Dwight Look College of
ENGINEERING
TEXAS A&M UNIVERSITY

Team 28: The NCID Defender Final Presentation

Amy Chen
Scott Kevil-Yeager
Matthew Hebrado

Sponsor: Dr. Tod Cox
TA: Rohith Kumar

Project Summary

The Problem

- $39.5 billion lost to phone scams
- Target: elderly American citizens
- Timeframe: can last for months

Our Project

- Voice signature matching
- Captures and records call data
- Integrates and augments Network Caller ID (NCID)

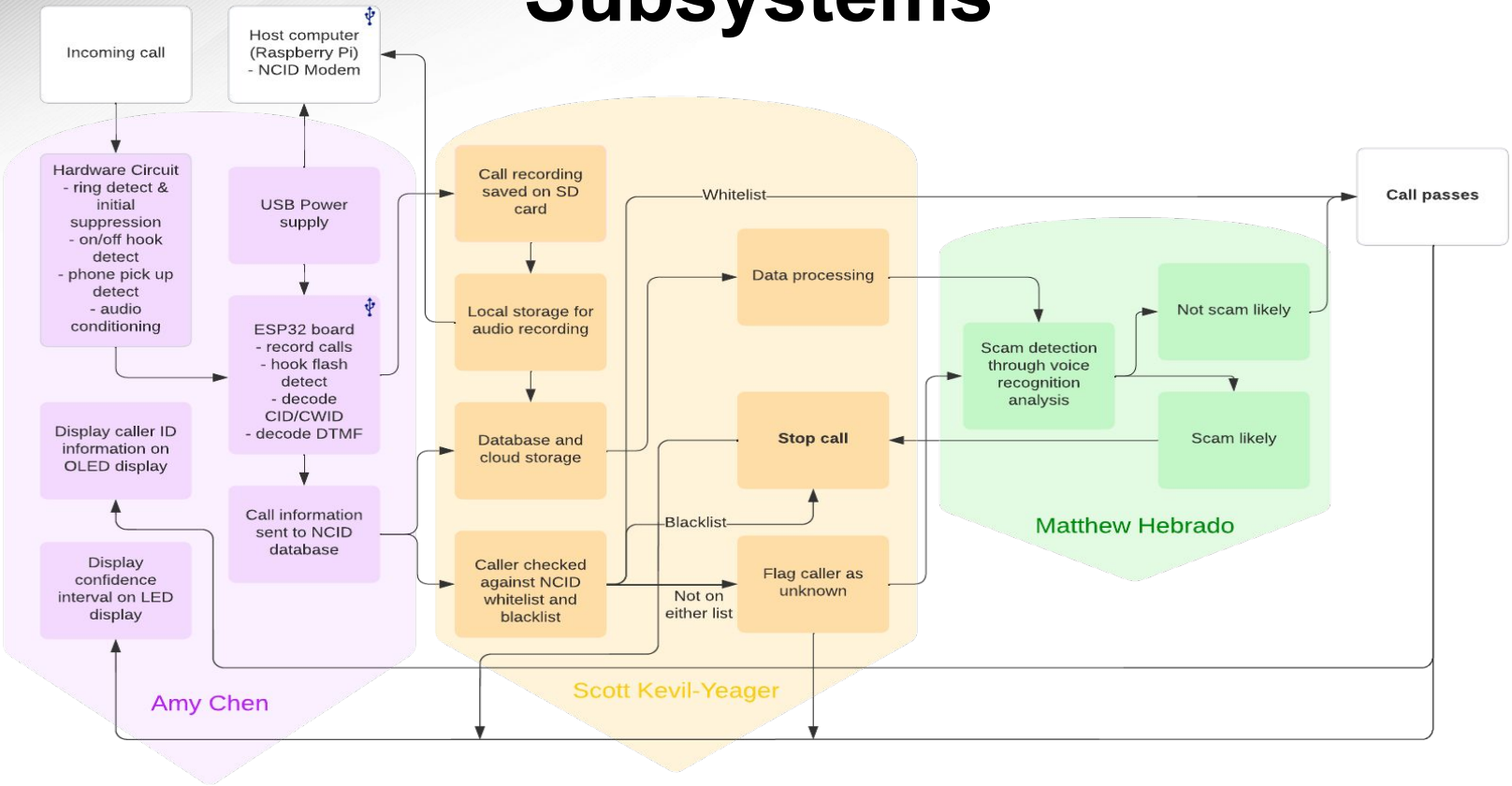
System Level Requirements

- Receive incoming landline telephone call
- Enable suppression of the first ring
- Provide a means for automated hangup/playback of recordings
- Capture caller ID and audio recording of the incoming call that can be stored locally (microSD)

