UNIVERSITY OF CALIFORNIA SAN DIEGO

Deep Learning Applications in Wireless Networks: Signal Classification and Privacy-Preserving Adversarial Attacks

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Electrical Engineering (Communication Theory and Systems)

by

Byungjun Kim

Committee in charge:

Professor Peter Gerstoft, Chair
Professor Tajana Rosing
Professor Xinyu Zhang
Professor Xiaolong Wang

2025

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Chapter 2, in full, is a reprint of the material as it appears in Kim, B., Mecklenbräuker, C., and Gerstoft, P. “Deep Learning-based Modulation Classification of Practical OFDM Signals for Spectrum Sensing”, in Proc. International Conference on Acoustics, Speech, and Signal Processing Workshop 2023. The dissertation author was the primary investigator and author of this paper. The coauthors listed in this publication directed and supervised the research.

Chapter 3, in full, is a reprint of the material as it appears in Kim, B., Mecklenbräuker, C., and Gerstoft, P. “Blind Modulation Classification of Wi-Fi 6 and 5G signals for Spectrum Sensing”, in Proc. ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems 2023. The dissertation author was the primary investigator and author of this paper. The coauthors listed in this publication directed and supervised the research.

Chapter 4, in full, is a reprint of the material as it appears in Kim, B., Mecklenbräuker, C., and Gerstoft, P. “Deep Learning-based Modulation Classification of Practical OFDM Signals for Spectrum Sensing”, in Proc. IEEE International Conference on Computer Communications 2024. The dissertation author was the primary investigator and author of this paper. The coauthors listed in this publication directed and supervised the research.

Chapter 5, in full, is a reprint of the material as it appears in proceeding of Kim, B., Panchagatti, A., and Gerstoft, P. “Real-time Adversarial Attack to Deep Learning-based Wi-Fi Human Activity Recognition”, in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing 2025. The dissertation author was the primary investigator and author of this paper. The coauthors listed in this publication directed and supervised the research.

Chapter 6, in part, has been double-blindly submitted for publication of the material as it may appear in proceeding of Kim, B., Panchagatti, A., Zhang, X., and Gerstoft,

P. “Remote Adversarial Attacks against Wi-Fi-based HAR for Privacy Protection”. The dissertation author was the primary investigator and author of this paper. The coauthors listed in this publication directed and supervised the research.

VITA

2017	B.S. in Electrical Engineering, Pohang University of Science and Technology
2017–2019	Research Assistant, Seoul National University
2019	M.S. in Electrical and Computer Engineering, Seoul National University
2019–2025	Research Assistant, University of California, San Diego
2025	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

B. Kim, A. Panchagatti, and P. Gerstoft, "Real-time Adversarial Attack to Deep Learning-based Wi-Fi Human Activity Recognition", Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2025

B. Kim, C. Mecklenbräuker, and P. Gerstoft, "Deep Learning-based Modulation Classification of Practical OFDM Signals for Spectrum Sensing", Proc. of IEEE International Conference on Computer Communications, 2024

A. Panchagatti, B. Kim, C. Mecklenbräuker, and P. Gerstoft, "Channel Estimation in Time-Varying Ocean Environments using OTFS Modulation", Proc. of IEEE Asilomar Conference on Signals, Systems & Computers, 2024

B. Kim, C. Mecklenbräuker, and P. Gerstoft, "Blind Modulation Classification of Wi-Fi 6 and 5G signals for Spectrum Sensing", Proc. of ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems, 2023

B. Kim, C. Mecklenbräuker, and P. Gerstoft, "Deep Learning-based OFDM Modulation Classification without Symbol-level Synchronization", Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing Workshop, 2022

ABSTRACT OF THE DISSERTATION

Deep Learning Applications in Wireless Networks: Signal Classification and Privacy-Preserving Adversarial Attacks

by

Byungjun Kim

Doctor of Philosophy in Electrical Engineering (Communication Theory and Systems)

University of California San Diego, 2025

Professor Peter Gerstoft, Chair

This dissertation investigates deep learning-based signal processing techniques for wireless communication system about two research topics: modulation classification in orthogonal frequency division multiplexing (OFDM) signals for intelligent spectrum sensing, and remote adversarial attacks on Wi-Fi-based human activity recognition (HAR) systems for privacy.

Deep learning-based modulation classification of OFDM signals following Wi-Fi 6 and 5G downlink specifications for spectrum sensing is studied. Since intelligent spectrum sensing often targets diverse wireless technologies, protocol-specific preambles or

channel allocation might not be available. To reliably estimate modulation under this constraint, this dissertation proposes algorithm to estimate essential OFDM parameters and extract a feature characterizing modulation to make it suitable to be an input to deep learning-based classifier. The OFDM parameters, symbol duration and cyclic prefix length, are estimated utilizing the cyclic autocorrelation properties of OFDM waveforms. Based on these estimates, a feature characterizing modulation of OFDM signals is designed to mitigate synchronization errors arising from unknown symbol boundaries. The extracted features are represented as two-dimensional histograms of amplitude and phase, which serve as inputs to a convolutional neural network (CNN)-based classifier. The classifier's performance, evaluated using synthetic and real-world measured over-the-air (OTA) datasets, achieves a minimum accuracy of 97% accuracy with OTA data when SNR is above the required SNR for reliable data transmission.

Remote adversarial attacks on neural network-based HAR classifier that utilize Wi-Fi channel state information (CSI) has been investigated. The capability of Wi-Fi routers to perform human activity recognition through CSI raises privacy concerns. To address this issue, this dissertation proposes a novel remote adversarial attack deploying geneative adversarial imitation learning (GAIL). The proposed method degrades the accuracy of HAR classifiers deployed at Wi-Fi routers by manipulating the channel estimation signals transmitted from user devices. Unlike gradient-based attacks, which require complete knowledge of the target model's inputs or future CSI, the GAIL-based algorithm generates effective perturbation signals without explicit knowledge of future CSI or details of the target HAR models. Comprehensive experimental evaluations performed across seven distinct environments and six target HAR models–including both deep learning and non-deep learning classifiers–demonstrate the versatility of the proposed adversarial method. The GAIL-based approach reduces HAR classifier accuracy to approximately 50%, requiring only a minimal increase (0.5 dB) in average perturbation amplitude compared to gradient-based methods based on impractical assumptions.

Chapter 1

Introduction

Next-generation wireless communication systems aim to achieve higher throughput and to fulfill diverse visions including incorporating intelligence for cognitive radio as well as joint sensing and communications [1, 2, 3]. Cognitive radios are envisioned to intelligently manage resources, optimize network performance, and detect anomalies autonomously [2]. Achieving these objectives requires reliable and accurate identification of signal types to enable more efficient spectrum utilization and improved coexistence among heterogeneous wireless technologies [4].

Due to the ubiquity of wireless signals, wireless-based sensing has become an extensively studied research area [5]. Typical applications include Wi-Fi-based localization [6], radar-based gesture recognition [7], and human activity recognition [8]. Additionally, the integration of deep learning into these applications has significantly enhanced their performance and accuracy [9]. However, this rapid technological advancement raises growing concerns about privacy leakage, especially given the pervasive nature of wireless signals in daily life [10]. Consequently, as wireless sensing technologies continue to advance, it is increasingly important to address and thoroughly investigate associated security and privacy issues.

The objective of this dissertation is to present two deep learning applications using modern wireless communication signals:

1. Among various signal characteristics, modulation information provides essential information about wireless signal transmission, including channel conditions between a transmitter (Tx) and a receiver (Rx). The identification of modulation schemes employed in state-of-the-art orthogonal frequency division multiplexing (OFDM) signals for data communication is studied. As the proposed system is intended to contribute to intelligent spectrum sensing, it operates without prior knowledge of wireless communication protocols. Instead, the system relies solely on fundamental OFDM structures—OFDM symbols and their cyclic prefixes (CP)—to extract features characterizing modulation. Key OFDM parameters, subcarrier spacing and cyclic prefix length, are estimated to provide the preliminary information to support this processing step using OFDM signal's cyclostationarity.

2. To address privacy leakage concern about Wi-Fi sensing, remote adversarial attack against Wi-Fi-based human activity recognition (HAR) is studied here. Wi-Fi users may have legitimate concerns about being monitored by Wi-Fi routers, which are generally outside their control. To mitigate this issue, the proposed attack aims to degrade the performance of deep learning-based classifier at a Wi-Fi router by manipulating channel estimation signals transmitted by a user device.

1.1 Basic Concepts

1.1.1 OFDM

The procedure by which OFDM signals are generated is illustrated in Fig. 1.1 and Fig. 1.2. The parameters in Fig. 1.1 are from non-high throughput (non-HT) format 20 MHz bandwidth OFDM-based Wi-Fi, which is supported by Wi-Fi 2 (IEEE 802.11a) to Wi-Fi 6 (IEEE 802.11ax) [11]. Fig. 1.1 describes how time samples carry the information. IFFT is taken over 64 symbols. These consist of 24+24=48 data symbols, 2+2=4 symbols on pilot subcarriers, and 64-27·2=12 inactive symbols on null subcarriers. Pilot subcarriers

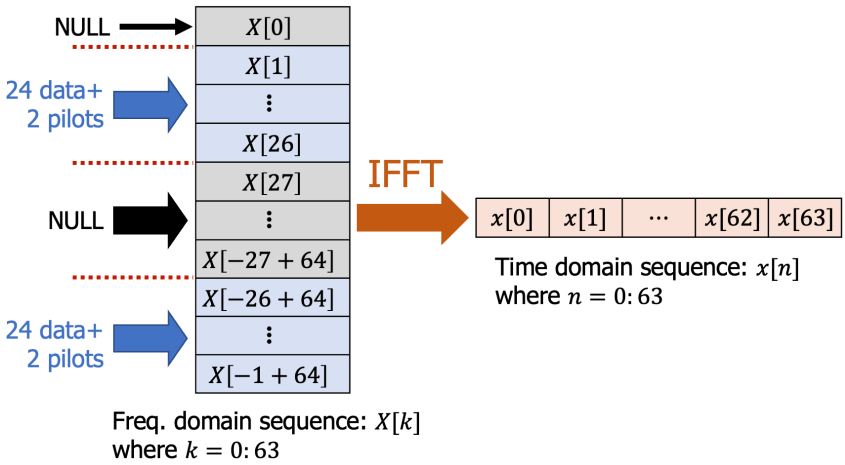


Figure 1.1. Conversion of frequency domain symbols to signal samples in time domain to generate OFDM signals

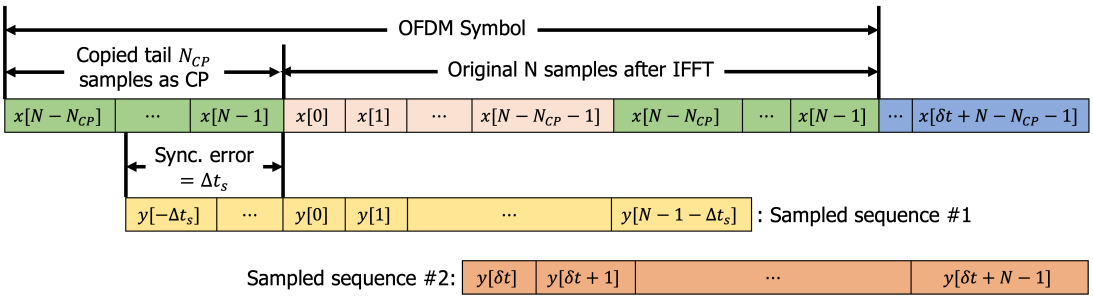


Figure 1.2. Insertion of CP and sampling sequences to generate OFDM signals.

are used to estimate residual CFO and sampling rate offset (SRO) [12]. Null subcarriers serve as a guard band to avoid DC leakage and interference to adjacent channels. Symbols in OFDM are included in the frequency domain, therefore identifying modulation features from time samples is not straightforward.

The CP is appended to the time sequence generated with IFFT, see Fig. 1.2. In OFDM systems, the last N_{CP} time samples are copied just before the IFFT sequence. This avoids inter-symbol and inter-carrier interference and simplifies frequency domain calculation, as CP makes channel output a circular convolution between a transmitted sequence and a channel response.

Carrier frequency offset (CFO) occurs when the local oscillator in the Rx does not synchronize with the carrier in the received signal. This phenomenon has two main causes: frequency mismatch between the Tx and the Rx oscillators and time-variance of the communication channel, e.g. due to Doppler effects in the signal propagation scenario. In the presence of CFO, the signal demodulated from the carrier frequency, $y[n]$ is related with $x[n]$ as

$$\begin{aligned} y[n] &= \left(x[n]\mathrm{e}^{j2\pi f_c n T_s}\right) \mathrm{e}^{-j(2\pi(f_c+\Delta f_c)nT_s)} \\ &= x[n]\mathrm{e}^{-j(2\pi\Delta f_c n T_s)}, \end{aligned} \tag{1.1}$$

where T_s, f_c, and Δf_c denote sampling period, Tx carrier frequency and deviation of Rx carrier frequency from f_c. With constant Δf_c, CFO causes a phase drift which varies linearly over time. In protocol-compliant transmission of OFDM signals, Δf_c is estimated using preambles and pilot subcarriers. The CFO effect is compensated by converting the received sequence by complex numbers whose phases are opposite of those caused by CFO.

1.1.2 Neural Network-based Classifiers

Convolutional Neural Networks (CNNs) have emerged as powerful tools for classification tasks due to their ability to effectively capture spatial and hierarchical features

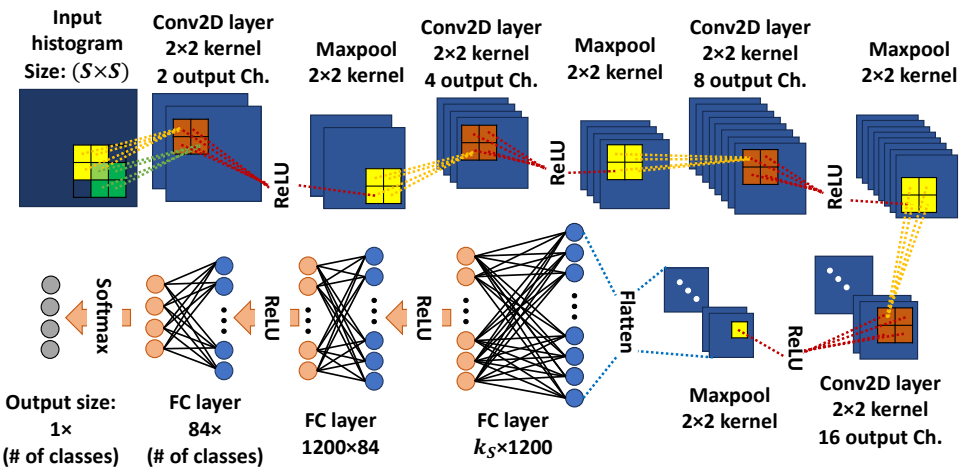


Figure 1.3. An CNN-based classifier structure for modulation classification. k_S denotes an input dimension of the first FC layer.

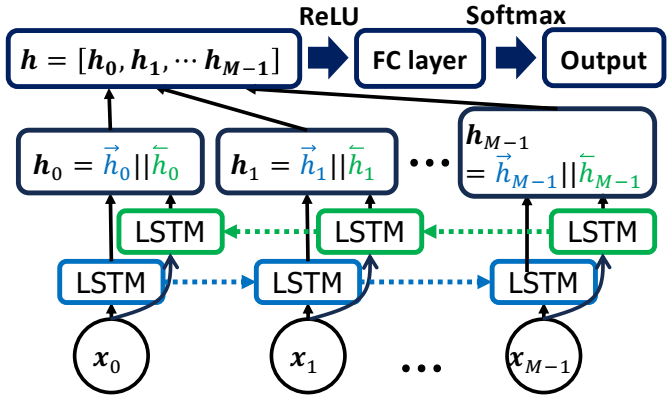


Figure 1.4. A Bi-LSTM-based classifier.

from input data [9]. A CNN-based classifier illustrated in Fig. 1.3, which is used Chapter 3, comprises multiple convolutional and pooling layers followed by fully-connected (FC) layers [13]. The convolutional layers apply spatial filters, known as kernels, to detect local patterns and produce feature maps. For example, the first convolutional layer depicted employs a 2×2 kernel to produce two output channels, each highlighting characteristics of the input data.

The architecture processes an input representation X characterized as a two dimensional histogram capturing amplitude and phase information. The outputs of the convolutional layers are flattened into a single feature vector z, which is then processed

by fully connected (FC) layers activated by Rectified Linear Units (ReLU). Finally, a softmax activation function is applied to produce the output probability vector. The input-output relationship of the CNN-based classifier f_C is expressed as:

$$y = f_C(X;\theta), \tag{1.2}$$

where y denotes the predicted probability vector corresponding to the class labels, and θ represents the set of trainable parameters (weights and biases) within the CNN model. The output vector y has dimensions equal to the number of target classes, with each element indicating the predicted probability that the input representation X belongs to the corresponding class.

Subsequent max-pooling layers reduce the spatial dimensionality, enhancing computational efficiency and robustness against spatial shifts. The network extracts higher-level features through consecutive convolutional layers, each with an increasing number of channels, as demonstrated in Fig. 1.3. Non-linear activation functions, such as Rectified Linear Units (ReLU), introduce non-linearity into the network. After feature extraction, the multi-dimensional feature maps are flattened into a vector, taken as the inputs of the following FC layers. These FC layers integrate learned features to perform the final classification task. The FC layer typically outputs logits corresponding to each class, and the final softmax layer converts these logits into normalized probabilities, predicting the class label with the highest probability.

Figure 1.4 illustrates a Long Short-Term Memory (LSTM)-based neural network architecture. LSTMs are a specialized type of Recurrent Neural Network (RNN) designed to effectively model sequential data by capturing temporal dependencies and mitigating the vanishing gradient problem encountered by standard RNNs [14]. As shown, the architecture comprises multiple LSTM cells arranged sequentially, processing inputs $X = \{x_0, x_1, \ldots, x_{M-1}\}$ at each timestep. Each cell generates hidden state representations

$h_0, h_1, \ldots, h_{M-1}$, which are concatenated into a combined representation vector h. This vector subsequently passes through an FC layer activated by a ReLU, followed by a softmax function to produce the final output. The input and output of the LSTM-classifier, f_C can be represented as:

$$y = f_C(X; h_0), \tag{1.3}$$

where y denotes the predicted probability vector corresponding to the class labels, and h_0 denotes the initial hidden state of the LSTM. The output vector y has dimensions equal to the number of target classes, and each element of y indicates the probability that the input sequence X belongs to the corresponding class.

One notable capability of LSTM networks is their ability to handle input sequences of variable length. Due to the recurrent nature of the LSTM structure, input sequences are processed iteratively, with each timestep updating and maintaining an internal memory state. This mechanism allows the LSTM to naturally accommodate sequences of varying lengths without the need for modifications to the architecture. The capability makes LSTMs advantageous in applications involving variable-length sequences, including language modeling, speech recognition, and time-series classification and prediction [9, 14], enabling the model to dynamically adjust its computations based on the given input sequence length.

1.2 Dissertation Overview

Chapter 2 [15] investigates deep learning-based modulation classification of practical OFDM signals without symbol-level synchronization. The proposed system identifies modulation schemes without information on boundary of OFDM symbols. A preprocessing algorithm has been developed to enhance deep learning-based classification performance under these constrained conditions.

Chapter 3 [16] investigates OFDM parameter estimation and modulation classifi-

cation for Wi-Fi 6 and 5G signals. Specifically, two essential parameters shaping OFDM signals, subcarrier spacing and CP length, are estimated. The parameter estimates are deployed to preprocess OFDM signals. Physical (PHY) layer signals corresponding to modern communication standards, notably 5G and Wi-Fi 6, implemented utilizing MATLAB toolboxes, are used for evaluation.

Chapter 4 [17] investigates the reliability and practicality of our algorithm of OFDM parameter estimation and modulation classification through experimental demonstrations employing software defined radio (SDR). Its capability to reliably identify modulation schemes in practical OFDM signals is highlighted.

Chapter 5 [18] investigates real-time adversarial attack to deep learning-based Wi-Fi human activity recognition. The proposed system addresses the privacy leakage issue of human activity recognition (HAR) systems utilizing Wi-Fi signals by proposing an attack scheme. Specifically, it investigates methods to degrade HAR performance through the manipulation of pilot signals transmitted by user devices. To overcome the limitation of gradient-based attacks, which requires prior knowledge of future channel state information, we propose a generative adversarial imitation learning (GAIL)-based algorithm. This algorithm effectively mimics gradient-based attack, successfully compromising the accuracy of HAR models and highlighting vulnerabilities of Wi-Fi-based sensing systems.

Chapter 6 examines the versatility of the proposed GAIL-based attack through comprehensive evaluation across various target models. The research studies beyond scenarios with unknown target model weights and structures to the cases where the sampling rate or duration of the target classifier is unknown. Furthermore, the chapter evaluates how GAIL-based attacks can effectively degrade non-deep learning models as well, demonstrating the broad applicability of this approach.

1.3 References

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5g," IEEE Commun. Mag., vol. 52, no. 2, pp. 74–80, 2014.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," IEEE J. Sel. Areas in Commun., vol. 23, no. 2, pp. 201–220, 2005.

[3] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, "Joint radar and communication design: applications, state-of-the-art, and the road ahead," IEEE Trans. Commun., vol. 68, no. 6, pp. 3834–3862, 2020.

[4] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," IEEE Commun. Surveys Tuts., vol. 11, no. 1, pp. 116–130, 2009.

[5] Z. Zhou, Q. Yang, J. Wang, Y. Ma, and J. Wan, "Internet of things sensing and data fusion: Future applications for next generation networks," IEEE Wireless Commun., vol. 25, no. 1, pp. 30–37, 2018.

[6] J. Xiong and K. Jamieson, "ArrayTrack: A fine-grained indoor location system," in Proceedings of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI), pp. 71–84, 2013.

[7] M. G. Amin, Z. Zeng, and T. Shan, "Hand gesture recognition based on radar micro-doppler signature envelopes," in Proceedings of IEEE Radar Conf. (radarconf), pp. 1–6, IEEE, 2019.

[8] Y. Wang, J. Liu, M. Wang, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in Proceedings of the 20th Annual International Conference on Mobile Computing and Networking (MobiCom), pp. 617–628, 2014.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[10] J. Liu, Y. Wang, Q. Liu, Y. Zhang, Y. Guo, W. Gao, and X. Li, "Behavior privacy preserving in rf sensing," IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 1, pp. 784–796, 2022.

[11] IEEE 802.11ax, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 1: enhancements for high-efficiency WLAN," May 2021.

[12] J. K. Tan, "An adaptive orthogonal frequency division multiplexing baseband modem for wideband wireless channels," Master's thesis, Massachusetts Institute of Technology, 2006.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems (NeurIPS), pp. 1097–1105, 2012.

[14] A. Graves, Supervised Sequence Labelling with Recurrent Neural Networks. Berlin, Heidelberg: Springer, 2012.

[15] B. Kim, V. Sathyanarayanan, C. Mecklenbräuker, and P. Gerstoft, "Deep learning-based modulation classification for ofdm systems without symbol-level synchronization," in 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), pp. 1–5, IEEE, 2023.

[16] B. Kim, C. F. Mecklenbräuker, and P. Gerstoft, "Blind modulation classification of wi-fi 6 and 5g signals for spectrum sensing," in Proceedings of the Int'l ACM Conference on Modeling Analysis and Simulation of Wireless and Mobile Systems, pp. 137–145, 2023.

[17] B. Kim, C. Mecklenbräuker, and P. Gerstoft, "Deep learning-based modulation classification of practical OFDM signals for spectrum sensing," in Proc. of IEEE Conference on Computer Communications, IEEE, 2024.

[18] B. Kim, A. Panchagatti, and P. Gerstoft, "Real-time adversarial attack to deep learning-based Wi-Fi human activity recognition," in Proc of. IEEE Conference on Acoustics, Speech, and Signal Processing, IEEE, 2025. to appear.

Chapter 2

Deep Learning-based OFDM Modulation Classification without Symbol-level Synchronization

2.1 Introduction

Deep learning (DL) has drawn lots of interest in wireless communications including spectrum sensing [1], channel coding [2], and channel prediction [3]. Complex scenario-by-scenario analysis in wireless communication studies become unnecessary by deploying DL. Recognizing modulation is crucial in spectrum sensing to perceive transmission types. By recognizing modulation with DL, a spectrum sensing detector can obtain essential transmission information without complex signal processing [4].

Orthogonal frequency division multiplexing (OFDM) is widely used in wireless communication protocols like Wi-Fi and 5G. In OFDM signals, message bits are modulated to digital symbols with modulation such as QPSK and the symbols are carried in data subcarriers. To recognize transmission type precisely, the modulation of OFDM signals should be determined. However, the modulation classifier for single-carrier signals [5, 6] cannot be directly applied to OFDM signals due to the OFDM structure. Each transmitted OFDM time sample contains partial information about multiple symbols stacked in the frequency domain. Due to this property, received time-domain IQ samples do not

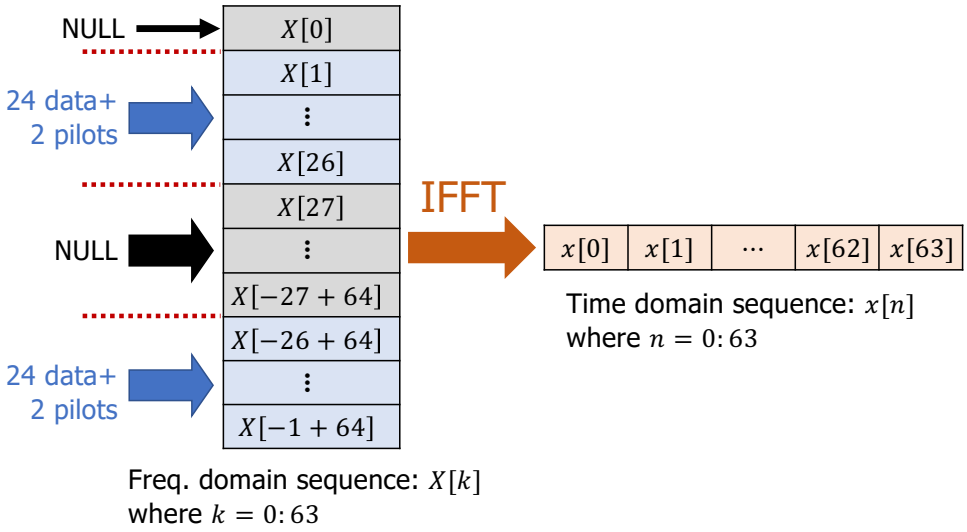


Figure 2.1. Conversion of frequency domain symbols to signal samples in time domain to generate OFDM signals.

explicitly feature modulation of OFDM signals. Therefore, additional processing is required than taking raw time IQ samples as an input for proper modulation classification of OFDM signals.

Different from generic OFDM transmission, spectrum sensing detectors should estimate modulation without detecting preamble. In a Wi-Fi system, preamble makes a receiver (Rx) synchronize with a transmitter (Tx) as well as notifies modulation being used [7]. Thus, for spectrum sensing, OFDM modulation classification needs to be addressed without the assumption of symbol-level synchronization.

In this paper, we propose a novel preprocessing algorithm to build a modulation feature robust to synchronization error, which can be caused by the Rx not knowing exactly when an OFDM symbol starts. Though there have been many efforts to classify modulation of OFDM signals [8], this is the first work to classify modulation of OFDM signals without symbol-level synchronization using DL. We verify that a CNN-based classifier which takes the proposed feature as an input outperforms existing work assuming symbol-level synchronization between a Tx and an Rx. The evaluation has been done with two datasets including hardware-generated data.

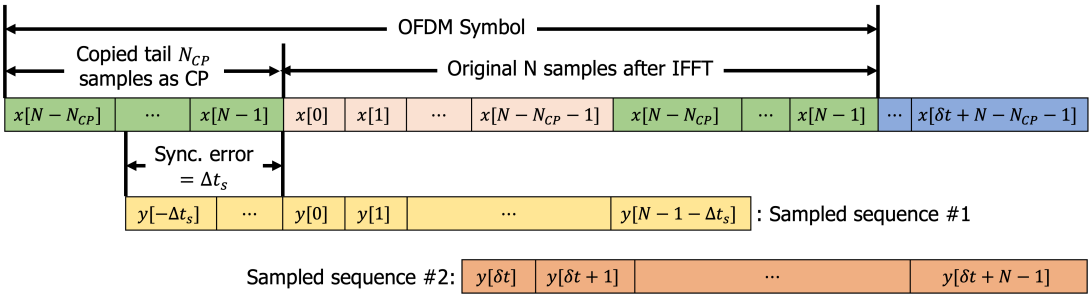


Figure 2.2. Insertion of CP and sampling sequences to generate OFDM signals.

The procedure by which OFDM signals are generated is illustrated in Fig. 2.1 and Fig. 2.2. The parameters in Fig. 2.1 and Fig. 2.2 are from non-high throughput (non-HT) mode OFDM-based Wi-Fi. Fig. 2.1 describes how time samples carry the information. IFFT is taken over 64 symbols, which consist of 48 data symbols, 4 symbols on pilot subcarriers, and 12 inactive symbols on null subcarriers. Pilot subcarriers are used to estimate residual CFO and sampling rate offset (SRO) [9]. Symbols in OFDM are included in the frequency domain, therefore identifying modulation features from time samples is not straightforward. The CP is appended to the time sequence generated with IFFT, see Fig. 2.2. In OFDM systems, the last N_{CP} time samples are copied just before the IFFT sequence to prevent Rx from inter-symbol interference.

