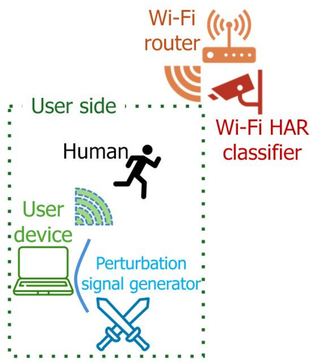
Adversarial Example against Wi-Fi-Based Sensing

Brian Chacon (UG), Oskar Iwaniuk (UG),
Advisors: Professor Minsung Kim,
Byungjun Kim

RUTGERS
WINLAB | Wireless Information Network Laboratory

Motivation

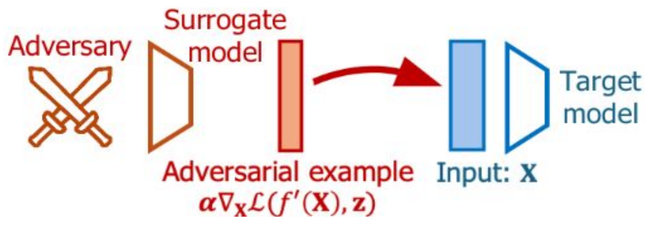
As wifi technology advances, their range of applications grows. Today, Wi-Fi technology can be used for human and spectrum sensing to detect and classify human activity.



While active recognition can be used for good application, there is a security concern regarding the protection and privacy of people in large public spaces

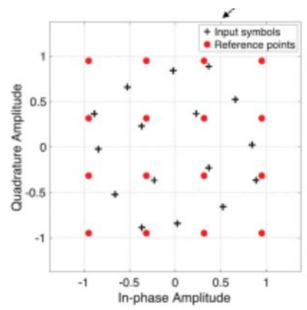
Background

For this study, we focus on two types of Wi-Fi based sensing: human activity recognition (HAR) and fingerprinting.
WiFi-based HAR detects human activity by analyzing how WiFi signals reflect off people moving between devices and the router.
Changes in these signals reveal physical activity patterns.
Wifi Fingerprinting uses unique signals that are emitted from a person's device. Cellular devices contain unique hardware imperfections in their analog components that transmits these unique signals.
We aim to pertube these unique signals to prevent device identification.



Design

Wifi Fingerprinting: We use a convolutional neural network as our wifi fingerprinter. The model is designed to take samples of raw 16 QAM IQ constellations (our signal data). These samples are then fed to our CNN model.



IQ samples with reference points (red dots) and the actual read caused by hardware imperfections (black dots)

Wifi HAR: This model is an LSTM-based neural network designed to classify human activity using sliding windows of Wi-Fi Channel State Information (CSI). The model was tested with two types of attacks: Fast Gradient Sign Method (FGSM) and Gaussian noise. FGSM creates small, targeted changes to the input that confuse the model, based on how it calculates its error. Gaussian noise adds random signal variations to test how well the model handles natural disturbances.

$$x^{adv} = x + \varepsilon \cdot \mathrm{sign}(\nabla_x J(x, y_{true})),$$

where

x is the input (clean) image,

x^{adv} is the perturbed adversarial image,

J is the classification loss function,

y_{true} is true label for the input x.

A user device sends signals to interfere with the Wi-Fi HAR classifier without altering normal data communication FGSM creates is a method that adds a small perturbation in a way that increases the model's loss most,

$$x_{\text{noisy}} = x_{\text{original}} + \mathcal{N}(\mu, \sigma^2)$$

$x_{original}$ is your original clean signal

$N(\mu,\sigma^2)$ is random noise drawn from a normal distribution with mean μ and standard deviation σ

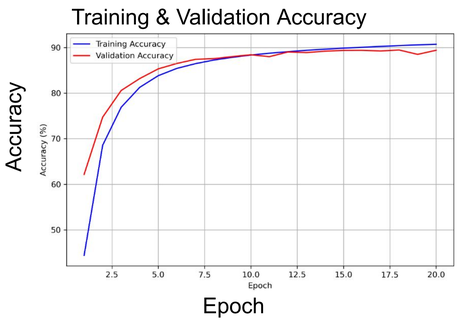
x_{noisy} is the resulting noisy signal

Random Gaussian signal noise is a type of distortion added to a signal where the variations follow a bell-shaped (normal) distribution, meaning most changes are small while larger deviations happen less frequently

Results

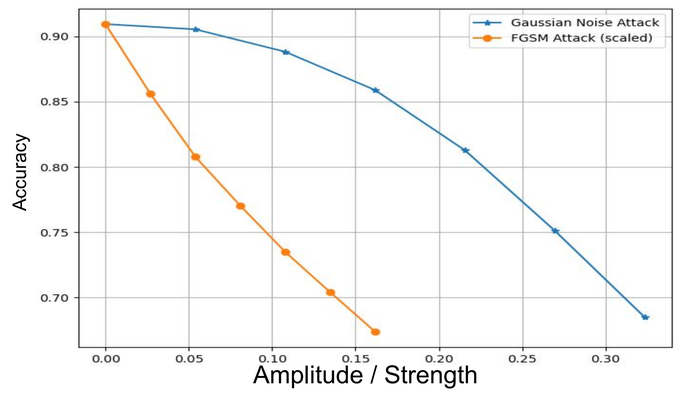
Wifi Fingerprinting:

- Model attained an accuracy of ~90% among 4 devices
- Used X310 USRP Radios



Wifi HAR:

- Accuracy decreases under both FGSM and Gaussian random noise attack, FGSM knocks accuracy down much faster because it's designed to trick the model, while random Gaussian noise only affects accuracy slightly until it gets very strong



Future Work

- Improve Wifi-Fingerprinting accuracy
- Expand data and classification to a larger variety of devices for the fingerprinting cnn
- Dive into developing remote adversarial attacks against the models

References

1. Kim, B. (2025). Machine learning for noise-robust wireless communication systems (Doctoral dissertation). University of California, San Diego. https://noiselab.ucsd.edu/group/Thesis/ByungjunKim2025.pdf
2. Sankhe, K., Belgiovine, M., Zhou, F., Riyaz, S., Ioannidis, S., & Chowdhury, K. (2019). ORACLE: Optimized Radio Classification through Convolutional Neural Networks. IEEE International Conference on Computer Communications (INFOCOM). https://ieeexplore.ieee.org/abstract/document/8737463/authors#authors

WINLAB