (SDR) and Signal Processing**: SDR shows high representation in tampering and information disclosure, with each having 18 datasets. Spoofing follows closely with 10 datasets, while repudiation has a moderate representation at 8 datasets. Denial of service is minimally addressed, with only 2 datasets. This suggests that while SDR research covers multiple threats, denial of service may require further emphasis to address potential vulnerabilities.
- **6G:** In 6G technology, spoofing, tampering, and information disclosure are particularly well-documented, with 43, 42, and 41 datasets, respectively. Repudiation has substantial coverage as well, with 27 datasets, while denial of service is the least represented, with just 3 datasets. The strong focus on spoofing, tampering, and information disclosure underscores these as primary concerns within 6G, but the low coverage for denial of service suggests an area for further exploration to enhance resilience.

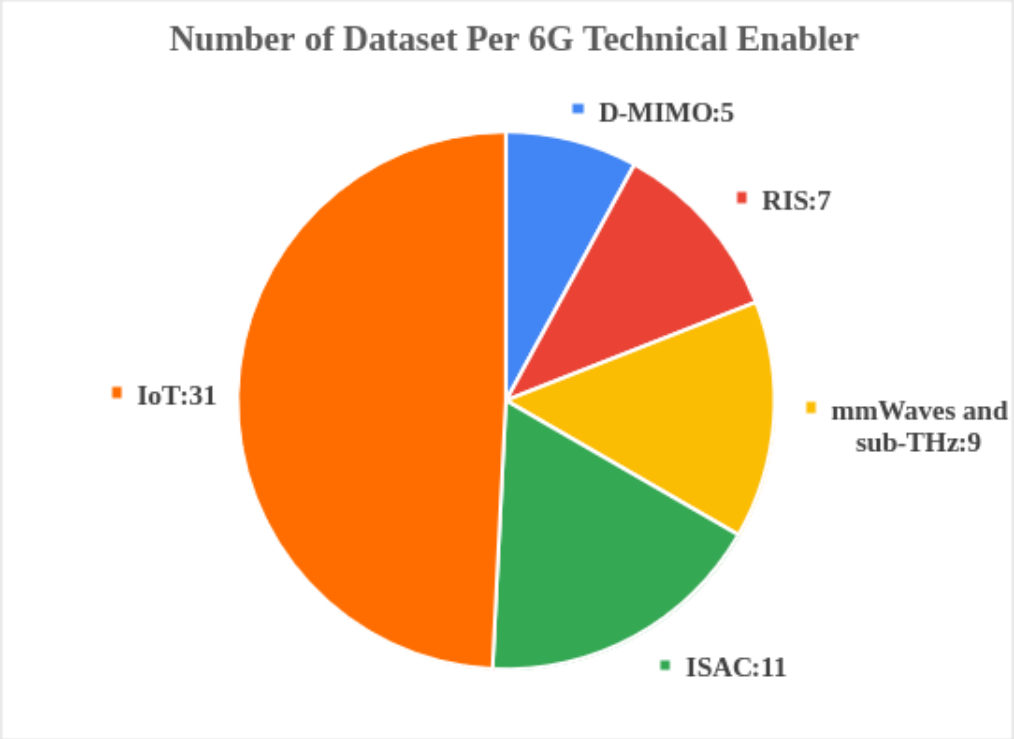

Figure 4-4: Distribution of Datasets by 6G-Technical Enabler

Figure 4-4 illustrates the distribution of datasets across 6G technical enablers, highlighting the areas of focus within current research effort:

Internet of Things (IoT): IoT has the highest representation, with 31 datasets. This strong focus reflects the anticipated central role of IoT in 6G applications, supporting a vast range of interconnected devices and systems. Also, it reflects that there is a considerable interest for PLS analysis in IoT technology. The prominence of IoT in dataset coverage suggests that it is a key driver in 6G research, emphasizing the need for robust, scalable, and secure communication solutions as IoT continues to expand.

Integrated Sensing and Communication (ISAC): ISAC is represented by 11 datasets, making it the second most covered enabler. ISAC combines communication and sensing functions, which is crucial for enabling advanced 6G capabilities such as location-aware services and enhanced situational awareness. The dataset count for ISAC indicates significant research interest, particularly in enhancing 6G's ability to integrate these dual functions effectively.

Millimetre Waves (mmWaves) and sub-THz: With 9 datasets, mmWaves and sub-THz frequencies are another focal point in 6G research. These frequencies are key to achieving the ultra-high data rates envisioned for 6G, as well as supporting high-capacity networks. The moderate dataset count shows ongoing exploration into the challenges and opportunities presented by these high-frequency bands, such as propagation characteristics and hardware development.

Reconfigurable Intelligent Surfaces (RIS): RIS is represented by 7 datasets, highlighting its emerging role as a transformative technology in 6G. RIS can improve signal strength, reduce interference, and enhance energy efficiency by dynamically adjusting the propagation environment.

The dataset count indicates active research, though further exploration could deepen understanding and improve practical implementations.

Distributed Multiple Input Multiple Output (D-MIMO): D-MIMO has the fewest datasets, with only 5, suggesting that while it is a recognized enabler, it may currently receive less research focus compared to other technologies. D-MIMO involves spatially distributed antennas, which could improve coverage, capacity, and resilience in 6G networks. Expanding dataset coverage in this area could advance understanding of D-MIMO's potential to enhance network performance in challenging environments.

4.3 Evaluating Threat Coverage Across 6G Enablers

In this section, we present an in-depth evaluation of dataset coverage across key 6G technological enablers. The focus is to identify strengths and gaps by cross-analysing datasets with the primary physical layer threats outlined in the D2.1 6G Threat Analysis Report: Spoofing, Tampering, Repudiation, Information Disclosure, and Denial of Service (DoS). This analysis helps pinpoint which enablers—such as Distributed Multiple-Input Multiple-Output (D-MIMO) and Reconfigurable Intelligent Surfaces (RIS)—require additional dataset coverage to ensure robust physical layer security (PLS).

Figure 4-5 shows the distribution of datasets in relation to 6G technical enablers and the associated threats. This visual representation highlights which threats are well-supported by the available datasets and where further data development is needed. The mapping shows that while D-MIMO is covered in areas like Spoofing, Tampering and Information Disclosure, it has limited representation in datasets addressing of Repudiation and no coverage in Denial of Service. Similarly, RIS datasets primarily focus on Tampering, Information Disclosure with only one less for Spoofing, no coverage for Repudiation and Denial of Service. The graph provides insights into the areas where targeted data collection efforts are necessary to ensure comprehensive coverage of threats for each technical enabler.