2.2 Proposed Algorithm

Figure 2.2 illustrates the scenario where we sample the sequences. Sequences of length N, the number of subcarriers, are sampled, so they might be contained in a single OFDM symbol (sampled sequence #1) or spans two OFDM symbols (sampled sequence #2). The term OFDM symbol denotes a sequence composed of an IFFT sequence and a CP, and a symbol means a complex number used to carry bits.

The motivating observation for our proposed algorithm is that if a sampled time-domain sequence is contained in a single OFDM symbol, the FFT of that sequence gives

the original symbols with phase drift scaling linearly with subcarrier index k

$$\begin{aligned} Y^i_{\Delta t_s}[k] \triangleq \mathscr{F}\left(y^i[n-\Delta t_s]\right) &= \sum_{n=0}^{N-1} y^i[n-\Delta t_s]\mathrm{e}^{-j2\pi nk/N} \\ &= Y^i[k]\mathrm{e}^{-j2\pi\Delta t_s k/N}, \end{aligned} \tag{2.1}$$

where $Y^i[k]$ denotes the received symbol in subcarrier k of the ith OFDM symbol, $y^i[n]$ the received time-domain IFFT sequence of the ith OFDM symbol, and Δt_s the real-valued time difference measured in time sample unit between an IFFT sequence and a sampled sequence. Input of $y^i[n]$ ranges over $n \in [-N_{CP}, N-1]$ and $y^i[n]$ where $n \in [-N_{CP}, -1]$ corresponds to CP. Equation (2.1) shows that synchronization error Δt_s causes phase drift proportional to Δt_s and k. To deploy this property in building a feature characterizing modulation, two objectives should be addressed: sampling a sequence contained in a single OFDM symbol and removing the phase drift caused by synchronization error.

2.2.1 Modulation feature extraction

Due to CP, the sequences are repeated at both ends of the symbols in every OFDM symbol. Since both the length of and the distance between repeated sequences are known[1], the position of CP can be found using the autocorrelation,

$$R_{yy}(n,N) = \frac{1}{N_{CP}} \sum_{i=0}^{N_{CP}-1} y[n+i]y^*[n+i+N], \tag{2.2}$$

which has peaks when n is the first index of CP. To find a peak, we find a sample whose amplitude is larger than both adjacent samples and the minimum distance between two adjacent peaks is set to 90% of OFDM symbol duration, 72-time sample indices. In the sampled sequence, from the remainders of acquired peak-indices divided by OFDM symbol duration, 80, we determine the mode among those remainders as the first index

[1]It is assumed that the OFDM parameters are known to the classifier. We leave the estimation of those parameters for future work.

of OFDM symbol, denoted as p.

Due to noise and varying amplitudes of time samples, the estimated CP position might not be accurate. However, our objective is not to find the exact first time sample of the OFDM symbol, but the sequence contained in a single OFDM symbol. Therefore, by sampling the sequence $\{y[p+N_{CP}/2],\ y[p+N_{CP}/2+1],\ \ldots,\ y[p+N_{CP}/2+N-1]\}$, we can sample sequences contained in a single OFDM symbol even though there is a minor error in finding first index of an OFDM symbol.

We have shown that $Y^i_{\Delta t_s}[k]$ is $Y^i[k]$ with phase drift and the amplitude of $Y^i_{\Delta t_s}[k]$ is the same with that of $Y^i[k]$. To remove $\mathrm{e}^{-j2\pi\Delta t_s k/N}$ term in (2.1), phase differences between the same subcarrier symbols in two consecutive symbol duration are deployed as

$$\begin{aligned}
&\Delta\angle Y^i_{\Delta t_s}[k] \triangleq \angle Y^{i+1}_{\Delta t_s}[k] - \angle Y^i_{\Delta t_s}[k] \\
&= \angle\left\{Y^{i+1}[k]\mathrm{e}^{-j2\pi\Delta t_s k/N}\right\} - \angle\left\{Y^i[k]\mathrm{e}^{-j2\pi\Delta t_s k/N}\right\} \\
&= \angle Y^{i+1}[k] - \angle Y^i[k].
\end{aligned} \tag{2.3}$$

Equation (2.3) shows that the phase differences between the same subcarrier symbols from sampled sequences are the same as the differences from the received IFFT sequences. Despite the unknown exact Δt_s value, sequences with the same Δt_s can be sampled by setting an interval between starting index of two sample sequences as one OFDM symbol. $|Y_{\Delta t_s}[k]|\mathrm{e}^{j\Delta\angle Y^i_{\Delta t_s}[k]}$ is used as a feature specifying modulation type. In addition, the null subcarrier symbols are removed by detaching symbols with N_{null} smallest amplitudes. Both cases have been evaluated: with and without null subcarrier symbols.

2.2.2 CFO correction

To estimate CFO without symbol-level synchronization, we need to use pilot subcarriers as in residual CFO estimation [9] because the preamble is not accessible. To estimate pilot subcarrier indices without a preamble, we use the property that the iden-

tical symbols are repeatedly transmitted in pilot subcarriers. The CFO-induced phase difference of pilot subcarrier symbols for adjacent OFDM symbols, $\Delta\angle Y^i_{\Delta t_s}[k]$ is

$$\begin{aligned} Y^i_{\Delta t_s}[k] &= \textstyle\sum_{n=0}^{N-1} y^i[n] \mathrm{e}^{-j2\pi k(n+\Delta t_s)/N} \\ &= \textstyle\sum_{n=0}^{N-1} \left(x^i[n] \mathrm{e}^{-j(2\pi\Delta f_c(n+(i-1)(N+N_{CP}))T_s)} \right) \mathrm{e}^{-j2\pi k(n+\Delta t_s)/N} \\ &= X^i[k+\Delta f_c(N+N_{CP})T_s] \mathrm{e}^{-j(2\pi(\Delta f_c(i-1)(N+N_{CP})T_s + k\Delta t_s/N)} \\ &\approx X^i[k] \mathrm{e}^{-j2\pi(\Delta f_c(i-1)(N+N_{CP})T_s + k\Delta t_s/N)} \\ &\Rightarrow \Delta\angle Y^i_{\Delta t_s}[k_p] = -2\pi\Delta f_c(N+N_{CP})T_s \ , \end{aligned} \tag{2.4}$$

where k_p denotes the subcarrier index of pilot subcarriers.

For a Wi-Fi link operating at $f_c = 5\,\mathrm{GHz}$ and a frequency tolerance of 1 ppm for commercial-off-the-shelf temperature-compensated crystal oscillators (TXCO) [10] on both sides of the link, the worst-case CFO is $\Delta f_c = 2 f_c \cdot 10^{-6} = 10$ kHz. The CFO-induced angular error on $\Delta\angle Y^i_{\Delta t_s}[k_p]$ due to CFO at ± 10 kHz is upper bounded by $10^4 \cdot 80/(20 \cdot 10^6) \cdot 360^\circ \approx 14.4^\circ$. Using those values, $X^i[k+\Delta f_c(N+N_{CP})T_s]$ is approximated to $X^i[k]$ since the worst case $\Delta f_c(N+N_{CP})T_s$ is 0.04, which is much smaller than one, the minimum unit of k. This only works for pilot subcarrier symbols since data subcarrier symbols change randomly with the data bits.

Using the pilot subcarriers' property, CFO is estimated with pilot subcarriers:

$$\begin{aligned} \Delta\angle Y^i_{\Delta t_s}[k] &= -2\pi\Delta f_c(N+N_{CP})T_s \\ \Rightarrow \Delta f_c &= -\Delta\angle Y^i_{\Delta t_s}[k]/(2\pi(N+N_{CP})T_s) \end{aligned} \tag{2.5}$$

We consider CFO as the average of Δf_c from (2.5) evaluated at null subcarriers. To correct CFO effect, we multiply time samples $y[n]$ by the term, $\mathrm{e}^{2\pi\Delta f_c n T_s}$ where $n = \{0,1,2,\cdots\}$, which is negative of the phase caused by CFO.

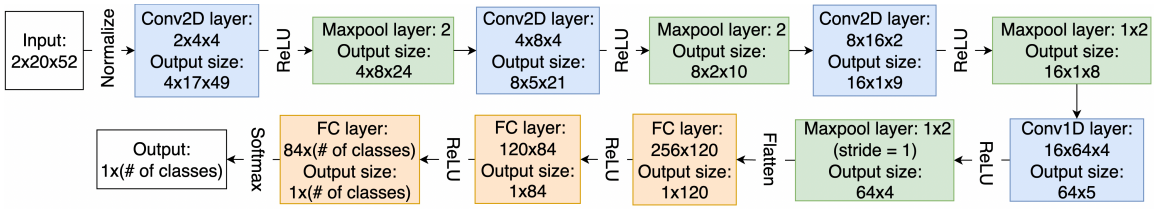


Figure 2.3. CNN-based modulation classifier structure.

Table 2.1. DL parameters

Batch size	32	Loss	Cross-entropy
Learning rate	$5 \cdot 10^{-5}$	Epochs	200

2.2.3 Convolutional neural network

Fig. 2.3 and Table 2.1 describe the overall structure of the DL model for the classifier, which is based on CNN. We use four convolutional layers followed by three fully-connected (FC) layers with ReLU as an activation function. Kernel size and stride of each max pooling layer are adaptively chosen by the output shape of each layer. Input is normalized so that the average amplitude of each input subcarrier sample is 1.

2.3 Evaluation

2.3.1 Evaluation Environments

Simulated and over-the-air (OTA) data are employed in our evaluation with parameters in Table 2.2 corresponding to Non-HT mode 20 MHz bandwidth Wi-Fi. OTA data are generated with two N310 USRP in OTA transmission settings, see in Fig. 2.4.

Table 2.2. Data generation parameters

Bandwidth	20 MHz
Carrier frequency	2.4 GHz
SNR	[2, 20] dB in steps of 2 dB
$\{N, N_{CP}\}$	$\{64, 16\}$
Input shape	Proposed feature: $2 \cdot (64/52) \cdot 20$ Time IQ: $2 \cdot 1600$ STFT feature: $2 \cdot 5 \cdot 512$

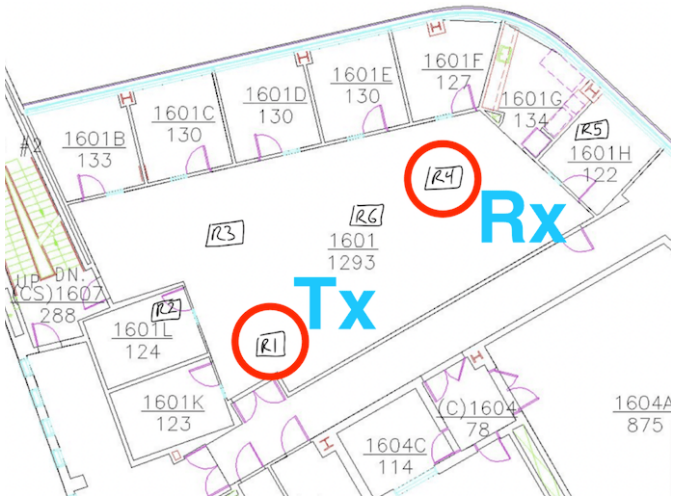


Figure 2.4. The evaluation environment map.

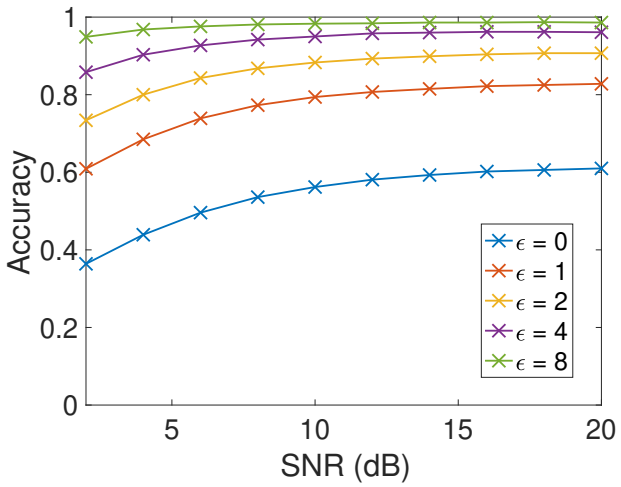


Figure 2.5. Accuracy for choosing the first index of CP with acceptable error ε.

The distance between Tx and Rx is 8.84 m. For simulated data, AWGN is utilized as a channel.

One classifier input requires 1600+80+80 samples. The first additional 80 samples are needed since the starting index of an OFDM symbol is unknown. We sample 1680 samples starting from index $i_s \in [0,\ N-1]$, deploying the procedure in Sec. 2.2. One more OFDM symbol corresponding to the next 80 samples is needed since phase differences between OFDM symbols are used as our features.

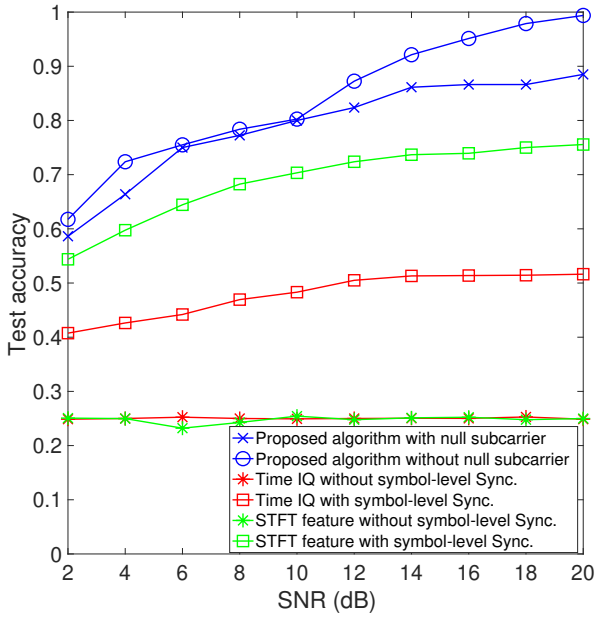


Figure 2.6. Test accuracy vs. test data SNR with simulated data.

For comparison schemes, using raw time IQ samples and complex STFT features as inputs are evaluated. For a fair comparison, the same number of OFDM symbols is used for both inputs. Time IQ samples use 1600 time samples so that the input dimension is $2 \cdot 1600$ (a channel for both real and imaginary). For STFT features, FFT size = 512 and frame overlapping = 50% are deployed as FFT parameters.

Fig. 2.6 shows the classification accuracy performance with simulated data. In every SNR, the proposed feature outperforms the time IQ and STFT feature regardless of symbol-level synchronization. Since the number of modulation classes is four, 25% accuracy achieved with time IQ and STFT without symbol-level synchronization corresponds to that of the uninformed random classifier. Removing the null subcarriers increases classification performance in every case. At 20 dB SNR, it is increased from 89% to 99% by deleting null subcarriers. Both accuracies are higher at 20 dB SNR than the 87% and 79% achieved by [11, 12], both assume symbol-level synchronization and test on simulated data.

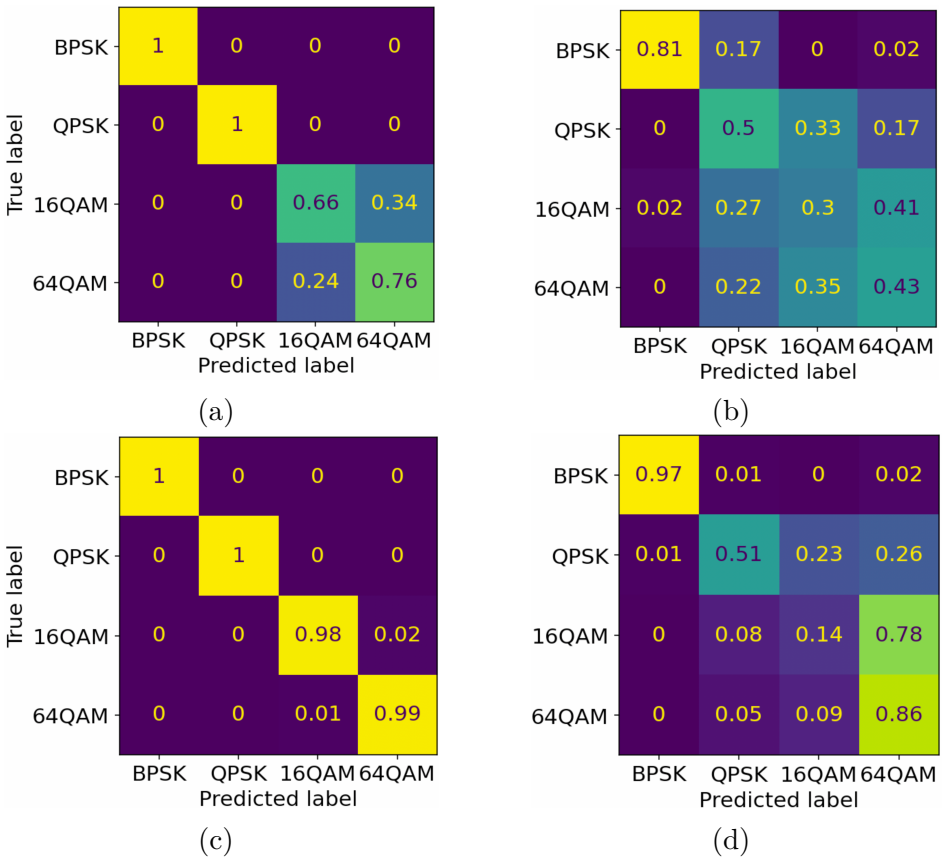


Figure 2.7. Confusion matrices for classification results with OTA data: (a) 20 dB SNR with null subcarrier, (b) 2 dB SNR with null subcarrier, (c) 20 dB SNR after removing null subcarrier, and (d) 2 dB SNR after removing null subcarrier.

Fig. 2.5 illustrates the accuracy of correctly choosing the first index with the method in Sec. 2.2 with error, ε. The accuracy means that the estimated first sample is at most ε time samples away from the ground-truth. In Fig. 2.5, accuracy to find the exact first time sample of an OFDM symbol is below 60% at 20 dB SNR, but accuracy accepting at most $N_{CP}/2 = 8$ time samples error is 95% at 2 dB SNR. The sampling method in Sec. 2.2 lets a sequence contained in a single OFDM symbol reliably sampled even at low SNR.

The proposed algorithm outperforms the case where time IQ samples or STFT features are used on OTA data as well as removing the null subcarrier improves the performance. In particular, at 2 dB SNR and 20 dB SNR, test accuracy is improved from

53 to 59% and 85 to 99%, respectively. Comparing Fig. 2.7a and Fig. 2.7c, removing null subcarrier makes the features of 16QAM and 64QAM more distinct at high SNR.

2.4 Related Work

Many have built DL-based modulation classifier for OFDM signals [11, 12, 13, 14, 15, 16, 17, 18] and all of [11, 12, 13, 14, 15, 16, 17, 18] have achieved at least 75% classification accuracy at 20 dB SNR. However, none of the studies [11, 12, 13, 14, 15, 16, 17, 18] has done the evaluation with hardware-generated data.

The algorithm in [13] deploys correlation both within a symbol and among different symbols, thus the classifier knows the exact first indices of OFDM symbols; i.e., symbol-level synchronization. Modulation classifiers has been implemented based on CNN [11, 14, 15] or long short term memory network (LSTM) [12] all with over 80% classification accuracy at 20 dB SNR. Their input comprises time samples for two OFDM symbols after removing CP, which requires the classifier synchronized at the symbol-level. Modulation classifiers in [17, 18] take signals after removing cyclic prefix (CP) and using FFT as an input. Therefore, it is assumed that the classifier is synchronized at the symbol-level. The work in [16] studies modulation classification under multipath channel, but only deals with the single-carrier signals. The authors of [19] classify wireless signals, but recognizes wireless protocols, not modulations.

There are papers on OFDM modulation classification without symbol-level synchronization based on mathematical modeling [20, 21] without using DL. However, their classifier structure depends on the modulation set, so the structure must be redesigned when classifying signals with a modulation not in the set [20, 21]. Their algorithms [20, 21] can identify OQPSK and MSK, but neither of their evaluations includes high-order QAM like 64QAM, used in practical Wi-Fi.

2.5 Conclusion

OFDM modulation classification is addressed without symbol-level synchronization. We propose a preprocessing for extracting features invariant to synchronization error. Fine-grained preprocessing include the estimated CP position and CFO correction. The proposed CNN-based classifier based on those features classifies the modulation of OFDM in evaluations with simulated and hardware-generated data with maximum 99% classification accuracy at 20 dB SNR. Best test accuracy is achieved by the proposed CNN-based classifier with null subcarrier removal.

2.6 Acknowledgements

Chapter 2, in full, is a reprint of the material as it appears in Kim, B., Mecklenbräuker, C., and Gerstoft, P. "Deep Learning-based Modulation Classification of Practical OFDM Signals for Spectrum Sensing", in Proc. International Conference on Acoustics, Speech, and Signal Processing Workshop 2023. The dissertation author was the primary investigator and author of this paper. The coauthors listed in this publication directed and supervised the research.

2.7 References

[1] J. Gao, X. Yi, C. Zhong, X. Chen, and Z. Zhang, "Deep learning for spectrum sensing," IEEE Wireless Commun. Letters, vol. 8, no. 6, pp. 1727–1730, 2019.

[2] H. Kim, Y. Jiang, S. Kannan, S. Oh, and P. Viswanath, "Deepcode: Feedback codes via deep learning," in Proc. Adv. Neural Inf. Process. Syst., vol. 31, pp. 9436–9446, 2018.

[3] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5g wireless communications: A deep learning approach," IEEE Trans. Netw. Sci. Eng., vol. 7, no. 1, pp. 227–236, 2018.

[4] F. Meng, P. Chen, L. Wu, and X. Wang, "Automatic modulation classification: A deep learning enabled approach," IEEE Trans. Veh. Technol., vol. 67, no. 11, pp. 10760–10772, 2018.

[5] V. Sathyanarayanan, M. Wagner, and P. Gerstoft, "Over the air performance of deep learning for modulation classification across channel conditions," in Proc. IEEE Asilomar Conf. Signals, Syst. Comput., pp. 157–161, Nov. 2020.

[6] V. Sathyanarayanan, A. Jolly, and P. Gerstoft, "Novel training methodology to enhance deep learning based modulation classification," in Proc. IEEE Asilomar Conf. Signals, Syst. Comput., pp. 356–360, Oct. 2021.

[7] IEEE 802.11ax, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 1: enhancements for high-efficiency WLAN," May 2021.

[8] A. Kumar, S. Majhi, G. Gui, H.-C. Wu, and C. Yuen, "A survey of blind modulation classification techniques for ofdm signals," Sensors, vol. 22, no. 3, p. 1020, 2022.

[9] J. K. Tan, "An adaptive orthogonal frequency division multiplexing baseband modem for wideband wireless channels," Master's thesis, Massachusetts Institute of Technology, 2006.

[10] G. E. Ltd, "GTXO-203T | 1.8V∼3.6V SM TCXO | Golledge," 2025.

[11] S. Hong, Y. Zhang, Y. Wang, H. Gu, G. Gui, and H. Sari, "Deep learning-based signal modulation identification in OFDM systems," IEEE Access, vol. 7, pp. 114631–114638, Aug. 2019.

[12] Z. Zhang, H. Luo, C. Wang, C. Gan, and Y. Xiang, "Automatic modulation classification using cnn-lstm based dual-stream structure," IEEE Trans. Veh. Technol., vol. 69, pp. 13521–13531, Nov. 2020.

[13] T. Huynh-The, Q.-V. Pham, T.-V. Nguyen, X.-Q. Pham., and D.-S. Kim, "Deep learning-based automatic modulation classification for wireless ofdm communications," in Proc. IEEE ICTC, pp. 47–49, Oct. 2021.

[14] J. Shi et al., "Deep learning-based automatic modulation recognition method in the presence of phase offset," IEEE Access, vol. 8, pp. 42841–42847, 2020.

[15] S. Hong, Y. Wang, Y. Pan, H. Gu, M. Liu, J. Yang, and G. Gui, "Convolutional neural network aided signal modulation recognition in OFDM systems," in Proc. IEEE VTC, pp. 1–5, May 2020.

[16] J. Venalainen, L. Terho, and V. Koivunen, "Modulation classification in fading multipath channel," in Proc. IEEE Asilomar Conf. Signals, Syst. Comput., vol. 2, pp. 1890–1894, 2002.

[17] D. H. Al-Nuaimi, N. A. M. Isa, M. F. Akbar, and I. S. Z. Abidin, "AMC2-pyramid: Intelligent pyramidal feature engineering and multi-distance decision making for automatic multi-carrier modulation classification," IEEE Access, vol. 9, pp. 137560–137583, Sept. 2021.

[18] Z. Zhao, et al., "Modulation format recognition based on transfer learning for visible light communication systems," in Proc. Optoelectronics Commun. Conf., p. JS2B.12, July 2021.

[19] S. R. Shebert, A. F. Martone, and R. M. Buehrer, "Wireless standard classification using convolutional neural networks," in Proc. IEEE GLOBECOM, pp. 1–6, 2021.

[20] R. Gupta, S. Kumar, and S. Majhi, "Blind modulation classification for asynchronous ofdm systems over unknown signal parameters and channel statistics," IEEE Trans. Veh. Technol., vol. 69, pp. 5281–5292, Mar. 2020.

[21] A. K. Pathy, A. Kumar, R. Gupta, S. Kumar, and S. Majhi, "Design and implementation of blind modulation classification for asynchronous mimo-ofdm system," IEEE Trans. Instrum. Meas., vol. 70, pp. 1–11, Sept. 2021.

Chapter 3

Blind Modulation Classification of Wi-Fi 6 and 5G signals for Spectrum Sensing

3.1 Introduction

The growth of wireless communication technologies has necessitated efficient usage of the radio spectrum, a challenge that is being addressed with cognitive radio. A critical component of cognitive radio is intelligent spectrum sensing, which allows for a more precise characterization of spectrum usage and aids in better decision-making for spectrum allocation. Spectrum sensing encompasses signal detection [1], predicting future spectrum [2], and identifying modulation schemes. In this study, we focus on the classification of modulations of practical orthogonal frequency division multiplexing (OFDM) signals. This also enables applications such as channel quality estimation between a transmitter (TX) and a receiver (RX) for a spectrum sensor and wireless network troubleshooting.

OFDM has become essential in modern wireless communication systems, such as Wi-Fi 6 and 5G. In these systems, message bits are converted to digital symbols using modulation schemes such as quadrature phase shift keying (QPSK) and transmitted via data subcarriers. In OFDM, the multiple symbols are stacked in subcarriers within the frequency domain, so each OFDM time sample contains only a fraction of the information on the multiple frequency domain symbols. As a result, the modulation classifiers designed

for single-carrier signals [3] cannot be applied directly to OFDM signals. Therefore, a precise modulation classification of Wi-Fi 6 and 5G signals requires additional processing beyond using raw time-domain samples as inputs.

A spectrum sensor must handle OFDM signals with diverse configurations without prior information, in contrast to user equipment connected to a wireless network. In Wi-Fi 6 and 5G systems, information about the user data transmission, including the modulation, is provided to the RX. However, since a spectrum sensor does not have prior knowledge of the type of signals it detects, it cannot deploy a protocol-specific procedure to obtain information about user data transmission. The parameters shaping OFDM signals, fast Fourier transform (FFT) size to generate inverse fast Fourier transform (IFFT) sequence, and cyclic prefix (CP) length, might be different even among OFDM signals with the same modulation scheme. Moreover, the carrier frequency configurations in 5G become increasingly diverse and data transmission might occupy only a part of channel bandwidth. As a result, estimation of these carrier frequency configurations is becoming increasingly difficult using transmission bandwidth and center frequency alone. Thus, a modulation classifier for spectrum sensing should estimate the modulation scheme using only the observed user data transmission without knowledge of carrier frequency.

We present a system to classify the modulation of the signals in Wi-Fi 6 [4] and 5G [5] for a spectrum sensing system. Without knowledge of the transmitter (TX) carrier frequency, Wi-Fi preamble, or 5G control information, our system deploys only the basic OFDM structure, IFFT sequence, and CP. The system includes the estimation of OFDM parameters: CP length and subcarrier spacing (SCS), which is directly related to the FFT size of the IFFT sequence. We focus on identifying modulation schemes used in the payload of Wi-Fi 6 signals and the physical downlink shared channel (PDSCH) of 5G signals. Signals studied in this paper are single-input single-output (SISO). For 5G, they are in the frequency range 1 (FR1), whose frequency band is below 7.125 GHz.

For IFFT sequence and CP length estimation, the cyclic autocorrelation function

(CAF) is deployed. The capability of CAF detecting repeated sequences as well as repetition periods enables the estimation of those parameters. We observe that symbol-level synchronization is not perfect if autocorrelation using CP is utilized only. Our preprocessing removes the effect of the synchronization error by using phase differences between phases of two adjacent OFDM symbols. The modulation classifier for Wi-Fi 6 and 5G signals should recognize high-order modulations such as 256 quadrature amplitude modulation (QAM) and 1024QAM since these state-of-the-art protocols include those schemes. We change the feature format to a histogram representing the distribution of the features so that the classifier can effectively capture high-order modulations characteristics. Related work on modulation classification: Many papers address modulation classification for wireless communication signals [3, 6, 7, 8, 9, 10, 11, 12, 13]. The works in [6, 7, 8, 9, 10, 11] study modulation classification of OFDM signals and achieve at least 78% accuracy at 20 dB SNR for an AWGN channel. It is assumed that the inputs start from the first sample of OFDM symbol duration [6, 7, 8, 11], which is only possible by detecting Wi-Fi preamble or 5G synchronization signals properly. To apply this idea to a spectrum sensor, the sensor should follow protocol-specific procedures.