System Level Requirements

- Upload call to database by sending audio file through serial port to host computer
- Receive call in database and remove silence to improve machine learning algorithm
- Machine learning compares audio to whitelist and blacklist and returns a confidence interval
- Sends alert to the end-user and notifies family members if the call is suspicious

Subsystems


Incoming call
Host computer (Raspberry Pi) - NCID Modem
Hardware Circuit - ring detect & initial suppression - on/off hook detect - phone pick up detect - audio conditioning
USB Power supply
ESP32 board - record calls - hook flash detect - decode CID/CWID - decode DTMF
Display caller ID information on OLED display
Call information sent to NCID database
Display confidence interval on LED display
Amy Chen
Call recording saved on SD card
Local storage for audio recording
Database and cloud storage
Caller checked against NCID whitelist and blacklist
Whitelist
Data processing
Stop call
Blacklist
Flag caller as unknown
Not on either list
Scott Kevil-Yeager
Call passes
Scam detection through voice recognition analysis
Not scam likely
Scam likely
Matthew Hebrado

Hardware Subsystem

Amy Chen

Requirements for hardware:

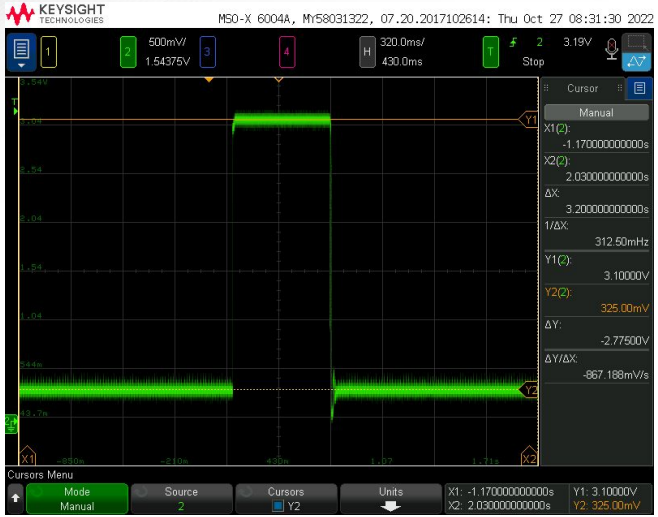
- Interface with POTS line
- Initial ring suppression
- Detect ringing and on/off hook
- Retrieve incoming analog phone signal

Requirements for state machine in firmware:

- Decode caller ID & caller waiting ID
- Detect hook flash
- Record beginning and end call times
- Detect DTMF tones

