Forensic Insights from Electromagnetic Radiation

Workshop at DFRWS APAC 2023

Dr. Asanka P. Sayakkara

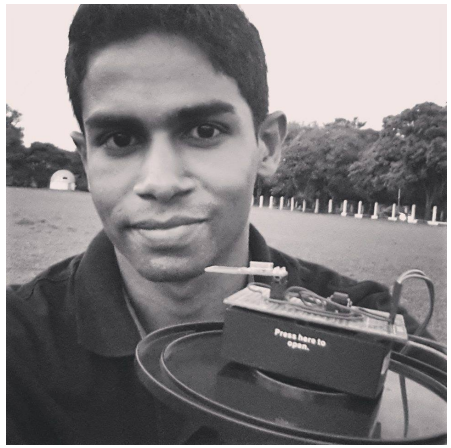
(asa@ucsc.cmb.ac.lk)

Suntec Convention & Exhibition Centre,
Singapore.

17th October, 2023

Asanka P. Sayakkara

- BSc in Computer Science, University of Colombo School of Computing (UCSC), 2012.
- PhD in Computer Science from University College Dublin, Ireland, 2020.
- Forensic & Security Research (ForSec) group of University College Dublin, Ireland, 2017–2020.
- Senior lecturer at University of Colombo School of Computing (UCSC), Sri Lanka.
- Coordinator of the MCS/MSc in CS degree programs.



- Introduction to Computing (FoS), Digital Forensics, Embedded Systems, and Operating Systems II.
- Running *Signal Insights* research lab.
- `https://ucsc.cmb.ac.lk/profile/asa`
  `https://www.asayakkara.org`

Signal Insights Research Lab

- Explore the potential of exploiting various kinds of signals originating from various sources.
- Signal sources: artificial, as well as biological sources (bioacoustics).
- Electromagnetic side-channels and covert channels.
- Radio tomographic imaging.
- Passive acoustic monitoring (of elephants).
- `https://www.asayakkara.org/signal-insights-lab.html`

Workshop Agenda

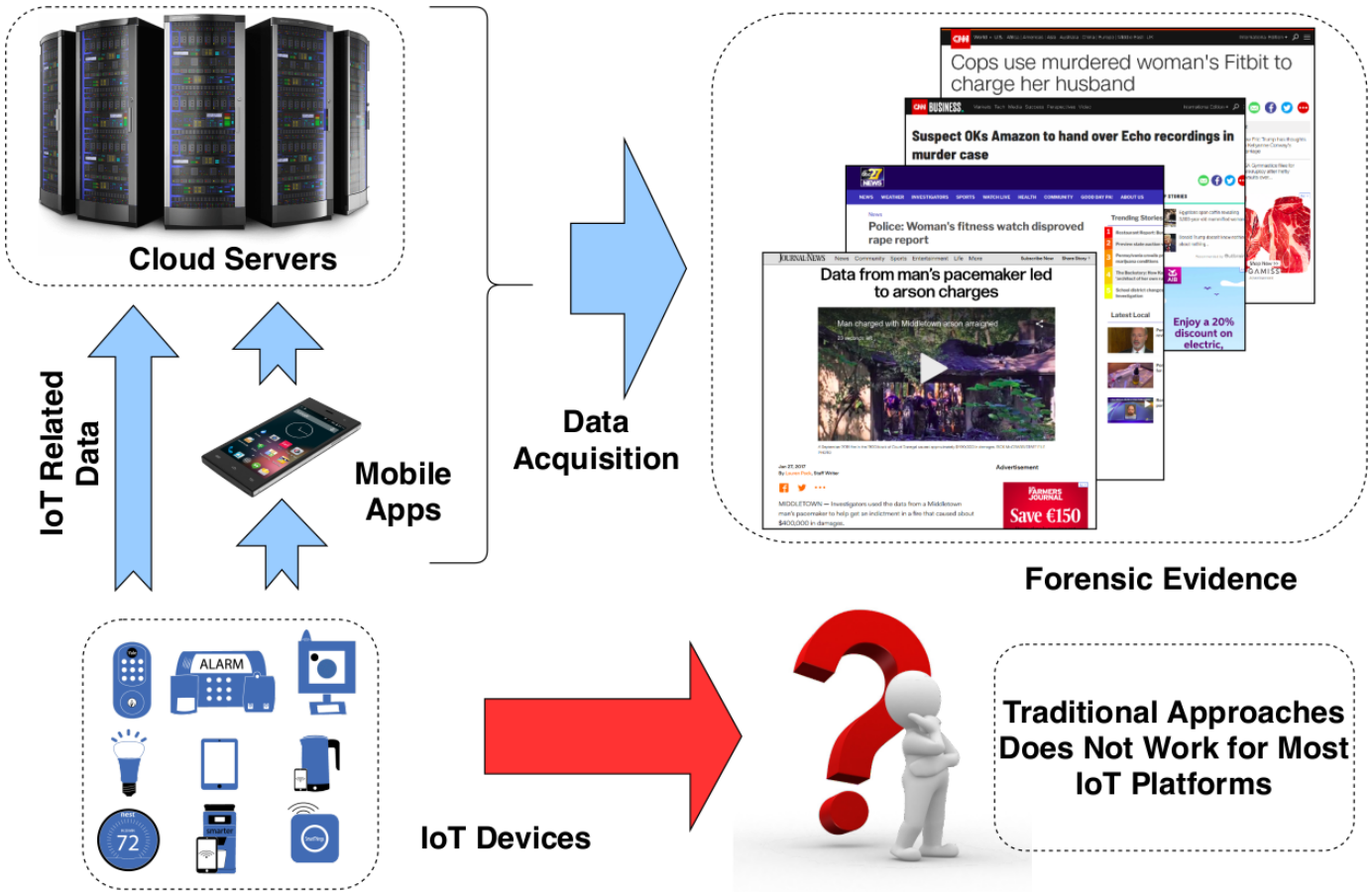
- **Part 1:**
  - Introduction and background
  - SDR hardware.
  - SDR software.
  - Capturing EM side-channel radiation.

  Break

- **Part 2:**
  - Analysing EM datasets using Python.
  - Build your own Arduino program classifier.
  - Discussion and conclusion

Introduction and Background

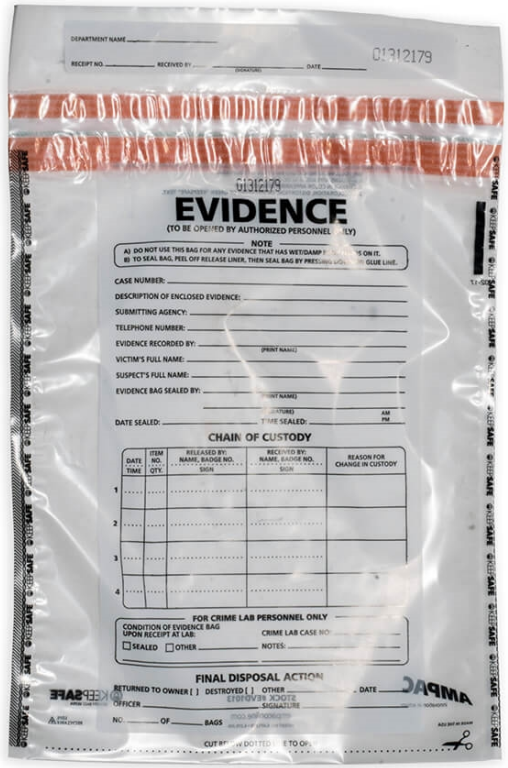
Challenge of IoT & Smart Device Forensics



Challenge of IoT & Smart Device Forensics (cont.)