There are papers on OFDM modulation classification without the symbol-level synchronization assumption [9, 10, 12, 13] and the algorithms [9, 10, 13] are evaluated with hardware-generated data. However, their algorithms [9, 10, 13] are not evaluated with high-order modulations such as 256QAM or 1024QAM, as used in Wi-Fi 6. Moreover, since their classifier structures [9, 10] is designed to recognize a fixed set of modulations, the overall structure needs redesign to identify a new modulation scheme. The work [12] proposes the system to estimate SCS of OFDM signals and modulation of single-carrier signals jointly. Nonetheless, it does not estimate the modulation of OFDM signals. The neural network-based modulation classifier [3] studies how environmental change affects classification performance for only the single-carrier signals, not OFDM signals.

Related work on sniffing OFDM signals: For spectrum sensing, modulation might

be identified by sniffing control data used to notify RX. The work [14, 15, 16, 17] tried to overhear Long Term Evolution (LTE) signals. LTEye [14] and OWL [15] decodes PHY DL control channel (PDCCH) data for LTE network monitoring. LTESniffer [17] decode sniffed both user and control data using PDCCH decoder FALCON [16]. FALCON overcomes the limitation of LTEye and OWL, which require more than 97% decoding accuracy. In LTE, a starting symbol of PDCCH in a slot is always the first symbol in a slot and it is different from 5G, where the PDCCH starting symbol in a slot can be any symbol in a slot and its information is notified with radio resource control (RRC) signals. Accordingly, it is not straightforward to generalize LTE PDCCH sniffer to 5G. Eavesdropping PDCCH data of 5G signals [18] can deal with the signal with diverse 5G configurations, but is vulnerable to configuration changes since it takes a few minutes to learn a new PDCCH configuration. The authors of [19] study sniffing Wi-Fi probe request packets, which is for mobile devices to broadcast the existence of themselves. They build a hardware model for a sniffer and test with real Wi-Fi probe request packets. However, the probe request packets are simpler than those for user data communication thus not straightforward to deploy this system for our target signal.

To summarize, the main contributions of the paper are:

- OFDM parameter estimation for up-to-date protocols: We have applied the OFDM parameter estimation method with CAF [20] to Wi-Fi 6 and 5G signals to estimate SCS and CP length.

- Feature extraction without symbol-level synchronization: Only with estimated values of SCS and CP length, our system builds the features characterizing modulation of OFDM signals. The proposed feature extraction algorithm is designed to be resilient to symbol-level synchronization errors caused by using CP only.

- Modulation classification without control information: For spectrum sensing, control information might not be accessible. We show that the proposed classification

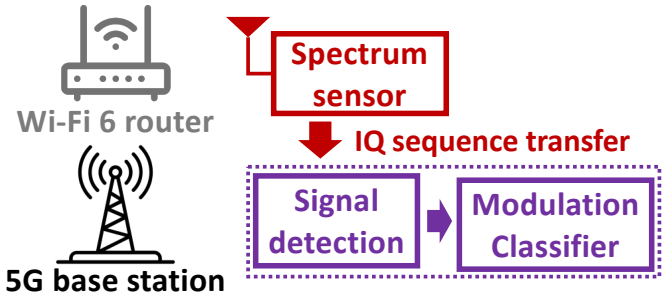


Figure 3.1. System model to capture DL Wi-Fi 6 and 5G signals.

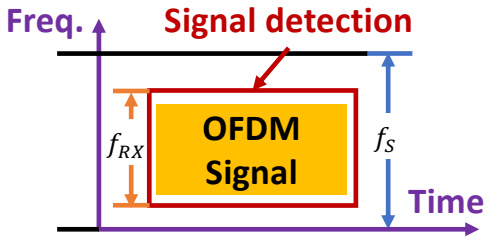


Figure 3.2. Signal detection scenario.

system robustly works with diverse configurations with the evaluation of hardware-generated data without knowledge of the information.

3.2 System Objective

We aim to build a modulation classifier using in-phase and quadrature (IQ) samples of SISO Wi-Fi 6 and FR1 5G DL signal for spectrum sensing. The system scenario is described in Fig. 3.1. There is a Wi-Fi 6 or 5G TX transmitting its signal to an RX. The spectrum sensor continuously senses the spectrum by generating IQ samples with sampling rate f_S and transfers those samples to a signal detection algorithm. Using IQ samples captured by a receiver antenna, the signal detection algorithm detects the duration and frequency band where the OFDM signal is located and extracts IQ samples corresponding to the detected OFDM signal, described as the blue rectangle in Fig. 3.2. We assume the accurate signal detection of Wi-Fi 6 or 5G signals and a single modulation scheme is used for data communication in one detected OFDM signal.

The IQ samples from the spectrum sensor sampled with rate f_S are resampled to

Table 3.1. Variable definitions

Variable	Definition (unit)
f_{TX}	TX sampling rate (Hz)
f_{RX}	Sampling rate of a system input sequence (Hz)
Δf_{SCS}	Subcarrier spacing (Hz)
T_{IFFT}	IFFT sequence duration (s)
N_{FFT}	FFT size used to generate IFFT sequence
T_{CP}	CP duration (s)
N_{CP}	Number of time samples in CP for one OFDM symbol
$y[n]$	Received time-domain sequence after resampling to 20 MHz
$y'[n]$	5G time-domain sequence after resampling to 30.72 MHz
$y^i[n]$	Received time-domain IFFT sequence for the ith OFDM symbol
$Y^i[k]$	Received symbol in subcarrier k for the ith OFDM symbol
$(\mathscr{S} \times \mathscr{S})$	Number of bins of a classifier input

f_{RX}, 20 MHz. We only consider Wi-Fi 6 signals with 20 MHz channel bandwidth and 5G signals with a PDSCH bandwidth from 15 to 20 MHz. Thus, 20 MHz sampling rate can let the resampled IQ sequence encompass the OFDM signal in our scenario. Extending the analysis to different transmission bandwidth ranges is straightforward. These resampled IQ samples, denoted by $y[n]$, are taken as inputs of the feature extraction algorithm, as elaborated in Sec. 3.3 in detail.

3.2.1 Wi-Fi 6 PHY layer

Wi-Fi 6 supports the high-efficiency (HE) transmission format as well as earlier formats, such as non-high throughput (non-HT), high throughput (HT), and very high throughput (VHT) formats. Table 3.2 summarizes the parameters that configure the payload of the Wi-Fi frame for each Wi-Fi format. In HE format, the number of subcarriers is increased because the subcarrier spacing (SCS, denoted as Δf_{SCS}) is one-fourth of that

Table 3.2. Parameters for different formats of Wi-Fi

	Non-HT format	HT format
T_{IFFT}	$3.2\,\mu$s	$3.2\,\mu$s
T_{CP}	$0.8\,\mu$s	$\{0.4, 0.8\}\,\mu$s
Modulations	BPSK, QPSK, 16QAM, 64QAM	BPSK, QPSK, 16QAM, 64QAM
	VHT format	HE format
T_{IFFT}	$3.2\,\mu$s	$12.8\,\mu$s
T_{CP}	$\{0.4, 0.8\}\,\mu$s	$\{0.8, 1.6, 3.2\}\,\mu$s
Modulations	BPSK, QPSK, 16QAM, 64QAM, 256QAM	BPSK, QPSK, 16QAM, 64QAM, 256QAM, 1024QAM

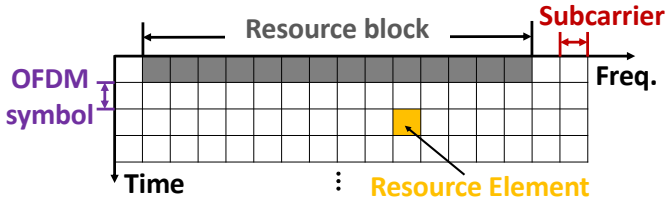


Figure 3.3. Example 5G resource grid.

of the previous transmission formats. Over time, the Wi-Fi standard has evolved and several options for the CP duration are available.

3.2.2 5G DL PHY layer

The 5G downlink (DL) resource structure and its associated terminology is illustrated in Fig. 3.3. A resource element (RE) represents the smallest unit which carries data, encompassing a single OFDM symbol in the time domain and a single subcarrier in

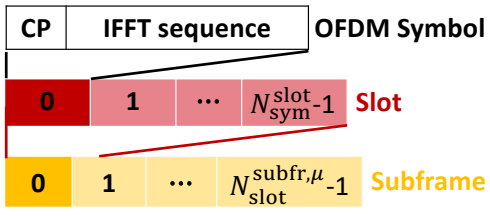


Figure 3.4. 5G subframe structure.

Table 3.3. 5G frame structure parameters

{SCS (kHz), CP option}	{60, Normal}	{60, Extended}
T_{IFFT}	$16.17\,\mu$s	$16.67\,\mu$s
{Short, long} T_{CP}	$\{1.17, 1.69\}\,\mu$s	$\{4.17, -\}\,\mu$s
N_{FFT} when $f_{\text{TX}} = 30.72$ MHz	512	512
{Short, long} N_{CP} when $f_{\text{TX}} = 30.72$ MHz	{36, 52}	128
{SCS (kHz), CP option}	{30, Normal}	{15, Normal}
T_{IFFT}	$33.33\,\mu$s	$66.67\,\mu$s
{Short, long} T_{CP}	$\{2.34, 2.86\}\,\mu$s	$\{4.69, 5.21\}\,\mu$s
N_{FFT} when $f_{\text{TX}} = 30.72$ MHz	1024	2048
{Short, long} N_{CP} when $f_{\text{TX}} = 30.72$ MHz	{72, 88}	{144, 160}

the frequency domain. A resource block (RB) is the smallest radio resource that can be allocated and refers to one OFDM symbol in the time domain and 12 subcarriers in the frequency domain.

Figure 3.4 shows the 5G subframe structure in the time domain. An OFDM symbol in 5G is comprised of both a CP and an inverse fast Fourier transform (IFFT) sequence. The number of symbols within a single slot ($N_{\text{sym}}^{\text{slot}}$) varies in accordance with the CP length. There are a normal and an extended CP option in the transmission format. When a normal CP is used then $N_{\text{sym}}^{\text{slot}} = 14$, otherwise $N_{\text{sym}}^{\text{slot}} = 12$. The SCS, the distance between two adjacent subcarries in OFDM systems, denoted by μ, determines the number of slots within a single subframe, $N_{\text{slot}}^{\text{subfr},\mu}$. There are five SCS options in 5G, but we consider only three cases, namely 15, 30, 60 kHz, which are available in FR1. These SCS values correspond to $\mu = 0, 1, 2$, respectively, and the number of slots in a subframe for each SCS is computed as $N_{\text{slot}}^{\text{subfr},\mu} = 2^{\mu}$.

The structural parameters which define the 5G frame are listed in Table 4.3. The length of an IFFT sequence, T_{IFFT}, is:

$$T_{\text{IFFT}} = N_{\text{FFT}}/f_{\text{TX}} = 1/\Delta f_{\text{SCS}}. \tag{3.1}$$

Table 3.4. Modulations used for 5G physical channels

Physical channel	PDSCH	PSS/SSS	PDCCH	CSI-RS
Modulation	QPSK, 16QAM, 64QAM, 256QAM, 1024QAM	BPSK	QPSK	QPSK
Physical channel	PBCH	PDSCH-PTRS	PDSCH-DMRS	
Modulation	QPSK	QPSK	QPSK	

There is a one-to-one correspondence between T_{IFFT} and Δf_{SCS} (3.1). Under the normal CP option, CP is longer than that in other symbols, every 0.5 ms, or equivalently, $7 \cdot 2^{\mu}$ OFDM symbols in OFDM symbol unit, called long CP. There is no long CP in the extended CP option, so T_{CP} is uniform. The transmission rate of 5G signals is a power of 2 times 15 kHz and 30.72 MHz is an example of 5G transmission rate. N_{FFT} and N_{CP} values are arranged when f_{TX} is 30.72 MHz, the value used in our evaluation.

In addition to PDSCH, there exist other physical (PHY) channels that but serve specific functions although not carrying user data. For instance, PDCCH conveys downlink control information (DCI), which contains information required to decode PDSCH data such as modulation and coding scheme (MCS). Each of these channels utilizes predefined single-type modulation, see Table 3.4.

Compared to Wi-Fi, which has a predefined configuration of data, pilot, and null subcarriers, 5G resource configuration for PHY channels is flexible. Instead, the 5G system has a network dedicated to exchanging information on how data packets are forwarded, called the control plane, in addition to the network for data transmission, called the user plane. An example of data transferred over the control plane is RRC signals. Information on the starting OFDM symbol of PDCCH and channel state information-reference signal (CSI-RS) is notified to an RX with RRC signals via control plane [5].

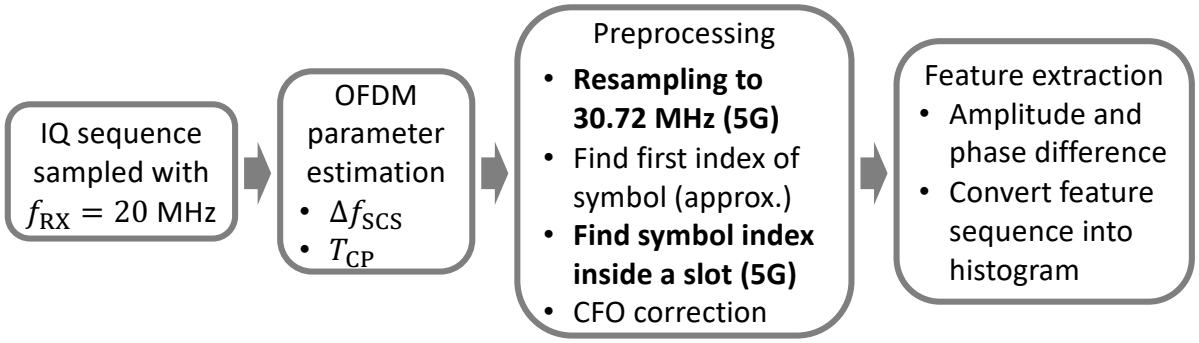


Figure 3.5. Flow chart of proposed feature extraction algorithm.

3.3 Proposed Algorithm

High-level procedures to build features characterizing the modulations of Wi-Fi 6 and 5G signals are illustrated in Fig. 3.5 and explained in Sec. 3.3.1 and 3.3.2. The output of this flowchart is taken as an input to the neural network model, described in Sec. 3.3.3.

3.3.1 OFDM parameter estimation

Prior to building the features which characterize modulation, it is necessary to estimate two essential OFDM parameters of OFDM signals, SCS and CP length. To estimate these parameters, we use CAF, a Fourier-series cofficient of autocorrelation function.

$$\mathscr{R}_{yy}(\alpha,\tau) = \sum_{n=-\infty}^{\infty} \mathscr{R}_{yy}(n,\tau)\mathrm{e}^{-j2\pi\alpha n}. \tag{3.2}$$

CAF is used to extract a repeated pattern presented in wireless signals [20, 21, 22]. A variant of the CAF estimator presented in [20] is deployed here,

$$\hat{\mathscr{R}}_{yy}(\alpha,\ell) = \frac{1}{\mathscr{N}} \sum_{n=0}^{\mathscr{N}-1} \left\{ \sum_{i=0}^{l-1} y(n+i)y^*(n+i+\ell) \right\} \mathrm{e}^{-j2\pi\alpha n}. \tag{3.3}$$

One sample of our estimator is computed as the autocorrelation with delay ℓ. It differs from the estimator in [20], where only two samples are used to compute one estimator sample. This modification aims to make peaks more distinct. We set $l = 8$ corresponding

to the shortest CP length.

CP in OFDM symbols causes a sequence to be repeated at both ends of each symbol. The distance between starting indices of the two repeated sequences located at both ends of an OFDM symbol is T_{IFFT} or $N_{\text{FFT}}(f_{\text{RX}}/f_{\text{TX}}) = f_{\text{RX}}/\Delta f_{\text{SCS}}$, depending on whether it is in time units or time sample units, respectively. This repetition makes the CAF estimator at $\alpha = 0$ have a peak at $\ell = f_{\text{RX}}/\Delta f_{\text{SCS}}$. T_{CP} is also estimated with the CAF estimator, $\hat{\mathscr{R}}_{yy}(\alpha, f_{\text{RX}}/\Delta f_{\text{SCS}})$. Since $\sum_{i=0}^{l-1} y(n+i)y^*(n+i+\tau)$ in Eq. (3.3) has peaks with the period of $f_{\text{RX}} \cdot (T_{\text{CP}} + 1/\Delta f_{\text{SCS}})$, it is expected of $\hat{\mathscr{R}}_{yy}(\alpha, f_{\text{RX}}/\Delta f_{\text{SCS}})$ to have a large amplitude at $\alpha = 1/\{f_{\text{RX}} \cdot (T_{\text{CP}} + 1/\Delta f_{\text{SCS}})\}$.

In our scenario, there are five candidates for ℓ values, $\ell_C = \{64, 256, 333, 667, 1333\}$, each of which corresponds to an IFFT sequence length for a given SCS at $f_{\text{RX}} = 20$ MHz. IFFT sequence length is estimated as:

$$T_{\text{IFFT}} = \ell'/f_{\text{RX}} \text{ s.t. } \ell' = \underset{\ell \in \ell_C}{\arg\max} \left| \hat{\mathscr{R}}_{yy}(0, \ell) \right| \tag{3.4}$$

When the estimated T_{IFFT} corresponds to that of Wi-Fi 6 or 60 kHz SCS NR, where multiple CP options are available, CP length is further estimated as:

$$T_{\text{CP}} = \frac{1}{f_{\text{RX}}} \left(\frac{1}{\alpha'} - \ell' \right) \text{ s.t. } \alpha' = \underset{\alpha \in \alpha_C^{\ell'}}{\arg\max} \left| \hat{\mathscr{R}}_{yy}(\alpha, \ell') \right| \tag{3.5}$$

where $\alpha_C^{\ell'}$ denotes a set of possible values of $\alpha = 1/\{\ell' + (f_{\text{RX}} \cdot T_{\text{CP}})\}$, given ℓ'.

3.3.2 Feature extraction

The motivation behind our proposed feature extraction lies in the observation that when a sampled time-domain sequence is contained within a single OFDM symbol, the Fast Fourier Transform (FFT) of that sequence yields the original symbols with a phase drift that scales linearly with subcarrier index k and synchronization error Δn, as shown

in

$$
\begin{aligned}
Y^i_{\Delta n}[k] \triangleq \mathscr{F}\left(y^i[n-\Delta n]\right) &= \sum_{n=0}^{N_{\mathrm{FFT}}-1} y^i[n-\Delta n]\mathrm{e}^{-j2\pi nk/N_{\mathrm{FFT}}} \\
&= Y^i[k]\mathrm{e}^{-j2\pi\Delta nk/N_{\mathrm{FFT}}}.
\end{aligned} \tag{3.6}
$$

In order to build a feature characterizing modulation based on this property, two objectives must be achieved: first, sampling a sequence that is fully contained within an OFDM symbol, and second, removing the phase drift caused by synchronization errors.

Utilizing the knowledge of N_{CP} and N_{FFT}, the CP position is determined through autocorrelation analysis,

$$
R_{yy}(n,N_{\mathrm{FFT}}) = \frac{1}{N_{\mathrm{CP}}}\sum_{i=0}^{N_{\mathrm{CP}}-1} y[n+i]y^*[n+i+N_{\mathrm{FFT}}]. \tag{3.7}
$$

The position of CP is indicated by the peaks in $|R_{yy}(n,N)|$. To locate a peak, we search for a sample whose amplitude is larger than both of its neighboring samples while ensuring that the minimum distance between two adjacent peaks is 90% of the OFDM symbol duration (i.e., 288-time sample indices for HE format with 3.2 μs CP), to avoid selecting undesired local peaks. We compute their remainders divided by the sample number of the OFDM symbol duration ($N_{\mathrm{CP}}+N_{\mathrm{FFT}}$) over multiple OFDM symbols. The median of those remainders is determined as the first index of the OFDM symbol, denoted as p. Noise and varying amplitudes of time samples can introduce small errors in the estimated CP position. To reliably sample the sequences contained in a single OFDM symbol, we deploy the sequence $\{y[p+N_{\mathrm{CP}}/2], y[p+N_{\mathrm{CP}}/2+1], \cdots,\ y[p+N_{\mathrm{CP}}/2+N-1]\}$. This sequence is entirely within a single OFDM symbol unless the estimation error of p is larger than $N_{\mathrm{CP}}/2$.

We demonstrate that $Y^i_{\Delta n}[k]$ exhibits a phase drift modeled by $\mathrm{e}^{-j2\pi\Delta nk/N}$, while maintaining its amplitude $Y^i[k]$ (3.6). We remove this phase drift due to synchronization errors by computing the phase differences between successive symbols in the same

subcarrier k as:

$$\begin{aligned}\Delta\angle Y^i_{\Delta n}[k] &\triangleq \angle Y^{i+1}_{\Delta n}[k] - \angle Y^i_{\Delta n}[k] \\ &= \angle\left\{Y^{i+1}[k]\mathrm{e}^{-j2\pi\Delta nk/N}\right\} - \angle\left\{Y^i[k]\mathrm{e}^{-j2\pi\Delta nk/N}\right\} \\ &= \angle Y^{i+1}[k] - \angle Y^i[k].\end{aligned} \tag{3.8}$$

Despite the lack of knowledge about Δn, sequences with constant Δn can be obtained by adjusting the interval between the starting indices of two sampled sequences to be one OFDM symbol. The feature used to identify the modulation type is $Y^i_f[k] \triangleq |Y^i_{\Delta n}[k]|\mathrm{e}^{j\Delta\angle Y^i_{\Delta n}[k]}$. The null subcarrier symbols is eliminated by discarding symbols with the N_{null} smallest amplitudes.

In protocol-compliant reception, the Wi-Fi preamble and 5G PDSCH-phase tracking reference signal (PDSCH-DMRS) are deployed for CFO estimation. However, since they are not accessible to a spectrum sensor, the CP in each OFDM symbol is used for CFO estimation, i.e.,

$$\angle\left(\frac{y(p+N_{\text{FFT}}+i)}{y(p+i)}\right) = 2\pi\Delta f_c/\Delta f_{\text{SCS}}, \tag{3.9}$$

where $y(p+i)$ is in CP. We use $i \in \{\lfloor N_{\text{CP}}/4\rfloor, \cdots, \lceil 3N_{\text{CP}}/4\rceil\}$ so that the sequence $y(p+i)$ are entirely within CP unless estimation error of p is larger than $N_{\text{CP}}/4$. If the absolute value of the CFO is larger than $\Delta f_{\text{SCS}}/2$, CFO cannot be accurately estimated due to aliasing. It is discussed in Sec. 3.3.2 in detail.

Additional procedures for 5G signal

To build a modulation feature for 5G, 5G characteristics distinct from those of Wi-Fi, including a different transmission rate, long CP, and flexible usage of subcarriers, should be considered. First, the transmission rate of 5G signals is not $f_{\text{RX}} = 20$ MHz, but is the form of a power of 2 times 15 kHz. Hence, if the signal is classified as 5G, we resample the sequence to 30.72 MHz $= 2048 \cdot 15$ kHz, the smallest sampling frequency above 20 MHz. N_{FFT} and N_{CP} with 30.72 MHz sampling rate for each Δf_{SCS} are arranged

in the last two rows in Table 3.3.

In the case of the normal CP option, there is a long CP every 0.5 ms, which is slightly longer than that of other OFDM symbols. Long CP breaks the assumption of the uniform OFDM symbol durations, which is required by the method to find the first indices of OFDM symbols and to build $Y_f^i[k]$. Specifically in building $Y_f^i[k]$, maintaining the fixed interval does not guarantee the constant Δn over multiple OFDM symbols. Therefore, long CP also should be located when finding the first index of the OFDM symbol.

Algorithm 1: Finding first index of long CP

Data: ($y'[n]$ of length (3 ms + 3 OFDM symbols)), μ

Result: firstIndexLongCP$= q +$ symLongCP$\cdot (N_{\text{FFT}} + N_{\text{CP}})$

1 $m = 7 \cdot 2^{\mu}$, $N_{\text{FFT}} = 512 \cdot 2^{2-\mu}$, $N_{\text{CP}} = 18 \cdot 2^{2-\mu}$, $i = 0$;

2 while $i \leq 5$ do

3 $\quad y'_i[n] = \{y'[(30.72 \cdot 10^6) \cdot (0.5 \cdot 10^{-3}) \cdot i],\ \cdots,\ y'[(30.72 \cdot 10^6) \cdot (0.5 \cdot 10^{-3}) \cdot (i+1) + 3 \cdot (N_{\text{FFT}} + N_{\text{CP}})]\}$;

4 $\quad$ Find peaks $\{p'_{i0}, \cdots, p'_{i(m+1)}\}$ with $y'_i[n]$ using autocorrelation $|R_{y'_i y'_i}(n, N_{\text{FFT}})|$ and peak locating function explained in Sec. 4.3.2;

5 $\quad p_{ij} = \text{mod}(p'_{ij}, N_{\text{FFT}} + N_{\text{CP}})$, $i = i + 1$;

6 end

7 $\Delta p_j = \text{Mean}(\{p_{0(j+1)} - p_{0(j-1)}, \cdots, p_{5(j+1)} - p_{5(j-1)}\})$;

8 $\{\Delta p_{k_0}, \cdots, \Delta p_{k_{m-1}}\} = \text{sortDescending}(\{\Delta p_j\})$;

9 symLongCP $= \arg\max_{k_q} \text{Var}(\{p_{0k_q}, \cdots, p_{5k_q}\})$ where $q \in \{0, 1\}$;

10 $q_{ij} = \begin{cases} p_{ij} & \text{if } j \leq \text{symLongCP} \\ p_{ij} - 16 & \text{otherwise} \end{cases}$ $q_j = \text{Median}(\text{Mean}(\{q_{0j}, \cdots q_{5j}\}))$ where $j \in \{0, 1, \cdots, m-1\} - \{\text{symLongCP}\}$;

Algorithm 1 explains the detailed steps to estimate the first index of OFDM symbol with long CP. $y'_i[n]$ in line 3 is a sequence cropped to be as long as (0.5 ms + 3 OFDM symbols). In line 4, we find $m+2$ peaks from $y'_i[n]$ using autocorrelation $|R_{y'_i y'_i}(n, N_{\text{FFT}})|$, where m denotes the number of OFDM symbols in 0.5 ms given μ. The difference between the remainders of two peaks separated by two OFDM symbols divided by OFDM symbol duration, Δp_j, is computed as the average of $p_{i(j+1)} - p_{i(j-1)}$ over i. We expect that Δp_j is the largest when p_j corresponds to long CP. For a more reliable estimation of a long CP,

we add one additional criterion. In line 8, we choose the two candidates k_0 and k_1 that give Δp_{k_i} the two largest values. We select k_q where the samples $\{p_{0k_q}, \cdots p_{5k_q}\}$ has the larger variance between two candidates of k_q. This is because we expect that $\{p_{0j}, \cdots, p_{5j}\}$ has the largest variance if p_{ij} corresponds to long CP since long CP makes $|R_{y'_i y'_i}(n, N_{\text{FFT}})|$ a plateau with certain width, not one sharp peak caused by non-long CP. Using estimated firstIndexLongCP, we put an additional 16 samples delay at the OFDM symbol with long CP while extracting the feature $Y_f^i[k]$ to maintain uniform Δn. The number of 16 samples comes from the difference between long CP and non-long CP with 30.72 MHz sampling rate.

In contrast to Wi-Fi 6 signals, some subcarriers might not be used for transmission in the midst of transmission. If no transmission is made in $Y^i[k]$ or $Y^{i+1}[k]$, their phases are random, and $\Delta \angle Y_{\Delta n}^i[k]$ cannot be represented as the phase difference. Therefore, we set the threshold for the amplitude, denoted as β, to check whether the PE is being used for transmission. Only when the amplitudes of both subcarrier symbols in adjacent OFDM symbols are higher than the threshold, this feature is used.

The discrepancy between the center frequency of TX and that of received IQ samples of 5G signals might be much larger than for Wi-Fi. This is because payload in Wi-Fi covers the entire channel bandwidth unless OFDMA is used. In contrast, PDSCH in 5G might use only the part of channel bandwidth so the center frequency of PDSCH might be different from that used for transmission. Thus, the discrepancy is solely from hardware imperfection in Wi-Fi. For a Wi-Fi link operating at $f_c = 5\,\text{GHz}$ and a frequency tolerance of 1 ppm for commercial-off-the-shelf temperature-compensated crystal oscillators [23] on both sides of the Wi-Fi link, the worst-case CFO is $\Delta f_c = 2 f_c \cdot 10^{-6} = 10$ kHz. However, in 5G, CFO can escalate to an MHz scale if we consider the center frequency of transmission bandwidth to be carrier frequency. If the method presented earlier in this subsection is employed, the difference could result in an inaccurate estimation of CFO due to aliasing. Even in absence of noise, it is only possible to measure Δf_c accurately up to $\Delta f_{\text{SCS}}/2$,

since $\Delta f_c + j \cdot \Delta f_{\text{SCS}}$ cannot be distinguished from each other, where $j \in \mathbb{Z}$. The provided algorithm makes the corrected CFO a multiple of Δf_{SCS}, not a zero.