Hardware Subsystem

On/Off Hook Detection Subcircuit

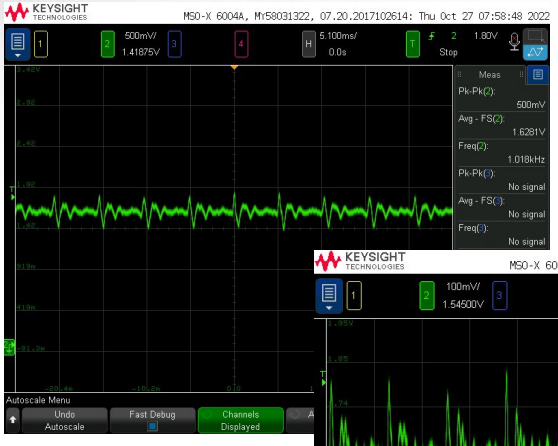


On hook: 3.1 V
Off Hook: 325 mV
Hook Flash: ~ 0.6 s

Display Subcircuit

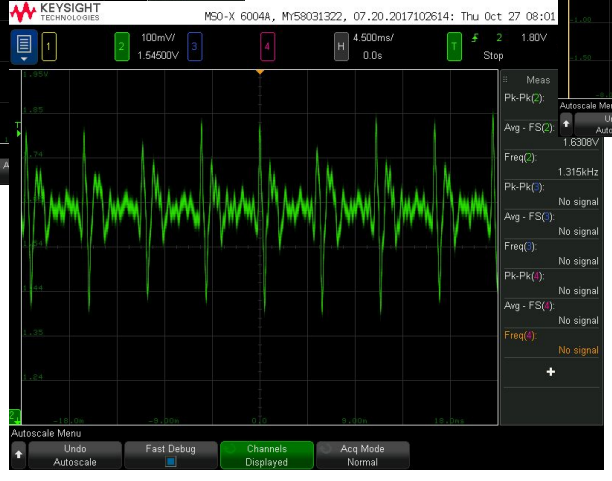
Hardware Subsystem

Audio Conditioning Subcircuit

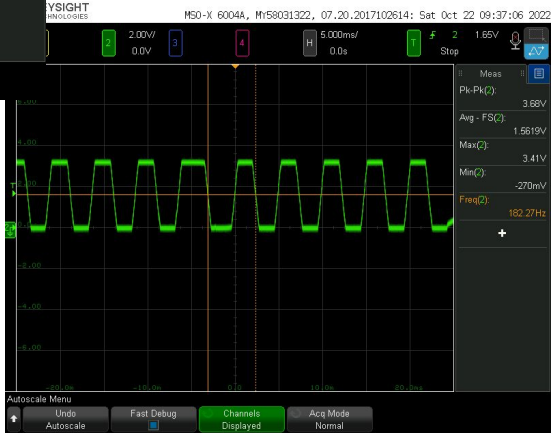


From the ADC pin

DC Bias 1.6 V (Vcc/2)

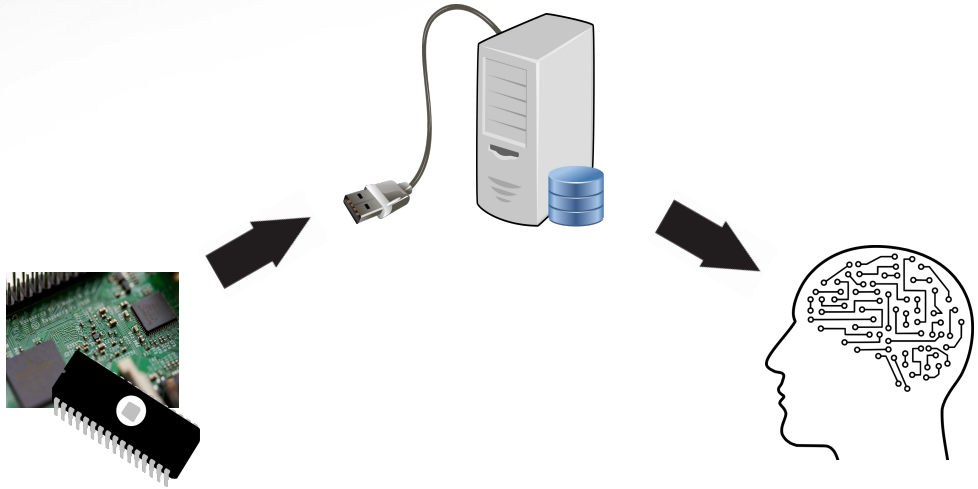


From the DAC pin



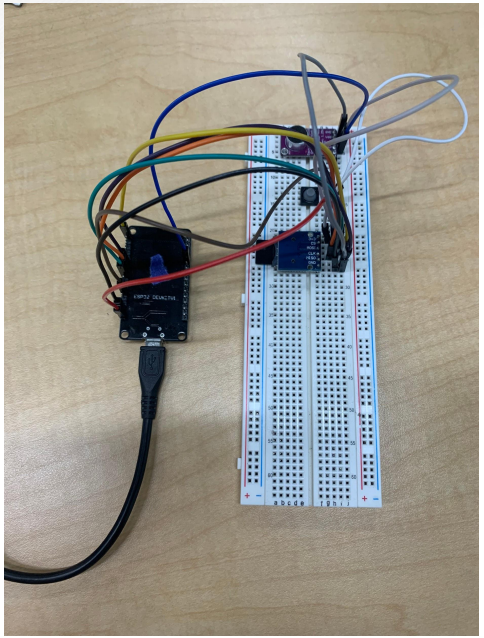
Database and Data Processing Subsystem Overview