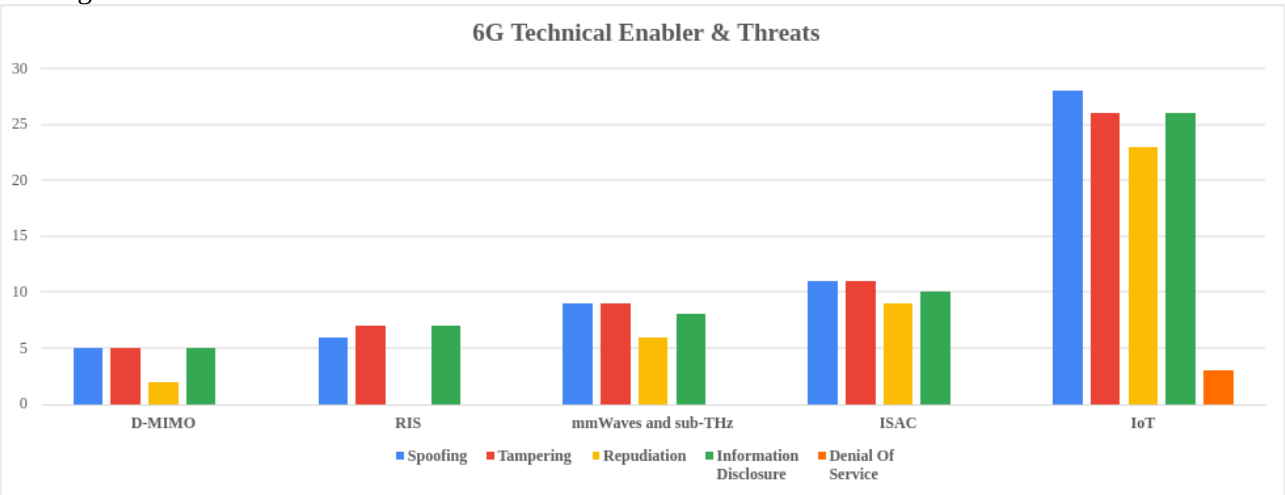

Figure 4-5: Dataset Alignment with 6G Technical Enabler & Threats

The chart presents the distribution of datasets addressing various threat categories across key 6G technical enablers. This breakdown provides insights into the security concerns associated with each enabler and highlights areas where additional research may be beneficial.

Distributed Multiple Input Multiple Output (D-MIMO): D-MIMO exhibits balanced attention across spoofing, tampering, and repudiation, with each threat represented by five datasets. Information disclosure has two datasets, while denial of service has no representation. This indicates that security research in D-MIMO is fairly broad, covering multiple threat areas, though more attention may be needed for denial of service to ensure comprehensive threat coverage.

Reconfigurable Intelligent Surfaces (RIS): RIS has a strong focus on tampering and repudiation threats, with seven datasets each, while spoofing is represented by six datasets. There is no representation for information

disclosure or denial of service, suggesting potential gaps in addressing data confidentiality and availability within RIS research. Expanding dataset coverage in these areas could help mitigate these vulnerabilities.

Millimetre Waves (mmWaves) and sub-THz: In the mmWaves and sub-THz category, spoofing and tampering are each represented by nine datasets, indicating a priority in addressing these threats. Repudiation has six datasets, while information disclosure is covered by eight datasets. Denial of service has no representation. This focus suggests an emphasis on data integrity and authentication in high-frequency communications, though more research on service availability (DoS) could strengthen resilience in this area.

Integrated Sensing and Communication (ISAC): ISAC displays a balanced distribution across spoofing (11 datasets), tampering (nine datasets), and information disclosure (10 datasets), showing significant coverage of these threats. Repudiation has five datasets, while denial of service is unrepresented. The dataset distribution suggests that ISAC research covers multiple security aspects, although denial of service remains an area for potential growth to ensure robust service continuity.

Internet of Things (IoT): IoT has the highest dataset counts across all threat categories, with spoofing (28 datasets), tampering (26 datasets), repudiation (23 datasets), and information disclosure (26 datasets) all showing extensive representation. Denial of service, however, is only represented by three datasets. This strong emphasis on IoT security reflects its critical role in 6G, addressing a broad range of threats, though additional attention to denial of service could further reinforce IoT resilience.

Specifically, the following trends and gaps were identified:

- **RIS-related Threats:** The datasets addressing threats associated with Reconfigurable Intelligent Surfaces (RIS) focus primarily on tampering and repudiation, each with 7 datasets, and spoofing with 6 datasets. However, there is no dataset representation for information disclosure or denial of service (DoS), leaving potential gaps in RIS security research. RIS introduces unique vulnerabilities, such as the potential for multi-path manipulation and surface misconfiguration, which could disrupt signal integrity or enable adversarial attacks. To mitigate these risks effectively, additional datasets are needed to examine these complex threat scenarios and develop targeted security strategies for RIS-enabled environments.
- **D-MIMO-related Threats:** Similarly, Distributed Multiple Input Multiple Output (D-MIMO) technology, critical for enabling high-capacity, low-latency 6G networks, also shows limited dataset coverage. The focus is primarily on spoofing, tampering, and repudiation threats, each with 5 datasets, while information disclosure has only 2 datasets and denial of service (DoS) has no representation. D-MIMO systems require secure synchronization across distributed antennas and resilience against jamming or tampering attacks. Given the limited dataset coverage in areas like information disclosure and DoS, additional research and data are essential to develop robust countermeasures, ensuring secure, reliable operation in D-MIMO deployments.
- **Lack of DoS Datasets:** The dataset analysis reveals that DoS threats are minimally or not represented at across key 6G enablers, including D-MIMO, RIS, mmWaves and sub-THz, and ISAC. Only the IoT enabler shows limited representation with 3 datasets, highlighting a significant gap in resilience research for continuous service availability across other enablers.

4.3.1 Cross-Analysis of Threats and Enablers

The analysis highlighted areas where the coverage of datasets aligns well with the identified threats and where deficiencies remain:

- **Spoofing:** Datasets on spoofing threats are well-represented across most enablers, indicating strong research interest in addressing identity-based attacks. However, additional coverage could benefit areas like D-MIMO to improve resilience.
- **Tampering:** Tampering threats show substantial dataset coverage, especially in IoT and ISAC, suggesting recognition of data integrity risks. ISAC and mmWaves and sub-THz have moderate tampering dataset representation, though continued effort is necessary to address specific tampering risks in these technologies.

- **Repudiation:** Repudiation has moderate dataset coverage, particularly within IoT and ISAC, reflecting a focus on authentication and accountability. However, gaps remain in RIS and D-MIMO, where improved dataset representation could enhance non-repudiation measures.
- **Information Disclosure:** Information disclosure is well-documented across all enablers, emphasizing confidentiality in high-priority areas. However, D-MIMO lacks sufficient dataset coverage, signalling a need for further research to protect sensitive information across all enablers.
- **Denial of Service:** DoS threats are the least represented across all enablers, with only minimal datasets in IoT and none in others. This underrepresentation highlights a critical gap in DoS research, underscoring the need for dedicated efforts to ensure service availability in 6G networks.