However, the CFO correction algorithm is still deployed for feature extraction. This is because even though this method cannot find the exact CFO, it can recover the orthogonality among subcarriers. The CFO effect in our feature can be represented as:

$$\begin{aligned} Y^i_{\Delta n}[k] &= \sum_{n=0}^{N_{\text{FFT}}-1} y[n-\Delta n] \mathrm{e}^{-j2\pi n(\Delta f_c/f_{\text{TX}} + k/N_{\text{FFT}})} \\ &= Y^i\left[k + N_{\text{FFT}}\Delta f_c/f_{\text{TX}}\right] \mathrm{e}^{-j2\pi\Delta n(k/N_{\text{FFT}} + \Delta f_c/f_{\text{TX}})} \\ Y^{i+1}_{\Delta n}[k] &= Y^{i+1}\left[k + N_{\text{FFT}}\Delta f_c/f_{\text{TX}}\right] \times \\ &\quad \mathrm{e}^{-j2\pi(\Delta n k/N_{\text{FFT}} + (\Delta n + (N_{\text{FFT}}+N_{\text{CP}})\Delta f_c/f_{\text{TX}})} \\ \Rightarrow \Delta\angle Y^i_{\Delta n}[k] &= \angle Y^{i+1}\left[k + \Delta f_c/\Delta f_{\text{SCS}}\right] - \angle Y^i\left[k + \Delta f_c/\Delta f_{\text{SCS}}\right] \\ &\quad - 2\pi\Delta f_c(1/\Delta f_{\text{SCS}} + T_{\text{CP}}). \end{aligned} \tag{3.10}$$

To maintain orthogonality of $\angle Y^i_{\Delta n}[k]$ across k, $\Delta f_c/\Delta f_{\text{SCS}}$ should be an integer. We have demonstrated that after the CFO correction using CP, the CFO value can be expressed as $j \cdot \Delta f_{\text{SCS}}$, which renders $\Delta f_c/\Delta f_{\text{SCS}}$ to be an integer. Consequently, the phase of our feature becomes the sum of a phase difference of originally transmitted symbols and a phase caused by CFO. Since $\Delta\angle Y^i_{\Delta n}[k]$ in (3.10) contains T_{CP} term, the CFO effect on $\Delta\angle Y^i_{\Delta n}[k]$ is different when OFDM symbol $i-1$ is an OFDM symbol with long CP. To make the CFO effect uniform in the feature, $\Delta\angle Y^i_{\Delta n}[k]$ where OFDM symbol $i-1$ is not used for building the feature.

The features may contain the effect of other PHY channels which use modulations other than those used by PDSCH. It is impossible to perfectly filter out the effect because information about which REs were used for which PHY channels is not accessible. However, since the modulations of other PHY channels are either BPSK or QPSK, the constellation diagram of the features is only affected by PDSCH modulation. Thus, the

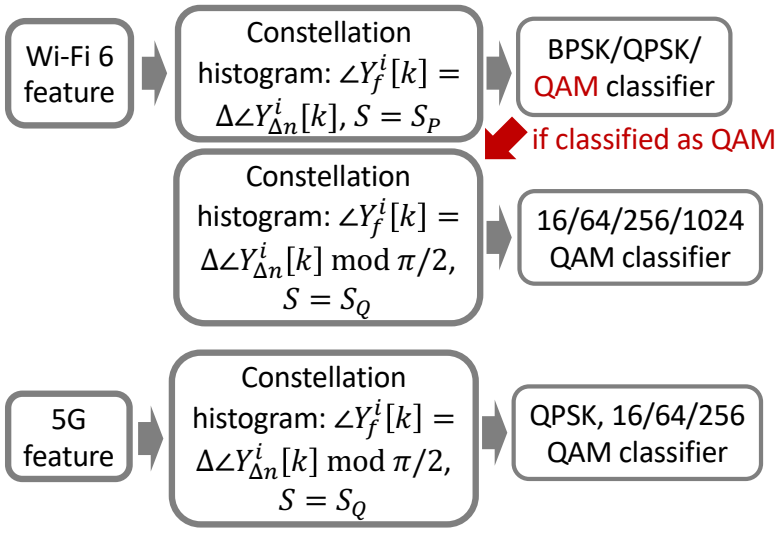


Figure 3.6. Flow chart of proposed classifier system.

Table 3.5. DL model parameters

Batch size	32
Learning rate	$5 \cdot 10^{-5}$
Epochs	200
Loss	Cross-entropy

distribution of phase differences is still an intrinsic characteristic of PDSCH modulation.

3.3.3 Input of neural network classifier

The obtained feature $Y_f^i[k]$ goes through two preprocessing steps to become input to the classifier. 1) instead of $\Delta\angle Y_{\Delta n}^i[k]$, $\Delta\angle Y_{\Delta n}^i[k]$ modulo $\pi/2$ is used as a phase of $Y_f^i[k]$. A constellation diagram of every target modulation and corresponding features $Y_f^i[k]$ without noise are symmetric with $\pi/2$. Thus, $\Delta\angle Y_{\Delta n}^i[k]$ modulo $\pi/2$ is used as a phase of our feature to characterize a modulation. For Wi-Fi 6 signals, BPSK cannot be distinguished from QPSK if $\Delta\angle Y_{\Delta n}^i[k]$ modulo $\pi/2$ is used. Thus, an additional classifier with the original phase as an input is used to distinguish BPSK and QPSK from the high-order QAM modulations, see Fig. 3.6. 2) A 2D histogram of the normalized amplitude of the features $|Y_f^i[k]|/|Y_f^i[k]|_{\text{p99}}$, where $|Y_f^i[k]|_{\text{p99}}$ denotes 99% percentile of $|Y_f^i[k]|$ in a single

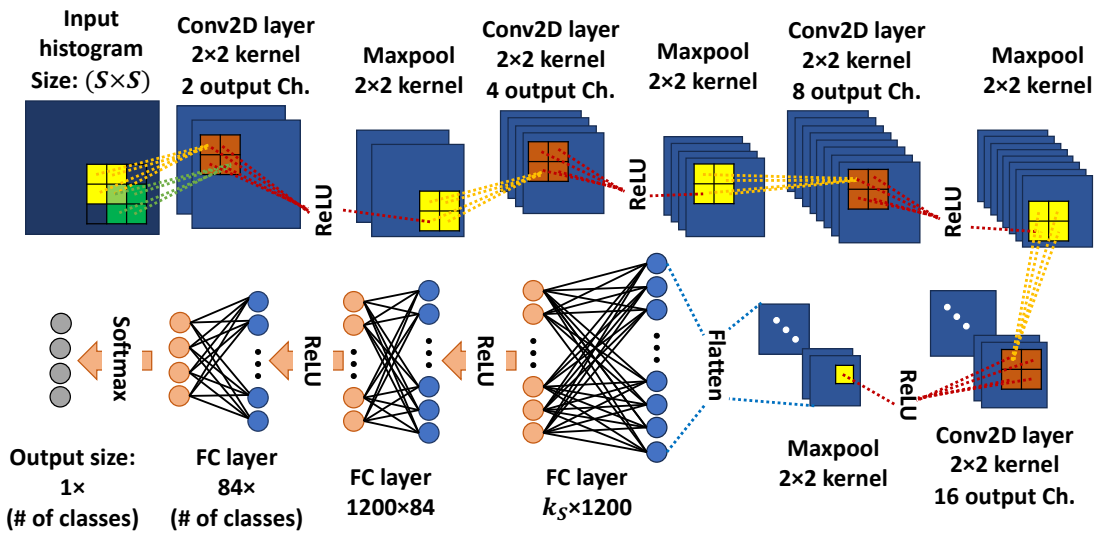


Figure 3.7. CNN-based modulation classfier structure.

data, and the phases $\angle Y_f^i[k]/2\pi$, as an input for the classifier. Each bin in the histogram is normalized by dividing it by the number of features in a single data. To remove outliers, $Y_f^i[k]$ whose amplitude is larger than $|Y_f^i[k]|_{\text{p99}}$ was not included in the histogram.

The overall structure and the parameter of the classifier with the histogram as input are summarized in Fig. 3.6 and Table 3.5. The neural network structure used for each classifier is described in Fig. 3.7. To identify BPSK and QPSK, $\mathscr{S} = \mathscr{S}_P$, and the third Maxpool layer is not used.

Figure 3.8 shows a scatterplot of the IQ data of $Y_f^i[k]$ and Fig. 3.9 the corresponding 2D histogram for 5G 16QAM data with $\Delta\angle Y_{\Delta n}^i[k]$ modulo $\pi/2$. $\angle Y_f^i[k]$ on the red and black dashed lines are the noise-free phase differences between two 16QAM symbols. Blue dashed lines are from the phase differences between BPSK or QPSK symbols of the PHY channel other than PDSCH. The noise-free phase difference values are (odd integer) $\cdot\, \pi/4$ and shifted further by CFO. The red, blue, and black dashed lines in Fig. 3.8 correspond to the red, blue, and black dashed lines in Fig. 3.9, respectively. Fig. 3.8 and Fig. 3.9 show that symbols are densely located at the points in the dashed lines, which is consistent with our expectations.

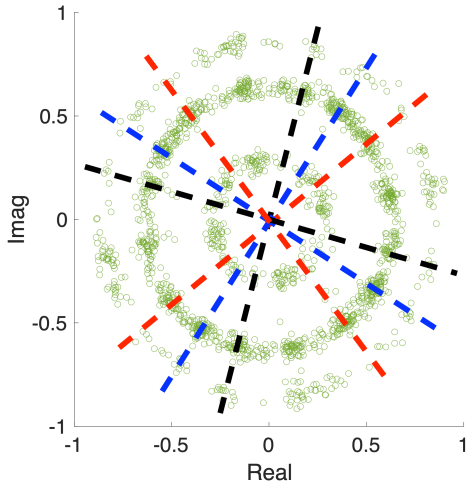


Figure 3.8. Scatterplot of $Y_f^s[k]$ of measured 16QAM features at SNR$=25$ dB with 5G OTA data.

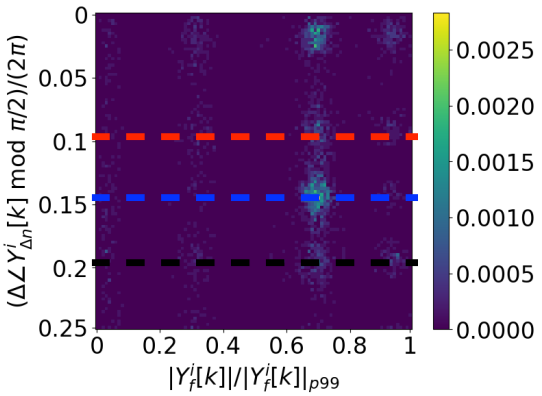


Figure 3.9. histogram of $|Y_f^i[k]|/|Y_f^i[k]|_{\text{p99}}$ and $(\Delta\angle Y^i_{\Delta n}[k] \bmod \pi/2)/2\pi$.

An advantage of using a histogram is that they are invariant to the length of $Y_f^i[k]$. This enables a neural network with a fixed structure to handle signals of any duration. This property is useful when dealing with 5G features where the number of samples of $Y_f^i[k]$ is unknown due to unused resources. Moreover, in a histogram input, the effect of CFO estimation error caused by aliasing (3.10) is a movement along y-axis of the histogram as far as orthogonality of $\angle Y^i_{\Delta n}[k]$ across k holds. The neural network can be trained to identify histogram movements along y-axis as a single class.

Table 3.6. Data generation parameters

SNR	[5, 40] dB in steps of 5 dB
Carrier frequency	2.4 GHz (Wi-Fi 6), 2.6 GHz (5G)
The number of {train, test} data	{800, 200} per each ($T_{\text{IFFT}}, T_{\text{CP}}$, modulation) case
$\{\mathscr{S}_P, \mathscr{S}_Q\}$	{15, 50}
Time duration of each data	$400\,\mu$s (Wi-Fi 6), 3.5 ms (5G)

3.4 Evaluation

3.4.1 Evaluation environments

Data collection

The proposed classifier is evaluated with synthetic data generated from AWGN channel simulations with the details in Table 4.6. MATLAB R2023a WLAN and 5G toolbox [24] are deployed to generate the synthetic AWGN dataset. Wi-Fi HT [25] and HE format [4] are used to generate data with $T_{\text{IFFT}} = 3.2\,\mu$s and $12.8\,\mu$s in Wi-Fi 6. For 5G data, every SCS option in FR1, $\mu \in \{0,1,2\}$, is tested. All PHY channels listed in Table 3.4 are included in every 5G data item.

To evaluate whether the performance of the proposed system remains invariant across varying 5G PHY channel configurations, the parameters for allocating REs to PHY channels are set for each data type. For example in PDCCH, symbol duration, aggregation level, and starting symbol number are randomly selected. PHY broadcast channel (PBCH), primary synchronization signal (PSS), and secondary synchronization signal (SSS) are included only when $\mu \in \{0,1\}$ since they are not available for $\mu = 2$. The other 5G PHY channel parameters are from FR1 test models in [26, 27].

Building classifier input

First, to avoid using the Wi-Fi preamble, we remove the first 2000 samples from each data. If the estimated T_{IFFT} corresponds to those of Wi-Fi 6, an IQ sequence whose

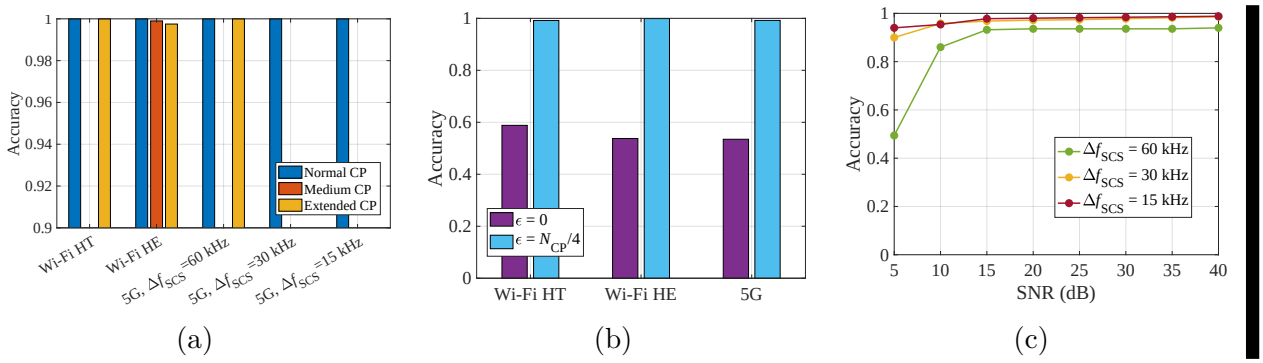
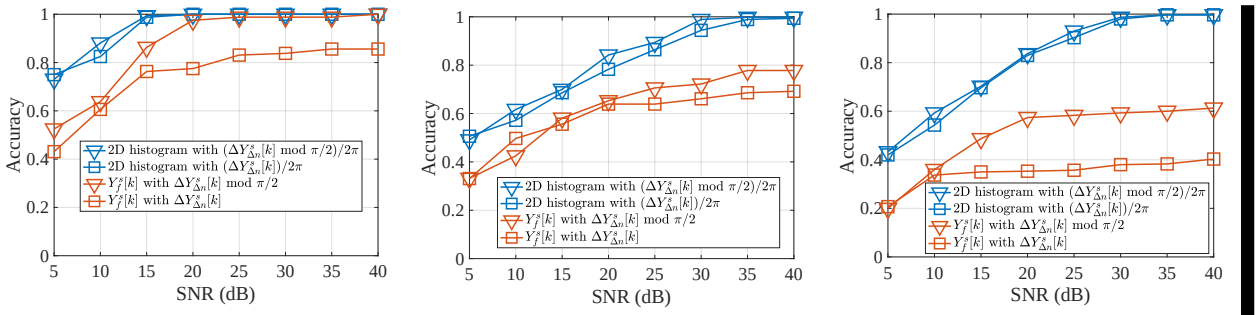


Figure 3.10. OFDM parameter estimation results: (a) Accuracy for estimating T_{IFFT} and T_{CP}, (b) Accuracy for choosing the first index of CP with acceptable error ε, and (c) Accuracy for finding an OFDM symbol with long CP of 5G signals.

length corresponds to 40+2 or 10+2 OFDM symbols is deployed to build the feature, $Y_f^i[k]$, starting with a random sample. We need an additional OFDM symbol due to the unknown starting index of an OFDM symbol sequence, $p \in [0, N_{\text{FFT}} - 1]$. Furthermore, one more extra OFDM symbol is required to evaluate phase differences between those of the last OFDM symbol and the next one. N_{null} is set to 8 and 32 for Wi-Fi HT and HE, respectively. If the estimated T_{IFFT} refers to 5G, $y'[n]$ of length (3 ms + 3 OFDM symbols) is used to estimate p and firstIndexLongCP. For 5G, the sequence of 14 OFDM symbols is utilized. We also evaluate the case using $Y_f^i[k]$ values as an input to assess how much the histogram input contributes to the performance. In this case, one data input consists of 2240 samples for Wi-Fi 6 or 7900 samples for 5G. The average number of feature elements in a single piece of 5G histogram data is 7858. We use fixed-duration data for a fair comparison, but the classifier can take the variable length data as input as the obtained feature is processed to a histogram using the algorithms in Sec. 3.3.3. For both cases of input formats, an input with both phases of $\angle Y_{\Delta n}^i[k]$ modulo $\pi/2$ and $\angle Y_{\Delta n}^i[k]$ are evaluated.



(a) (b) (c)

Figure 3.11. Classification accuracy for modulations vs. SNR: (a) Wi-Fi HT, (b) Wi-Fi HE, (c) 5G.

3.4.2 Evaluation results

Results in Fig. 3.10 are obtained with synthetic AWGN channel data. Fig. 3.10a shows estimation accuracy of the OFDM parameters $\{T_{\text{CP}}, T_{\text{IFFT}}\}$. Normal CP and Extended CP in the legend refer to the shortest and longest option for T_{CP}, respectively, given Δf_{SCS}. Medium CP of Wi-Fi HE refers to $1.6\,\mu$s T_{CP}. In every case, accuracy is over 99%. In Fig. 3.10b, the estimation accuracy of correctly finding the starting index of an OFDM symbol is shown for the method in Sec. 3.3.2. Correctly finding means that the starting index time is within ε tolerance of the true time. In Fig. 3.10b, we note that the estimation accuracy for identifying the starting index of an OFDM symbol falls below 60% for both Wi-Fi 6 formats and 5G. When the tolerance is relaxed to $N_{\text{CP}}/4$ time samples, the reported estimation accuracy increases to 99%.

The accuracy of estimating an OFDM symbol with long CP is shown in Fig. 3.10c. Aside from $\Delta f_{\text{SCS}} = 60$ kHz, the performance is over 90% even at low SNR of 5 dB. Accuracy at $\Delta f_{\text{SCS}} = 60$ kHz is low because the duration of an OFDM symbol with long CP is larger than the others. The large number of symbols that the peak detection function needs to detect also negatively affects the peak detection performance. At $\Delta f_{\text{SCS}} = 60$ kHz, there are 30 peaks that should be identified in line 4 of Algorithm 1, which is considerably larger than the 9 or 16 peaks at $\Delta f_{\text{SCS}} = 15$ kHz and $\Delta f_{\text{SCS}} = 30$ kHz.

Table 3.7. SNR required for data communication with each modulation

Modulation	BPSK	QPSK	16QAM
SNR for Wi-Fi 6 (dB)	5	10	16
SNR for 5G (dB)	-	15	18
Modulation	64QAM	256QAM	1024QAM
SNR for Wi-Fi 6 (dB)	22	30	35
SNR for 5G (dB)	21	27	-

Figure 3.11 shows modulation classification accuracy with AWGN channel data. The proposed algorithm with a histogram input with the phases $\Delta\angle Y^i_{\Delta n}[k]$ modulo $\pi/2$ outperforms in all considered cases, except for Wi-Fi 6 at 5 dB SNR. The performance gap between using the histogram as classifier input as opposed to using the feature value input increases in Wi-Fi HE and even more so in 5G. This is because the histogram input helps the classifier to discriminate the detailed symbol constellation of high-order modulations.

In Table 3.8, the accuracy of each modulation format is shown when the SNR satisfies the minimum requirement for standard-compliant data communication. We deploy error vector magnitude (EVM) levels required for data communication with each modulation for Wi-Fi 6 and 5G documentations [4, 27]. Required SNR values are calculated using the relation between EVM and SNR presented in [28]. SNR values required for the smallest coding rate are chosen for each modulation and chosen values are arranged in Table 3.7. For every modulation with both Wi-Fi 6 formats and 5G, accuracy is at least 98%.

3.5 Conclusion

Modulation classification of Wi-Fi 6 and 5G signals for spectrum sensing is studied. Our system deploys CAF to estimate SCS and CP length and achieve 99% accuracy. Without control information, our proposed preprocessing algorithm extracts features characterizing modulation schemes insensitive to synchronization errors. The preprocessing

Table 3.8. Accuracy when SNR is over the minimum requirements for standard-compliant data communication

Modulation	BPSK	QPSK	16QAM
Wi-Fi HT	100%	100%	100%
Wi-Fi HE	100%	100%	100%
5G	-	99%	100%
Modulation	64QAM	256QAM	1024QAM
Wi-Fi HT	100%	-	-
Wi-Fi HE	100%	100%	98%
5G	100%	100%	-

stage also estimates the CP position and the symbol with long CP of 5G signals. The form of the features is converted to be more suitable as inputs for the CNN-based classifier, which contributes to performance improvement in identifying high-order modulation. With data under various protocol configurations, our system identifies modulations of OFDM signals with 98% classification accuracy when SNR is higher than the value required for data transmission given a modulation. We are planning to extend this study to multiple-input multiple-output (MIMO) and orthogonal frequency division multiple access (OFDMA) scenario, to make more general transmission cases covered.

3.6 Acknowledgements

Chapter 3, in full, is a reprint of the material as it appears in Kim, B., Mecklenbräuker, C., and Gerstoft, P. “Blind Modulation Classification of Wi-Fi 6 and 5G signals for Spectrum Sensing”, in Proc. ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems 2023. The dissertation author was the primary investigator and author of this paper. The coauthors listed in this publication directed and supervised the research.

3.7 References

[1] J. Gao, X. Yi, C. Zhong, X. Chen, and Z. Zhang, "Deep learning for spectrum sensing," IEEE Wireless Commun. Letters, vol. 8, no. 6, pp. 1727–1730, 2019.

[2] L. Yu, J. Chen, and G. Ding, "Spectrum prediction via long short term memory," in Proc. IEEE ICCC, pp. 643–647, 2017.

[3] D. Liu, K. Ergun, and T. S. Rosing, "Towards a robust and efficient classifier for real world radio signal modulation classification," in Proc. IEEE ICASSP, pp. 1–5, 2023.

[4] IEEE 802.11ax, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 1: enhancements for high-efficiency WLAN," May 2021.

[5] 3GPP TR 38.331, "NR; Radio Resource Control (RRC); Protocol specification," Mar. 2023. ver 17.4.0.

[6] S. Hong, Y. Zhang, Y. Wang, H. Gu, G. Gui, and H. Sari, "Deep learning-based signal modulation identification in OFDM systems," IEEE Access, vol. 7, pp. 114631–114638, Aug. 2019.

[7] D. H. Al-Nuaimi, N. A. M. Isa, M. F. Akbar, and I. S. Z. Abidin, "AMC2-pyramid: Intelligent pyramidal feature engineering and multi-distance decision making for automatic multi-carrier modulation classification," IEEE Access, vol. 9, pp. 137560–137583, Sept. 2021.

[8] Z. Zhang, H. Luo, C. Wang, C. Gan, and Y. Xiang, "Automatic modulation classification using cnn-lstm based dual-stream structure," IEEE Trans. Veh. Technol., vol. 69, pp. 13521–13531, Nov. 2020.

[9] R. Gupta, S. Kumar, and S. Majhi, "Blind modulation classification for asynchronous ofdm systems over unknown signal parameters and channel statistics," IEEE Trans. Veh. Technol., vol. 69, pp. 5281–5292, Mar. 2020.

[10] A. K. Pathy, A. Kumar, R. Gupta, S. Kumar, and S. Majhi, "Design and implementation of blind modulation classification for asynchronous mimo-ofdm system," IEEE Trans. Instrum. Meas., vol. 70, pp. 1–11, Sept. 2021.

[11] S. Hong, Y. Wang, Y. Pan, H. Gu, M. Liu, J. Yang, and G. Gui, "Convolutional neural network aided signal modulation recognition in OFDM systems," in Proc. IEEE VTC, pp. 1–5, May 2020.

[12] M. C. Park and D. S. Han, "Deep learning-based automatic modulation classification with blind OFDM parameter estimation," IEEE Access, vol. 9, pp. 108305–108317, 2021.

[13] A. Kumar, K. K. Srinivas, and S. Majhi, "Automatic modulation classification for adaptive OFDM systems using convolutional neural networks with residual learning," IEEE Access, vol. 11, pp. 61013–61024, Jun. 2023.

[14] S. Kumar, E. Hamed, D. Katabi, and L. Erran Li, "Lte radio analytics made easy and accessible," Proc. ACM SIGCOMM, vol. 44, pp. 211–222, Aug. 2014.

[15] N. Bui and J. Widmer, "Owl: A reliable online watcher for lte control channel measurements," in Proc. 5th Workshop on All Things Cellular: Operations, Applications and Challenges, (New York (NY), USA), pp. 25–30, Association for Computing Machinery, 2016.

[16] R. Falkenberg and C. Wietfeld, "FALCON: An accurate real-time monitor for client-based mobile network data analytics," in Proc. IEEE GLOBECOM, (Waikoloa (HI), USA), pp. 1–7, IEEE, 2019.

[17] T. D. Hoang, C. Park, M. Son, T. Oh, S. Bae, J. Ahn, B. Oh, and Y. Kim, "Ltesniffer: An open-source lte downlink/uplink eavesdropper," in Proc. 16th ACM Conf. on Security and Privacy in Wireless and Mobile Networks (WiSec), (Guildford, UK), pp. 43–48, May 29–Jun. 1 2023.

[18] N. Ludant, P. Robyns, and G. Noubir, "From 5g sniffing to harvesting leakages of privacy-preserving messengers," in 2023 IEEE Symposium on Security and Privacy (SP), (Los Alamitos, CA, USA), pp. 1919–1934, IEEE Computer Society, may 2023.

[19] Y. Li, J. Barthelemy, S. Sun, P. Perez, and B. Moran, "A case study of wifi sniffing performance evaluation," IEEE Access, vol. 8, pp. 129224–129235, 2020.

[20] A. Punchihewa, V. K. Bhargava, and C. Despins, "Blind estimation of OFDM parameters in cognitive radio networks," IEEE Trans. Wireless Commun., vol. 10, pp. 733–738, Mar. 2011.

[21] S. S. Hong and S. R. Katti, "Dof: A local wireless information plane," in Proc. ACM SIGCOMM, pp. 230–241, 2011.

[22] W. A. Gardner and C. M. Spooner, "The cumulant theory of cyclostationary time-series. I. Foundation," IEEE Trans. Signal Process., vol. 42, pp. 3387–3408, Dec. 1994.

[23] G. E. Ltd, "GTXO-203T | 1.8V∼3.6V SM TCXO | Golledge," 2025.

[24] MathWorks, "MATLAB Products," 2023.

[25] IEEE 802.11n, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 5: enhancements for higher throughput," Oct. 2009.

[26] 3GPP TR 38.521-4, "NR; User Equipment (UE) conformance specification; Radio transmission and reception; Part 4: Performance requirements," 2023. ver 17.2.1.

[27] 3GPP TR 38.141-1, "NR; Radio Resource Control (RRC); Base Station (BS) conformance testing; Part 1: Conducted conformance testing," 2023. ver. 18.1.0.

[28] R. A. Shafik, M. S. Rahman, and A. H. M. R. Islam, "On the extended relationships among EVM, BER and SNR as performance metrics," in Proc. IEEE ICECE, pp. 408–411, 2006.

Chapter 4

Deep Learning-based Modulation Classification of Practical OFDM Signals for Spectrum Sensing

4.1 Introduction

The growth of wireless technologies in the scarce radio spectrum has strongly prioritized spectral efficiency: A challenge that is being addressed by, e.g., (massive) MIMO technology, joint radar communications, and cognitive radio [1, 2, 3]. Here, we focus on an essential component of cognitive radio, namely intelligent spectrum sensing, which allows for real-time characterization of radio spectrum usage and aids in online decision-making for spectrum allocation. Spectrum sensing encompasses signal detection [4], predicting available spectrum [5], and identifying modulation schemes. In this study, we focus on the classification of modulations of state-of-the-art wireless orthogonal frequency division multiplexing (OFDM) signals.

OFDM transmission has become foundational in current wireless communication systems, such as Wi-Fi 6 and 5G. In these systems, message bits are first encoded and subsequently mapped to digital symbols using quadrature amplitude modulation (QAM) on individual subcarriers. Many QAM symbols are modulated onto many subcarriers, so each time sample contains only a small fraction of the information carried by an OFDM

symbol. As a result, the modulation classifiers designed for single-carrier signals [6, 7] are not directly applicable to OFDM signals. Therefore, an accurate modulation classifier for Wi-Fi 6 and 5G signals requires additional processing beyond using raw time-domain samples as inputs.