Scott Kevil-Yeager

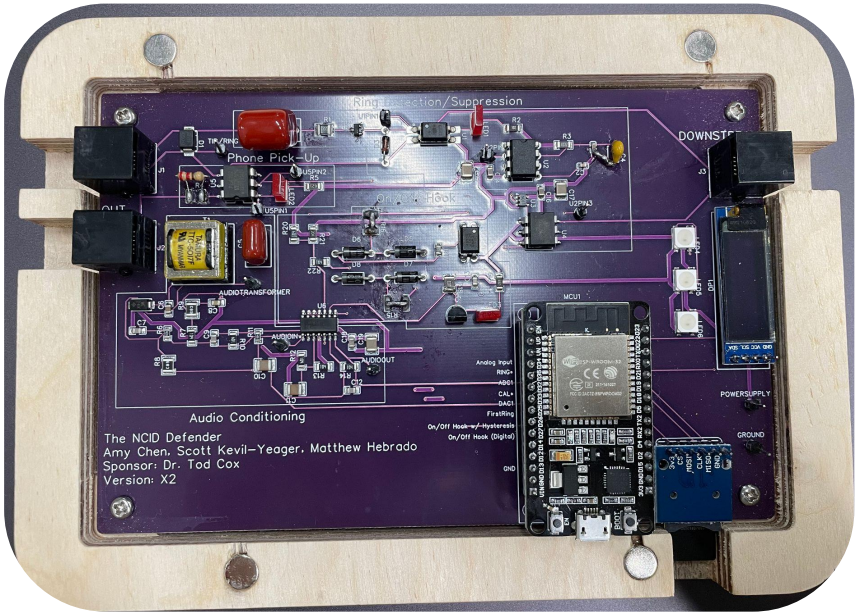


- The database and data processing subsystem acts as a bridge between the hardware and machine learning subsystems
- ESP32 sends that data through serialized JSON packets to the host computer
- JSON packet is decoded and audio data is sent to the database to be stored

Database and Data Processing

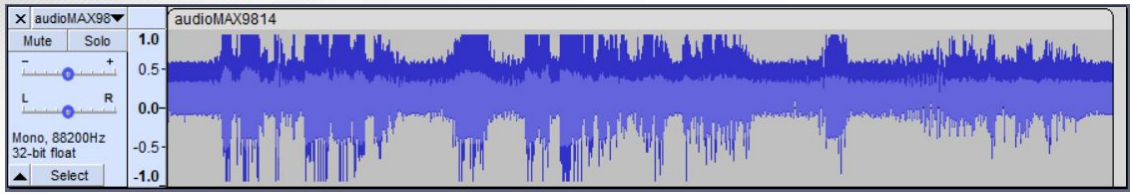


Subsystem prior to integration

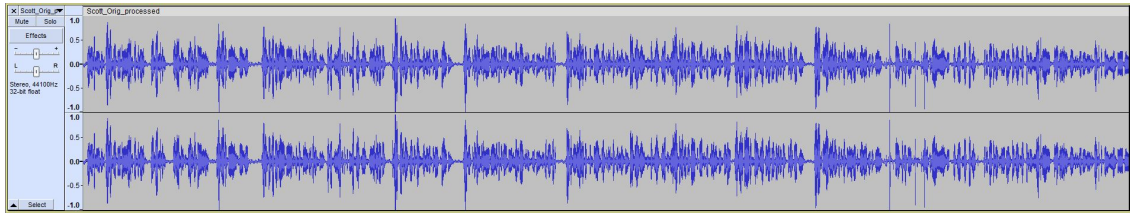


Subsystem after integration

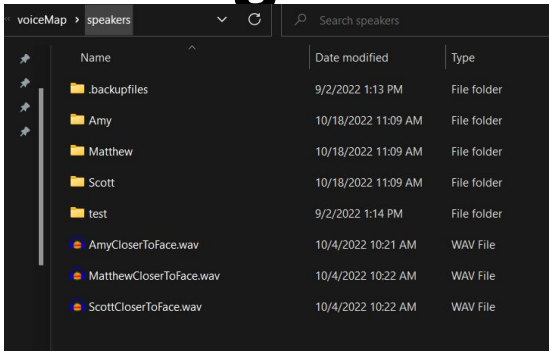
Database and Data Processing



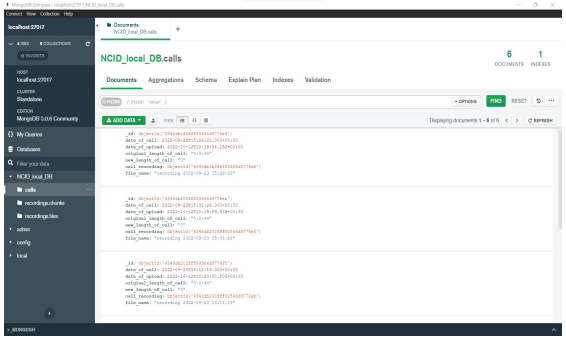
Audio recorded through subsystem prior to integration



Example of audio file with silence removed



Matthew's ML testing directory



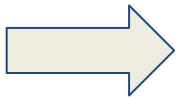
Database shown through MongoDB Atlas

Machine Learning Overview

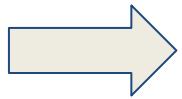
Matthew Hebrado

- User sets up a whitelist/blacklist
- Recordings in the whitelist/blacklist gets processed then classified
- Silence removed recording from handset gets passed through
- Test files are compared against the classification file
- Returns the greatest accuracy percentage based on all of the entries in each list

