IEEE 802.11g/n Channel State Information (CSI) Datasets

Nora Basha and Bechir Hamdaoui

EECS, Oregon State University
Version 1 - April 2025

1 General Description

CSI datasets collected from Pycom Lopy, Luatos ESP32, XIAO ESP32, and USRP B210 devices under different scenarios are described and provided in this document. These datasets are used for the analysis of the proposed Wavelet-based reciprocity enhancement and secret key generation presented in [1, 2] and the analysis of the proposed CSI authentication in [3].

2 CSI Datasets Using ESP32-based Devices

This section provides CSI datasets collected using ESP32-based development boards/devices: Pycom (LoPy model 4 devices connected to PySense extension boards), LuatOS ESP32 boards, and XIAO ESP32 boards using the CSI toolkit presented in [4].

2.1 Two-Node CSI Collection Setup

Two Pycom devices, one serving as an Access Point (AP) with MAC address 3E:71:BF:87:6D:94 and one as station (STA) with MAC address 3C:71:BF:87:6D:A4, communicate using IEEE 802.11n WiFi protocol at 2.427 GHz. AP and STA exchange packets at a rate of 10 packets per second (in an alternating manner; AP sends 10 packets, then STA sends 10 packets, then AP sends 10 packets, etc.) and collect CSI (Channel State Information) using the ESP32 CSI toolkit[4]. Each packet yields one CSI sample point. The CSI datasets were collected on the same building floor, while considering the following 3 different location scenarios:

- **LoS-ShortRange:** A line-of-sight (LoS) scenario, where both AP and STA are located in a room of size 9 meters $\times$ 9.6 meters and separated by a distance of 6.5 meters.
- **NLoS-ShortRange:** A non-line of sight (NLoS) scenario, where AP is located in a room and STA is located in an adjacent corridor with scarce human movement.
- **NLoS-LongRange:** A NLoS scenario, where AP and STA are in two different rooms, and the two rooms are about 13 meters apart with frequent human movement.

The following three links provide access to datasets collected in each of the three location scenarios. Each link contains 8 files, four corresponding to the AP and four to the STA, with each of the four files corresponding to a different experiment. Each filename indicates the device type (AP/STA), the location scenario, and the day of the experiment.

- LoS-ShortRange Scenario
- NLoS-ShortRange Scenario
- NLoS-LongRange Scenario

The following link provides access to datasets collected in a LoS location scenario but at 7 different time blocks (each of duration 10 minutes) over two days. Each filename indicates the device type (AP/STA), the day of data collection in year-month-day, and the time of the day in hours and minutes.

- LoS-ShortRange Scenario at Different Time Blocks

Table 1 explains what each of the 26 columns in the file refers to. Column 26 is a vector of the estimated channel response for the OFDM subcarriers represented as the real part of the estimated channel for a subcarrier followed by the imaginary part of the estimated channel for the same subcarrier. For example, the first and second elements of the CSI vector are the real and the imaginary parts of the estimated channel response for the first subcarrier, respectively; the third and fourth elements correspond to the real and imaginary for second subcarrier, etc. More information on the subcarriers indexing is provided at Wi-Fi Channel State Information.

Column	Description
Column 1	CSI_DATA
Column 2	Device role: AP or STA
Column 3	Source MAC address of the received frame
Column 4	Received signal strength indicator (RSSI) in dBm
Column 5	PHY rate encoding of the packet. Only valid for non-HT(11bg) frames.
Column 6	Signal mode: 0 for non HT(11bg) packet; 1 for HT(11n) frames.
Column 7	Modulation Coding Scheme.
Column 8	Channel bandwidth. 0 for 20MHz; 1 for 40MHz.
Column 10	Sounding frames. Takes the value 0 for sounding frames. Sounding frames are non-HT frames in AP data only.
Column 12	Space Time Block Code (STBC). Value is 0 as no STBC is used.
Column 13	low-density parity check frames. Only valid for HT (n) frames.
Column 14	Short Guide Interval (SGI). 0 for Long GI; 1 for Short GI.
Column 15	Radio frequency module noise floor in dBm.
Column 17	WiFi channel. The channels 4 and 6 are used in data collection.
Column 18	The offset of the secondary channel. Takes the value 0 for non-HT (bg) and 1 for HT (n).
Column 19	The local time when this packet is received in microseconds.
Column 21	Length of the frame including Frame Check Sequence (FCS).
Column 22	Number of errors in the packet.
Column 25	Length of CSI vector. Takes the value 128 for non-HT (bg) and 384 for HT (n).
Column 26	CSI vector

Table 1: Columns description for data collected using Lopy devices.