In contrast to a dedicated receiver (RX) as a node in a wireless network, a spectrum sensor must be able to handle OFDM signals with diverse subcarrier configurations without access to prior information about the transmission format. In Wi-Fi 6 and 5G systems, information about the user data transmission, including the modulation, is provided to the RX through a protocol-specific procedure. However, since a spectrum sensor does not have prior knowledge of the type of signals it detects, it cannot deploy the procedure to obtain user data transmission information. The parameters shaping OFDM signals, fast Fourier transform (FFT) size to generate inverse fast Fourier transform (IFFT) sequence, and cyclic prefix (CP) length, might be different even among OFDM signals with the same modulation scheme. The diverse parameter options complicate the Wi-Fi preamble structure and in the recent Wi-Fi 6 these become more diverse. This makes spectrum sensing harder only with Wi-Fi preamble to identify the modulation scheme, even though the preamble structure is known. Moreover, the carrier frequency configurations in 5G become increasingly diverse and data transmission might occupy only a part of channel bandwidth. As a result, estimation of these carrier frequency configurations is becoming increasingly difficult using transmission bandwidth and center frequency alone. Thus, a modulation classifier for spectrum sensing should estimate the modulation scheme using only the observed user data transmission without knowledge of the OFDM signal parameters including FFT size, CP length, and carrier frequency.

We propose and analyze a modulation classifier for Wi-Fi 6 [8] and 5G [9] for a spectrum sensing system. Without knowledge of the transmitter (TX) carrier frequency, Wi-Fi preamble, or 5G control information, the classifier exploits only the basic OFDM structure, IFFT sequence, and CP. This includes the estimation of OFDM parameters:

CP length and subcarrier spacing (SCS), which is directly related to the FFT size of the IFFT sequence. We focus on identifying modulation schemes used in the payload of Wi-Fi 6 signals and the physical downlink shared channel (PDSCH) of 5G signals. Signals studied in this paper are single-input single-output (SISO). For 5G, they are in the frequency range 1 (FR1), whose frequency band is below 7.125 GHz.

For the SCS and CP length estimation, the cyclic autocorrelation function (CAF) is deployed. Thrfule capability of CAF to detect intervals of repeated sequences and repetition periods enables the estimation of those parameters. We observe that symbol-level synchronization is not perfect if autocorrelation using CP is utilized only. Our preprocessing removes the effect of the synchronization error by using phase differences between phases of two adjacent OFDM symbols. The modulation classifier for Wi-Fi 6 and 5G signals should recognize high-order modulations such as 256QAM and 1024QAM since these state-of-the-art protocols include those schemes. We change the feature format to a histogram representing the distribution of the features so that the classifier can effectively capture high-order modulation characteristics.

Related work on modulation classification: Many papers address modulation classification for wireless communication signals [6, 7, 10, 11, 12, 13, 14, 15, 16, 17]. The works in [10, 11, 12, 13, 14, 15] study modulation classification of OFDM signals and achieve at least 78% accuracy at 20 dB SNR for an AWGN channel. It is assumed that the inputs start from the first sample of the OFDM symbol duration [10, 11, 12, 15], which requires detecting the timing of the Wi-Fi preamble or 5G synchronization signals. To apply this approach to a spectrum sensor, the sensor needs to follow protocol-specific procedures. Further, neither of these works is evaluated on real-world measured data.

Previous works on OFDM modulation classification without symbol-level synchronization [13, 14, 16, 17] and the algorithms [13, 14, 17] are evaluated with hardware-generated data. However, their algorithms [13, 14, 17] are not evaluated with high-order modulations such as 256QAM or 1024QAM, as used in Wi-Fi 6 and 5G. Moreover, since

their classifier structures [13, 14] are designed to recognize only a fixed set of modulations, the overall structure needs to be redesigned to identify a new modulation scheme. The work [16] proposes the system to estimate SCS of OFDM signals and modulation of single-carrier signals jointly. Nonetheless, it does not estimate the modulation of OFDM signals. The neural network-based modulation classifiers [6, 7] study how environmental change affects classification performance for only the single-carrier signals, not OFDM signals.

Related work on sniffing OFDM signals: One approach to modulation identification for spectrum sensing uses sniffing of control information which notifies the RX about modulation and coding formats. The work [18, 19, 20, 21] attempts to overhear Long Term Evolution (LTE) signals. LTEye [18] and OWL [19] decode PHY DL control channel (PDCCH) data for LTE network monitoring. LTESniffer [21] decodes sniffed both user and control data using the PDCCH decoder FALCON [20]. FALCON overcomes the limitation of LTEye and OWL, which require more than 97% decoding accuracy. In LTE, the starting symbol of the PDCCH is always the first symbol in a slot. This is different from 5G, where the PDCCH starting symbol can be any symbol in a slot and its information is notified by radio resource control (RRC) signaling. Accordingly, it is not straightforward to modify the LTE PDCCH sniffer for 5G. Eavesdropping PDCCH data of 5G signals [22] applies to 5G signals with diverse configurations. Still, it is vulnerable to configuration changes since it takes a few minutes to learn a new PDCCH configuration. The authors of [23] study sniffing Wi-Fi probe request packets, which is for mobile devices to broadcast the existence of themselves. They build a hardware model for a sniffer and test with real Wi-Fi probe request packets. However, the probe request packets are simpler in format than those for user data communication. Thus, it is not straightforward to deploy this approach to our setting.

To summarize, the main contributions of the paper are:

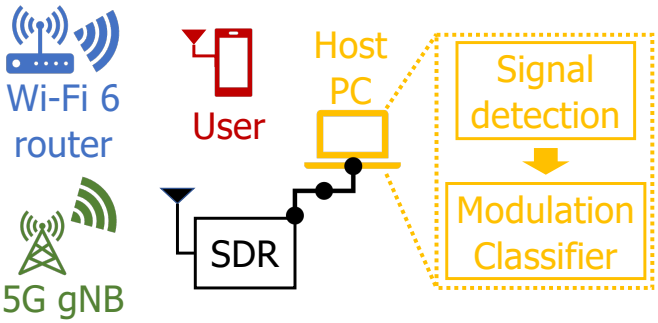


Figure 4.1. System capturing DL Wi-Fi 6 and 5G signals.

- OFDM parameter estimation for up-to-date protocols: We have applied the OFDM parameter estimation method with CAF [24] to Wi-Fi 6 and 5G signals to estimate SCS and CP length.

- Feature extraction without symbol-level synchronization: Only with estimated values of SCS and CP length, our system builds the features characterizing modulation of OFDM signals. The proposed feature extraction algorithm is designed to be resilient to symbol-level synchronization errors caused by using CP only.

- Modulation classification without control information: For spectrum sensing, control information might not be accessible. We show that the proposed classification system robustly works with diverse configurations with the evaluation of hardware-generated data without knowledge of the information.

4.2 System Objective

We aim to build a modulation classifier using IQ samples of SISO Wi-Fi 6 and FR1 5G DL signal for spectrum sensing. The system scenario is described in Fig. 4.1. There is a Wi-Fi 6 or 5G TX transmitting its signal to an RX. SDR continuously senses the spectrum by generating IQ samples with sampling rate f_{SDR} and transfers those samples to the host PC. In the host PC, there is a signal detection algorithm and a modulation classifier. Using IQ samples generated from SDR, the signal detection algorithm detects

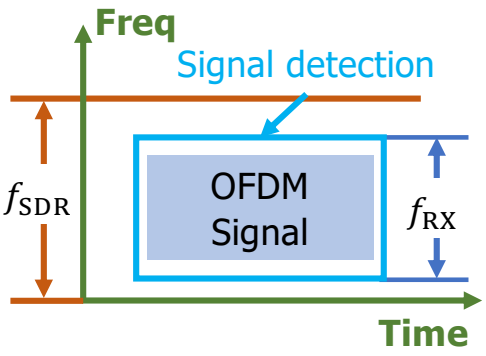


Figure 4.2. Spectrum sensing scenario using a software defined radio.

the duration and frequency band where the OFDM signal is located and extracts IQ samples corresponding to the detected OFDM signal, described as the blue rectangle in Fig. 4.2. We assume the accurate signal detection of Wi-Fi 6 or 5G signals and a single modulation scheme is used for data communication in one detected OFDM signal.

The IQ samples from SDR sampled with rate f_{SDR} are resampled to f_{RX}, 20 MHz. We only consider Wi-Fi 6 signals with 20 MHz channel bandwidth and 5G signals with a PDSCH bandwidth from 15 to 20 MHz. Thus, a 20 MHz sampling rate can let the resampled IQ sequence encompass the OFDM signal in our scenario. Extending the analysis to different transmission bandwidth ranges is straightforward. These resampled IQ samples, denoted by $y[n]$, are taken as inputs of the feature extraction algorithm, as elaborated in Sec. 4.3 in detail.

4.2.1 Wi-Fi 6 PHY Layer

Wi-Fi 6 supports the high-efficiency (HE) transmission format as well as earlier formats, which are non-high throughput (non-HT), high throughput (HT), and very high throughput (VHT) formats. Table 4.2 summarizes the parameters that configure the payload of the Wi-Fi frame for each Wi-Fi format. In HE format, given channel bandwidth, the number of subcarriers is increased because the SCS (denoted as Δf_{SCS}) is one-fourth of that of the previous transmission formats. Over time, the Wi-Fi standard has evolved

Table 4.1. Variable definitions

Variable	Definition (unit)
f_{TX}	TX sampling rate (Hz)
f_{RX}	Sampling rate of a system input sequence (Hz)
Δf_{SCS}	Subcarrier spacing (Hz)
T_{IFFT}	IFFT sequence duration (s)
N_{FFT}	FFT size used to generate IFFT sequence
T_{CP}	CP duration (s)
N_{CP}	Number of time samples in CP for one OFDM symbol
$y[n]$	Received time-domain sequence after resampling to 20 MHz
$y'[n]$	5G time-domain sequence after resampling to 30.72 MHz
$y^s[n]$	Received time-domain IFFT sequence for the sth OFDM symbol
$Y^s[k]$	Received symbol in subcarrier k for the sth OFDM symbol
$(\mathscr{S} \times \mathscr{S})$	Number of bins in a 2D histogram

Table 4.2. Parameters for different formats of Wi-Fi

	Non-HT format	HT format
T_{IFFT}	$3.2\,\mu$s	$3.2\,\mu$s
T_{CP}	$0.8\,\mu$s	$\{0.4, 0.8\}\,\mu$s
Modulations	BPSK, QPSK, 16QAM, 64QAM	BPSK, QPSK, 16QAM, 64QAM
	VHT format	HE format
T_{IFFT}	$3.2\,\mu$s	$12.8\,\mu$s
T_{CP}	$\{0.4, 0.8\}\,\mu$s	$\{0.8, 1.6, 3.2\}\,\mu$s
Modulations	BPSK, QPSK, 16QAM, 64QAM, 256QAM	BPSK, QPSK, 16QAM, 64QAM, 256QAM, 1024QAM

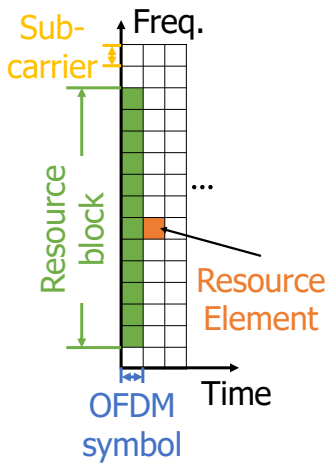


Figure 4.3. 5G resource grid structure.

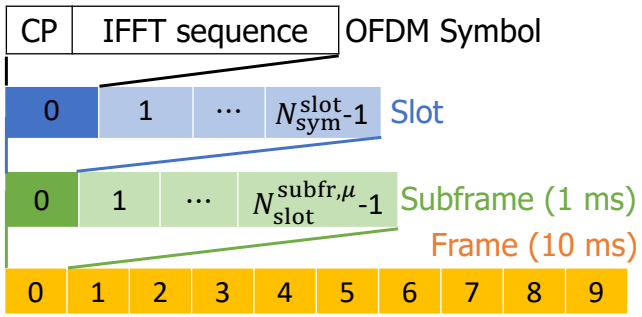


Figure 4.4. 5G frame structure.

and several options for the CP duration are available.

4.2.2 5G DL PHY Layer

The 5G downlink (DL) resource structure and its associated terminology is illustrated in Fig. 4.3 and Fig. 4.4. A resource element (RE), illustrated in Fig. 4.3, represents the smallest unit that carries data, encompassing a single OFDM symbol in the time domain and a single subcarrier in the frequency domain. A resource block (RB) is the smallest radio resource that can be allocated and refers to one OFDM symbol in the time domain and 12 subcarriers in the frequency domain.

Fig. 4.4 shows the 5G frame structure in the time domain. An OFDM symbol in 5G is comprised of both a CP and an IFFT sequence. The number of symbols within

Table 4.3. 5G frame structure parameters

{SCS (kHz), CP option}	{60, Normal}	{60, Extended}
T_{IFFT}	$16.17\,\mu$s	$16.67\,\mu$s
{Short, long} T_{CP}	$\{1.17, 1.69\}\,\mu$s	$\{4.17, -\}\,\mu$s
N_{FFT} when $f_{\text{TX}} = 30.72$ MHz	512	512
{Short, long} N_{CP} when $f_{\text{TX}} = 30.72$ MHz	{36, 52}	128
{SCS (kHz), CP option}	{30, Normal}	{15, Normal}
T_{IFFT}	$33.33\,\mu$s	$66.67\,\mu$s
{Short, long} T_{CP}	$\{2.34, 2.86\}\,\mu$s	$\{4.69, 5.21\}\,\mu$s
N_{FFT} when $f_{\text{TX}} = 30.72$ MHz	1024	2048
{Short, long} N_{CP} when $f_{\text{TX}} = 30.72$ MHz	{72, 88}	{144, 160}

a single slot ($N_{\text{sym}}^{\text{slot}}$) varies following the CP length. There are normal and extended CP options in the transmission format. When a normal CP is used then $N_{\text{sym}}^{\text{slot}} = 14$, otherwise $N_{\text{sym}}^{\text{slot}} = 12$. The SCS, the distance between two adjacent subcarriers in OFDM systems, determines the number of slots within a single subframe, $N_{\text{slot}}^{\text{subfr},\mu}$. μ represents an SCS option and corresponds to $\Delta f_{\text{SCS}} = 15 \times 2^{\mu}$ kHz. There are five SCS options in 5G, but we consider only three cases, namely 15, 30, and 60 kHz, which are available in FR1. The number of slots in a subframe for each SCS is computed as $N_{\text{slot}}^{\text{subfr},\mu} = 2^{\mu}$. Finally, one frame of duration 10 ms consists of ten subframes.

The structural parameters that define the 5G frame are listed in Table 4.3. The length of an IFFT sequence, T_{IFFT}, is:

$$T_{\text{IFFT}} = N_{\text{FFT}}/f_{\text{TX}} = 1/\Delta f_{\text{SCS}}. \tag{4.1}$$

There is a one-to-one correspondence between T_{IFFT} and Δf_{SCS} (4.1). Under the normal CP option, CP is longer than that in other symbols, every 0.5 ms, or equivalently, $7 \cdot 2^{\mu}$ OFDM symbols in OFDM symbol unit, called long CP. There is no long CP in the extended CP option, so T_{CP} is uniform. The transmission rate of 5G signals is a power of 2 times 15 kHz and 30.72 MHz is an example of a 5G transmission rate. N_{FFT} and N_{CP}

Table 4.4. Modulations used for 5G physical channels

Physical channel	PDSCH	PSS/SSS	PDCCH	CSI-RS
Modulation	QPSK, 16QAM, 64QAM, 256QAM, 1024QAM	BPSK	QPSK	QPSK
Physical channel	PBCH	PDSCH-PTRS	PDSCH-DMRS	
Modulation	QPSK	QPSK	QPSK	

values are arranged when f_{TX} is 30.72 MHz, the value used in our evaluation.

In addition to PDSCH, there exist other physical (PHY) channels that serve specific functions although not carrying user data. For instance, PDCCH conveys downlink control information (DCI), which contains information required to decode PDSCH data such as modulation and coding scheme (MCS). Each of these channels utilizes predefined single-type modulation, see Table 4.4.

Compared to Wi-Fi, which has a predefined configuration of data, pilot, and null subcarriers, 5G resource configuration for PHY channels is flexible. Instead, the 5G system has a network dedicated to exchanging information on how data packets are forwarded, called the control plane, in addition to the network for data transmission, called the user plane. An example of data transferred over the control plane is RRC signals. Information on the starting OFDM symbol of PDCCH and channel state information-reference signal (CSI-RS) is notified to an RX with RRC signals via control plane [9].

4.3 Proposed Algorithm

High-level procedures to build features characterizing the modulations of Wi-Fi 6 and 5G signals are illustrated in Fig. 4.5 and explained in Sec. 4.3.1 and 4.3.2 with additional processing for 5G signals in Sec. 4.3.3. The 2D histogram is then taken as an input to the neural network model, described in Sec. 4.3.4.

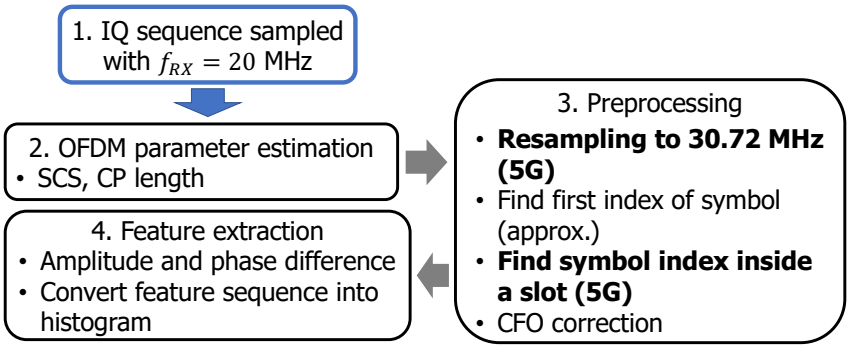


Figure 4.5. Flow chart of proposed modulation classification algorithm.

4.3.1 OFDM Parameter Estimation

Before building the features that characterize modulation, it is necessary to estimate two essential OFDM parameters of OFDM signals, SCS and CP length. To estimate these parameters, we use CAF, a Fourier-series coefficient of the autocorrelation function,

$$\mathscr{R}_{yy}(\alpha,\tau)=\sum_{n=-\infty}^{\infty}\mathscr{R}_{yy}(n,\tau)\mathrm{e}^{-j2\pi\alpha n}. \tag{4.2}$$

CAF is used to extract a repeated pattern presented in wireless signals [24, 25, 26]. A variant of the CAF estimator presented in [24] is deployed here,

$$\hat{\mathscr{R}}_{yy}(\alpha,\ell)=\frac{1}{\mathscr{L}-l-\ell+1}\sum_{n=0}^{\mathscr{L}-l-\ell}\left\{\sum_{i=0}^{l-1}y[n+i]y^{*}[n+i+\ell]\right\}\times\mathrm{e}^{-j2\pi\alpha n}, \tag{4.3}$$

where α is a cycle frequency and $\mathscr{L}$ is the length of $y[n]$. One sample of our estimator is computed as the autocorrelation with delay ℓ. It differs from the estimator in [24], which corresponds to $l=1$ in (4.3). The increase in the length of a sample sequence $y[n+i]$ aims to make peaks more distinct. We set $l=8$ corresponding to the shortest CP length in our scenario.

CP in OFDM symbols causes a sequence to be repeated at both ends of each symbol.

The distance between starting indices of the two repeated sequences at both ends of an OFDM symbol is T_{IFFT} in time units or $N_{\text{FFT}}(f_{\text{RX}}/f_{\text{TX}}) = f_{\text{RX}}/\Delta f_{\text{SCS}}$ in time sample units. This repetition makes the CAF estimator at $\alpha = 0$ have a peak at $\ell = f_{\text{RX}}/\Delta f_{\text{SCS}}$. T_{CP} is also estimated with the CAF estimator, $\hat{\mathscr{R}}_{yy}(\alpha, f_{\text{RX}}/\Delta f_{\text{SCS}})$. Since $\sum_{i=0}^{l-1} y[n+i]y^*[n+i+\ell]$ in (4.3) has peaks at period of $f_{\text{RX}}(T_{\text{CP}} + 1/\Delta f_{\text{SCS}})$, it is expected of $\hat{\mathscr{R}}_{yy}(\alpha, f_{\text{RX}}/\Delta f_{\text{SCS}})$ to have a large amplitude at $\alpha = 1/\{f_{\text{RX}}(T_{\text{CP}} + 1/\Delta f_{\text{SCS}})\}$.

In our scenario, there are five candidates ℓ values, $\ell_C = \{64, 256, 333, 667, 1333\}$, each corresponding to an IFFT sequence length for a given SCS at $f_{\text{RX}} = 20$ MHz. IFFT sequence length is estimated as:

$$T_{\text{IFFT}} = \ell'/f_{\text{RX}} \text{ s.t. } \ell' = \underset{\ell \in \ell_C}{\arg\max} \left|\hat{\mathscr{R}}_{yy}(0, \ell)\right| \tag{4.4}$$

When the estimated T_{IFFT} corresponds to that of Wi-Fi 6 or 60 kHz SCS NR, where multiple CP options are available, CP length is further estimated as:

$$T_{\text{CP}} = \frac{1}{f_{\text{RX}}}\left(\frac{1}{\alpha'} - \ell'\right) \text{ s.t. } \alpha' = \underset{\alpha \in \alpha_C^{\ell'}}{\arg\max} \left|\hat{\mathscr{R}}_{yy}(\alpha, \ell')\right| \tag{4.5}$$

where $\alpha_C^{\ell'}$ denotes a set of possible values of $\alpha = 1/\{\ell' + (f_{\text{RX}} \cdot T_{\text{CP}})\}$, given ℓ'.

4.3.2 Feature Extraction

The motivation behind our proposed feature extraction lies in the observation that when a sampled time-domain sequence is contained within a single OFDM symbol s, the FFT of that sequence yields the original symbols with a phase drift that scales linearly

with subcarrier index k and synchronization error Δn, as shown in

$$\begin{aligned} Y^s_{\Delta n}[k] &\triangleq \mathscr{F}\left(y^s[n-\Delta n]\right) \\ &= \sum_{n=0}^{N_{\text{FFT}}-1} y^s[n-\Delta n]\mathrm{e}^{-j2\pi nk/N_{\text{FFT}}} \\ &= Y^s[k]\mathrm{e}^{-j2\pi\Delta nk/N_{\text{FFT}}}. \end{aligned} \tag{4.6}$$

To build a feature characterizing modulation based on this property, two objectives must be achieved: first, sampling a sequence fully contained in an OFDM symbol, and second, removing the phase drift caused by synchronization errors.

Utilizing the knowledge of N_{CP} and N_{FFT}, the CP position is determined through autocorrelation analysis,

$$R_{yy}(m,N_{\text{FFT}}) = \frac{1}{N_{\text{CP}}}\sum_{i=0}^{N_{\text{CP}}-1} y[m+i]y^*[m+i+N_{\text{FFT}}], \tag{4.7}$$

where m is the first index of original sequence of autocorrelation $R_{yy}(m,N_{\text{FFT}})$. The position of CP is indicated by the peaks in $|R_{yy}(m,N_{\text{FFT}})|$ since it is expected that $|R_{yy}(m,N_{\text{FFT}})|$ peaks when m is the first index of CP. To locate a peak, we search for a sample whose amplitude is larger than both of its neighboring samples while ensuring that the minimum distance between two adjacent peaks is 90% of the OFDM symbol duration (i.e., $(256+64)\times 0.9 = 288$-time samples for HE format with 3.2 μs CP), to avoid selecting undesired local peaks. The indices of peaks are denoted as $\{p'_0,\cdots,p'_{S-1}\}$ for S potential OFDM symbols. Using those peaks, the first index of the OFDM symbol is estimated:

$$p = \text{Median}_i\{\text{mod}(p'_i, N_{\text{FFT}}+N_{\text{CP}})\}, \tag{4.8}$$

where $i \in \{0,\cdots j-1\}$. Noise and varying amplitudes of time samples can introduce small errors in the estimated CP position. To reliably sample the sequences contained in a

single OFDM symbol, we deploy the sequence $\{y[p+N_{\text{CP}}/2], y[p+N_{\text{CP}}/2+1], \cdots, y[p+N_{\text{CP}}/2+N-1]\}$. This sequence is entirely within a single OFDM symbol for estimation error of p below $N_{\text{CP}}/2$.

We demonstrated (4.6) that $Y^s_{\Delta n}[k]$ exhibits a phase drift, $\mathrm{e}^{-j2\pi\Delta nk/N}$, while maintaining amplitude $Y^s[k]$. We compute the phase differences between successive potential symbols s and $s+1$ in subcarrier k to build the feature invariant of this phase drift due to synchronization errors as:

$$\begin{aligned}
&\Delta\angle Y^s_{\Delta n}[k] \triangleq \angle Y^{s+1}_{\Delta n}[k] - \angle Y^s_{\Delta n}[k] \\
&= \angle\left\{Y^{s+1}[k]\mathrm{e}^{-j2\pi\Delta nk/N}\right\} - \angle\left\{Y^s[k]\mathrm{e}^{-j2\pi\Delta nk/N}\right\} \\
&= \angle Y^{s+1}[k] - \angle Y^s[k].
\end{aligned} \tag{4.9}$$

Despite Δn unknown, sequences with constant Δn are obtained by adjusting the interval between the starting indices of two sampled sequences to be one OFDM symbol. The feature used to identify the modulation type is

$$Y^s_f[k] \triangleq |Y^s_{\Delta n}[k]|\mathrm{e}^{j\Delta\angle Y^s_{\Delta n}[k]}. \tag{4.10}$$

For Wi-Fi 6, the null subcarrier symbols are eliminated by discarding symbols with the N_{null} smallest average amplitudes.

In protocol-compliant reception, the Wi-Fi preamble and 5G PDSCH-phase tracking reference signal (PDSCH-DMRS) are deployed for CFO estimation. However, since not accessible to a spectrum sensor, the CP in each OFDM symbol is used for CFO estimation Δf_c, i.e.,

$$\angle(y[p+N_{\text{FFT}}+i]\cdot y^*[p+i]) = 2\pi\Delta f_c/\Delta f_{\text{SCS}}, \tag{4.11}$$

where $y[p+i]$ is in CP. We use $i \in \{\lfloor N_{\text{CP}}/4\rfloor, \cdots, \lceil 3N_{\text{CP}}/4\rceil\}$ so that the sequence

$y[p+i]$ are entirely within CP unless estimation error of p exceeds $N_{\mathrm{CP}}/4$. We determine CFO as the average of Δf_c (4.11) evaluated over multiple OFDM symbols. If the absolute value of the CFO is larger than $\Delta f_{\mathrm{SCS}}/2$, the CFO cannot be accurately estimated due to aliasing. It is discussed in Sec. 4.3.3.

4.3.3 Additional Procedures for 5G Signal

To build a modulation feature for 5G, 5G characteristics distinct from those of Wi-Fi, including a different transmission rate, long CP, and flexible usage of subcarriers, should be considered. First, the transmission rate of 5G signals is not $f_{\mathrm{RX}} = 20\,\mathrm{MHz}$, but is a power of 2 times 15 kHz. Hence, for the signal classified as 5G, we resample the sequence to $f_{5\mathrm{G}} = 30.72\,\mathrm{MHz} = 2048 \cdot 15\,\mathrm{kHz}$, the smallest sampling frequency above 20 MHz. N_{FFT} and N_{CP} with 30.72 MHz sampling rate for each Δf_{SCS} are arranged in the last two rows in Table 4.3.

In the case of the normal CP option, there is a long CP every $T_{\mathrm{LCP}} = 0.5\,\mathrm{ms}$, which is slightly longer than that of other OFDM symbols. Long CP breaks the assumption of uniform OFDM symbol duration, which is required by the method to find the first indices of OFDM symbols and estimate CFO. Moreover, in building $Y_f^s[k]$, maintaining the fixed interval does not guarantee the constant Δn over multiple OFDM symbols. Therefore, long CP also should be located when finding the first index of the OFDM symbol.

Algorithm 2 explains the detailed steps to estimate the first index of OFDM symbol with long CP. $\mathbf{y_i'}$ in line 3 is a sequence cropped to be as long as (0.5 ms + 2 OFDM symbols + T_{CP}).

In line 4, we find $M+2$ peaks from $\mathbf{y_i'}$ using autocorrelation $|R_{\mathbf{y_i'}\mathbf{y_i'}}(m, N_{\mathrm{FFT}})|$, where M denotes the number of OFDM symbols in T_{LCP} given μ and we also compute the autocorrelation at the two symbols at each end. The M average differences between the remainders of two peaks separated by two OFDM symbols modulo OFDM symbol duration, Δp_j, are computed in line 7. We expect that Δp_j is the largest when p_j corresponds

Algorithm 2: Finding first index of long CP in 5G

Data: ($y'[n]$ of length (3 ms + 3 OFDM symbols)), μ

1 $M = 7 \cdot 2^{\mu}$, $N_{\text{FFT}} = 512 \cdot 2^{2-\mu}$, $N_{\text{CP}} = 18 \cdot 2^{2-\mu}$;

2 for $i = 0:5$ do

3 $\quad \mathbf{y'_i} \triangleq \{y'[f_{5\text{G}} T_{\text{LCP}} \cdot i],\ \cdots,\ y'[f_{5\text{G}} T_{\text{LCP}}(i+1) + 2(N_{\text{FFT}} + N_{\text{CP}}) + N_{\text{CP}} - 1]\}$;

4 $\quad$ Find peaks $\{p'_{i0}, \cdots, p'_{i(m+1)}\}$ with $\mathbf{y'_i}$ using $|R_{\mathbf{y'_i y'_i}}(m, N_{\text{FFT}})|$ and peak locating function in Sec. 4.3.2;

5 $\quad p_{ij} = \text{mod}(p'_{ij}, N_{\text{FFT}} + N_{\text{CP}})$;

6 end

7 $\Delta p_j = (\sum_{k=0}^{5} \{p_{k(j+1)} - p_{k(j-1)}\})/6$ where $j \in \{1, 2, \cdots, M\}$;

8 $\{\Delta p_{r_0}, \cdots, \Delta p_{r_{M-1}}\} = \text{sortDescending}(\{\Delta p_j\})$;

9 $\text{symLongCP} = \arg\max_{r_q} \text{Var}(\{p_{0r_q}, \cdots, p_{5r_q}\})$ where $q \in \{0, 1\}$;

10 $q_{ij} = \begin{cases} p_{ij} & \text{if } j \leq \text{symLongCP} \\ p_{ij} - 16 & \text{otherwise} \end{cases}$ $q = \text{Median}_j(\sum_{k=0}^{5} q_{kj}/6)$;

Result: IndexLongCP= q+symLongCP(N_{FFT}+N_{CP})

to long CP. For a more reliable estimation of a long CP, we add a criterion.