- New devices are emerging in the market too frequently.
- The internal components, storage, and interfaces of the devices varies drastically.
- Analysing devices at too close to the hardware level — such as chip-off forensics — is susceptibe to irreversible mistakes that can destroy a device entirely.
- It would be ideal if we can inspect a device from a safe distance.

Evidence vs Insights



- Digital **evidence** are information that may be presented to a court of law.
- They need to be concrete enough to be relied upon at the courts.
- The field of digital forensics is aimed at providing this reliability as much as possibile.
- In some situations, where evidence are not available, some **insights** can a lifeline for an investigator.
- Insights are — most likely — not reproducible, but they can provide useful hints and directions to go and locate reliable evidence through other means.

Forensic Insights from IoT and Smart Devices

- Is this device running the official firmware from the manufacturer?
- Has a malware been injected to the memory of this device?
- Is this device doing something it is not supposed to be doing right now; such as wiping the storage or encrypting it, instead of shutting down?

Electromagnetic Side-Channel Analysis

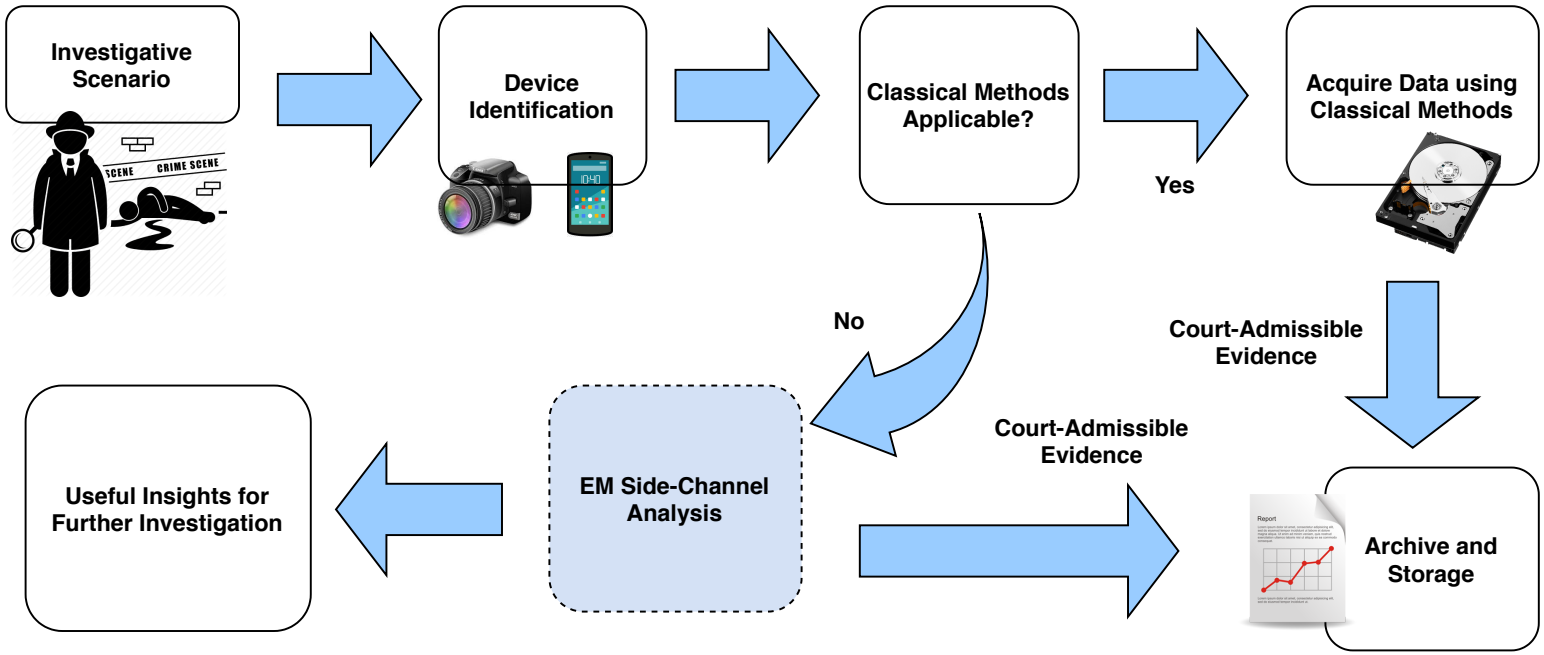
- Time-varying electrical currents are generating electromagnetic (EM) radiation.
- Our electronic equipment are a source of strong EM radiation.
- The EM radiation of computers (specifically, the processors) is shown to be correlating with the software running on them, i.e., the exact instructions and their execution pattern.
- EM Side-Channel Analysis (EM-SCA) is the exploitation of these radiation to eavesdrop on computers:
  - Software behaviour dection.
  - Malicious firmware modification detection.
  - Cryptographic key retrieval.

Electromagnetic Side-Channel Analysis (cont.)

A Few Demonstrations

- Radiation from the laptop screen/graphic card:
  `https://www.youtube.com/watch?v=YtolwTPDBwk`
- Remote servelliance of video displays:
  `http://www.youtube.com/watch?v=8OlkywZBJGU`
- Data exfiltration through EM covert channel on an Ethernet cable:
  `https://www.youtube.com/watch?v=ciM4M5h3q0w`