2.2 Four-Node Network CSI Collection Setup Under Different Environments

This setting consists of a WiFi network of 4 devices with the following hardware configurations:

- Two **Pycom** devices, one serving as AP with MAC address 3E:71:BF:87:6D:94 and another as STA with MAC address 3C:71:BF:87:6D:A4. The Pycom devices are development boards based on the Espressif ESP32 System on Chip (SoC).
- One **LuatOS** device serving as STA with MAC address EC:DA:3B:D1:20:34. LuatOS boards are based on the Espressif ESP32-C3 SoC.
- One Seeed Studio **XIAO** device serving as STA with MAC address 64:E8:33:86:E5:D4. Seeed Studio XIAO devices are compact development board based on the Espressif ESP32-C3 SoC with an external antenna to increase the signal strength.

AP and STAs communicate using IEEE 802.11n WiFi protocol at 2.427 GHz. AP exchanges with each STA packets at a rate of 10 packets per second. The AP and STAs collect CSI data using the ESP23 CSI toolkit, while considering two different RF-rich environments:

- **Indoor, RF-rich environment:** The devices are located in a room of size 9 m $\times$ 9.6 m in the RF-rich environment of Kelley Engineering Center at Oregon State University (OSU), within the range of OSU's other wireless network services (including Bluetooth building management and WiFi system).
- **Outdoor, RF-rich environment:** The devices are located outdoors on OSU campus in an RF-rich environment within the range of the wireless service of OSU and other RF interference sources. The outdoor environment is exposed to sunlight and at 30C.

For both environments, STAs and AP are placed at different distances. For each environment, we collected several raw CSI datasets from each device in the network. Each of the following links provides access to the CSI datasets for the indoor and outdoor scenarios. Each filename represents the device type (AP/STA), the device hardware configurations, and the time (block number) of the experiment; all experiments are taken in one day.

- Indoor Scenario
- Outdoor Scenario

3 CSI Datasets Using USRP B210 Devices

This section provides CSI datasets collected using two USRP devices (Sec.3.1) and CSI datasets collected using two USRP devices under a replay attack launched by a HackRF One device (Sec.3.2).

3.1 CSI Datasets Using USRPs

Two USRP B210 devices (serving as an AP and an STA) utilize IEEE 802.11g protocol to exchange packets at 2.427 GHz and estimate CSI for 1 minute under a LoS setting. The following link provides access the CSI

estimated at AP and STA datasets for 4 experiments. Each filename indicates the device type (AP/STA) and the experiment day (each experiment is taken on a different day).

- USRPs CSI at AP and STA

3.2 CSI Datasets Under Replay Attack

Two USRP B210 and one HackRF One devices serve as an AP, an STA, and an attacker, respectively in an MiTM attack scenario. Utilizing IEEE 802.11g, the USRP devices exchange packets at a rate of 1 packet per second at 2.427 GHz and estimate CSI for 1 minute under a LoS setting. The HackRF attacker records RF signals exchanged between AP and STA and later attempts to replay signals to AP to launch the replay attack and authenticate to AP. The details of the attack is described in [3]. The provided datasets below include the magnitude of the CSI obtained at AP by exchanging packets with a STA as well as the CSI obtained at AP when a STA replays previously recorded signals to AP. Each of the links below provides access to 5 files: (a) one file named "ap_CSI.bin" which includes CSI that AP estimated by exchanging packets with a current legitimate STA willing to connect, as well as (b) 4 files named "ap_CSI_replay_attempt.bin" which include CSI obtained at AP when a STA tries to connect to AP and replays previously recorded signals to AP. The attempt number represents STA's 4 attempts to connect to AP by replaying old recorded signals to AP. Different experiments provide CSI data for the attack scenario at the same location on different days.

- Experiment 1
- Experiment 2
- Experiment 3
- Experiment 4

The following is a Python code to convert the binary files into CSI magnitudes.

```
filename = 'csi_reciprocal_USRP1_2_5.bin'
f = np.fromfile(open(filename), dtype=np.float32)
#parsing CSI magnitude samples for 52 subcarriers:
parsed_samples = np.zeros([int(len(f)/52),52], dtype= np.float32)
for i in range(0,int(len(f)/52)):
    parsed_samples[i,:] = f[i:i+52]
```

References

[1] N. Basha, B. Hamdaoui, and D. A. Al-Fuqaha, "Enhancing Wireless Secret-Key Generation Through Time-Frequency Analysis Using Wavelet Coherence," *IEEE Global Communications Conference*, 2024.

[2] N. Basha and B. Hamdaoui, "Wavelet-based csi reconstruction for improved wireless security through channel reciprocity," *Computers Security*, vol. 154, p. 104423, 2025.

[3] N. Basha, B. Hamdaoui, A. Erbad, and M. Guizani, "On the Detection of Replay Authentication Attacks Through Channel State Information Analysis," *IEEE Global Communications Conference*, 2024.

[4] S. M. Hernandez and E. Bulut, "Lightweight and Standalone IoT Based WiFi Sensing for Active Repositioning and Mobility," in *21st International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM) (WoWMoM 2020)*, June 2020.