In line 10, we choose the two candidates k_0 and k_1 that give Δp_{k_i} the two largest values. We select k_q where the set $\{p_{0k_q}, \cdots p_{5k_q}\}$ has the larger variance between two candidates of k_q. This is because we expect that $\{p_{0j}, \cdots, p_{5j}\}$ has the largest variance if p_{ij} corresponds to long CP since long CP makes $|R_{\mathbf{y'_i y'_i}}(m, N_{\text{FFT}})|$ a plateau with some width. Using estimated IndexLongCP, we put an additional 16 samples delay at the OFDM symbol with long CP while extracting the feature $Y_f^s[k]$ to maintain uniform Δn. The number of 16 samples comes from the difference between long CP and non-long CP with a 30.72 MHz sampling rate.

In contrast to Wi-Fi 6 signals, some subcarriers might not be used for transmission amid transmission. If no transmission is made in $Y^s[k]$ or $Y^{s+1}[k]$, their phases are random, and $\Delta\angle Y_{\Delta n}^s[k]$ cannot be the phase difference between two constellation points. Therefore, we set the threshold for the amplitude, denoted as β, to check whether the RE is being used for transmission. Only when $|Y^s[k]|$ and $|Y^{s+1}[k]|$ are higher than β, $Y^s[k]$ is used.

The discrepancy between the center frequency of TX and that of received IQ

samples of 5G signals might be much larger than for Wi-Fi. In contrast to Wi-Fi, which covers the entire channel bandwidth unless OFDMA is used, PDSCH in 5G might use only the part of channel bandwidth so the center frequency of PDSCH might be different from that used for transmission. Thus, the discrepancy is solely from hardware imperfection in Wi-Fi. For a Wi-Fi link operating at $f_c = 5\,\text{GHz}$ and a frequency tolerance of 1 ppm for commercial-off-the-shelf temperature-compensated crystal oscillators [27] on both sides of the Wi-Fi link, the worst-case CFO is $\Delta f_c = 2f_c \cdot 10^{-6} = 10$ kHz. However, in 5G, the CFO can escalate to an MHz scale if we consider the center frequency of transmission bandwidth to be carrier frequency. If the method presented earlier in this section is employed, the difference could result in an inaccurate estimation of CFO due to aliasing. Even in the absence of noise, it is only possible to measure Δf_c accurately up to $\Delta f_{\text{SCS}}/2$, since $\Delta f_c + z\Delta f_{\text{SCS}}$ cannot be distinguished from each other, where $z \in \mathbb{Z}$. The algorithm makes the corrected CFO a multiple of Δf_{SCS}, not a zero.

However, the CFO correction is still deployed for feature extraction. This is because even though this method cannot find the exact CFO, it can recover the orthogonality among subcarriers. The CFO effect in our feature is represented as:

$$\begin{aligned} Y^s_{\Delta n}[k] &= \sum_{n=0}^{N_{\text{FFT}}-1} y[n-\Delta n]\mathrm{e}^{-j2\pi n(\Delta f_c/f_{\text{TX}}+k/N_{\text{FFT}})} \\ &= Y^s\left[k+N_{\text{FFT}}\Delta f_c/f_{\text{TX}}\right]\times \mathrm{e}^{-j2\pi\Delta n(k/N_{\text{FFT}}+\Delta f_c/f_{\text{TX}})} \\ Y^{s+1}_{\Delta n}[k] &= Y^{s+1}\left[k+N_{\text{FFT}}\Delta f_c/f_{\text{TX}}\right]\times \mathrm{e}^{-j2\pi(\Delta nk/N_{\text{FFT}}+(\Delta n+(N_{\text{FFT}}+N_{\text{CP}})\Delta f_c/f_{\text{TX}})} \end{aligned}$$

$$\Rightarrow \Delta\angle Y^s_{\Delta n}[k] = \angle Y^{s+1}\left[k+\Delta f_c/\Delta f_{\text{SCS}}\right] - \angle Y^s\left[k+\Delta f_c/\Delta f_{\text{SCS}}\right] - 2\pi\Delta f_c(1/\Delta f_{\text{SCS}}+T_{\text{CP}}). \tag{4.12}$$

To maintain orthogonality of $\angle Y^s_{\Delta n}[k]$ across k, $\Delta f_c/\Delta f_{\text{SCS}}$ should be an integer. We have demonstrated that after the CFO correction using CP, the CFO is expressed

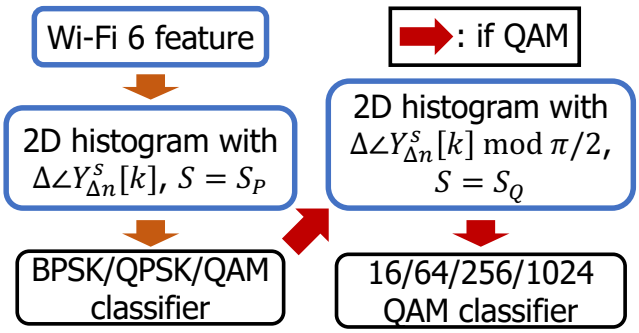


Figure 4.6. Flow chart of proposed Wi-Fi 6 classifier system.

Table 4.5. DL model parameters

Batch size	32	Learning rate	$5 \cdot 10^{-5}$
Epochs	200	Loss	Cross-entropy

as $z \cdot \Delta f_{\text{SCS}}$, which renders $\Delta f_c / \Delta f_{\text{SCS}}$ to be an integer. Consequently, the phase of our feature becomes the sum of a phase difference of originally transmitted symbols and a phase caused by the CFO. Since $\Delta\angle Y^s_{\Delta n}[k]$ in (4.12) contains T_{CP} term, the CFO effect on $\Delta\angle Y^s_{\Delta n}[k]$ is different when OFDM symbol $s+1$ is an OFDM symbol with long CP. To make the CFO effect uniform in the feature, $\Delta\angle Y^s_{\Delta n}[k]$ where OFDM symbol $s+1$ is an OFDM symbol with long CP is not used for building the feature.

The features may contain the effect of other PHY channels that use modulations other than those used by PDSCH. It is impossible to perfectly filter out the effect because information about which REs were used for which PHY channels is not accessible for spectrum sensors. However, since the modulations of other PHY channels are either BPSK or QPSK, the constellation diagram of the features is only affected by changes in PDSCH modulation. Thus, the distribution of phase differences is still an intrinsic characteristic of PDSCH modulation.

4.3.4 Neural Network Classifier

The obtained feature $Y^s_f[k]$ goes through two preprocessing steps to become input to the classifier:

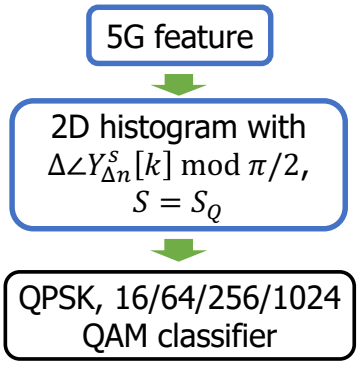


Figure 4.7. Flow chart of proposed 5G classifier system.

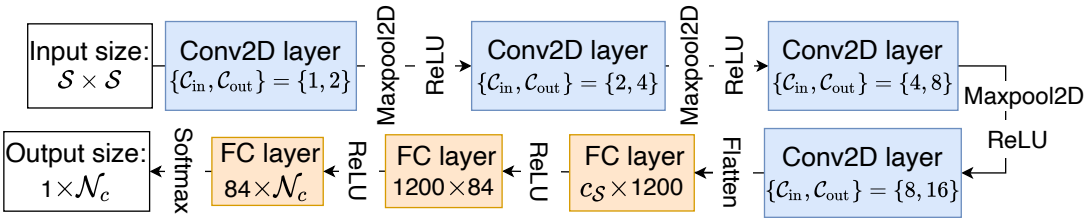


Figure 4.8. CNN-based modulation classifier structure. $\mathcal{N}_c$ is the number of modulations a classifier aims to recognize.

1) instead of $\Delta\angle Y^s_{\Delta n}[k]$, $\Delta\angle Y^s_{\Delta n}[k]$ modulo $\pi/2$ is used as a phase of $Y^s_f[k]$. A constellation diagram of every target modulation and corresponding features $Y^s_f[k]$ without noise are rotationally symmetric with $\pi/2$. Thus, $\Delta\angle Y^s_{\Delta n}[k]$ modulo $\pi/2$ is used as a phase of our feature to characterize a modulation. For Wi-Fi 6 signals, BPSK cannot be distinguished from QPSK if $\Delta\angle Y^s_{\Delta n}[k]$ modulo $\pi/2$ is used. Thus, an additional classifier with the original phase as an input is used to distinguish BPSK and QPSK from the high-order QAM modulations, see Fig. 4.6.

2) A 2D histogram of the normalized amplitude of the features $|Y^s_f[k]|/|Y^s_f[k]|_{\text{p99}}$, where $|Y^s_f[k]|_{\text{p99}}$ denotes 99th percentile of $|Y^s_f[k]|$ in a single data, and the phases $\angle Y^s_f[k]/2\pi$,

as an input for the classifier. The histogram value of each bin is computed as:

$$\begin{aligned} Z(u,v) = &\text{ The number of } Y_f^s[k] \text{ s.t.} \\ &u/\mathscr{S} \le |Y_f^s[k]|/|Y_f^s[k]|_{\text{p99}} \le (u+1)/\mathscr{S} \text{ and} \\ &v/\mathscr{S} \le \Delta\angle Y_{\Delta n}^s[k]/\phi \le (v+1)/\mathscr{S}. \end{aligned} \tag{4.13}$$

If $\Delta\angle Y_{\Delta n}^s[k]$ modulo $\pi/2$ is used, ϕ is $\pi/2$, otherwise 2π. We normalize histogram value to be classifier input:

$$Z'(u,v) = Z(u,v)/\mathscr{Z}, \tag{4.14}$$

where $\mathscr{Z}$ denotes the number of valid $Y_f^s[k]$ in one data. To remove outliers, $Y_f^s[k]$ whose amplitude is larger than $|Y_f^s[k]|_{\text{p99}}$ was not included in the histogram.

The overall structure and the parameter of the classifier with the histogram input are summarized in Fig. 4.6, Fig. 4.7, and Table 4.5. The neural network structure used for each classifier is described in Fig. 4.8. $\mathscr{C}_{\text{in}}$ and $\mathscr{C}_{\text{out}}$ in Conv2D layers correspond to the number of input and output depth. A 2×2 size kernel is used in every Conv2D and Maxpool2D layer. $\mathscr{N}_c$ is the number of modulations that a classifier aims to recognize. For the classifier to identify BPSK and QPSK, the third Maxpool layer is not used, $\mathscr{S} = \mathscr{S}_P$, and $\mathscr{N}_c = 3$. The classifier for 5G and for identifying the QAM types for Wi-Fi 6 use $\mathscr{N}_c = 5, 4$, respectively.

For 5G 16QAM real-world measured over-the-air (OTA) data, Fig. 4.9 shows a scatterplot of the IQ data of $Y_f^s[k]$ and Fig. 4.10 the corresponding 2D histogram with $\Delta\angle Y_{\Delta n}^s[k]$ modulo $\pi/2$. $\angle Y_f^s[k]$ on the red and black dashed lines are the sum of the noise-free phase differences between two 16QAM constellation points and the phase shift caused by CFO. Blue dashed lines are from the phase differences between BPSK or QPSK symbols of the PHY channel other than PDSCH and the shift by CFO. The red, blue, and black dashed lines in Fig. 4.9 correspond to the red, blue, and black dashed lines in Fig. 4.10, respectively. Fig. 4.9 and Fig. 4.10 show that symbols are densely located at

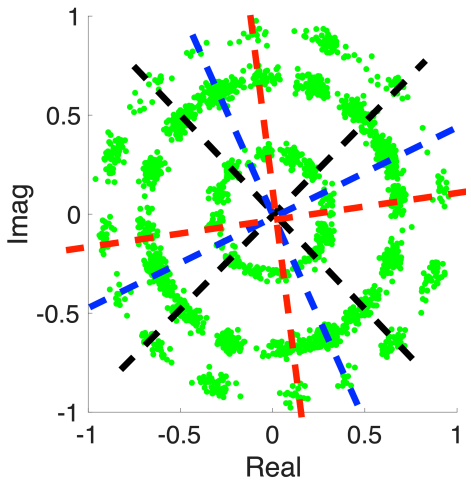


Figure 4.9. Scatterplot of $Y_f^s[k]$ of measured 16QAM features at SNR$=25\,$dB with 5G OTA data.

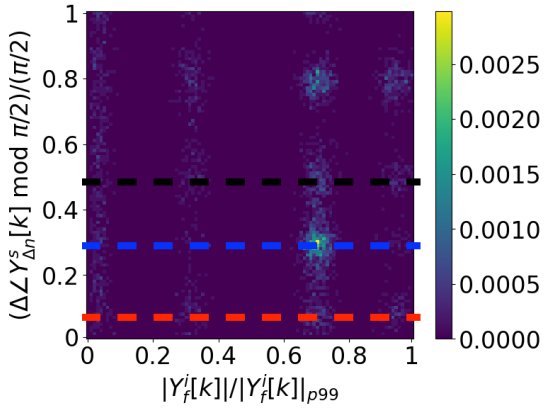
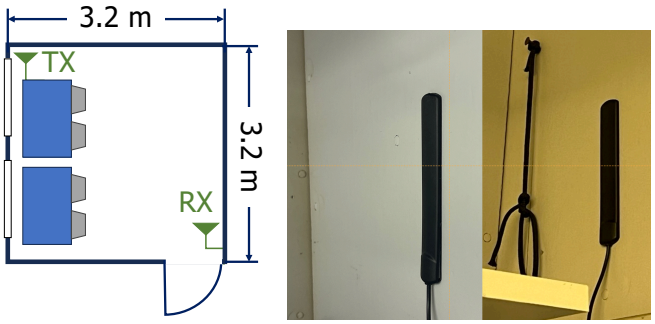


Figure 4.10. Histogram of $|Y_f^s[k]|/|Y_f^s[k]|_{\text{p99}}$ and $(\Delta\angle Y^s_{\Delta n}[k] \bmod \pi/2)/(\pi/2)$ of measured 16QAM features at SNR$=25\,$dB with 5G OTA data.

the points in the dashed lines, which is consistent with our expectations.

An advantage of using a histogram is that they are invariant to the length of $Y_f^s[k]$. This enables a neural network with a fixed structure to handle signals of any duration. This property is useful when dealing with 5G features where the number of samples of $Y_f^s[k]$ is unknown due to unused resources. Moreover, in a histogram input, the effect of CFO estimation error caused by aliasing (4.12) is a movement along the y-axis of the histogram as far as orthogonality of $\angle Y^s_{\Delta n}[k]$ across k holds. The neural network can be trained to identify histogram movements along the y-axis as a single class.



(a) (b)

Figure 4.11. OTA data propagation environment: (a) map with TX/RX locations, (b) vertically polarized antennas for TX (left) and RX (right), attached to the wall.

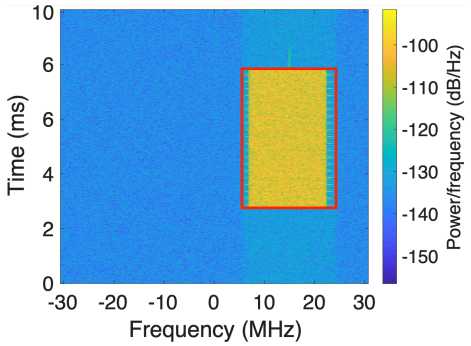


Figure 4.12. Scatterplot of $Y_f^s[k]$ of measured 16QAM features at SNR$=25$dB with 5G OTA data.

4.4 Evaluation

4.4.1 Data Collection

The proposed classifier is evaluated with synthetic data generated from AWGN channel simulations and real-world measured OTA data with the details in Table 4.6. MATLAB R2023a WLAN and 5G toolbox [28] are deployed to generate the synthetic AWGN dataset. Wi-Fi HT [29] and HE format [8] are used to generate data with $T_{\text{IFFT}} = 3.2\mu$s and 12.8μs in Wi-Fi 6. For 5G data, every SCS option in FR1, $\mu \in \{0,1,2\}$, is tested. All PHY channels listed in Table 4.4 are included in every 5G data item.

To evaluate whether the performance of the proposed system remains invariant across varying 5G PHY channel configurations, the parameters for allocating REs to PHY

Table 4.6. Data generation parameters

SNR	AWGN data: [5, 40] dB in steps of 5 dB OTA data: [4, 32] dB in steps of 4 dB
Carrier frequency	2.4 GHz (Wi-Fi 6), 2.6 GHz (5G)
The number of {train, test} data	{800, 200} per each ($T_{\text{IFFT}}, T_{\text{CP}}$, modulation) case
$\{\mathscr{S}_P, \mathscr{S}_Q\}$	{15, 50}
Time duration of each data	$400\,\mu$s (Wi-Fi 6), 5 ms (5G)

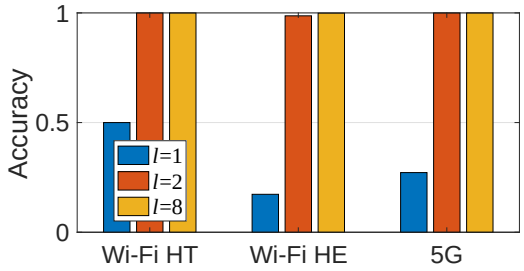


Figure 4.13. Accuracy for estimating T_{IFFT} and T_{CP} evaluatd with synthetic AWGN channel data.

channels are set for each data. For example in PDCCH, symbol duration, aggregation level, and starting symbol number are randomly selected. PHY broadcast channel (PBCH), primary synchronization signal (PSS), and secondary synchronization signal (SSS) are included only when $\mu \in \{0, 1\}$ since they are not available for $\mu = 2$. The other 5G PHY channel parameters are from FR1 test models in [30, 31].

Figure 4.11 documents the propagation environment where OTA data are measured. We deploy two networked software-defined radios, USRP N310 [32], for transmitting and receiving signals OTA. Both TX and RX are in the same room and the distance between TX and RX is $4.52\,$m, see Fig. 4.11a. TX and RX antenna are attached to the wall, see Fig. 4.11b. Fig. 4.12 shows a spectrogram with a 5G signal detected. Utilizing the assumed accurate signal detection, an IQ sequence corresponding to a detected signal (red box in Fig. 4.12) is extracted. After resampling to 20 MHz ($y[n]$), the sequence is taken as an input of the OFDM parameter estimator.

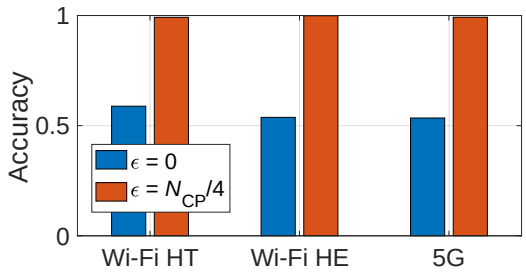


Figure 4.14. Accuracy for choosing the first index of CP with acceptable error ε with synthetic AWGN channel data.

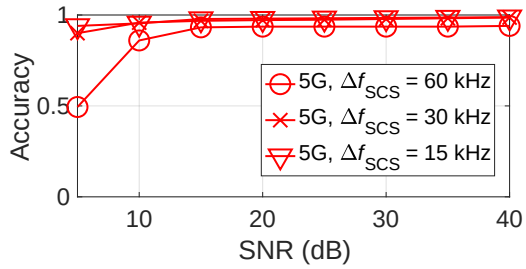
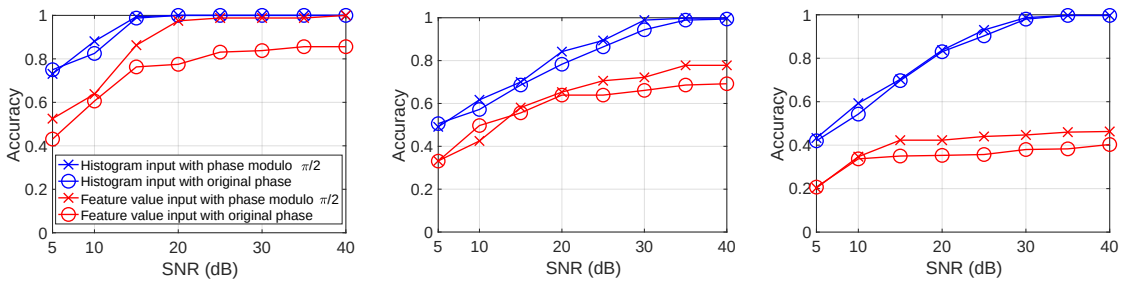


Figure 4.15. Accuracy for finding an OFDM symbol with long CP of 5G signals with synthetic AWGN channel data.

4.4.2 Building Classifier Input

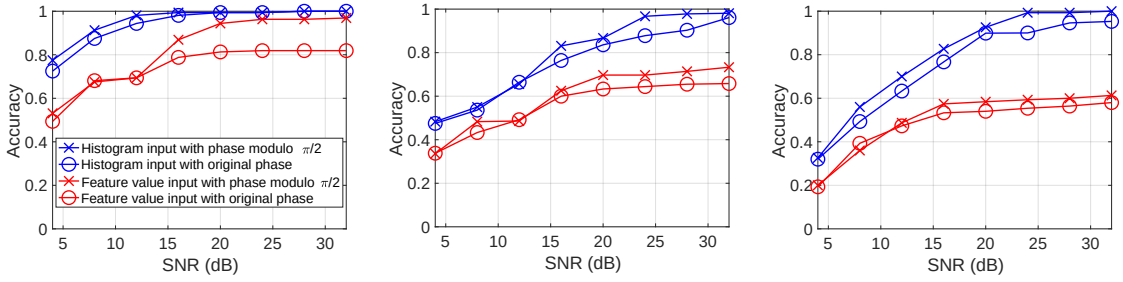
First, to avoid using the Wi-Fi preamble, we remove the first 2000 samples from each data. If the estimated T_{IFFT} corresponds to those of Wi-Fi 6, an IQ sequence whose length corresponds to 40+2 or 10+2 OFDM symbols is deployed to build $Y_f^s[k]$, starting with a random sample. We need an additional OFDM symbol due to the unknown starting index of an OFDM symbol sequence, $p \in [0, N_{\text{FFT}} + N_{\text{CP}} - 1]$. One more symbol is required since phase differences between those of every OFDM symbol and the next one should be computed. N_{null} is set to 8 and 32 for Wi-Fi HT and HE, respectively. If the estimated T_{IFFT} refers to 5G, $y'[n]$ of length (3 ms + 3 OFDM symbols) is used to estimate p and IndexLongCP.

For 5G, the sequence of 14 OFDM symbols is utilized for a classifier input. β is set to $|Y_f^s[k]|_{\text{p99}}/10$ in each input. We also evaluate $Y_f^s[k]$ values as an input to assess how much the histogram input contributes to the performance. In this case, one data input



(a) (b) (c)

Figure 4.16. Classification accuracy for modulations vs. SNR with synthetic data: (a) Wi-Fi HT, (b) Wi-Fi HE, (c) 5G.



(a) (b) (c)

Figure 4.17. Classification accuracy for modulations vs. SNR with OTA data. (a) Wi-Fi HT, (b) Wi-Fi HE, (c) 5G signals.

consists of 2240 samples for Wi-Fi 6 or 7900 samples for 5G, which is the average number of feature elements in a single 5G histogram data. We use fixed-duration data for a fair comparison, but the classifier can take the variable length data as input as the obtained feature is processed to a histogram using the algorithms in Sec. 4.3.4. For both input formats, an input with both phases of $\angle Y^s_{\Delta n}[k]$ modulo $\pi/2$ and $\angle Y^s_{\Delta n}[k]$ are evaluated.

4.4.3 Evaluation Results

AWGN channel data

Results in Fig. 4.13, Fig. 4.14, and Fig. 4.15 are obtained with synthetic AWGN channel data. Fig. 4.13 shows estimation accuracy of the OFDM parameters $\{T_{\text{CP}}, T_{\text{IFFT}}\}$ over different l, the length of $y[n+i]$ in CAF estimator (4.3). Using $l = 2,4$ achieves 99% accuracy for both Wi-Fi 6 formats and 5G and outperforms $l = 1$ as used in [24]. In

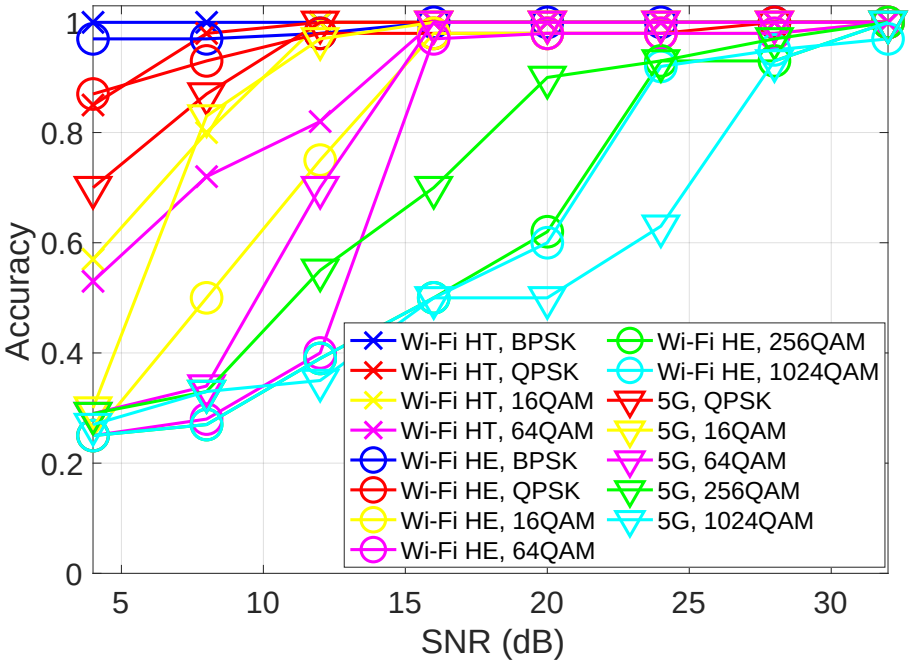


Figure 4.18. Classification accuracy for modulation with OTA data for each modulation format separately.

Fig. 4.14, the estimation accuracy of correctly finding the starting index of an OFDM symbol is shown for the method in Sec. 4.3.2. Correctly finding means that the starting index time is within ε samples tolerance of the true time. In Fig. 4.14, we note that the estimation accuracy for identifying the starting index of an OFDM symbol falls below 60% for both Wi-Fi 6 formats and 5G. When the tolerance is relaxed to $N_{\text{CP}}/4$ time samples, the reported estimation accuracy increases to 99%.

The accuracy of estimating an OFDM symbol with long CP is shown in Fig. 4.15. Aside from $\Delta f_{\text{SCS}} = 60$ kHz, the performance is over 90% even at low SNR of 5 dB. Accuracy at $\Delta f_{\text{SCS}} = 60$ kHz is low because the period of an OFDM symbol with long CP is larger than the others. The degraded peak detection performance due to the large number of symbols that the peak detection function needs to detect also negatively affects the estimation performance. At $\Delta f_{\text{SCS}} = 60$ kHz, 30 peaks should be identified in line 4 of Algorithm 2, which is considerably larger than the 9 or 16 peaks at $\Delta f_{\text{SCS}} = 15$ kHz and $\Delta f_{\text{SCS}} = 30$ kHz.

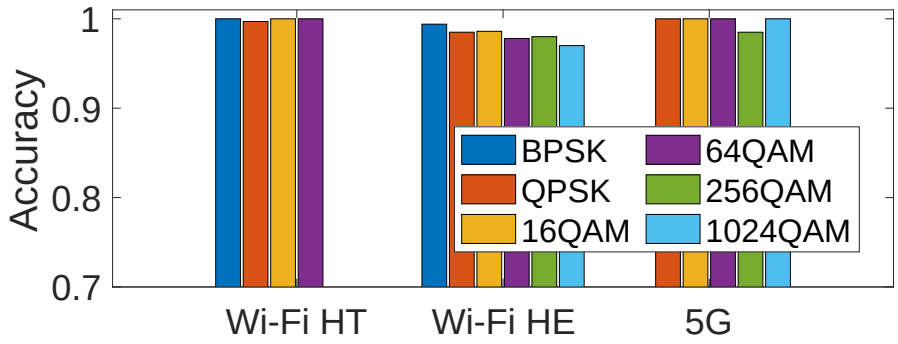


Figure 4.19. Classifier accuracy with OTA data when SNR exceeds the minimum requirements required for standard-compliant data communication.