5 Conclusions and Future Work

The document D5.1 Library of Known PHY Attacks and PLS Dataset consolidates prior research on Physical Layer Security (PLS) with new contributions from the ROBUST-6G project to create a comprehensive dataset library. This effort aims to address existing vulnerabilities, test mitigation strategies, and support the development of robust security solutions tailored to 6G networks.

5.1 Summary of Key Findings

The assessment of datasets reveals several key insights and critical gaps in Physical Layer Security (PLS) for 6G networks. One of the primary findings is the strong coverage of Spoofing and Tampering threats. Technologies such as Short-Range Communications, Cellular Technologies, and Software Defined Radio (SDR) have well-represented datasets, offering valuable insights into mitigation strategies. These datasets provide researchers with the tools needed to develop solutions that address unauthorized access and signal manipulation, which are vital for ensuring secure communication.

However, the analysis also identifies significant gaps in datasets for Repudiation and Denial of Service (DoS) threats. Repudiation, which ensures accountability and non-repudiation in communications, remains underrepresented across most technologies. This scarcity of datasets poses challenges for the development of robust mechanisms to verify and authenticate communications. Similarly, datasets for DoS threats—which are critical for maintaining service continuity in LPWAN, IoT, and mmWave/sub-THz technologies—are deemed insufficient. This gap is particularly concerning given the importance of these technologies in ultra-reliable low-latency communication (URLLC) scenarios, where service availability is paramount.

Another key insight is the underrepresentation of emerging 6G technologies, such as Distributed Multiple-Input Multiple-Output (D-MIMO) and Reconfigurable Intelligent Surfaces (RIS). While some datasets cover Spoofing threats for these technologies, there is minimal support for other essential threats, including Tampering, Information Disclosure, and DoS. This lack of comprehensive datasets limits the ability to fully understand and mitigate the security challenges posed by these innovations. Additionally, mmWave and sub-THz technologies, which are crucial for achieving the high data rates expected in 6G networks, are underrepresented in terms of datasets for threat analysis.

While existing datasets focus primarily on Spoofing and Tampering threats, they do not adequately address other critical vulnerabilities, such as Information Disclosure and DoS threats. Given the expected proliferation of IoT devices and LPWAN deployments in future 6G networks, addressing these gaps is crucial to ensure their security and resilience. Without comprehensive datasets, these technologies could remain vulnerable to attacks, jeopardizing the success of 6G applications across diverse sectors.

5.2 Recommendations for Future Research

Addressing the identified gaps in Physical Layer Security (PLS) datasets is critical to ensuring that 6G networks are secure and resilient. One key area for future research is the development of datasets focused on underrepresented threats. In particular, datasets addressing Repudiation and Denial of Service (DoS) are essential for creating robust security mechanisms. The lack of such datasets limits researchers' ability to develop non-repudiation protocols and DoS mitigation strategies, which are vital for applications such as ultra-reliable low-latency communications (URLLC). Expanding these datasets across key 6G technologies, including D-MIMO, RIS, and mmWave, will provide a solid foundation for tackling these critical threats.

Another area requiring attention is the expansion of datasets for LPWAN technologies. Given their expected prevalence in 6G networks, it is imperative to ensure that these technologies are comprehensively covered.

Existing datasets primarily address almost all threats but lack sufficient focus on DoS threats. Developing more diverse datasets for LPWAN will enable better threat modelling and improved protection strategies, particularly for use cases in smart cities, healthcare, and agriculture.

Future research should also focus on the enhancement of multi-modal and multi-technology datasets. As 6G networks integrate diverse technologies, there is a growing need for datasets that combine multiple data sources, such as RF signals, network traffic, environmental sensors, and positioning information. This multi-modal approach will allow for the development of holistic PLS solutions capable of detecting and mitigating complex, cross-layer attacks, improving the overall security of 6G deployments.

In addition, researchers should prioritize collecting real-world datasets from practical environments such as urban areas, industrial sites, and vehicular networks. Many existing datasets are derived from controlled simulations, which may not fully capture the complexities of real-world conditions. By collecting data in real-world settings, researchers can develop more accurate models for threat detection and mitigation, ensuring that PLS mechanisms perform effectively under realistic conditions.

To facilitate better collaboration and interoperability, there is a need for standardization of dataset formats and documentation. Establishing consistent metadata requirements, dataset structures, and labelling conventions will ensure that datasets are easily shareable and reusable across research initiatives. This will promote collaboration within the research community and support the integration of datasets into comprehensive security studies.

Furthermore, future research should place greater emphasis on targeted datasets for emerging 6G technologies, including D-MIMO, RIS, ISAC, mmWave and sub-THz frequencies. These technologies bring new security challenges, such as vulnerabilities in beamforming and signal manipulation, which require dedicated datasets for in-depth analysis. Addressing these challenges will ensure the secure deployment and operation of these innovative technologies within the 6G ecosystem.

Finally, leveraging Artificial Intelligence (AI) and Machine Learning (ML) for dataset generation and analysis will accelerate progress in this domain. AI/ML algorithms can be used to simulate diverse attack scenarios, generate synthetic data, and identify hidden vulnerabilities. These techniques can also enhance the analysis of existing datasets by enabling pattern recognition, anomaly detection, and the development of intelligent security mechanisms tailored to 6G environments.

References

[AAW23] Tomer Avrahami, Ofer Amrani, and Avishai Wool. "Let's Shake on It: Extracting Secure Shared Keys from Wi-Fi CSI." arXiv preprint arXiv:2307.05423v1, 2023.

[ABV+19] M. Aernouts, R. Berkvens, K. Van Vlaenderen, and M. Weyn, "Sigfox and LoRaWAN Datasets for Fingerprint Localization in Large Urban and Rural Areas (1.3) [Data set]," Zenodo, 2019. DOI: 10.5281/zenodo.3904158.

[ACK+21] C. Ayyildiz, R. Cetin, Z. Khodzhaev, T. Kocak, E. G. Soyak, V. Ç. Güngör, and G. K. Kurt, "Physical layer authentication for extending battery life," Ad Hoc Networks, vol. 123, 2021.

[ACO+23] Ahmed Alkhateeb, Gouranga Charan, Tawfik Osman, Andrew Hredzak, Joao Morais, Umut Demirhan, and Nikhil Srinivas. "DeepSense 6G: A Large-Scale Real-World Multi-Modal Sensing and Communication Dataset." IEEE Communications Magazine, 2023.

[Alk19] Ahmed Alkhateeb. "DeepMIMO: A Generic Deep Learning Dataset for Millimeter Wave and Massive MIMO Applications." Proceedings of the Information Theory and Applications Workshop (ITA), San Diego, CA, USA, February 2019.

