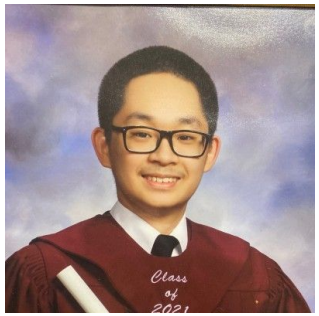
Security in AI

By: Ethan Lung, Damon Lin, Jacob Morin, Rut Mehta

Team Members

Jacob Morin (UG)



Damon Lin (UG)

Ethan Lung (HS)

Rut Mehta (UG)

Prof. Chen (Advisor)

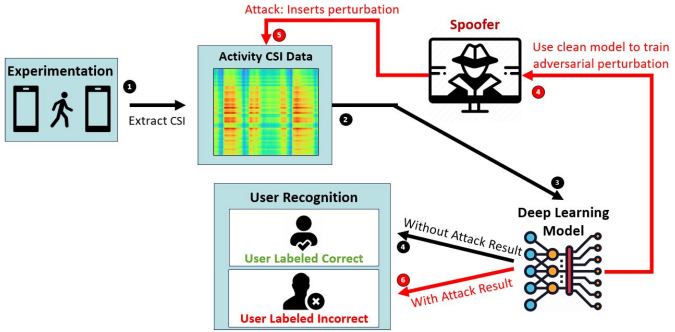
Tianfang Zhang (Mentor)

Changming Li (Mentor)

Honglu Li (Mentor)

Overview and Goal

- Study the Security of WiFi sensing systems under adversarial attack
- Utilize mobile device to extract channel state information (CSI) to train deep learning model for recognition tasks
  - Human Activity Recognition and User Authentication
- Develop a type of adversarial attack algorithm to generate perturbation that can deceive deep learning model



Motivation

- New WiFi sensing techniques use a learning-based approach for **accurate and efficient** recognition tasks
  - Activity Recognition and User Authentication
  - Applied to various commodity devices (i.e., smart phone, laptop, routers)
  - Passive engagement from legitimate user

- Learn-based techniques (i.e., Deep learning) are **vulnerable** against adversarial attack
  - Input data can be easily manipulated
  - Model can be fooled by the attacker to cause false recognition results

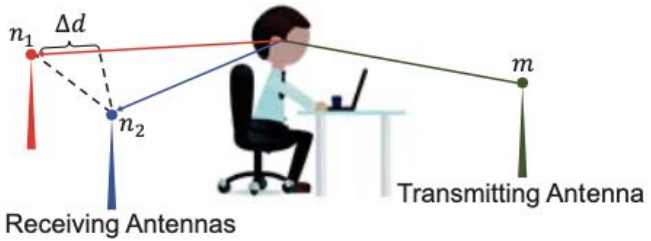
- Research in adversarial attacks against WiFi sensing system can **reveal security issues**
  - Sensing system can be targeted and result serious privacy and security concerns
  - For instance, legitimate user is blocked by his/her property due to the attack

IoT Devices

Internet of Things (IoT) Devices

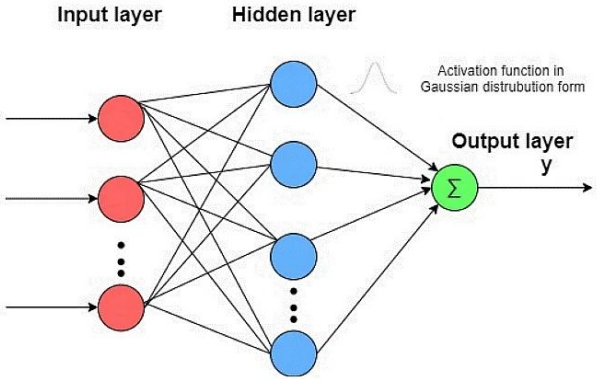
- Connected to internet
- Gather information
- Can pick up interference from
  - behavioral characteristics
    - body movements
- Send and receive data through wireless channels