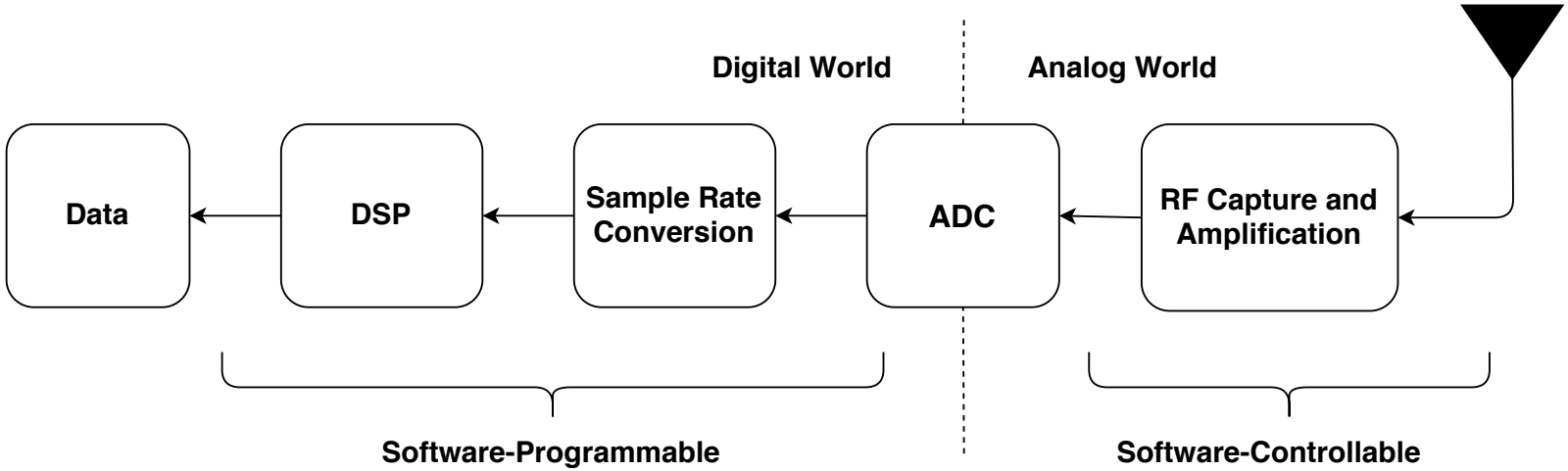
Forensic Insights through EM-SCA



Software Defined Radios

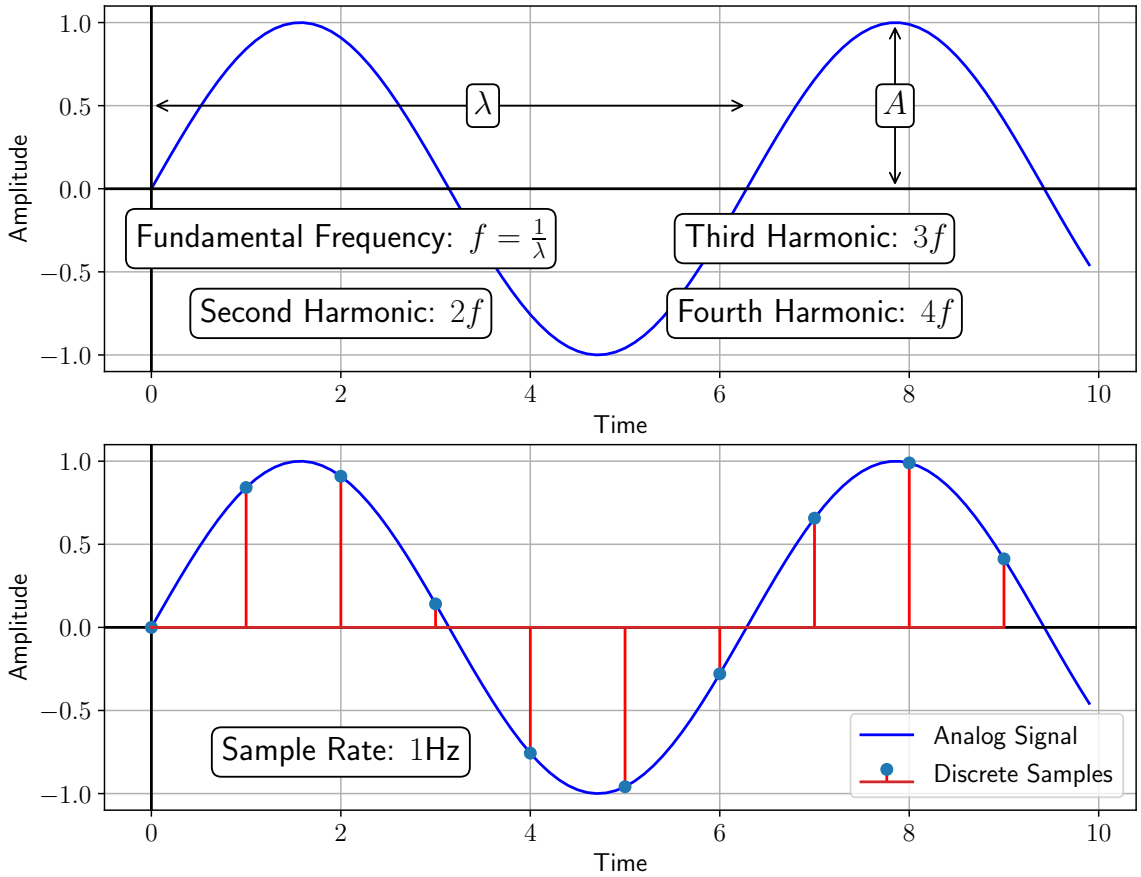
Software Defined Radios

- Moving most of radio functions from analog domain into the digital domain.
- Requires a generic hardware radio interface including a very fast analog-to-digital converter (ADC).
- Need sophisticated and optimised software implementations for digital signal processing (DSP).



Software Defined Radios (cont.)

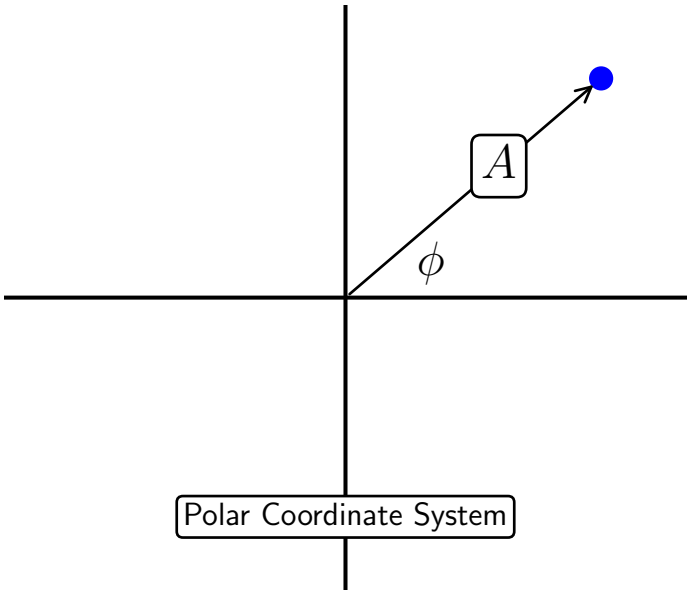
Converting analog signals to digital signals require **sampling** and **quantisation**.

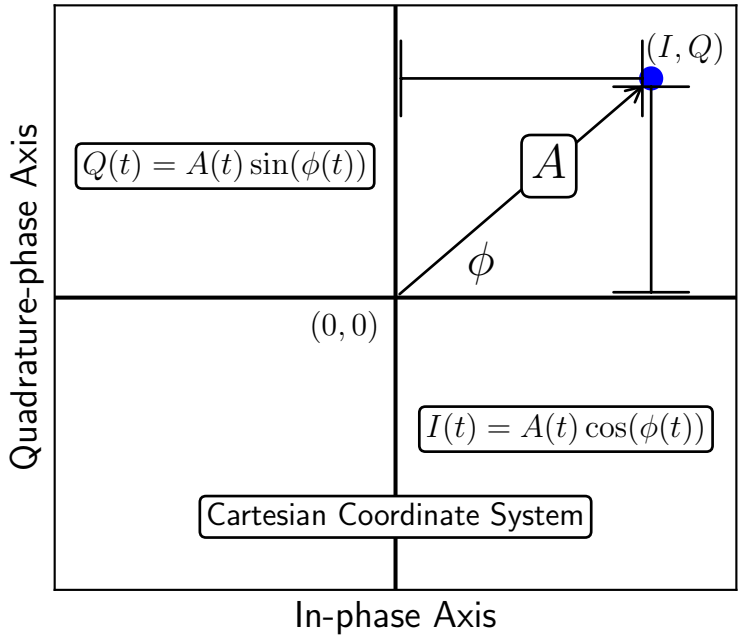


Real-valued sampling faces the Nyquist limit: difficult to capture high frequencies.

Software Defined Radios (cont.)

- SDRs are using complex **In-phase/Quadrature-phase (I/Q)** sampling.
- Each sample taken at a given time instance consists of two values; hence, each sample is a complex number.
- A EM signal captured by an SDR is basically an array of complex numbers.
- Sampling rate is equal to the bandwidth of the captured data.





Software Defined Radios (cont.)