[AP23] Christos Anagnostopoulos, Nikos Piperigkos. "Cooperative Localization using CARLA-SUMO-Artery simulators." IEEE Dataport, June 28, 2023. https://dx.doi.org/10.21227/511y-4s83.

[AR21] Ammar Rafique. "Reconfigurable Intelligent Surface (RIS) Benchmarking Results and Simulation Code." IEEE Dataport, August 15, 2021. DOI: 10.21227/eg0q-y563.

[ASD+23] Qing An, Santiago Segarra, Chris Dick, Ashutosh Sabharwal, Rahman Doost-Mohammady. "A Deep Reinforcement Learning-Based Resource Scheduler for Massive MIMO Networks," IEEE Transactions on Machine Learning in Communications and Networking, 2023, doi: 10.1109/TMLCN.2023.3313988

[ASO22] S. Alhazbi, S. Sciancalepore, and G. Oligeri, "A Dataset of IQ samples in Indoor Jamming Scenarios," IEEE Consumer Communications & Networking Conference (CCNC2023), Las Vegas, 2022. Zenodo. Available at: https://doi.org/10.5281/zenodo.7119040.

[ATL+24] K. I. Ahmed, M. Tahir, S. L. Lau, M. H. Habaebi, A. Ahad, and I. M. Pires, "Dataset for authentication and authorization using physical layer properties in indoor environment," Data in Brief, vol. 55, 2024, doi:10.1016/j.dib.2024.110589.

[BSB+21] Mauro Belgiovine, Kunal Sankhe, Carlos Bocanegra, Debashri Roy, and Kaushik R. Chowdhury. "Deep Learning at the Edge for Channel Estimation in Beyond-5G Massive MIMO." IEEE Wireless Communications, vol. 28, no. 2, pp. 19-25, April 2021. DOI: 10.1109/MWC.001.2000322.

[BWR+24] B. Baker, J. Woods, M.J. Reed, and M. Afford, "A Survey of Short-Range Wireless Communication for Ultra-Low-Power Embedded Systems," Journal of Low Power Electronics and Applications, vol. 14, no. 27, pp. 1-20, May 2024. doi: 10.3390/jlpea14020027.

[CB20] L. Chettri and R. Bera, "A Comprehensive Survey on Internet of Things (IoT) Toward 5G Wireless Systems," in IEEE Internet of Things Journal, vol. 7, no. 1, pp. 16-32, Jan. 2020. DOI: 10.1109/JIOT.2019.2948888.

[CBK+22] A. Chorti, A. N. Barreto, S. Köpsell, M. Zoli, M. Chafii, P. Sehier, and G. Fettweis, "Context-Aware Security for 6G Wireless: The Role of Physical Layer Security," IEEE Communications Standards Magazine, vol. 6, no. 1, pp. 102-110, March 2022.

[CBV+23] Z. Chen, C. N. Barati, J. Veihl, C. Shepard, and A. Sabharwal, "LensFD: Using Lenses for Improved Sub-6 GHz Massive MIMO Full-Duplex," IEEE Transactions on Vehicular Technology, 2023, pp. 1-13. Available at: https://doi.org/10.1109/TVT.2023.3240558.

[CHS+20] S. Chen, J. Hu, Y. Shi, L. Zhao, and W. Li, "A Vision of C-V2X: Technologies, Field Testing, and Challenges With Chinese Development," in IEEE Internet of Things Journal, vol. 7, no. 5, pp. 3872-3881, May 2020. DOI: 10.1109/JIOT.2020.2974823.

[CNA24] R. Chataut, M. Nankya, and R. Akl, "6G Networks and the AI Revolution—Exploring Technologies, Applications, and Emerging Challenges," Sensors, vol. 24, no. 1888, pp. 1-29, Mar. 2024.

[CTC+23] Colpaert, Achiel; Thys, Cel; Cui, Zhuangzhuang; Pollin, Sofie, 2023. "MaMIMO-UAV 3D Channel State Information Dataset." KU Leuven RDR, V1. DOI: 10.48804/0IMQDF.

[DBZ+18] R. Doost-Mohammady, O. Bejarano, L. Zhong, J. R. Cavallaro, E. Knightly, Z. M. Mao, W. W. Li, X. Chen, and A. Sabharwal, "RENEW: Programmable and Observable Massive MIMO Networks," 2018 52nd Asilomar Conference on Signals, Systems, and Computers, 2018, pp. 1654-1658. Available at: https://doi.org/10.1109/ACSSC.2018.8645391.

[DGW+23] B. J. B. Deutschmann, M. Graber, T. Wilding, and K. Witrisal, "Bistatic MIMO Radar Sensing of Specularly Reflecting Surfaces for Wireless Power Transfer," 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), Rhodes Island, Greece, 2023, pp. 1-4, doi: 10.1109/ICASSPW59220.2023.10193617.

[DP21] Sibren De Bast, Sofie Pollin. "Ultra Dense Indoor MaMIMO CSI Dataset." IEEE Dataport, February 9, 2021. Available at: https://dx.doi.org/10.21227/nr6k-8r78.

[DRC+20] M. Di Renzo, et al., "Smart Radio Environments Empowered by Reconfigurable Intelligent Surfaces: How It Works, State of Research, and The Road Ahead," in IEEE Journal on Selected Areas in Communications, vol. 38, no. 11, pp. 2450-2525, Nov. 2020. DOI: 10.1109/JSAC.2020.3007211.

[DS21] Du, Xu, and Sabharwal, Ashutosh. "Massive MIMO Channels with Inter-User Angle Correlation: Open-Access Dataset, Analysis and Measurement-Based Validation." IEEE Transactions on Vehicular Technology, 2021. doi:10.1109/TVT.2021.3131606.

[DZS22] Rahman Doost-Mohammady, Mehdi Zafari, and Ashutosh Sabharwal. "Robustness of Distributed Multi-User Beamforming: An Experimental Evaluation." IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2022), 2022, pp. 146-150.

[EAC+21] M. Ericson, M.S.H. Abad, M. Condoluci, O. Haliloglu, P. Rugeland, M. Saimler, S. Wänstedt, and L. Feltrin, "6G Architectural Trends and Enablers," in IEEE 5G World Forum, October 2021.

[EEA+20] Martins Ezuma, Fatih Erden, Chethan K. Anjinappa, Ozgur Ozdemir, and Ismail Guvenc. "Drone Remote Controller RF Signal Dataset." IEEE Dataport, November 25, 2020. DOI: 10.21227/ss99-8d56.

[EGD+21] Euchner, Florian, Marc Gauger, Sebastian Dörner, and Stephan ten Brink. "A Distributed Massive MIMO Channel Sounder for 'Big CSI Data'-driven Machine Learning." In WSA 2021; 25th International ITG Workshop on Smart Antennas, 2021.