- **Channel State Information** (CSI) has the communications between devices



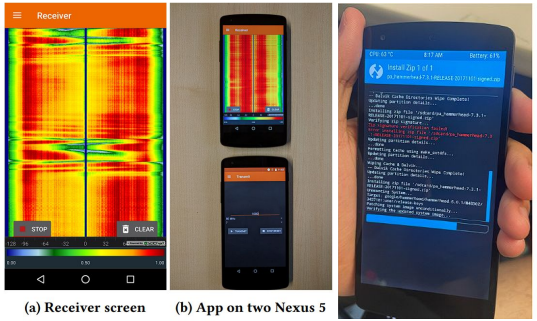
Deep Learning Models

- Deep Learning Models possess the capability to extract features from input data (individuals)
- These models are able to categorize the input data into specific labels, such as daily activities
  - Walking, Kicking, Raising Arm, Squatting, Sitting



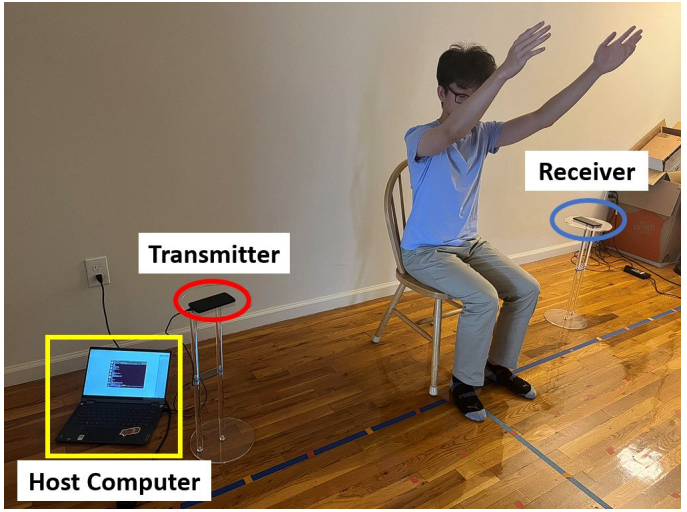
Setting Up Experiments

- Setting up a Linux virtual machine through VirtualBox
  - Ubuntu ISO (disk image file)
  - Increased familiarity with Linux terminal
- Setting up phones on Virtual machine
- Preparing to collect CSI data



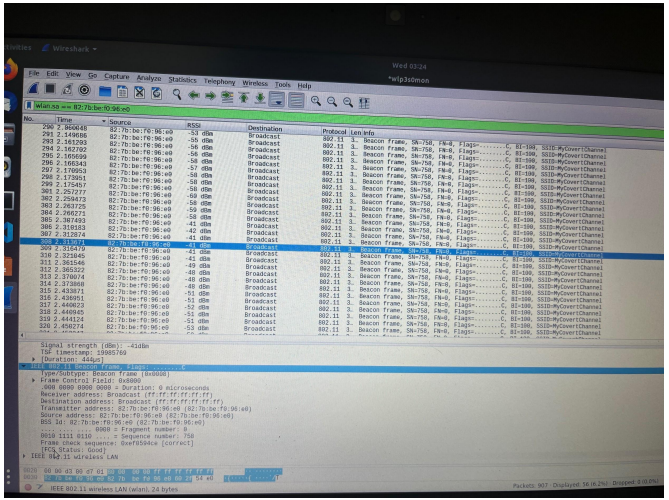
(a) Receiver screen (b) App on two Nexus 5

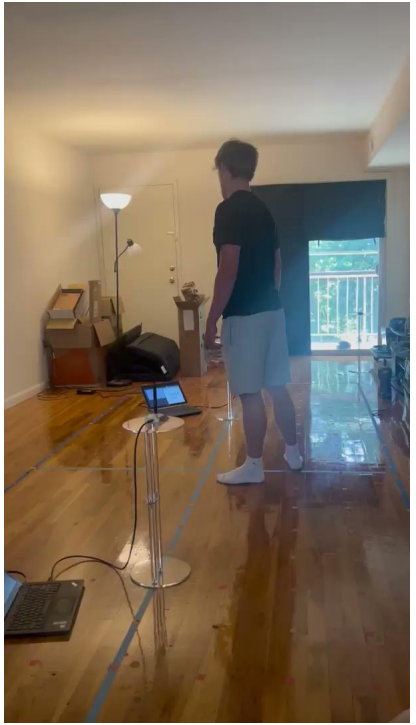
```
                #     # #####    #    #     #     # #    # #     #
                #     # #       # #   #     #     # #    # #     #
                #     # #  #### #   #  #     #     #  ###  #     #

                The C-based Firmware Patching Framework


                                !!! WARNING !!!
    Our software may damage your hardware and may void your hardware's
     warranty! You use our tools at your own risk and responsibility


  COLLECTING STATISTICS read /home/sudouser/nexmon/STATISTICS.md for more infor
mation
  CREATING DIRECTORIES obj, gen, log
  COMPILING src/version.c => obj/version.o (details: log/compiler.log)
  PREPARING gen/nexmon.pre => gen/nexmon2.pre
  GENERATING LINKER FILE gen/nexmon.pre => gen/nexmon.ld
  GENERATING LINKER FILE gen/nexmon.pre => gen/flashpatches.ld
  LINKING OBJECTS => gen/patch.elf (details: log/linker.log, log/linker.err)
  GENERATING MAKE FILE gen/nexmon.pre => gen/nexmon.mk
  GENERATING MAKE FILE gen/nexmon.pre => gen/flashpatches.mk
  APPLYING FLASHPATCHES gen/flashpatches.mk => fw_bcmdhd.bin (details: log/flas
hpatches.log)
  APPLYING PATCHES gen/nexmon.mk => fw_bcmdhd.bin (details: log/patches.log)
  REMOUNTING /system
  COPYING TO PHONE fw_bcmdhd.bin => /sdcard/fw_bcmdhd.bin
  COPYING /sdcard/fw_bcmdhd.bin => /vendor/firmware/fw_bcmdhd.bin
  RELOADING FIRMWARE
```