RTL-SDR

- A digital TV tuner repurposed as an SDR.
- The cheapest possible SDR you may find.
- A wide variety of manufacturers; hence different variations of capabilities.
- Sample rate is about 3.2 MHz
- Tunable frequency range: 22 MHz – 1 GHz.
- Receive only; no transmission.

Software Defined Radios (cont.)

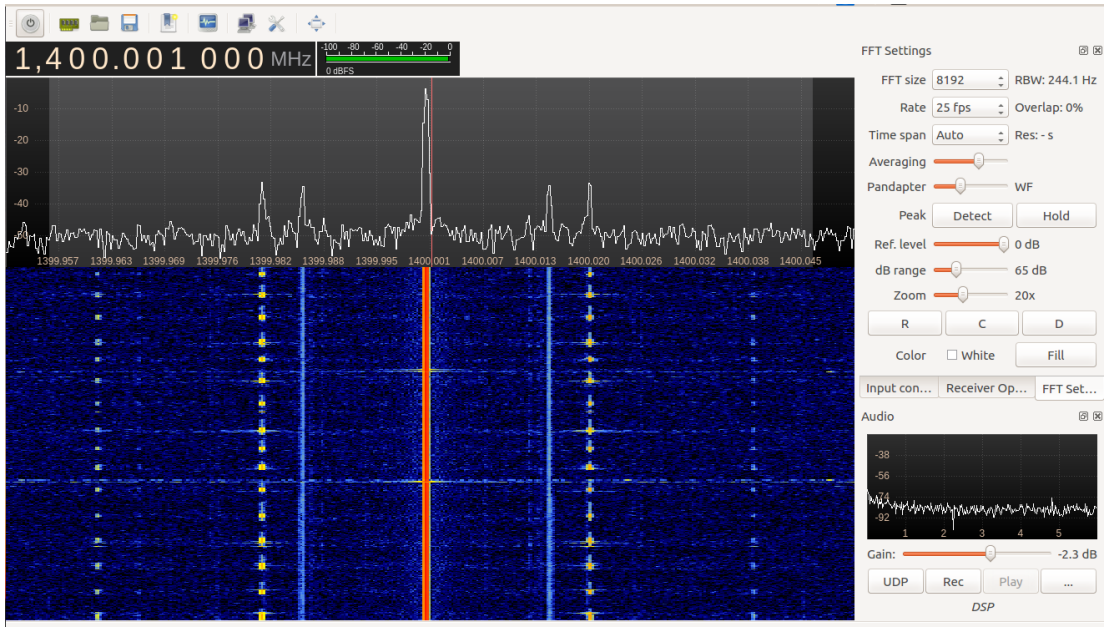
HackRF One

- A purpose-built SDR device.
- A mid-range price tag.
- Sample rate: upto 20 MHz.
- Tunable frequency range: 1 MHz – 6 GHz.
- A wide range of antennas can be connect through the SMA connector.
- Half-duplex: either transmit or receive at a given time.
- Possible to time-synchronise with another device through clock input or output.

Software Defined Radios (cont.)

GQRX

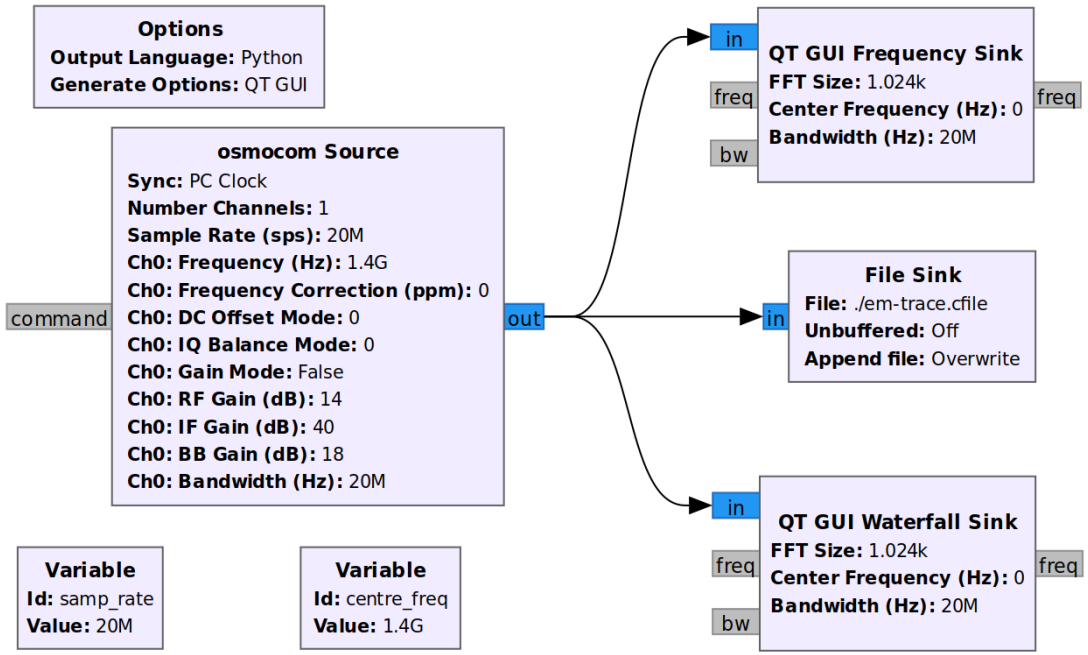
- Easily view signals at different frequencies.
- Facilitates live processing and saving observed signals.
- Uses *GNURadio* library underneath.



Software Defined Radios (cont.)

GNURadio Companion (GRC)

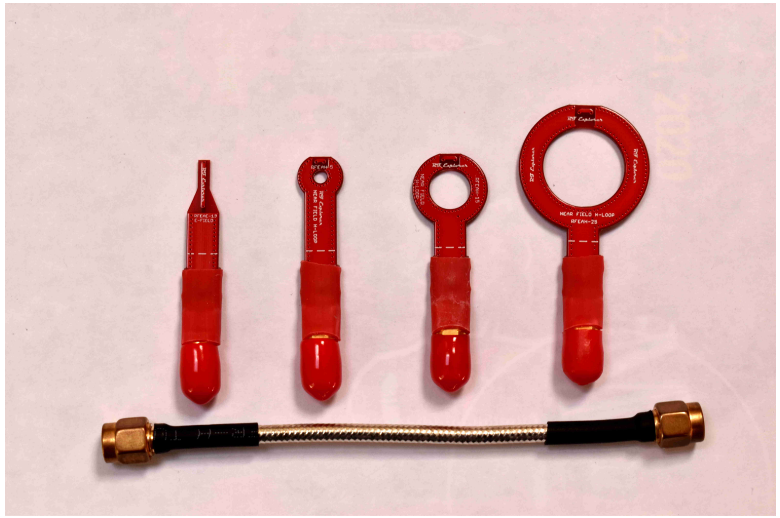
- Various signal processing blocks.
- Custom build any application by creating flow graph using blocks.
- Generates executable Python scripts for flow graphs, using *GNURadio* library.