[EH21a] A. Elmaghbub and B. Hamdaoui, "LoRa Device Fingerprinting in the Wild: Disclosing RF Data-Driven Fingerprint Sensitivity to Deployment Variability," in IEEE Access, vol. 9, pp. 142893-142909, 2021.

[EH21b] A. Elmaghbub and B. Hamdaoui, "Comprehensive RF Dataset Collection and Release: A Deep Learning-Based Device Fingerprinting Use Case," 2021 IEEE Globecom Workshops (GC Wkshps), Madrid, Spain, 2021, pp. 1-7.

[ESZ+16] Evan Everett, Clay Shepard, Lin Zhong, and Ashutosh Sabharwal, "SoftNull: Many-antenna Full-duplex Wireless via Digital Beamforming," IEEE Transactions on Wireless Communications, vol. 15, no. 12, pp. 8077–8092, Dec. 2016.

[FP24] Hua Fu, Linning Peng. "Bidirectional CSI Measurement for V2X Communications," IEEE Dataport, June 14, 2024. DOI: 10.21227/3mkx-aq02.

[GS23] Merkebu Girmay and Adnan Shahid, “Dataset: IQ samples of LTE, 5G NR, Wi-Fi, ITS-G5, and C-V2X PC5,” IEEE Dataport, May 16, 2023, doi: https://dx.doi.org/10.21227/72qq-z464.

[Gua23] J. Gu, "e-FLASH," IEEE Dataport, August 2, 2023. Available at: https://dx.doi.org/10.21227/qz96-yh44.

[GWL+22] Gao, K., Wang, H., Lv, H., & Liu, W. "Toward 5G NR High-Precision Indoor Positioning via Channel Frequency Response: A New Paradigm and Dataset Generation Method." IEEE Journal on Selected Areas in Communications, vol. 40, no. 7, pp. 2233-2247, July 2022. DOI: 10.1109/JSAC.2022.3157397

[HAK22] S. A. Abdel Hakeem, H. H. Hussein, and H. Kim, "Security Requirements and Challenges of 6G Technologies and Applications," Sensors, vol. 22, no. 1969, pp. 1-43, Mar. 2022.

[HBK+24] M. Harvanek, J. Bolcek, J. Kufa, L. Polak, M. Simka, and R. Marsalek, “Survey on 5G Physical Layer Security Threats and Countermeasures,” Sensors, vol. 24, no. 5523, 2024. DOI: 10.3390/s24175523.

[HGP+23] R. Hernangómez, P. Geuer, A. Palaios, D. Schäufele, C. Watermann, and K. Taleb-Bouhemadi. "Berlin V2X: A Machine Learning Dataset from Multiple Vehicles and Radio Access Technologies." 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, 2023, pp. 1-5, doi: 10.1109/VTC2023-Spring57618.2023.10200750.

[HKC20] S. Hanna, S. Karunaratne, and D. Cabric, “Open Set Wireless Transmitter Authorization: Deep Learning Approaches and Dataset Considerations,” IEEE Transactions on Cognitive Communications and Networking, pp. 59–72, 2020. DOI: 10.1109/TCCN.2020.3043332.

[HKC22] S. Hanna, S. Karunaratne, and D. Cabric, "WiSig: A Large-Scale WiFi Signal Dataset for Receiver and Channel Agnostic RF Fingerprinting," in IEEE Access, vol. 10, pp. 22808-22818, 2022, doi: 10.1109/ACCESS.2022.3154790.

[HPW+22] Rodrigo Hernangomez, Alexandros Palaios, Cara Watermann, Daniel Schäufele, Philipp Geuer, Rafail Ismayilov, Mohammad Parvini, Anton Krause, Martin Kasparick, Thomas Neugebauer, Oscar D. Ramos-Cantor, Hugues Tchouankem, Jose Leon Calvo, Bo Chen, Slawomir Stanczak, Gerhard Fettweis. "AI4Mobile Industrial Wireless Datasets: iV2V and iV2I+." IEEE Dataport, October 31, 2022. DOI: 10.21227/04ta-v128.

[HWO+18] B. Hilburn, N. West, T. O’Shea, and T. Roy, “SigMF: The Signal Metadata Format,” in Proceedings of the GNU Radio Conference, vol. 3, no. 1, 2018.

[HYM+23] O. Haliloglu, H. Yu, C. Madapatha, H. Guo, F. E. Kadan, A. Wolfgang, R. Puerta, P. Frenger, and T. Svensson, “Distributed MIMO Systems for 6G,” 2023 European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), pp. 156-163, 2023.

[JJ22] A.Jagannath and J.Jagannath, "Embedding-Assisted Attentional Deep Learning for Real-World RF Fingerprinting of Bluetooth" TechRxiv. Preprint 2022. https://doi.org/10.36227/techrxiv.20767315.v1

[JKJ22] A. Jagannath, Z. Kane, and J. Jagannath, “RF Fingerprinting Needs Attention: Multi-task Approach for Real-World WiFi and Bluetooth,” in Proc. of IEEE Global Communications Conference (GLOBECOM), Rio de Janeiro, Brazil, December 2022.

[JKM23] J. M. Jornet, E. W. Knightly, and D. M. Mittleman, “Wireless Communications Sensing and Security Above 100 GHz,” Nature Communications, vol. 14, no. 841, 2023. DOI: 10.1038/s41467-023-36621-x.

[JLS+24] X. Jiang, P. Li, Y. Shang, Y. Zou, B. Li, and P. Yan, "Improving Physical Layer Security for Distributed Antenna Systems With a Friendly Jammer", IEEE Transactions on Communications, vol. 72, no. 8, pp. 4756-4773, August 2024.

[JR22] Marko Jacovic and Xaime Rivas Rey. "Mitigating RF Jamming Attacks at the Physical Layer with Machine Learning Dataset." Zenodo, 2022. DOI: 10.5281/zenodo.6304194.

[KBG+18] A. Klautau, P. Batista, N. González-Prelcic, Y. Wang, and R. W. Heath, “5G MIMO Data for Machine Learning: Application to Beam-Selection Using Deep Learning,” 2018 Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 2018, pp. 1-9. DOI: 10.1109/ITA.2018.8503086.

[KGH19] A. Klautau, N. González-Prelcic, and R. W. Heath, “LIDAR Data for Deep Learning-Based mmWave Beam-Selection,” IEEE Wireless Communications Letters, vol. 8, no. 3, pp. 909-912, June 2019. DOI: 10.1109/LWC.2019.2899571.

[Kih22] Billy Kihei. "IEEE 802.11p Wireless Congestion and Jamming Experiments." IEEE Dataport, October 31, 2022. DOI: https://dx.doi.org/10.21227/yaw2-2997.