Table 4.7. SNR required for data communication with each modulation

Modulation	BPSK	QPSK	16QAM
SNR for Wi-Fi 6 (dB)	5	10	16
SNR for 5G (dB)	-	15	18
Modulation	64QAM	256QAM	1024QAM
SNR for Wi-Fi 6 (dB)	22	30	35
SNR for 5G (dB)	21	27	30

Figure 4.16 shows modulation classification accuracy with synthetic AWGN channel data. The proposed algorithm with a histogram input with the phases $\Delta\angle Y^s_{\Delta n}[k]$ modulo $\pi/2$ outperforms in all considered cases, except for Wi-Fi 6 at 5 dB SNR. The performance gap between using the histogram as classifier input as opposed to using the feature value input increases in Wi-Fi HE and even more so in 5G. This is because the histogram input helps the classifier to discriminate the detailed symbol constellation of high-order modulations.

OTA data

The modulation classification accuracy with measured OTA data is in Fig. 4.17. The achieved OTA accuracy is similar to the synthetic AWGN channel data: a histogram input with the phases $\Delta\angle Y^s_{\Delta n}[k]$ modulo $\pi/2$ achieves the highest classification accuracy, except for Wi-Fi 6 at 5 dB SNR and a larger performance gap for Wi-Fi HE and 5G.

The classification accuracy of all considered modulation formats with OTA data is in Fig. 4.18. For a chosen accuracy, higher modulation orders require higher received SNR. E.g., Wi-Fi HE 16QAM signals have 90% accuracy if the SNR exceeds 16 dB, whereas Wi-Fi HE 256QAM requires 24 dB SNR. In Fig. 4.19, the accuracy of each modulation format is shown when the SNR satisfies the minimum requirement for standard-compliant data communication. We deploy error vector magnitude (EVM) levels required for data communication with each modulation for Wi-Fi 6 and 5G documentations [31, 8]. Required SNR values are calculated using the relation between EVM and SNR [33]. SNR required for the smallest coding rate are chosen for each modulation and chosen values are arranged in Table 4.7. For every modulation with both Wi-Fi 6 formats and 5G, accuracy is at least 97%.

4.5 Conclusion

Modulation classification of Wi-Fi 6 and 5G signals for spectrum sensing is studied. Simulations show that our classifier which uses SCS and CP length estimates based on the CAF achieves 99% accuracy. The classifier includes a preprocessing stage that is agnostic to control information, and extracts signal features characterizing modulation schemes insensitive to synchronization errors. For 5G signals, the preprocessing also estimates the symbol positions with a long CP. The features are converted to a more suitable form as inputs for the CNN-based classifier. This improves the classification of high-order modulation constellations. The modulation classifier identifies OFDM modulations with 97% accuracy when the SNR satisfies the requirements for standard-compliant data transmission for each modulation format with both synthetic AWGN channel data and measured OTA data.

4.6 Acknowledgements

Chapter 4, in full, is a reprint of the material as it appears in Kim, B., Mecklenbräuker, C., and Gerstoft, P. “Deep Learning-based Modulation Classification of Practical OFDM Signals for Spectrum Sensing”, in Proc. IEEE International Conference on Computer Communications 2024. The dissertation author was the primary investigator and author of this paper. The coauthors listed in this publication directed and supervised the research.

4.7 References

[1] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?," IEEE Trans. on Wireless Commun., vol. 15, no. 2, pp. 1293–1308, 2016.

[2] K. V. Mishra, M. Bhavani Shankar, V. Koivunen, B. Ottersten, and S. A. Vorobyov, "Toward millimeter-wave joint radar communications: A signal processing perspective," IEEE Signal Processing Mag., vol. 36, no. 5, pp. 100–114, 2019.

[3] N. Devroye, P. Mitran, and V. Tarokh, "Limits on communications in a cognitive radio channel," IEEE Commun. Mag., vol. 44, no. 6, pp. 44–49, 2006.

[4] J. Gao, X. Yi, C. Zhong, X. Chen, and Z. Zhang, "Deep learning for spectrum sensing," IEEE Wireless Commun. Letters, vol. 8, no. 6, pp. 1727–1730, 2019.

[5] L. Yu, J. Chen, and G. Ding, "Spectrum prediction via long short term memory," in Proc. IEEE ICCC, pp. 643–647, 2017.

[6] V. Sathyanarayanan, P. Gerstoft, and A. El Gamal, "Rml22: Realistic dataset generation for wireless modulation classification," IEEE Trans. on Wireless Commun., 2023.

[7] D. Liu, K. Ergun, and T. S. Rosing, "Towards a robust and efficient classifier for real world radio signal modulation classification," in Proc. IEEE ICASSP, pp. 1–5, 2023.

[8] IEEE 802.11ax, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 1: enhancements for high-efficiency WLAN," May 2021.

[9] 3GPP TR 38.331, "NR; Radio Resource Control (RRC); Protocol specification," Mar. 2023. ver 17.4.0.

[10] S. Hong, Y. Zhang, Y. Wang, H. Gu, G. Gui, and H. Sari, "Deep learning-based signal modulation identification in OFDM systems," IEEE Access, vol. 7, pp. 114631–114638, Aug. 2019.

[11] D. H. Al-Nuaimi, N. A. M. Isa, M. F. Akbar, and I. S. Z. Abidin, "AMC2-pyramid: Intelligent pyramidal feature engineering and multi-distance decision making for automatic multi-carrier modulation classification," IEEE Access, vol. 9, pp. 137560–137583, Sept. 2021.

[12] Z. Zhang, H. Luo, C. Wang, C. Gan, and Y. Xiang, "Automatic modulation classification using cnn-lstm based dual-stream structure," IEEE Trans. Veh. Technol., vol. 69, pp. 13521–13531, Nov. 2020.

[13] R. Gupta, S. Kumar, and S. Majhi, "Blind modulation classification for asynchronous ofdm systems over unknown signal parameters and channel statistics," IEEE Trans. Veh. Technol., vol. 69, pp. 5281–5292, Mar. 2020.

[14] A. K. Pathy, A. Kumar, R. Gupta, S. Kumar, and S. Majhi, "Design and implementation of blind modulation classification for asynchronous mimo-ofdm system," IEEE Trans. Instrum. Meas., vol. 70, pp. 1–11, Sept. 2021.

[15] S. Hong, Y. Wang, Y. Pan, H. Gu, M. Liu, J. Yang, and G. Gui, "Convolutional neural network aided signal modulation recognition in OFDM systems," in Proc. IEEE VTC, pp. 1–5, May 2020.

[16] M. C. Park and D. S. Han, "Deep learning-based automatic modulation classification with blind OFDM parameter estimation," IEEE Access, vol. 9, pp. 108305–108317, 2021.

[17] A. Kumar, K. K. Srinivas, and S. Majhi, "Automatic modulation classification for adaptive OFDM systems using convolutional neural networks with residual learning," IEEE Access, vol. 11, pp. 61013–61024, Jun. 2023.

[18] S. Kumar, E. Hamed, D. Katabi, and L. Erran Li, "Lte radio analytics made easy and accessible," Proc. ACM SIGCOMM, vol. 44, pp. 211–222, Aug. 2014.

[19] N. Bui and J. Widmer, "Owl: A reliable online watcher for lte control channel measurements," in Proc. 5th Workshop on All Things Cellular: Operations, Applications and Challenges, (New York (NY), USA), pp. 25–30, Association for Computing Machinery, 2016.

[20] R. Falkenberg and C. Wietfeld, "FALCON: An accurate real-time monitor for client-based mobile network data analytics," in Proc. IEEE GLOBECOM, (Waikoloa (HI), USA), pp. 1–7, IEEE, 2019.

[21] T. D. Hoang, C. Park, M. Son, T. Oh, S. Bae, J. Ahn, B. Oh, and Y. Kim, "Ltesniffer: An open-source lte downlink/uplink eavesdropper," in Proc. 16th ACM Conf. on Security and Privacy in Wireless and Mobile Networks (WiSec), (Guildford, UK), pp. 43–48, May 29–Jun. 1 2023.

[22] N. Ludant, P. Robyns, and G. Noubir, "From 5g sniffing to harvesting leakages of privacy-preserving messengers," in 2023 IEEE Symposium on Security and Privacy (SP), (Los Alamitos, CA, USA), pp. 1919–1934, IEEE Computer Society, may 2023.

[23] Y. Li, J. Barthelemy, S. Sun, P. Perez, and B. Moran, "A case study of wifi sniffing performance evaluation," IEEE Access, vol. 8, pp. 129224–129235, 2020.

[24] A. Punchihewa, V. K. Bhargava, and C. Despins, "Blind estimation of OFDM parameters in cognitive radio networks," IEEE Trans. Wireless Commun., vol. 10, pp. 733–738, Mar. 2011.

[25] S. S. Hong and S. R. Katti, "Dof: A local wireless information plane," in Proc. ACM SIGCOMM, pp. 230–241, 2011.

[26] W. A. Gardner and C. M. Spooner, "The cumulant theory of cyclostationary time-series. I. Foundation," IEEE Trans. Signal Process., vol. 42, pp. 3387–3408, Dec. 1994.

[27] G. E. Ltd, "GTXO-203T | 1.8V∼3.6V SM TCXO | Golledge," 2023.

[28] MathWorks, "MATLAB Products," 2023.

[29] IEEE 802.11n, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 5: enhancements for higher throughput," Oct. 2009.

[30] 3GPP TR 38.521-4, "NR; User Equipment (UE) conformance specification; Radio transmission and reception; Part 4: Performance requirements," 2023. ver 17.2.1.

[31] 3GPP TR 38.141-1, "NR; Radio Resource Control (RRC); Base Station (BS) conformance testing; Part 1: Conducted conformance testing," 2023. ver. 18.1.0.

[32] Ettus Research, "USRP N300/N310," in Knowledge Base, Ettus Research, 2018–2022. last accessed 2025-02-28.

[33] R. A. Shafik, M. S. Rahman, and A. H. M. R. Islam, "On the extended relationships among EVM, BER and SNR as performance metrics," in Proc. IEEE ICECE, pp. 408–411, 2006.

Chapter 5

Real-time Adversarial Attack to Deep Learning-based Wi-Fi Human Activity Recognition

5.1 Introduction

As Wi-Fi sensing employing channel state information (CSI) is used for a range of applications, including indoor localization [1, 2] and human activity recognition (HAR) [3, 4], concerns regarding privacy from widespread presence of Wi-Fi routers have escalated. Data-driven approaches, deployed in diverse wireless communication applications [5, 6, 7], also help to address the randomness of human bodies and activities, but the capability to identify diverse human activities using sequences of Wi-Fi CSI with deep learning (DL) raises privacy concerns. To address these issues in HAR, we propose a technique to obscure human activities from being detected by HAR classifiers. This method involves adding perturbation signals into the signals transmitted from the user device to the HAR classifier, aiming to degrade its performance.

We examine a setup involving a room with a Wi-Fi router and a user device, along with a human as in Fig. 5.1. The HAR classifier is located at a Wi-Fi router, which communicates with a user device, and identifies human activities using CSI estimated by the router. The signals used for CSI estimation in the Wi-Fi [8] is the long training field (LTF) transmitted by the user device. The estimated CSI is then used as input for the HAR classifier. On the user device, a perturbation signal generator manipulates the LTF

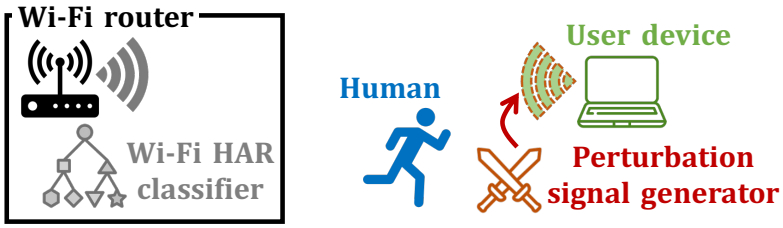


Figure 5.1. The scenario attacking Wi-Fi-based HAR from the user side.

to degrade the performance of the HAR classifier.

Previous research [9, 10] has adapted adversarial attack originally developed for deep learning (DL)-based image classification to degrade DL-based Wi-Fi HAR classifiers using methods like the fast gradient method (FGM) [11]. In FGM, classifier inputs should be manipulated; however, users typically cannot access the CSI collected by Wi-Fi routers. In this setting, the adversary knows the CSI sequence estimated only up to when it manipulates LTF. Subsequent CSI sequences, which could be part of inputs for the classifier, are not accessible at the point of manipulation. This paper presents a real-time adversarial attack as building signals that degrade the target model using the CSI available up to the moment the LTF is perturbed [12].

For real-time adversarial attacks, we employ reinforcement learning (RL) to find good decisions from the observed states. RL finds the actions, the adversarial examples in our context, that maximize the reward function given observed states. The reward function and the observed states correspond to the degraded performance of the HAR classifier and the observed CSI in our scenario. While typical RL provides reward function feedback [13], our case differs. Imitation learning (IL) offers an alternative, training only with expert state-action pairs–known to yield high rewards–rather than reward values, effectively addressing this challenge.

We propose an algorithm that constructs adversarial examples using only the CSI sequence estimated up to the point of their insertion. We employ a generative adversarial imitation learning (GAIL) algorithm. Notably, our algorithm does not necessitate knowl-

edge of the target model's structure. Evaluation of a dataset collected over eight days verifies that the GAIL model requires 1.0 dB lower perturbation signal amplitude than that of baseline schemes that rely on impractical assumptions.

Designing perturbation signals to degrade Wi-Fi-based HAR for privacy protection has been studied [14, 15, 10, 16, 17, 18]. Ref. [14, 15] suggest manipulation of the loss function of the Wi-Fi HAR classifier so that the proposed loss function is trained not to detect a prearranged set of activities. Finding the subcarriers and timesteps of the input sequence crucial to the HAR classifier performance is proposed in [10]. The attacking scenarios [14, 15, 10] require the adversary to know the weights of the target HAR classifier. Target channel and attack power determination for the adversary against Wi-Fi HAR considering the collision avoidance protocol is presented [17]. The authors of [18] present the black-box and universal adversarial attacks against HAR with mmWave radar. WiAdv [16] presents a targeted adversarial attack against Wi-Fi-based gesture recognition, which only considers the target gesture [19]. These proposed adversarial attacks [16, 17, 18] are built upon statistics of pre-observed surrounding wireless environments.

5.2 System objective

The Wi-Fi-based HAR classifier takes the sequence of CSI matrices as an input to determine the probability of each activity. Specifically, the orthogonal frequency-division multiplexing (OFDM)-MIMO system at ith timestamp and jth subcarrier with the transmitted and received signal vector, $\mathrm{x}_{ij} \in \mathbb{C}^{N_{\mathrm{TX}}}$ and $\mathrm{y}_{ij} \in \mathbb{C}^{N_{\mathrm{RX}}}$, is modeled as:

$$\mathrm{y}_{ij} = \mathrm{H}_{ij}\mathrm{x}_{ij} + \mathrm{n}_{ij}. \tag{5.1}$$

Here, $i \in \{1, \cdots, M\}$, $j \in \{1, \cdots, N_{\mathrm{SC}}\}$, and M denotes the human activity duration in the time unit. $\mathrm{H}_{ij} \in \mathbb{C}^{N_{\mathrm{RX}} \times N_{\mathrm{TX}}}$ and $\mathrm{n}_{ij} \in \mathbb{C}^{N_{\mathrm{RX}}}$ are the CSI matrix and the noise vector at ith

subcarrier and jth timestamp, respectively. A receiver (RX) deploys the known LTF (x_{ij}) to estimate the channel (H_{ij}) using the received signals (y_{ij}). The DL-based classifier is used for the classifier model [20, 21]. The classifier f_C takes as input a sequence of CSI matrices $\mathscr{H} \in \mathbb{R}^{M \times N_{\mathrm{SC}} \times N_{\mathrm{RX}} \times N_{\mathrm{TX}}} \triangleq \{|\mathbf{H}_{ij}|\}_{1 \le i \le M, 1 \le j \le N_{\mathrm{SC}}}$, where $|\mathbf{H}_{ij}|$ consists of the amplitudes of $\mathbf{H}_{ij}$'s elements. The classifier outputs the probability of each activity.

The adversary computes the adversarial example $\mathscr{A} \in \mathbb{R}^{M \times N_{\mathrm{SC}} \times N_{\mathrm{RX}} \times N_{\mathrm{TX}}}$, which degrades the target HAR classifier when added to the CSI sequence $\mathscr{H}$. Utilizing $\mathscr{A}$, the adversary generates perturbation signals, resulting in a modified CSI estimated as $\mathscr{H} + \mathscr{A}$ by the router. The adversary adds $\beta_{ij}^k \in \mathbb{C}$ to LTF transmitted by kth antenna on the user device $x_{ij}^k \in \mathbb{R}$. Using the known LTF, x_{ij}^k, the router estimates the CSI $\mathrm{h}_{ij}^k \in \mathbb{C}^{N_{\mathrm{RX}}}$. This estimation is distorted to $\bar{\mathrm{h}}_{ij}^k \in \mathbb{C}^{N_{\mathrm{RX}}}$:

$$\bar{\mathrm{h}}_{ij}^k = \mathrm{h}_{ij}^k (x_{ij}^k + \beta_{ij}^k)/x_{ij}^k = \mathrm{h}_{ij}^k (1 + \beta_{ij}^k / x_{ij}^k). \tag{5.2}$$

The purpose of β_{ij}^k is minimizing the accuracy of the target HAR classifier, f_{C}, with input $\bar{\mathscr{H}} \in \mathbb{R}^{M \times N_{\mathrm{SC}} \times N_{\mathrm{RX}} \times N_{\mathrm{TX}}}$ consists of $\bar{\mathrm{h}}_{ij}^k$.

5.3 Proposed Algorithm

5.3.1 Black-box FGM

Since the target HAR classifier is unknown to the adversary, we take a black-box attack approach to compute adversarial examples. The surrogate HAR classifier is trained using available CSI data and the adversarial which effectively degrades the surrogate model performance is computed. Bidirectional long-short-term memory (Bi-LSTM), which is used in multiple Wi-Fi-based HAR classifiers and is known to achieve high accuracy [20, 21, 22, 23], is adopted as the network model of the surrogate LSTM classifier, f'_{C}. We use the surrogate HAR with one Bi-LSTM layer and the following FC

layer. Using a surrogate model, black-box FGM $\hat{\mathscr{A}} \in \mathbb{R}^{M \times N_{\mathrm{SC}} \times N_{\mathrm{RX}} \times N_{\mathrm{TX}}}$ is computed as:

$$\hat{\mathscr{H}} = \mathscr{H} + \alpha \nabla_{\mathscr{H}} \mathscr{L}\left(f'_{\mathrm{C}}(\mathscr{H}), \mathrm{y}\right) = \mathscr{H} + \hat{\mathscr{A}}. \tag{5.3}$$

5.3.2 GAIL-based Real-time Adversarial Example Generator

The real-time adversarial example generator is designed to compute adversarial examples that degrade the target classifier at each time step i, denoted as $\mathrm{A}_i \in \mathbb{R}^{N_{\mathrm{SC}} \times N_{\mathrm{RX}} \times N_{\mathrm{TX}}}$, only using the sequence of CSI estimated up to time step i,

$$\mathscr{H}_i^{\ell} \triangleq [\mathbf{H}_{i-\ell}, \mathbf{H}_{i-\ell+1}, \cdots, \mathbf{H}_{i-1}] \in \mathbb{R}^{\ell \times N_{\mathrm{SC}} \times N_{\mathrm{RX}} \times N_{\mathrm{TX}}}, \tag{5.4}$$

where ℓ denotes the time sample length of the generator input. The real-time adversarial example generator is trained to find the policy function $\boldsymbol{\pi}(\mathrm{A}_i | \mathscr{H}_i^{\ell})$ such that $\mathscr{H} + \bar{\mathscr{A}}$ where $\bar{\mathscr{A}} \triangleq \{\bar{\mathrm{A}}_i\}_{i=1}^{M} \sim \boldsymbol{\pi}(\mathrm{A}_i | \mathscr{H}_i^{\ell})$ minimizes the target classifier accuracy.

We aim to train the generator, $\boldsymbol{\pi}$, to mimic adversarial examples that have been demonstrated to effectively impair the target classifier. For the training process, the black-box FGM $\hat{\mathscr{A}}$ (5.3) of CSI sequences, $\mathscr{H}$, which are collected apriori is used as the reference for imitation. The adversary computes $\hat{\mathscr{A}}$ using the available CSI sequence and the surrogate model f'_{C}, associating $\hat{\mathrm{A}}_i$ with $\mathscr{H}_i^{\ell}$. These pairs compose an expert trajectory guiding the training process of the generator.

GAIL [24] is an IL algorithm to extract the relation between $\mathscr{H}_i^{\ell}$ and $\hat{\mathrm{A}}_i$. GAIL imitates the policy of an expert only using its trajectories, $\{\mathscr{H}_i^{\ell}, \hat{\mathrm{A}}_i\}_{i=1}^{M}$. The architecture of GAIL is similar to that of a generative adversarial network (GAN). Like a GAN, GAIL consists of two network components, comprising a discriminator (D_w) and a policy function ($\boldsymbol{\pi}_{\boldsymbol{\theta}}$). The networks are trained with two types of trajectories, the expert trajectories, $\{\mathscr{H}_i^{\ell}, \hat{\mathrm{A}}_i\}_{i=1}^{M}$, and the learner trajectories, $\{\mathscr{H}_i^{\ell}, \mathrm{A}_i\}_{i=1}^{M}|_{\mathrm{A}_i \sim \boldsymbol{\pi}_{\boldsymbol{\theta}}(\cdot | \mathscr{H}_i^{\ell})}$, generated by $\boldsymbol{\pi}_{\boldsymbol{\theta}}$. The discriminator, similar to that in a GAN, is trained to distinguish between the expert and

Algorithm 3: GAIL-based real-time adversarial example generator

Data: Expert trajectories $\tau_E = \{\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i\}$ where $i = \{1, 2, \cdots, M\}$, initial parameters of discriminator w_0 and policy function θ_0

1 for $k = 0, 1, \cdots, K-1$ do

2 Sample trajectories with the policy of a learner $\tau_k \sim \pi_{\theta_k}(\mathrm{A}_i|\mathscr{H}_i^\ell)$;

3 Update discriminator parameters to increase the objective:
$w_{k+1} \leftarrow w_k + \nabla_{w_k} J(w_k)$ (5.6)

4 Update policy function parameters to decrease the objective:
$\theta_{k+1} \leftarrow \theta_k - \nabla_{\theta_k} K(\theta_k)$ (5.7)

5 end

Output: Trained policy network which can generate real-time adversarial attack $\pi_{\theta_K}(\mathrm{A}_i|\mathscr{H}_i^\ell)$

learner trajectories. The training objective of the discriminator is to maximize the function values evaluated at the expert trajectories, $D_w(\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i)$ and minimize those evaluated at the learner trajectories, $D_w((\mathscr{H}_i^\ell, \mathrm{A}_i)|_{\mathrm{A}_i \sim \pi_\theta(\cdot|\mathscr{H}_i^\ell)})$. Concurrently, the policy function is optimized to increase the discriminator values computed for the learner trajectories. Thus, the combined training objective that GAIL seeks to optimize can be summarized as follows:

$$\begin{aligned} &\min_{\pi_\theta} \max_{D_w} \; \mathbb{E}_{(\mathscr{H}_i^\ell, \mathrm{A}_i) \sim \pi_\theta(\mathrm{A}_i|\mathscr{H}_i^\ell)}[\log(D_w(\mathscr{H}_i^\ell, \mathrm{A}_i))] + \\ &\mathbb{E}_{(\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i)}[\log(1 - D_w(\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i)] - \lambda_G H(\pi). \end{aligned} \tag{5.5}$$

$\log(D_w)$ instead of D_w is optimized to address the constraint of the learnable function types [24]. The entropy of policy $H(\pi_\theta)$ is used as a regularizer function. Algorithm 3 presents the full procedure to optimize the objective.

In each iteration in Algorithm 3, learner trajectories $\tau = \{\mathscr{H}_i^\ell, \mathrm{A}_i\}$ are generated from CSI data available to the user device with A_i sampled with the policy function, $\pi_{\theta_k}(\cdot|\mathscr{H}_i^\ell)$. Both the discriminator and policy functions are optimized alternately using learner and expert trajectories. In line 3, the discriminator function is trained by adjusting its parameter, w, to minimize its values evaluated on expert trajectories and maximize

those on learner trajectories,

$$\begin{aligned}\nabla_w J(w) = \mathbb{E}_{(\mathscr{H}_i^\ell, \mathrm{A}_i)\sim\pi_{\theta_k}}[\nabla_w \log(D_w(\mathscr{H}_i^\ell, \mathrm{A}_i))] \\ +\mathbb{E}_{(\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i)}[\nabla_w \log(1 - D_w(\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i))].\end{aligned} \tag{5.6}$$

The policy gradient in line 4 aims to minimize the cost function evaluated on the learner trajectories and maximize the regularizer function value,

$$\nabla_\theta K(\theta) = \mathbb{E}_{\tau_k}[\nabla_\theta \log \pi_\theta(\mathrm{A}_i|\mathscr{H}_i^\ell) C(\mathscr{H}_i^\ell, \mathrm{A}_i)] - \lambda_G \nabla_\theta H(\pi_\theta) \tag{5.7}$$

where cost function $C(\mathscr{H}_i^\ell, \mathrm{A}_i) = \mathbb{E}_{\tau_k}[\log(D_{w_{i+1}}(\mathscr{H}_i^\ell, \mathrm{A}_i))]$. Since the learner trajectory $\tau_k = \{\mathscr{H}_i^\ell, \mathrm{A}_i\}_{i=1}^M$ is sampled using π, cost function depends on policy. To compute the gradient with regard to the policy, we deploy trust region policy optimization (TRPO) [25], a policy gradient algorithm used in [24], to compute the gradient of $K(\theta)$.

5.3.3 LTF Manipulation in MIMO System

The adversary adds perturbation signals to LTF so that perturbed CSI is estimated by the router as the sum of CSI and the adversarial example computed in Sec. 5.3.2. At each subcarrier, manipulation on a single LTF at kth antenna distorts CSI estimated by the router's every antenna. If multiple antennas are present on the router, a change in one LTF cannot induce arbitrary changes in multiple CSIs. To tackle this, we deploy a perturbation signal that minimizes the distance between perturbed CSI and the sum of CSI and the desired adversarial example.

For kth user device antenna, the adversary modifies LTF, $x_{ij}^k \in \mathbb{R}$, with an adversarial example, $\beta_{ij}^k \in \mathbb{R}$. It aims to make original CSI, $\mathrm{h}_{ij}^k \in \mathbb{R}^{N_{\mathrm{RX}}}$ to be inaccurately estimated as $\mathrm{h}_{ij}^k + \bar{\mathrm{a}}_{ij}^k$, where $\bar{\mathrm{a}}_{ij}^k \in \mathbb{R}^{N_{\mathrm{RX}}}$ represents an element of $\bar{\mathscr{A}}$, the output of the trained adversarial example generator. This element corresponds to ith timestamp, jth

Table 5.1. Dataset parameters

Channel bandwidth (MHz)	20	$\{N_{\text{TX}},\ N_{\text{RX}}\}$	$\{1,\ 3\}$
No. experiment participants	6	No. evaluation days	8
CSI sampling rate (Hz)	50	Carrier frequency (GHz)	2.4

subcarrier, and kth transmitter antenna. The objective is represented as:

$$\mathrm{h}_{ij}^{k}(x_{ij}^{k}+\beta_{ij}^{k})/x_{ij}^{k}=(\mathrm{h}_{ij}^{k}+\bar{\mathrm{a}}_{ij}^{k})\Rightarrow \mathrm{h}_{ij}^{k}\beta_{ij}^{k}/x_{ij}^{k}=\bar{\mathrm{a}}_{ij}^{k} \tag{5.8}$$

If $N_{\text{RX}}>1$, it is not always possible to find β_{ij}^{k} which exactly satisfies (5.8). Instead, we propose using $\bar{\beta}_{ij}^{k}$, which minimizes the distance between the two terms in (5.8), defined as:

$$\bar{\beta}_{ij}^{k}=\arg\min_{\beta_{ij}^{k}}|\mathrm{h}_{ij}^{k}\beta_{ij}^{k}/x_{ij}^{k}-\bar{\mathrm{a}}_{ij}^{k}|=\mathrm{h}_{ij}^{k}\cdot\bar{\mathrm{a}}_{ij}^{k}/|\mathrm{h}_{ij}^{k}|^{2}. \tag{5.9}$$

Instead of the original LTF, x_{ij}^{k}, the manipulated LTF, $x_{ij}^{k}+\bar{\beta}_{ij}^{k}$, is transmitted by the user device antenna k.