Capturing EM Side-Channel Radiation

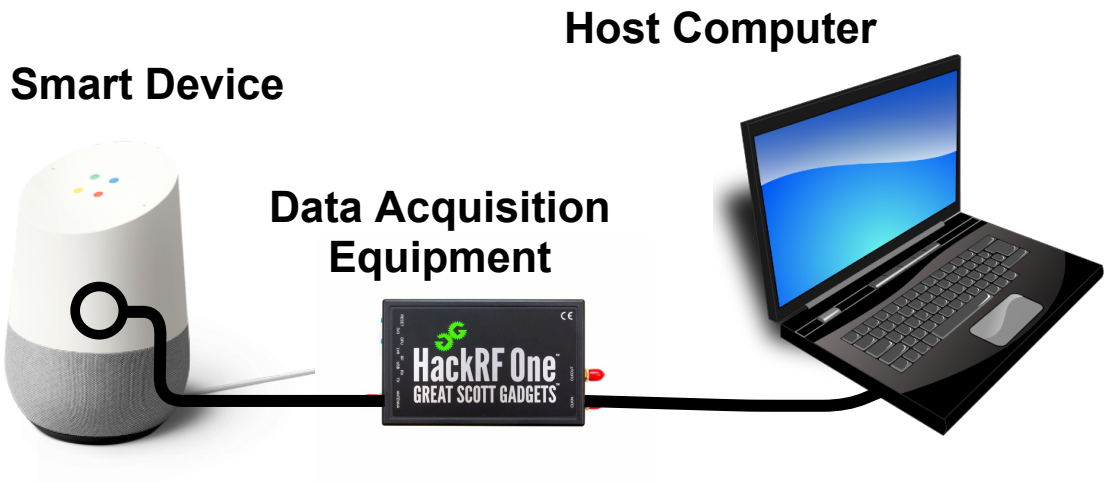
Capturing EM Side-Channel Radiation

- Our key focus is EM radiation from the processor/microcontroller/SoC.
- Strongest signals are in the clock frequency or its harmonics.
- Signal acquisition should be performed as closer to the target chip as possible.
- Magnetic H-loop antennas are more suitable for the job.



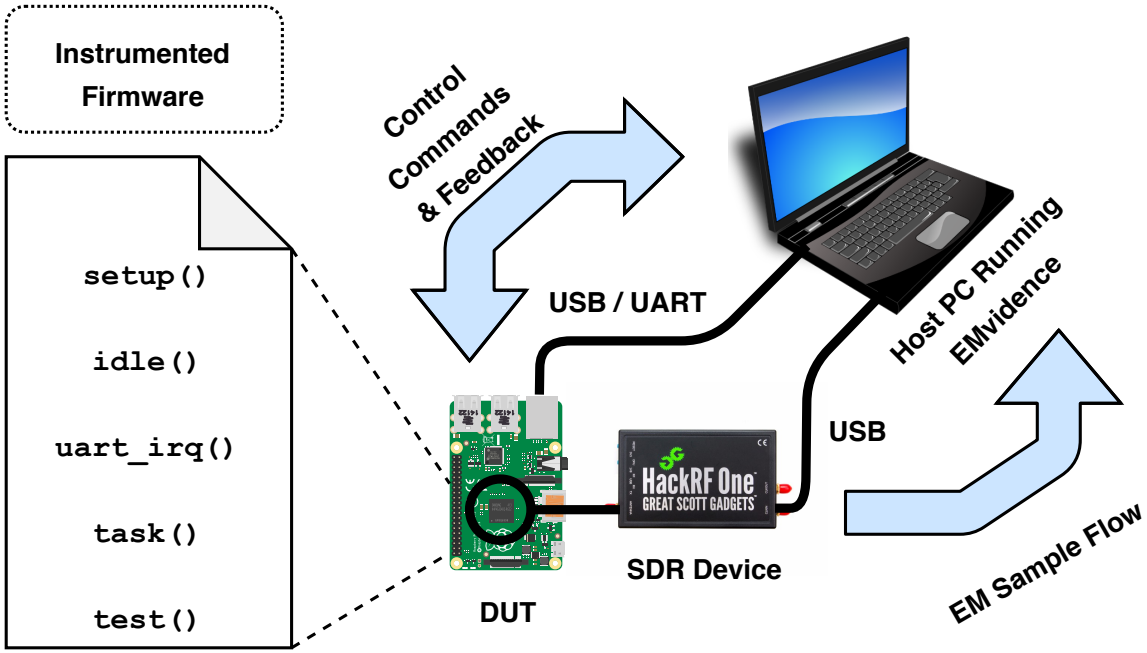
Capturing EM Side-Channel Radiation (cont.)

Passive acquisition:



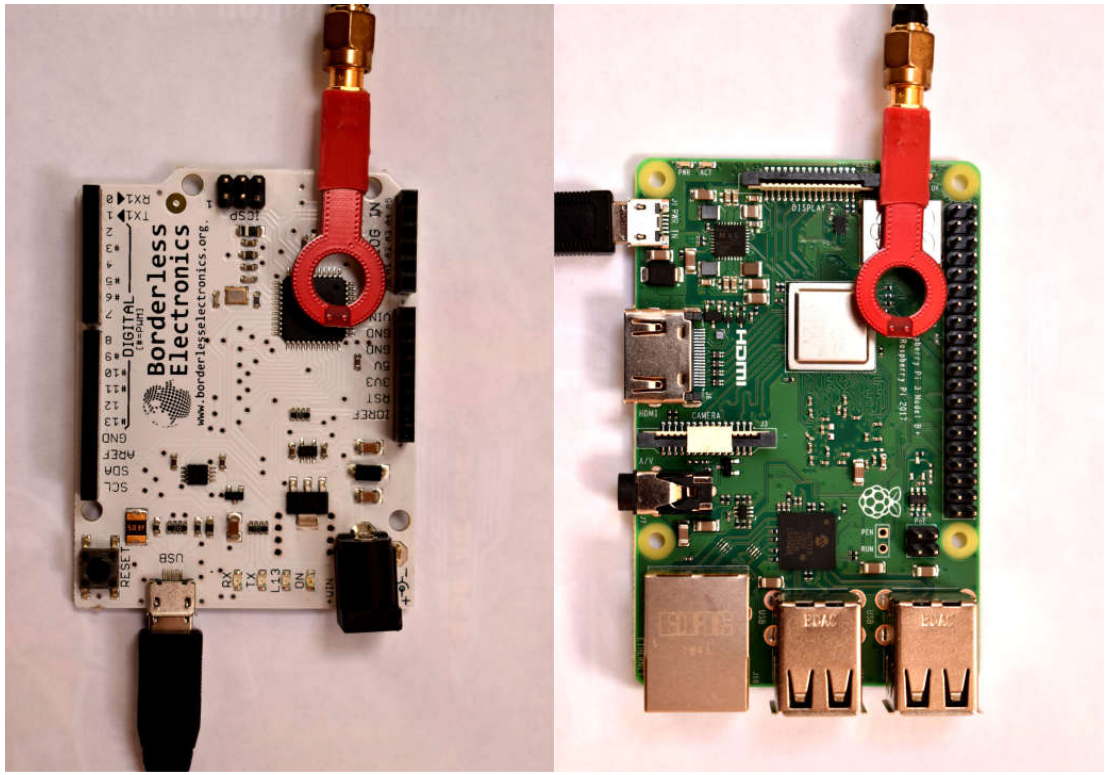
Capturing EM Side-Channel Radiation (cont.)