NCID Defender



ML Processing



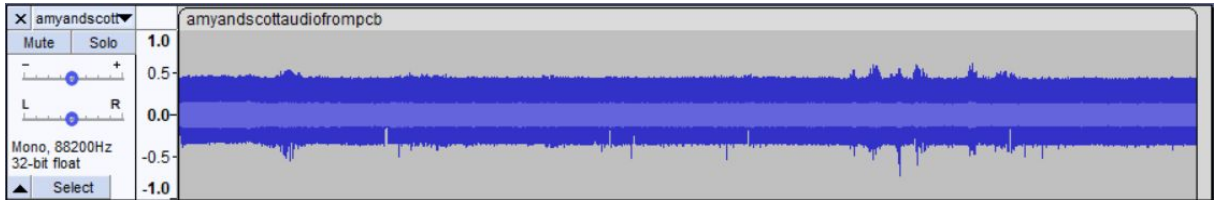
Machine Learning Data

	Optimal Conditions	Updated Training Files	Random Forest	Gradient Boost (1 sec splits)	Gradient Boost (0.5 sec splits)
Amy	96.8	85.1	50.8	76.3	68.6
Matthew	93.4	58.8	52.4	87.8	84.4
Scott	84.2	31.8	41.6	51.6	62.0

- Optimal conditions - recorded with desktop/laptop microphone using audacity
- Updated training files - using iPhone voice memo to get voice recording
- Random forest - testing a different classification method
- Gradient boost (1 sec splits) - testing a different classification method
- Gradient boost (0.5 sec splits) - decreasing the split interval in order to get more data to test

System and Validation Results

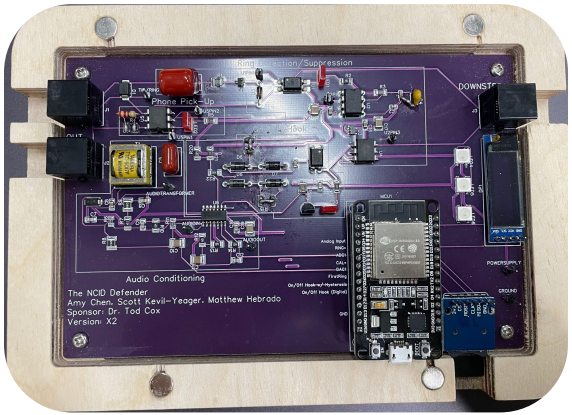
- Machine learning to Database integration
  - We have confirmed that files can be transferred over the database and onto the host machine
    - The current recording from the handset is very noisy
- Database to Hardware integration

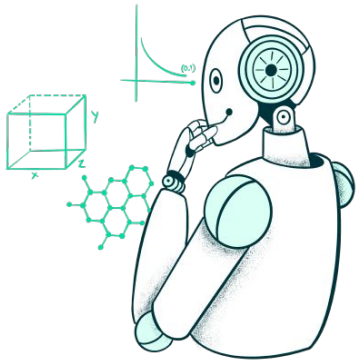


Conclusion

The NCID Defender Team built a device to:

- Capture incoming analog signals from a landline telephone
- Database to store calls for data processing and for reviewing recordings of calls
- Machine learning algorithm that takes in audio from the phone and compares it to stored files of known speakers





Conclusion

Accomplishments:

- Great teamwork
- Communicating with sponsor weekly
- Working through challenges
- Built a project that will be available to 100k active users
- Creating and revising and testing two versions of the PCB
- Built off of existing library to create ML voice mapping

Challenges:

- Unit testing code for database subsystem
- Finding documentation regarding landline telephones
- Coding in c++ for an arduino system; code dependency issues, lack of resources on the MCU, handling multi-core tasks, SD card library (mounting and writing to SD card from MCU)

Conclusion

What we learned:

- Operating as an engineering team
- Researching complex topics that may be outdated or difficult to find data on
- Integrating a complex project with many moving parts
- Presentation/Communication Skills
- Machine learning using Python libraries to extract features from a .wav file
- Coding for microcontrollers and communicating via serial communication
- Circuit design, testing, and revision
- Soldering skills

Project Mitigation

Things that need to be worked on:

- TCPIP communication with NCID
- Serial communication of files to host computer
- Validation of state machine
- Run Database and ML on Ras Pi

Thank You!