[KKA+21] D. Kosmanos, D. Karagiannis, A. Argyriou, S. Lalis, and L. Maglaras, “RF Jamming Classification Using Relative Speed Estimation in Vehicular Wireless Networks,” Security and Communication Networks, Hindawi, 2021. DOI: 10.1155/2021/9959310.

[KKS+16] C. Kolias, G. Kambourakis, A. Stavrou, and S. Gritzalis, “Intrusion Detection in 802.11 Networks: Empirical Evaluation of Threats and a Public Dataset,” IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 184-208, Firstquarter 2016. DOI: 10.1109/COMST.2015.2402161.

[KSK+23] S. M. Kashani, S. M. H. Sherazi, A. Khokhar, S. W. Kim, and F. Nait-Abdesselam, “BLE-WBAN: RF Real-World Dataset of BLE Devices in Human-Centric Healthcare Environments,” IEEE Dataport, September 22, 2023. DOI: https://dx.doi.org/10.21227/mtg7-eb43.

[KZA+24] Bharath Keshavamurthy, Yaguang Zhang, Christopher R. Anderson, Nicolò Michelusi, David J. Love, and James V. Krogmeier. "Statistical Characterization of 28GHz V2X Channels via Autonomous Beam-Steered Measurements." IEEE Dataport, February 29, 2024. doi: 10.21227/0r6d-z728.

[LDM+23] Zhecun Liu, Keerthi Dasala, Di Mu, Rahman Doost-Mohammady, Edward Knightly, "M3A: Multipath Multicarrier Misinformation to Adversaries," in Proceedings of the 29th Annual International Conference on Mobile Computing and Networking (ACM MobiCom), Oct. 2023.

[Liu24] Yuepei Li. "Radio Frequency Fingerprinting Exploiting Power Amplifier Active Load-pulling." IEEE Dataport, April 10, 2024. DOI: https://dx.doi.org/10.21227/vh95-rs64.

[MEL+22] Olusiji Medaiyese, Martins Ezuma, Adrian Lauf, Ayodeji Adeniran, "Cardinal RF (CardRF): An Outdoor UAV/UAS/Drone RF Signals with Bluetooth and WiFi Signals Dataset," IEEE Dataport, July 13, 2022. DOI: 10.21227/1xp7-ge95.

[Mir22] J. Miraglia, "Signal Discovery with Convolutional Neural Nets," The University of Utah, 2022.

[MJC+21] L. Mucchi, S. Jayousi, S. Caputo, E. Panayirci, S. Shahabuddin, J. Bechtold, I. Morales, R.-A. Stoica, G. Abreu, and H. Haas, “Physical-Layer Security in 6G Networks,” IEEE Open Journal of the Communications Society, vol. X, no. X, pp. X-X, August 2021. Digital Object Identifier 10.1109/OJCOMS.2021.3103735.

[MKL23] Moradbeikie, A., Keshavarz, A., & Lopes, S. (2023). A dataset for RSSI-based outdoor localization using LoRaWAN in a harbor as a harsh and industrial environment. Internet of Things. Zenodo. https://doi.org/10.5281/zenodo.10142174

[MMC+23] A. Mayya, M. Mitev, A. Chorti, and G. Fettweis, “Dataset for the Paper: A SKG Security Challenge: Indoor SKG Under an On-The-Shoulder Eavesdropping Attack,” IEEE Dataport, May 4, 2023. DOI: https://dx.doi.org/10.21227/kxnw-r386.

[MRT+20] G. M. Mendoza-Silva, P. Richter, J. Torres-Sospedra, E. S. Lohan, and J. Huerta, “Long-Term Wi-Fi Fingerprinting Dataset and Supporting Material (2.2) [Data set],” Zenodo, 2020. DOI: 10.5281/zenodo.3748719.

[MSM+18] M. H. Mahalat, S. Saha, A. Mondal, and B. Sen, “A PUF based Light Weight Protocol for Secure WiFi Authentication of IoT devices,” 2018 8th International Symposium on Embedded Computing and System Design (ISED), Cochin, India, 2018, pp. 183-187. DOI: 10.1109/ISED.2018.8703993.

[MSS+20] S. Mohanti, N. Soltani, K. Sankhe, D. Jaisinghani, M. D. Felice, and K. R. Chowdhury. "AirID: Injecting a Custom RF Fingerprint for enhanced UAV Identification using Deep Learning." IEEE GLOBECOM 2020, Taipei, Taiwan, Dec. 2020. DOI: 10.1109/GLOBECOM42002.2020.9322561.

[NDB22] Manish Nair, Shuping Dang, Mark Beach. "LoRa Sensor Data Sets for RF Fingerprinting via Self-Organizing Feature Maps." IEEE Dataport, January 30, 2023. DOI: 10.21227/63kx-sr47.

[PLL+23] Mengguan Pan, Shengheng Liu, Peng Liu, Wangdong Qi, Yongming Huang, Wang Zheng, Qihui Wu, Markus Gardill, December 29, 2022, "5G CFR/CSI dataset for wireless channel parameter estimation, array calibration, and indoor positioning", IEEE Dataport, doi: https://dx.doi.org/10.21227/k2f0-k132.

[PRS+24] S. Pradhan, D. Roy, B. Salehi, K. Chowdhury, "COPILOT: Cooperative Perception using LiDAR for Handoffs between Road Side Units," IEEE INFOCOM 2024, Vancouver, Canada, May 2024.

[PRT+23] Karel Pärlin, Taneli Riihonen, Matias Turunen, Vincent Le Nir, Marc Adrat, June 5, 2023, "Securing the Physical Layer of IEEE 802.15.4 Through Cooperative Jamming", IEEE Dataport, doi: https://dx.doi.org/10.21227/9mty-pf96.

[PZZ+22] Pan C, Zhou G, Zhi K, Hong S, Wu T, Pan Y, Ren H, Di Renzo M, Swindlehurst AL, Zhang R, Zhang AY. An overview of signal processing techniques for RIS/IRS-aided wireless systems. IEEE Journal of Selected Topics in Signal Processing. 2022;16(5):883-917. https://doi.org/10.1109/JSTSP.2022.3195671

[R] Remcom. "Wireless InSite." Available at: http://www.remcom.com/wireless-insite.

[RCT+23] Debashri Roy, Vini Chaudhury, Chinenye Tassie, Chad Spooner, and Kaushik Chowdhury. "ICARUS: Learning on IQ and Cycle Frequencies for Detecting Anomalous RF Underlay Signals." IEEE INFOCOM 2023 - IEEE Conference on Computer Communications, New York City, NY, USA, May 2023. DOI: 10.1109/INFOCOM53939.2023.10228929.