Instrumented acquisition:



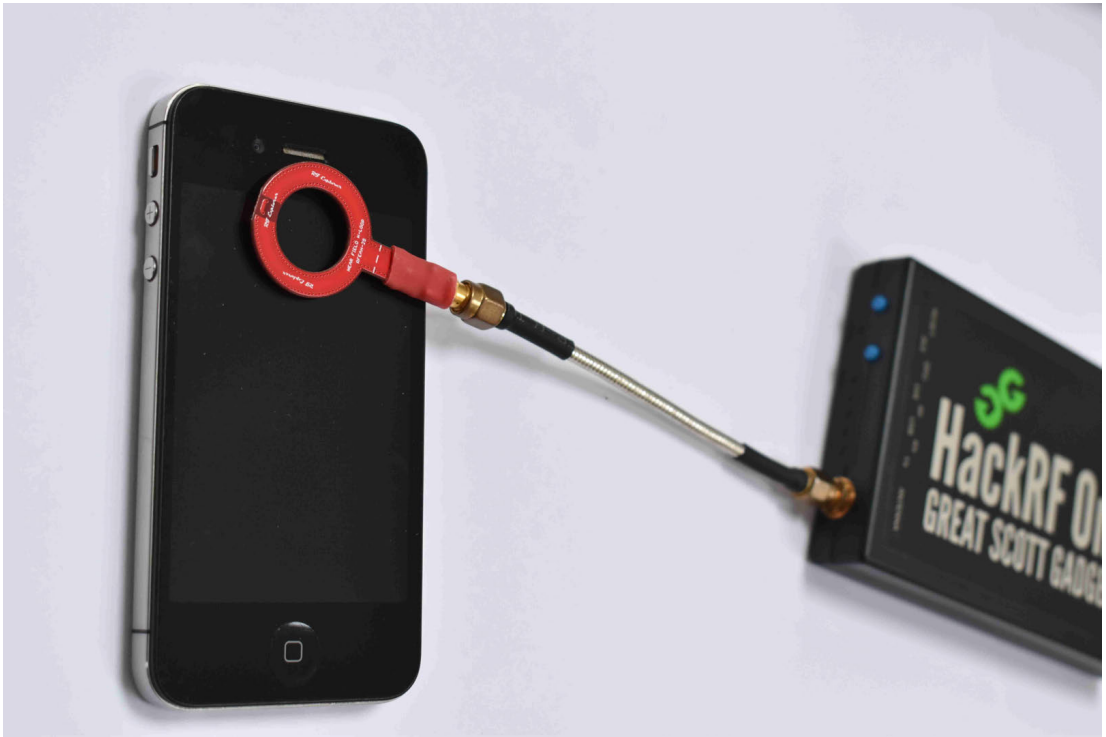
Capturing EM Side-Channel Radiation (cont.)

Arduino and Raspberry Pi



Capturing EM Side-Channel Radiation (cont.)

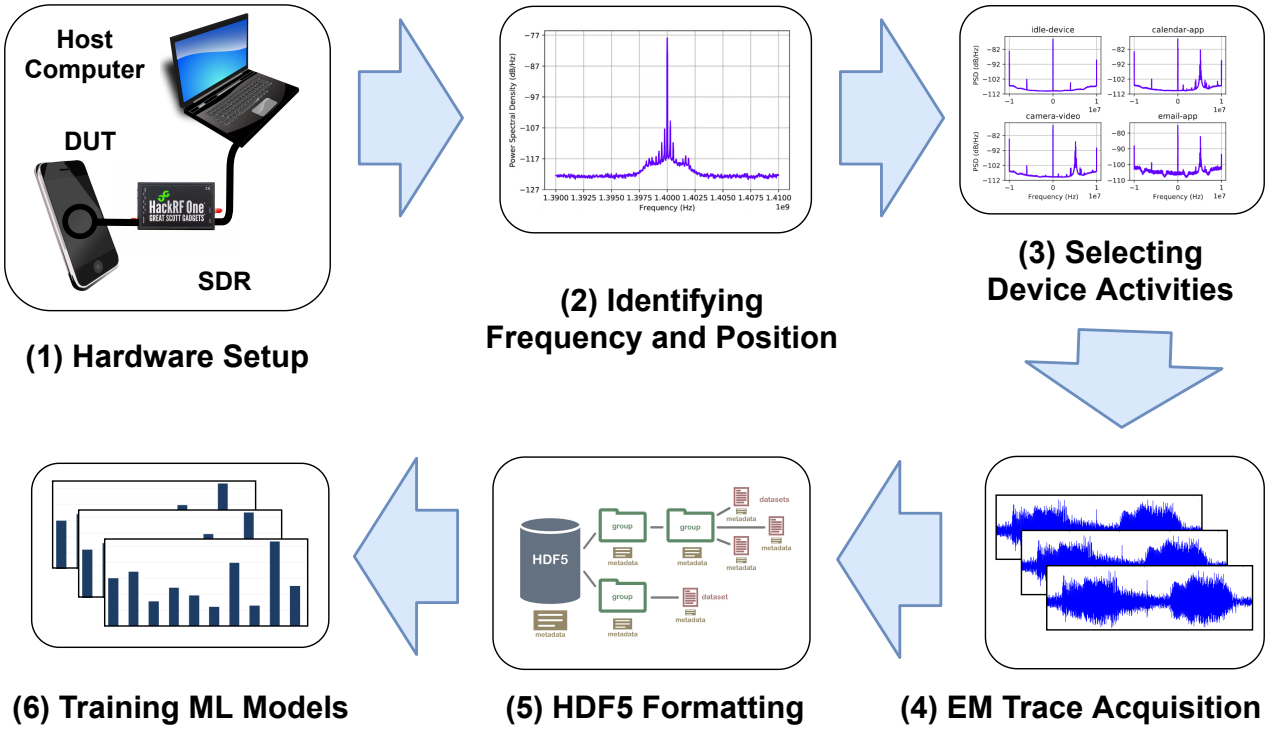
Smartphone



Analysing EM Dataset

Analysing EM Dataset

The pipeline from capturing EM data to analysis...



Analysing EM Dataset

Huge Size of the Data Files

- In GNU Radio library, two 32 bit (4 byte) floating point values are used to represent a complex I/Q sample.
- Therefore, each EM data sample is a 8 bytes long complex value.
- Consider if we sampled data at the maximum sample rate of HackRF One device (i.e., 20 MHz) using GNURadio library to save data.
- Size of data per second $= 8$ bytes $\times 20 \times 10^6 = 160$ MB
- Size of data for 10 seconds $= 160$ MB $\times 10 = 1.6$ GB