Collecting Data

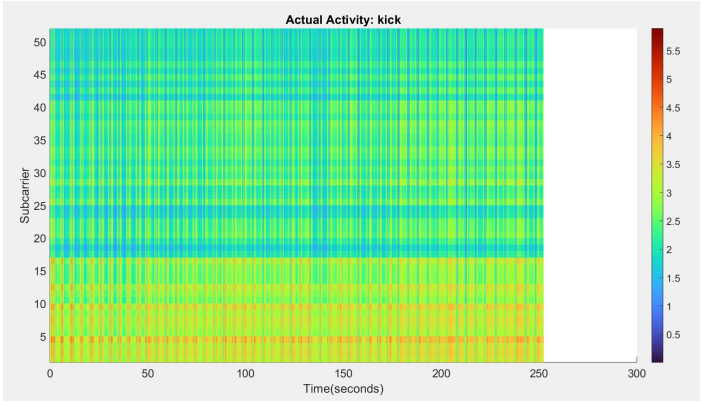
- Used WiFi Transmitters/Receivers to collect CSI data
- Performed daily movements (walking, squatting, etc)
- Data was then trained using deep learning model

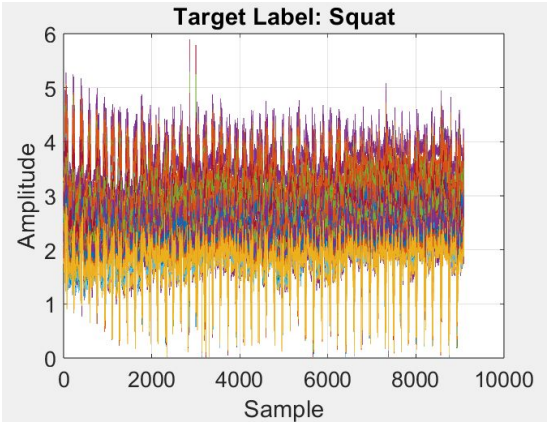


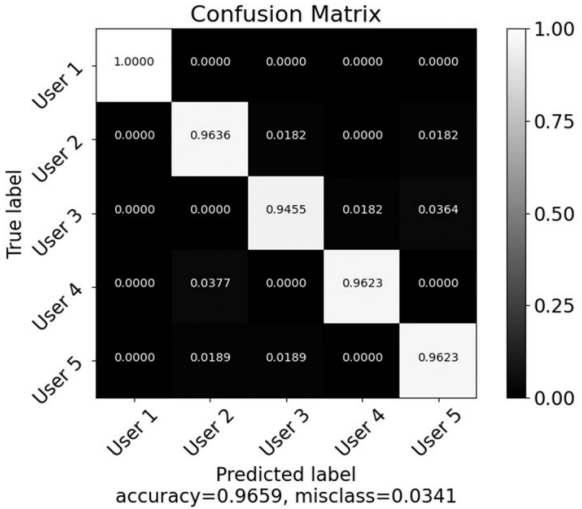


Results

- Used a Confusion Matrix to check accuracy
- Model is able to achieve recognition accuracy at 96% for User authentication
- Overall attack success rate can reach to 80%







Thank You!