[RFG+24] H. Rifà-Pous, V. Garcia-Font, C. Núñez-Gómez, and J. Salas, “Security, Trust and Privacy challenges in AI-driven 6G Networks,” Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya (UOC), Center for Cybersecurity Research of Catalonia (CYBERCAT), Barcelona, Spain, arXiv:2409.10337v1 [cs.CR], September 2024.

[RGS24] Rugeles, J., Guillen, E., & Sampaio Cardoso, L. (2024). WiFi 2.4 GHz Jamming attack scenario P2 measurements using ADALM Pluto and Maia SDR (V.1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10456777

[RH22] M. A. Rahman and M. S. Hossain, "A Deep Learning Assisted Software Defined Security Architecture for 6G Wireless Networks: IIoT Perspective," IEEE Wireless Communications, vol. 29, no. 2, pp. 52-59, Apr. 2022.

[Ric22] Tarence Rice. "Experimental Evaluation of AoA Estimation for UAV to Massive MIMO." Rice University, 2022.

[RJS+20] G. Reus-Muns, D. Jaisinghani, K. Sankhe, and K. R. Chowdhury. "Trust in 5G Open RANs through Machine Learning: RF Fingerprinting on the POWDER PAWR Platform." IEEE Globecom, 7-11 December 2020, Taipei, Taiwan.

[RMG+22] Marco Rossanese, Placido Mursia, Andres Garcia-Saavedra, Vincenzo Sciancalepore, Arash Asadi, and Xavier Costa-Perez. "Designing, Building, and Characterizing RF Switch-based Reconfigurable Intelligent Surfaces." arXiv preprint arXiv:2207.07121, 2022.

[RML+17] P. Robyns, E. Marin, W. Lamotte, P. Quax, D. Singelée, and B. Preneel, “Physical-Layer Fingerprinting of LoRa Devices Using Supervised and Zero-Shot Learning,” WiSec '17: 10th ACM Conference on Security & Privacy in Wireless and Mobile Networks (WiSec), Boston, USA, 2017. DOI: 10.5281/zenodo.583965.

[ROB24-D21] ROBUST-6G, “6G Threat Analysis”, ROBUST-6G, Project Deliverable D2.1, July 2024. [Online]. Available: robust-6g.eu

[RSC+19] T. S. Rappaport, et al., “Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond,” in IEEE Access, vol. 7, pp. 78729-78757, 2019. DOI: 10.1109/ACCESS.2019.2921522.

[RSO+05] D. Raychaudhuri, I. Seskar, M. Ott, S. Ganu, K. Ramachandran, H. Kremo, R. Siracusa, H. Liu, and M. Singh, “Overview of the ORBIT radio grid testbed for evaluation of next-generation wireless network protocols,” in Wireless Communications and Networking Conference, 2005 IEEE, vol. 3, pp. 1664–1669, IEEE, 2005.

[RTD+24] Rusins, Artis; Tiscenko, Deniss; Dobelis, Eriks; Blumbergs, Eduards; Nesenbergs, Krisjanis; Paikens, Peteris. "Wearable Device Bluetooth/BLE Physical Layer Dataset." Data, vol. 9, no. 4, pp. 53, 2024, MDPI.

[RUD+23] G. Reus-Muns, P. S. Upadhyaya, U. Demir, N. Stephenson, N. Soltani, V. K. Shah, and K. Chowdhury, "SenseORAN: O-RAN based Radar Detection in the CBRS Band," IEEE Journal on Selected Areas in Communications (JSAC), 2023.

[RXM+17] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, “Overview of Millimeter Wave Communications for Fifth-Generation (5G) Wireless Networks—With a Focus on Propagation Models,” in IEEE Transactions on Antennas and Propagation, vol. 65, no. 12, pp. 6213-6230, Dec. 2017. DOI: 10.1109/TAP.2017.2734243.

[SBA+19] K. Sankhe, M. Belgiovine, F. Zhou, L. Angioloni, F. Restuccia, S. D’Oro, T. Melodia, S. Ioannidis, and K. R. Chowdhury, "No Radio Left Behind: Radio Fingerprinting Through Deep Learning of Physical-Layer Hardware Impairments," IEEE Transactions on Cognitive Communications and Networking, Special Issue on Evolution of Cognitive Radio to AI-enabled Radio and Networks, 2019.

[SBF+24] G. Stanco, A. Botta, F. Frattini, U. Giordano, and G. Ventre, "On the Performance of IoT LPWAN Technologies: The Case of Sigfox, LoRaWAN, and NB-IoT," IEEE International Conference on Communications, pp. 2095-2100, May 2024. doi: 10.1109/ICC45855.2024.9839078.

[SBG+20] B. Salehi, M. Belgiovine, S. Garcia Sanchez, J. Dy, S. Ioannidis, and K. Chowdhury, "Machine Learning on Camera Images for Fast mmWave Beamforming," IEEE MASS, 10-13 December 2020, Delhi NCR, India.

[SCR+22] Nasim Soltani, Vini Chaudhary, Debashri Roy, and Kaushik Chowdhury. "Finding Waldo in the CBRS Band: Signal Detection and Localization in the 3.5 GHz Spectrum." IEEE GLOBECOM, December 2022, pp. 4570-4575.

[SDG+16] C. Shepard, J. Ding, R. E. Guerra, and L. Zhong, "Understanding real many-antenna MU-MIMO channels," 2016 50th Asilomar Conference on Signals, Systems and Computers, 2016, pp. 461-467. Available at: https://doi.org/10.1109/ACSSC.2016.7869082.

[SGR+22] B. Salehi, J. Gu, D. Roy, and K. Chowdhury, "FLASH: Federated Learning for Automated Selection of High-band mmWave Sectors," in IEEE INFOCOM 2022 - IEEE Conference on Computer Communications, London, United Kingdom, 2022, pp. 1719-1728, doi: 10.1109/INFOCOM48880.2022.9796865.

[Sha22] A. F. M. Shahen Shah, “A Survey From 1G to 5G Including the Advent of 6G: Architectures, Multiple Access Techniques, and Emerging Technologies,” 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2022, pp. 1117-1123. DOI: 10.1109/CCWC54503.2022.9720781.

[SHC+24] B. D. Son, N. T. Hoa, T. V. Chien, W. Khalid, M. A. Ferrag, W. Choi, and M. Debbah, "Adversarial Attacks and Defenses in 6G Network-Assisted IoT Systems", IEEE Internet of Things Journal, vol. 11, no. 11, pp. 19168-19187, June 2024.

[SHv+22] R. Solo de Zaldivar, A. Huertas Celdrán, J. von der Assen, P. M. Sánchez Sánchez, G. Bovet, G. Martínez Pérez, and B. Stiller, "MalwSpecSys: A Dataset Containing Syscalls of an IoT Spectrum Sensor Affected by Heterogeneous Malware," IEEE Dataport, May 16, 2022. Available at: https://dx.doi.org/10.21227/nvmb-eg69.