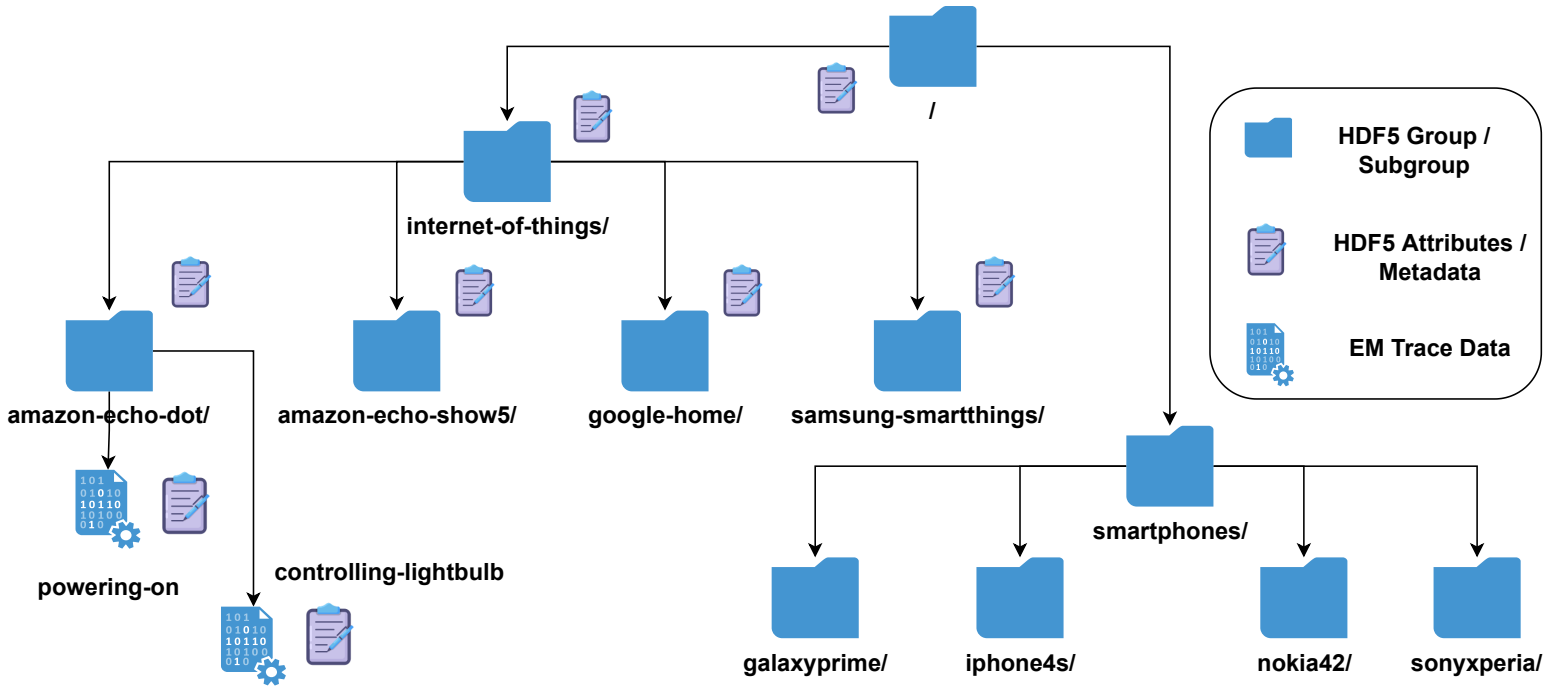
Analysing EM Dataset (cont.)

The specifications of devices in the dataset:

Smart Device	System-on-Chip	Architecture	CPU Frequency	Software Activities
Amazon Echo Show 5	MediaTek MT 8163	ARMv 8-A	1.5 GHz (4 cores)	(1) asking a definition, (2) asking for time, (3) asking to play radio, (4) controlling light-bulb, (5) device idle, (6) device resetting, (7) just wake up word, (8) powering off, (9) powering on.
Amazon Echo Dot (3rd Gen)	Mediatek MT 8516	ARMv 8-A	1.3 GHz (4 cores)	(1) asking a definition, (2) asking for time, (3) asking to play radio, (4) controlling light-bulb, (5) device idle, (6) device muted (7) device resetting, (8) just wakeup word, (9) powering on.
Google Home	Marvell 88DE3006 Armada 1500 Mini Plus	ARMv 7	1.2 GHz (2 cores)	(1) asking a definition, (2) asking for time, (3) asking to play radio, (4) controlling light bulb, (5) device idle, (6) device muted (7) device resetting, (8) just wake-up word, (9) powering on.
Samsung SmartThings Hub (v2)	MCIMX6L2DVN10AB	ARMv 7-A	1 GHz (1 core)	(1) controlling smart outlet, (2) device idle, (3) device powered off, (4) device powering on, (5) opening the app, (6) viewing arrival sensor, (7) viewing door sensor, (8) view motion sensor.
Apple iPhone 4S	Apple A5	ARMv 7-A	1 GHz (2 cores)	(1) calendar app, (2) camera photo, (3) camera video, (4) email app, (5) gallery app, (6) home screen, (7) device idle, (8) phone app, (9) SMS app, (10) web browser app.
Sony Xperia T	Qualcomm Snapdragon MSM8260A	ARM v7-A	1.5 GHz (2 cores)	(1) calendar app, (2) camera photo, (3) camera video, (4) email app, (5) gallery app, (6) home screen, (7) device idle, (8) phone app, (9) SMS app, (10) web browser app.
Samsung Galaxy Grand Prime	Qualcomm Snapdragon MSM8916	ARMv 8-A	1.2 GHz (4 cores)	(1) audio recording, (2) camera photo, (3) camera video, (4) email app, (5) gallery app, (6) home screen, (7) device idle, (8) phone app, (9) SMS app, (10) web browser app.
Nokia 4.2	Qualcomm Snapdragon SDM439	ARMv 8-A	1.95 GHz (4 cores), 1.45 GHz (4 cores)	(1) calendar app, (2) camera photo, (3) camera video, (4) email app, (5) gallery app, (6) home screen, (7) device idle, (8) phone app, (9) SMS app, (10) web browser app.

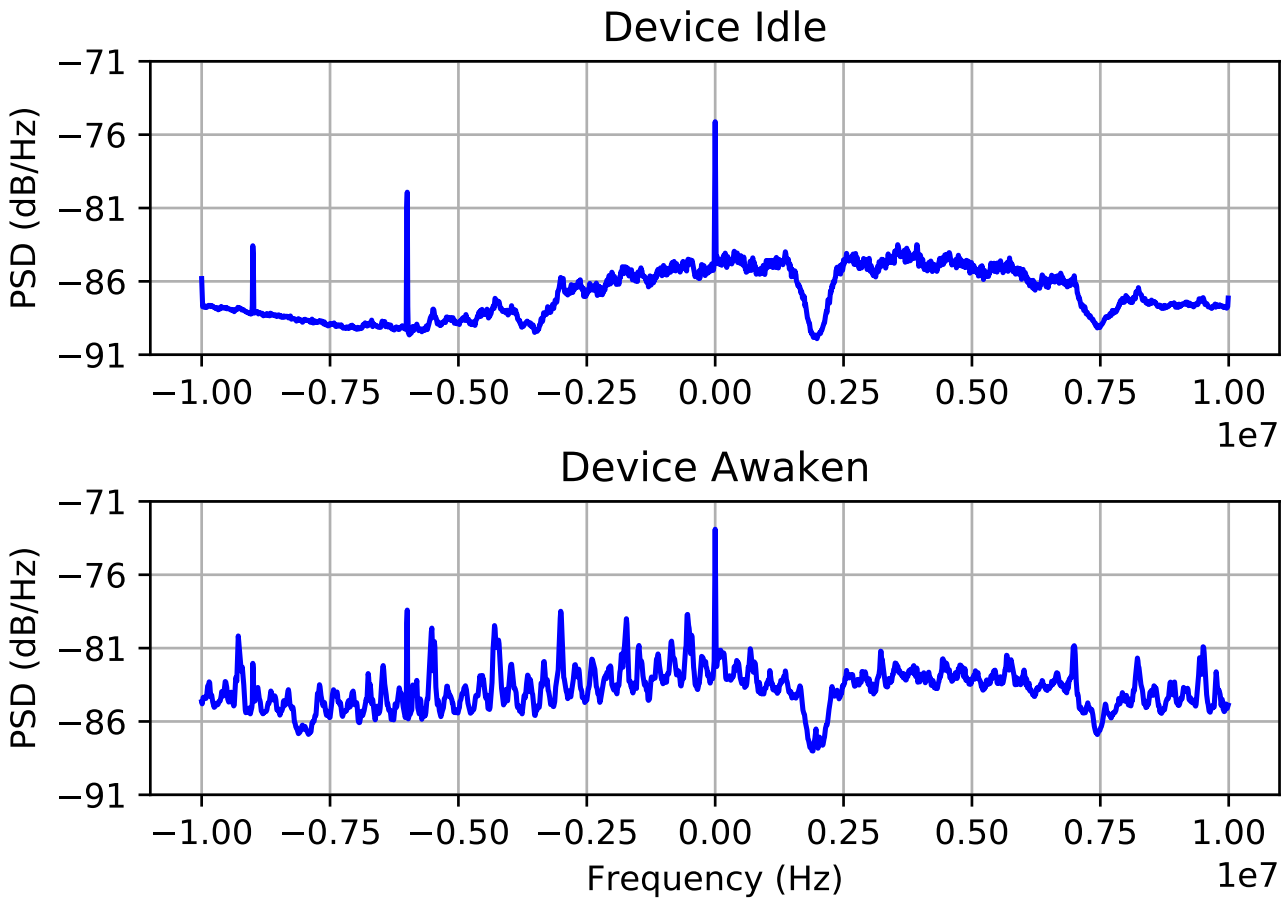
Analysing EM Dataset (cont.)

Structure of the dataset in HDF5 file format (em-dataset.h5):



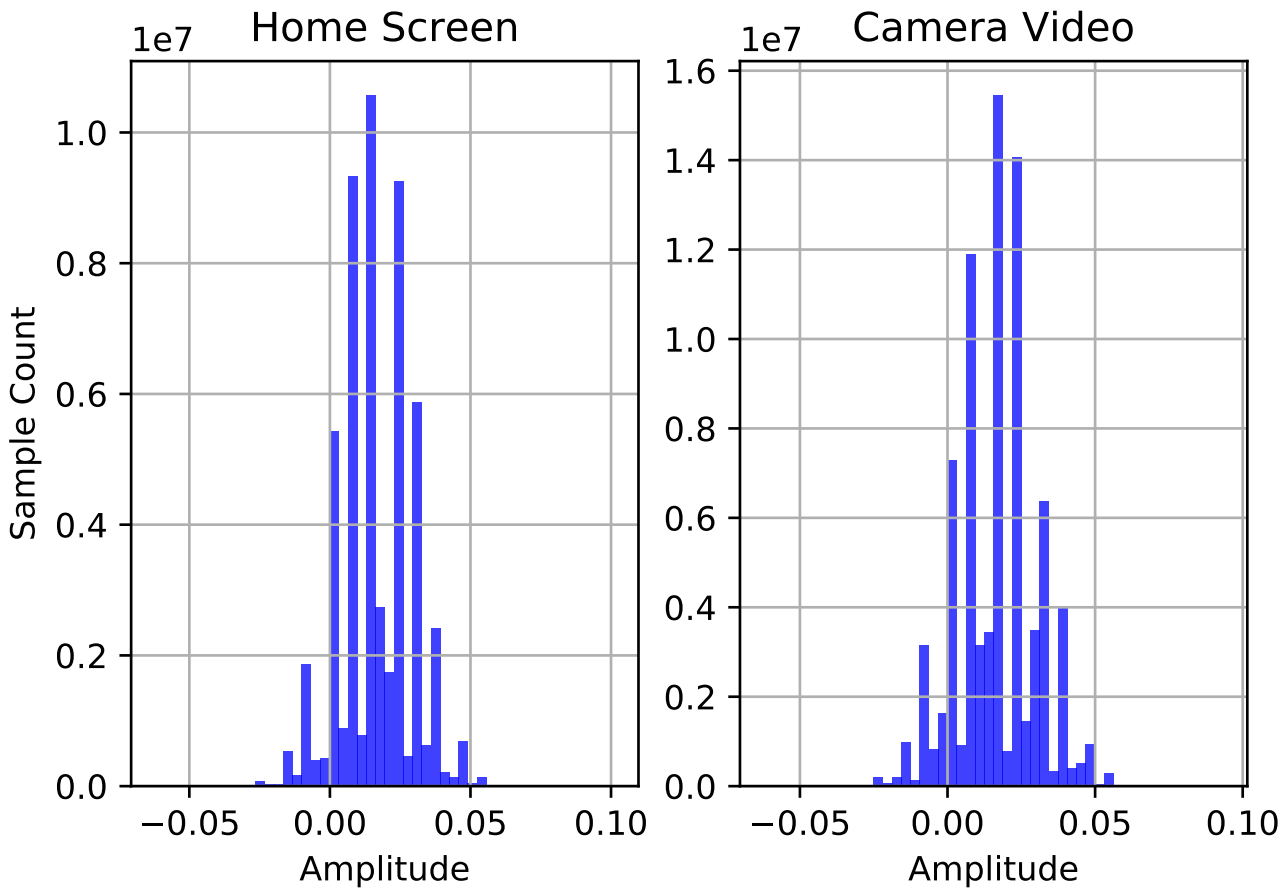
Analysing EM Dataset (cont.)

Amazon Echo Dot – Power Spectral Density (PSD) Plots



Analysing EM Dataset (cont.)

Nokia 4.2 – Histogram



Analysing EM Dataset (cont.)

- Go ahead and launch Jupyter Notebook inside the downloaded Git repository.
- We'll explore the code examples for the following things:
  1. Reading and basic visualisation of EM data.
  2. Preprocessing EM data for feature extraction.
  3. Simple machine learning classifiers to distinguish software behaviour.

Build Your Arduino Classifier

- Shall we collect our own data and build a classifier for software behaviour detection?
- We can program an Arduino Uno to do two different tasks and capture two EM data files representing each task.
- Let's see if we can build a simple classifier to distinguish between the two tasks.

Conclusion

- EM-SCA for digital forensic insight acquisition is still in its early days.
- Loads of technical and scientific problems remaining to be solves; great for research!
- *Cross-device portability* of trained models.
- No need to possess hardware equipment to conduct research in this area; datasets are available to work on (from our group and many others).
- Thank you for your participation. Feel free to get in touch:

  Asanka Sayakkara (asa@ucsc.cmb.ac.lk)

References

1. Asanka Sayakkara, Le-Khac, N-A., and Scanlon, M., "Electromagnetic Side-Channel Attacks: Potential for Progressing Hindered Digital Forensic Analysis", International Workshop on Speculative Side Channel Analysis (WoSSCA 2018), Amsterdam, Netherlands, July 2018.

2. Asanka Sayakkara, Le-Khac, N-A., and Scanlon, M., "A Survey of Electromagnetic Side-Channel Attacks and Discussion on their Case-Progressing Potential for Digital Forensics", Elsevier Digital Investigation, 2019.

3. Asanka Sayakkara, Le-Khac, N-A., and Scanlon, M., "Leveraging Electromagnetic Side-Channel Analysis for the Investigation of IoT Devices", DFRWS USA, Portland, OR, USA, July 2019.

4. Asanka Sayakkara and Nhien-An Le-Khac , "Electromagnetic Side-Channel Analysis for IoT Forensics: Challenges, Framework, and Datasets," in IEEE Access, vol 9, pp. 113585-113598, 2021.

Find more here: `https://www.asayakkara.org/publications.html`