[SJM+23] G. Shen, J. Zhang, and A. Marshall, “LORA_RFFI_DATASET_DIFFERENT_SPREADING_FACTORS,” IEEE Dataport, April 2, 2023. DOI: https://dx.doi.org/10.21227/5q6q-c107.

[SMR+22] P. Staat, S. Mulzer, S. Roth, V. Moonsamy, M. Heinrichs, R. Kronberger, A. Sezgin, and C. Paar, “IRShield: A Countermeasure Against Adversarial Physical-Layer Wireless Sensing [Data set],” 43rd IEEE Symposium on Security and Privacy (IEEE S&P), 2022. DOI: https://doi.org/10.5281/zenodo.6367411.

[SNY+20] N. Soltanieh, Y. Norouzi, Y. Yang, and N. C. Karmakar, “A Review of Radio Frequency Fingerprinting Techniques,” in IEEE Journal of Radio Frequency Identification, vol. 4, no. 3, pp. 222-233, Sept. 2020. DOI: 10.1109/JRFID.2020.2968369.

[SPT+23] Ivo Silva, Cristiano Pendão, Joaquín Torres-Sospedra, and Adriano Moreira. "Industrial Environment Multi-Sensor Dataset for Vehicle Indoor Tracking with Wi-Fi, Inertial and Odometry Data." Data, vol. 8, no. 157, 2023. https://doi.org/10.3390/data8100157

[SPY+21] Y. Siriwardhana, P. Porambage, M. Liyanage, and M. Ylianttila, “AI and 6G Security: Opportunities and Challenges,” 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Porto, Portugal, 2021, pp. 616-621. DOI: 10.1109/EuCNC/6GSummit51104.2021.9482503.

[SRS+20] Nasim Soltani, Guillem Reus-Muns, Batool Salehi, Jennifer Dy, Stratis Ioannidis, and Kaushik Chowdhury. "RF Fingerprinting Unmanned Aerial Vehicles with Non-standard Transmitter Waveforms." IEEE Transactions on Vehicular Technology, vol. 69, no. 12, pp. 15518–15531, Dec. 2020.

[SSD+20] Nasim Soltani, Kunal Sankhe, Jennifer Dy, Stratis Ioannidis, and Kaushik Chowdhury, "More Is Better: Data Augmentation for Channel-Resilient RF Fingerprinting," in IEEE Communications Magazine, vol. 58, no. 10, pp. 66-72, Oct. 2020. DOI: 10.1109/MCOM.001.2000180.

[SYZ13] Clayton Shepard, Hang Yu, and Lin Zhong. "ArgosV2: A Flexible Many-Antenna Research Platform," Proceedings of the 19th Annual International Conference on Mobile Computing & Networking (MobiCom '13), Miami, Florida, USA, 2013, pp. 163–166, doi: 10.1145/2500423.2505302.

[SZM+22] G. Shen, J. Zhang, and A. Marshall, “LoRa_RFFI_dataset,” IEEE Dataport, February 3, 2022. DOI: https://dx.doi.org/10.21227/qqt4-kz19.

[SZM+24] G. Shen, J. Zhang, A. Marshall, R. Woods, J. Cavallaro, and L. Chen. "Towards Receiver-Agnostic and Collaborative Radio Frequency Fingerprint Identification." IEEE Trans. Mobile Comput., vol. 23, no. 7, pp. 7618-7634, July 2024.

[TCG+23] C. Tassie, V. Chaudhary, A. Gaber, N. Soltani, M. Belgiovine, M. Loehning, V. Kotzsch, C. Schroeder, and K. R. Chowdhury, "Detection of Co-Existing RF Signals in CBRS Using ML: Dataset and API-Based Collection Testbed," IEEE Communications Magazine, vol. 61, no. 9, pp. 82-88, September 2023.

[THW+23] Simon Tewes, Markus Heinrichs, Kevin Weinberger, Rainer Kronberger, and Aydin Sezgin. "A Comprehensive Dataset of RIS-Based Channel Measurements in the 5GHz Band." 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, June 20-23, 2023. DOI: 10.1109/VTC2023-Spring57618.2023.10200973.

[TKK+23] Kürşat Tekbiyik, Güneş Karabulut Kurt, Ali Rıza Ekti, Halim Yanikomeroglu. "Dataset for Channel Estimation in RIS-Assisted Satellite IoT Communications." DOI: 10.21227/261x-mr59. February 19, 2023.

[TKN+24] B. C. Tedeschini, G. Kwon, M. Nicoli, and M. Z. Win, “Real-Time Bayesian Neural Networks for 6G Cooperative Positioning and Tracking,” in IEEE Journal on Selected Areas in Communications, vol. 42, no. 9, pp. 2322-2338, Sept. 2024. DOI: 10.1109/JSAC.2024.3413950.

[UDK20] E. Uzundurukan, Y. Dalveren, and A. Kara, “A Database for Radio Frequency Fingerprinting of Bluetooth Devices,” Data, vol. 5, no. 2, 2020. DOI: 10.3390/data5020055. DOI: 10.5281/zenodo.3876140.

[WKS21] P. Walther, R. Knauer, and T. Strufe, "Ultra-Wideband Channel State Information and Localization for Physical Layer Security," IEEE Dataport, February 16, 2021. Available at: https://dx.doi.org/10.21227/0wej-bc28.

[WSM+20] Jason Whelan, Thanigajan Sangarapillai, Omar Minawi, Abdulaziz Almehmadi, Khalil El-Khatib. "UAV Attack Dataset," IEEE Dataport, February 26, 2020. doi: https://dx.doi.org/10.21227/00dg-0d12.

[Xia23] Jian Xiao. "RIS_CE." IEEE Dataport, January 16, 2023. DOI: https://dx.doi.org/10.21227/3c2t-dz81.

[Yan24] W. Yan, “IEEE 802.15.4 Backscatter Radio Frequency Fingerprinting,” IEEE Dataport, April 2, 2024. DOI: https://dx.doi.org/10.21227/26n7-kx31.

[YL23] X. Yang and D. Li, "LTE_RFF_IDENTIFICATION_DATASET," IEEE Dataport, September 25, 2023. Available at: https://dx.doi.org/10.21227/yzef-4378.

[Zha23] X. Zha, "Toward receiver, modulation, carrier and symbol rate agnostic SEI Dataset," IEEE Dataport, August 10, 2023. Available at: https://dx.doi.org/10.21227/ebpb-mn26.

[ZZS18] Zhang, X., Zhong, L., and Sabharwal, A. "Directional Training for FDD Massive MIMO." IEEE Transactions on Wireless Communications, vol. 17, no. 8, pp. 5183-5197, Aug. 2018. doi:10.1109/TWC.2018.2838600.