5.4 Evaluation

For evaluation, we use the Wi-Fi HAR dataset [20], with details in Table 6.4. Since N_{RX} is 3, the perturbation signal minimizing the distance (5.9), $\bar{\beta}_{ij}^{k}$, is added to the LTF. The surrogate classifier and corresponding GAIL model are trained with the dataset downsampled from 1 kHz, the original CSI sampling rate of the dataset, to 50 Hz. This adjustment is designed to extend the time duration of GAIL input without altering the input time sample length, ℓ. This downsampling approach is justified as it does not compromise the performance of the trained models, since the sampling rate is above the maximum frequency caused by human movement, which is 30 Hz at 2.4 GHz carrier fre-

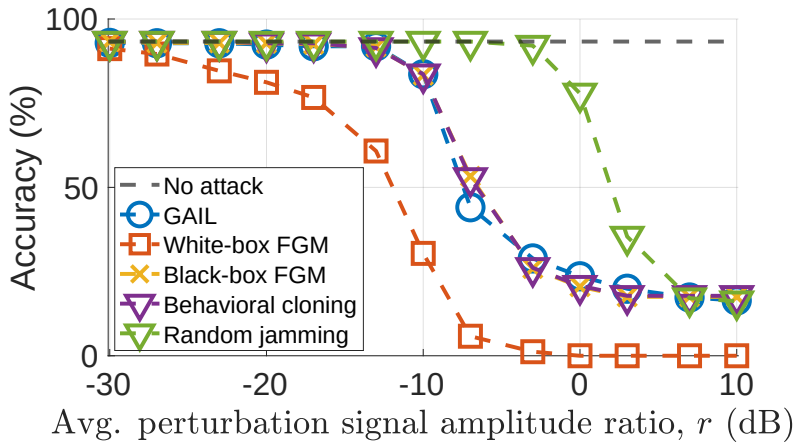


Figure 5.2. r values needed to degrade the accuracy of the Bi-LSTM-based target classifier across varying levels when training and test data are collected from different environments.

quency. For the target classifier, we utilize a Bi-LSTM model whose structure is identical to the surrogate.

5.4.1 Baseline schemes

The perturbation signal, $\bar{\beta}_{ij}^k$, is added to the LTF transmitted by the user device in every baseline scheme except the white-box attack. To compare each scheme's performance of degrading target classifier, we test different average perturbation signal amplitude ratios:

$$r = \mathbb{E}[|\beta_{ij}^k|/|x_{ij}^k|] \tag{5.10}$$

where the average is taken over $i \in \{1,\cdots,M\}$, $j \in \{1,\cdots,N_{\text{SC}}\}$, and $k \in \{1,\cdots,N_{\text{TX}}\}$.

White-box attack: White-box FGM (5.3) is computed with the target model f_{C}, instead of f'_{C}. To compute white-box FGM, the adversary should know the target classifier structures, its weights, and future CSI.

Black-box attack: Black-box FGM $\hat{\mathscr{H}}$ (5.3) is computed using the surrogate model f'_{C}. The adversary needs to know the future CSI at the point of adding perturbation signals.

Behavioral cloning: Behavioral cloning [26] is an IL algorithm used for real-time ad-

versarial attack. Similar to GAIL, black-box FGM is computed with the training dataset of the surrogate model. State-action pairs, $(\mathscr{H}_i^\ell, \mathrm{A}_i)$, are deployed as input features and output labels. In behavioral cloning, a Bi-LSTM classifier function $f_{\mathrm{BC}}(\mathscr{H}_i^\ell)$ is trained in a supervised learning manner using those pairs. The input length is set to $\ell = 5$, as in GAIL. Behavioral cloning works without information on the target classifier, future CSI, or action duration.

Random jamming: Random jamming, noise elements sampled with a Gaussian distribution, is evaluated. This method requires no prior knowledge of the target classifier.

5.4.2 Evaluation Results

Figure 5.2 presents the results of the proposed algorithm and the baseline attack schemes with the perturbation signal amplitude ratios, r. Here, the same Bi-LSTM-based structure and sampling rate are used for the surrogate and target model. Across all the datasets, white-box FGM yields the lowest accuracy, while random jamming results in the highest accuracy for the same amplitude of perturbation signals. With similar r, GAIL and behavioral cloning achieve performance degradation equivalent to that of black-box FGM. This is promising for GAIL and behavioral cloning as they require less information compared to black-box FGM.

The efficient perturbation signals effectively degrade the target classifier while minimizing the communication link impairment. Even when r values are the same, variations in the perturbation signal amplitude distribution can have differing effects on the link performance. Table 5.2 presents the ratio of LTF where the amplitude ratio, $|\beta_{ij}^k|/|x_{ij}^k|$, exceeds a threshold. r is set for the target degraded accuracy specified in each column. The threshold is set to 5 dB (0.301), the minimum SNR for data transfer over a wireless link [8]. If the $|\beta_{ij}^k|/|x_{ij}^k|$ surpasses this threshold, data transmission becomes impossible even under ideal channel conditions. The ratio of LTF where perturbation signal proportion is larger than 5 dB is comparable with black-box GAIL to degrade the classifier

Table 5.2. The ratio of LTF symbols larger than the threshold with target degraded accuracy.

Target accuracy	90%	50%	25%
GAIL	1.0%	12.8%	52.5%
White-box FGM	0.0%	2.5%	9.5%
Black-box FGM	0.7%	12.0%	20.7%
Behavioral cloning	0.8%	26.5%	85.5%

performance below 90% or 50%. To make classifier accuracy below 25%, the required ratio of LTF where communication is unavailable of GAIL is 48% of that of behavioral cloning.

5.5 Conclusion

The real-time adversarial attack against Wi-Fi-based HAR is studied. We deploy the emblematic IL algorithm, GAIL, to build a perturbation signal generator. With a Bi-LSTM-based surrogate classifier, our GAIL-based algorithm degrades the classifier accuracy to 50% with 0.5 dB higher than the other schemes based on impractical assumptions on average. This is achieved with only 12.8% of the resources distorted to the extent that data communication is impossible, only 48% of the comparison scheme under the same assumptions.

5.6 Acknowledgements

Chapter 5, in full, is a reprint of the material as it appears in proceeding of Kim, B., Panchagatti, A., and Gerstoft, P. “Real-time Adversarial Attack to Deep Learning-based Wi-Fi Human Activity Recognition”, in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing 2025. The dissertation author was the primary investigator and author of this paper. The coauthors listed in this publication directed and supervised the research.

5.7 References

[1] Z. Wang, B. Guo, Z. Yu, and X. Zhou, "Wi-fi csi-based behavior recognition: From signals and actions to activities," IEEE Commun. Mag., vol. 56, no. 5, pp. 109–115, 2018.

[2] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "Confi: Convolutional neural networks based indoor wi-fi localization using channel state information," IEEE Access, vol. 5, pp. 18066–18074, 2017.

[3] K. Ismail, R. Liu, Z. Qin, A. Athukorala, B. P. L. Lau, M. Shalihan, C. Yuen, and U.-X. Tan, "Efficient WiFi LiDAR SLAM for autonomous robots in large environments," in Proc. IEEE CASE, pp. 1132–1137, 2022.

[4] Y. Chapre, A. Ignjatovic, A. Seneviratne, and S. Jha, "Csi-mimo: Indoor wi-fi fingerprinting system," in Proc. IEEE LCN, pp. 202–209, IEEE, 2014.

[5] B. Kim, C. Mecklenbräuker, and P. Gerstoft, "Deep learning-based modulation classification of practical OFDM signals for spectrum sensing," in Proc. of IEEE Conference on Computer Communications, IEEE, 2024.

[6] H. Ye, L. Liang, G. Y. Li, and B.-H. Juang, "Deep learning-based end-to-end wireless communication systems with conditional gans as unknown channels," IEEE Trans. Wireless Commun., vol. 19, no. 5, pp. 3133–3143, 2020.

[7] B. Kim, C. F. Mecklenbräuker, and P. Gerstoft, "Blind modulation classification of Wi-Fi 6 and 5G signals for spectrum sensing," in Proc. ACM MSWiM, pp. 137–145, 2023.

[8] IEEE 802.11ax, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 1: enhancements for high-efficiency WLAN," May 2021.

[9] J. Yang, H. Zou, and L. Xie, "Securesense: Defending adversarial attack for secure device-free human activity recognition," IEEE Trans. Mobile Comput., vol. 23, no. 1, pp. 823–834, 2022.

[10] L. Xu, X. Zheng, X. Li, Y. Zhang, L. Liu, and H. Ma, "Wicam: Imperceptible adversarial attack on deep learning based wifi sensing," in Proc. IEEE SECON, pp. 10–18, 2022.

[11] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," arXiv preprint arXiv:1312.6199, 2013.

[12] Y. Gong, B. Li, C. Poellabauer, and Y. Shi, "Real-time adversarial attacks," in Proc. IJCAI, pp. 4672–4680, 2019.

[13] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," Machine learning, vol. 8, pp. 229–256, 1992.

[14] S. Zhou, W. Zhang, D. Peng, Y. Liu, X. Liao, and H. Jiang, "Adversarial wifi sensing for privacy preservation of human behaviors," IEEE Commun. Lett., vol. 24, no. 2, pp. 259–263, 2019.

[15] W. Zhang, S. Zhou, D. Peng, L. Yang, F. Li, and H. Yin, "Understanding and modeling of wifi signal-based indoor privacy protection," IEEE Internet of Things J., vol. 8, no. 3, pp. 2000–2010, 2020.

[16] Y. Zhou, H. Chen, C. Huang, and Q. Zhang, "Wiadv: Practical and robust adversarial attack against wifi-based gesture recognition system," Proc. ACM IMWUT, vol. 6, no. 2, pp. 1–25, 2022.

[17] J. Liu, Y. He, C. Xiao, J. Han, and K. Ren, "Time to think the security of wifi-based behavior recognition systems," IEEE Trans. on Dependable and Secure Comput., 2023.

[18] Y. Xie, R. Jiang, X. Guo, Y. Wang, J. Cheng, and Y. Chen, "Universal targeted adversarial attacks against mmwave-based human activity recognition," in Proc. IEEE INFOCOM, pp. 1–10, IEEE, 2023.

[19] Y. Zhang, Y. Zheng, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Widar3.0: Zero-effort cross-domain gesture recognition with wi-fi," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 11, pp. 8671–8688, 2021.

[20] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," IEEE Commun. Mag., vol. 55, no. 10, pp. 98–104, 2017.

[21] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "Wifi csi based passive human activity recognition using attention based blstm," IEEE Trans. Mobile Comput., vol. 18, no. 11, pp. 2714–2724, 2018.

[22] J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah, and J. Cheng, "Data augmentation and dense-lstm for human activity recognition using wifi signal," IEEE Internet of Things J., vol. 8, no. 6, pp. 4628–4641, 2020.

[23] P. Khan, B. S. K. Reddy, A. Pandey, S. Kumar, and M. Youssef, "Differential channel-state-information-based human activity recognition in iot networks," IEEE Internet of Things J., vol. 7, no. 11, pp. 11290–11302, 2020.

[24] J. Ho and S. Ermon, "Generative adversarial imitation learning," Proc. NeurIPS, vol. 29, 2016.

[25] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in Proc. ICML, pp. 1889–1897, 2015.

[26] F. Torabi, G. Warnell, and P. Stone, "Behavioral cloning from observation," in Proc. IJCAI, pp. 4950–4957, 2018.

Chapter 6

Remote Adversarial Attacks against Wi-Fi-based HAR for Privacy Protection

6.1 Introduction

As Wi-Fi routers become prevalent indoors, Wi-Fi sensing with Wi-Fi channel state information (CSI) has been employed for diverse purposes, including indoor localization [1, 2], radio fingerprinting [3, 4], and human mesh construction with millimeter-wave Wi-Fi [5]. One application of Wi-Fi sensing is human activity recognition (HAR) [6, 7], which aims to identify human activities using sequences of Wi-Fi CSI. Data-driven approaches for HAR have been introduced in the classification framework [8, 9] due to the randomness of human bodies and activities. The capacity of these methods to discern human activities from sequences of Wi-Fi CSI highlights significant privacy implications. We address the privacy concerns in HAR by proposing a method to prevent human activities from being detected by HAR classifiers. The proposed method adds perturbation signals into the signals transmitted from the user device to the router to degrade the classifier's accuracy.

Our study focuses on a scenario depicted in Fig. 6.1, involving a room with a Wi-Fi router and a user device that transmits and receives signals, along with a human, whose movements cause variations in CSI over time. A target Wi-Fi HAR classifier operates on a Wi-Fi router that receives signals from user devices and identifies human activities using CSI estimated by the router. The router estimates CSI using the long training field

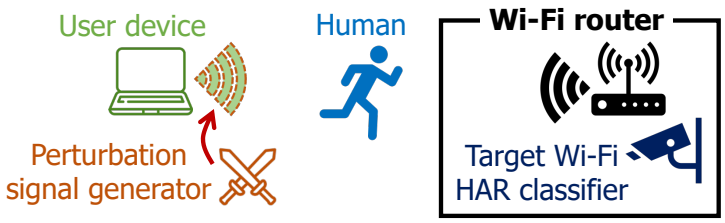


Figure 6.1. The scenario attacking Wi-Fi-based HAR by adding perturbation signals to Wi-Fi preamble from the user side.

(LTF) transmitted by the user device, as specified in Wi-Fi standards [?]. A perturbation signal generator on the user device modifies the LTF to degrade the HAR classifier with little impact on communication link quality.

Previous works on compromising Wi-Fi HAR classifier [10, 11] have explored the manipulation of HAR classifier inputs, adapting adversarial attacks originally designed for DL-based image classifiers. These approaches effectively disrupt HAR classifiers using the fast gradient method (FGM) [12]. However, such methods require the adversary to directly manipulate the CSI estimated at Wi-Fi routers, making them impractical since users might not be able to control these router-side measurements. In contrast, our approach is more practical since users modify their own transmitted signals to protect their privacy.

In adding perturbation signals to the HAR classifier input, we address two challenges to generating perturbation signals.

1. Unknown target HAR classifier: The adversary lacks information about the structure or training data of the target HAR classifier. The adversary constructs a surrogate HAR classifier on the same activities as the target classifier. The adversarial attacks designed for the surrogate HAR can transfer to the original model as well [13]—known as black-box attack. Moreover, in a remote attack against Wi-Fi-based HAR, the adversary does not know the CSI sequence sampling rate or which portion of the CSI sequence is used as input. Existing HAR classifiers take sliding windows of the CSI sequence for the entire activity duration [14, 15, 16]. However, the adversary at the user device cannot syn-

chronize the input sequence with the target classifier due to unknown start time, sequence length, and sampling rate.

2. Unknown future CSI: In contrast with attacks on an image classifier, in the remote attack on Wi-Fi HAR, perturbation signals should be computed only with the CSI sequence estimated up to the point when the LTF is manipulated. Future CSI sequences, which could be used as inputs to the classifier, are unknown at the time of the attack. This paper defines the real-time adversarial attack as generating signals degrading the target model with CSI available up to the point when LTF is manipulated [17].

For unknown future CSIs, reinforcement learning (RL)-based approaches find optimal decisions based on CSI estimated up to the current time. In RL, the decision performance is evaluated using the SIreward function which the agent maximizes with its actions. The reward function represents the degraded accuracy of the target HAR. Estimated CSI up to the current time and corresponding adversarial examples against the HAR classifier form a state-action pair. In RL, as REINFORCE [18], the learning agent interacts with the reward function for feedback on its actions. However, the adversary only determines the accuracy of the target classifier when processing the entire input. Imitation learning (IL) provides a solution to this limitation. The IL is trained only with expert trajectories, the sequences of state-action pairs, and not reward function values as in RL. The IL is trained to imitate the state-action pairs of experts, which achieve high rewards. Here, estimated CSI and pre-computed black-box FGM serve as expert state-action pairs at each time step.

The main contributions are summarized as follows:

Real-time adversarial attack against Wi-Fi-based HAR: We present a real-time adversarial attack against Wi-Fi HAR, which builds perturbation signals using only past CSI estimates. We utilize GAIL, an RL algorithm, to train the adversarial example generator.

Attacking scheme with limited knowledge available to the adversary: Our algo-

rithm does not require information on the target HAR structure or activity duration. Moreover, the adversary is not synchronized with the target model's input to generate perturbation signals.

Evaluation with diverse scenarios: Our GAIL-based real-time adversarial attack algorithm is evaluated on three Wi-Fi HAR datasets [14, 19, 20], collected across seven environments. We evaluate our attack algorithm for six target models, including DL-based classifiers and random forest (RF) [21].

6.2 Background

6.2.1 Wi-Fi-based HAR

Wi-Fi-based HAR classifier takes a sequence of CSI matrices as an input and outputs the probability of each activity. Specifically, the orthogonal frequency-division multiplexing (OFDM)-MIMO system for ith timestamp and jth subcarrier with the transmitted and received signal vector, $\mathrm{x}_{ij} \in \mathbb{C}^{N_{\mathrm{TX}}}$ and $\mathrm{y}_{ij} \in \mathbb{C}^{N_{\mathrm{RX}}}$, is modeled as:

$$\mathrm{y}_{ij} = \mathrm{H}_{ij}\mathrm{x}_{ij} + \mathrm{n}_{ij} \tag{6.1}$$

where $i \in \{1, \cdots, M\}$, $j \in \{1, \cdots, N_{\mathrm{SC}}\}$, and M is the length of human activity in the time sample unit. $\mathrm{H}_{ij} \in \mathbb{C}^{N_{\mathrm{RX}} \times N_{\mathrm{TX}}}$ and $\mathrm{n}_{ij} \in \mathbb{C}^{N_{\mathrm{RX}}}$ represent the CSI matrix and the noise vector of ith subcarrier and jth timestamp. A receiver (RX) uses apriori known knowledge on LTF (x_{ij}) for channel estimation (H_{ij}) using the received signals (y_{ij}). In this work, we denote the user device as the transmitter (TX) and the router as the receiver (RX).

LSTM is used for a neural network model of the HAR classifier in other Wi-Fi HAR studies [14, 15, 22, 23]. LSTM HAR classifier, f_C, takes as input a sequence of CSI matrices, $\mathscr{H} \in \mathbb{R}^{M \times N_{\mathrm{SC}} \times N_{\mathrm{RX}} \times N_{\mathrm{TX}}} \triangleq \{|\mathbf{H}_{ij}|\}_{1 \le i \le M, 1 \le j \le N_{\mathrm{SC}}}$, where $|\mathbf{H}_{ij}|$ is the matrix containing the amplitudes of $\mathbf{H}_{ij}$'s elements. The classifier outputs the vector of each element that represents the probability of each activity. Since different activities take

different durations, the time length of each input sequence is not fixed. The capability of LSTM to take inputs of flexible length is useful for processing CSI data of different lengths. The exact bidirectional-LSTM (Bi-LSTM)-based model used as the surrogate model is explained in detail in Sec. 6.4.

6.2.2 Adversarial Attacks against Neural Networks

FGM [24] computes adversarial examples that efficiently degrade the neural network. With the input of neural network X, FGM is:

$$\dot{\mathrm{X}} = \mathrm{X} + \alpha \nabla_{\mathrm{X}} \mathscr{L}\left(f(\mathrm{X}), \mathrm{z}\right) \tag{6.2}$$

where $\dot{\mathrm{X}}$ and $\mathrm{z} \in \mathbb{R}^{N_C}$ are a perturbed input and the one-hot encoded label corresponding to the input X, respectively. α is a parameter tuning the amplitude of FGM attack, $\mathscr{L}$ the loss function of f, and N_C the number of classes which f classifies.

When the adversary lacks access to the target classifier information, the FGM (6.2) cannot compute adversarial examples. This scenario, known as a black-box attack [13], occurs when the adversary has limited knowledge of the target model, including its network weights, architecture, training data labels, or training dataset. Table 6.1 describes the black-box attack scenario considered in this work. The adversary employs the surrogate model, $f'(\mathrm{X})$, instead of $f(\mathrm{X})$, to generate the adversarial examples using FGM (6.2). We specifically investigate cases where the adversary knows training data labels since the adversary on the user device is likely to know user activity types while lacking all other information detailed in Table 6.1. We evaluate scenarios where the surrogate and target models have both identical and different network structures.

6.2.3 Imitation Learning

We propose RL architecture to address real-time adversarial attack problems. In real-time adversarial attack scenarios, adversarial examples must be computed at each time step using the sequence of CSI estimated up to that point. This challenge aligns with the RL paradigm, where actions are determined based on observed states and influence subsequent state transitions. The RL problem is represented with a Markov decision process (MDP), MDP($\mathsf{S}, \mathsf{A}, T, R, \gamma$):

- S: A set of possible states
- A: A set of possible actions
- $T : \mathsf{S} \times \mathsf{A} \to \mathsf{S}$: State transition probability distribution
- $R : \mathsf{S} \times \mathsf{A} \times \mathsf{S} \to \mathbb{R}$: Reward function
- γ: Discount factor for future rewards.

A policy function $\pi(a_i|s_i) : \mathsf{S} \times \mathsf{A} \to [0,1]$ defines the learning agent's decision-making strategy, specifying the probability of selection action a_i given state s_i at each time step. Given MDP($\mathsf{S}, \mathsf{A}, T, R, \gamma$), RL aims to find the optimal policy function π which maximizes the cumulative reward function over a trajectory, $\tau = \{(s_1, a_1), (s_2, a_2), \cdots, (s_M, a_M)\}$, where a trajectory represents a sequence of state-action pairs, (s_i, a_i).

IL is a specialized form of RL, where a training agent learns to replicate expert behavior solely from given expert trajectories. Behavioral cloning [25] treats it as a supervised learning task where states serve as inputs and actions as labels. Behavioral cloning is prone to the environment not encountered during the training, called covariate shift [26]. Inverse RL (IRL) [27] also addresses the IL problem by learning the policy and the reward function, where the latter assigns higher scores to the trajectories similar to expert behavior. However, IRL is computationally expensive, as it requires executing complete RL

optimization at each time step. Apprenticeship learning is a computationally efficient alternative to IRL [28]. While it learns both reward function and expert-mimicking policies with reasonable computational cost, the class of functions it can learn is restrictive [29].

6.2.4 Related Work

Adversarial attack against Wi-Fi HAR: Designing perturbation signals to degrade Wi-Fi-based HAR for privacy protection has been studied [11, 30, 31, 32, 33, 34, 35, 36]. Ref. [30, 31] suggest modifying the HAR classifier's loss function to prevent the detection of specific activities. The approach in [11] presents a method that identifies and perturbs critical subcarriers and time steps in the input sequence that significantly influence classifier accuracy. However, these attacking scenarios in [30, 31, 11] assume the adversary has complete access to both the target HAR classifier and the exact inputs of the classifier. In contrast, our algorithm works only with CSI which the adversary can estimate.

The work [32] proposes Zigbee-like perturbation signals using statistics of pre-observed CSI to avoid denoising schemes. The study [34] develops an algorithm to determine jamming signal power that disrupts Wi-Fi HAR while evading detection under the Wi-Fi collision avoidance protocol. Black-box and universal adversarial attacks against HAR with mmWave radar are presented in [35]. WiAdv [33] introduces a targeted adversarial attack against Wi-Fi-based gesture recognition where perturbation signals depend solely on the target gesture [37]. Li et al. [36] explored the Carlini & Wagner (C&W) attack scheme to compromise Wi-Fi HAR by adding perturbation signals to the Wi-Fi preamble. While previous approaches [32, 33, 34, 35, 36] rely on pre-observed environment statistics, with the works [32, 33, 36] limited to single-antenna scenarios, our GAIL-based method generates perturbation signals using real-time CSI while considering practical multi-antenna receiver constraints.

Remote adversarial attacks against the neural network for wireless communications:

Several studies address remote adversarial attacks on neural network-based wireless systems [38, 39, 40, 41, 42]. These works address Wi-Fi-based indoor localization [38], device identification [39], and vulnerabilities in DL-based decoders [40, 41, 42]. Universal perturbation signals are proposed for modulation classification [42] and autoencoder-based decoders [40], while [41] presents a general attack scheme for DL-based decoders. However, some approaches [39, 40, 42] assume knowledge of input CSI sequence boundaries, which is not impractical in our scenario. Others [38, 41] only consider CSI inputs for a single OFDM symbol, unlike HAR classifiers that process variable-length sequences.

Real-time adversarial attack against neural networks: A real-time adversarial attack using behavioral cloning-based IL against neural networks processing time-series inputs is introduced in [17]. While universal perturbation signals [43] can be computed using pre-observed data, their input-invariant nature makes them unsuitable to attack HAR classifiers whose input length is unknown. Further studies on universal attacks for time-series neural networks are presented in [44, 45], with [44] proposing perturbation signals for audio inputs without needing synchronization. The predictive attack in [46] develops a neural network that both predicts upcoming audio streams and generates effective perturbations for these predictions.

6.3 System Objective

We develop an adversary at the user device that generates perturbation signals to degrade the accuracy of the HAR classifier at the Wi-Fi router as illustrated in Fig. 6.1. The adversary determines perturbation signals added to the LTF transmitted by the user device, causing the router to estimate CSI as the sum of unperturbed CSI and the adversarial examples against the target classifier.

Unlike approaches where the adversary is located at the router [10, 11, 30, 31], our system places the adversary at the user device. Compared to manipulation of classifier

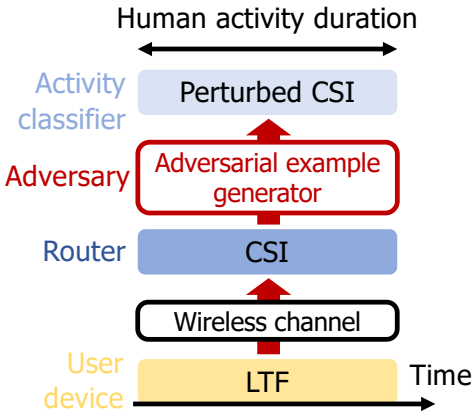


Figure 6.2. Conventional adversarial attack scenario.

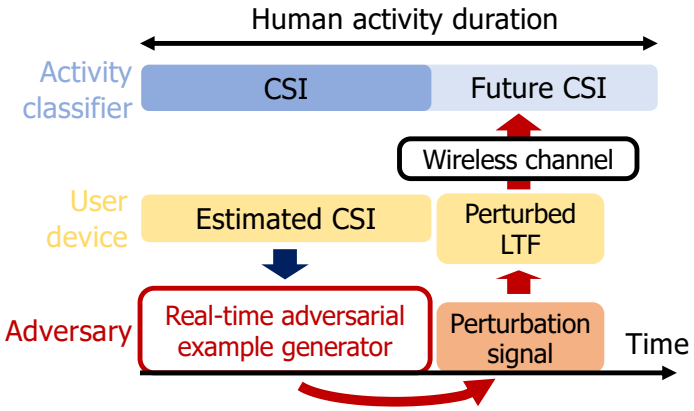


Figure 6.3. Real-time adversarial attack scenario.

inputs at the router, manipulation of LTF transmitted by the user device is more practical since users can directly control their signal transmission for privacy. While manipulating the inputs at the router could be feasible with open-source Wi-Fi router drivers, most router vendors (e.g., Broadcom [47]) adopt proprietary drivers. Although customizing Wi-Fi operations of user devices (e.g., in smartphones) is challenging, smartphone vendors could incorporate our algorithm into future firmware releases for their customers' privacy.

Fig. 6.2 illustrates the scenario where the adversary at the Wi-Fi router computes adversarial examples with knowledge of the entire-length CSI input sequence. Conventional adversarial example generation algorithms from other domains like image classification can be applied in this setting. However, as previously discussed, this approach is

impractical since it requires control over router-side CSI.

In contrast, our real-time approach, described in Fig. 6.3, places the adversary at the user device with control over transmitted signals. At time step i, the adversary computes perturbation signals, $\mathscr{B}_i \in \mathbb{R}^{N_{\mathrm{SC}} \times N_{\mathrm{TX}}}$, using only the CSI estimated at the user device up to the moment before LTF manipulation, $\mathscr{H}_i \triangleq [\mathbf{H}_0, \mathbf{H}_1, \cdots, \mathbf{H}_{i-1}]$, as it cannot utilize future CSI which may serve as classifier input. Note that $\mathscr{B}_i$ has no RX antenna dimension since each TX antenna transmits one single LTF to all RX antennas. For each time step i, subcarrier j, and kth antenna of the user device, the adversary adds perturbation $\beta_{ij}^k \triangleq [\mathscr{B}_i]_{jk} \in \mathbb{R}$ to LTF, $x_{ij}^k \in \mathbb{R}$. Using the known LTF, the router estimates the CSI for user device antenna k as $\mathrm{h}_{ij}^k \in \mathbb{C}^{N_{\mathrm{RX}}}$. This estimation is distorted to $\bar{\mathrm{h}}_{ij}^k \in \mathbb{C}^{N_{\mathrm{RX}}}$ as follows:

$$\bar{\mathrm{h}}_{ij}^k = \mathrm{h}_{ij}^k (x_{ij}^k + \beta_{ij}^k)/x_{ij}^k = \mathrm{h}_{ij}^k (1 + \beta_{ij}^k / x_{ij}^k). \tag{6.3}$$

The objective of $\mathscr{B} \triangleq [\mathscr{B}_0, \mathscr{B}_1, \cdots, \mathscr{B}_{M-1}] \in \mathbb{R}^{M \times N_{\mathrm{SC}} \times N_{\mathrm{TX}}}$ is to minimize the accuracy of the target HAR classifier, f_{C}:

$$\arg \min_{\mathscr{B}=(\beta_{ij}^k)} \mathbb{E}_{\{\mathscr{H},\, l\}} \left[\mathbb{1}(\{f_{\mathrm{C}}(\bar{\mathscr{H}})\} = l) \right], \tag{6.4}$$

where $\bar{\mathscr{H}} \in \mathbb{C}^{M \times N_{\mathrm{SC}} \times N_{\mathrm{RX}} \times N_{\mathrm{TX}}}$ contain elements $\bar{\mathrm{h}}_{ij}^k$ and l is the true activity label of CSI sequence $\mathscr{H}$.

We assume that the user device and router estimate CSI at least once each perturbation signal generation period, and that the CSI estimated by the user device matches that estimated by a Wi-Fi router within this period. To maintain these assumptions, we limit the maximum sampling period to 33.3 ms, corresponding to the channel coherence time of 2.4 GHz Wi-Fi under human movement [48]. Channel reciprocity ensured through Wi-Fi's time division duplex guarantees that CSI estimated by two devices re-

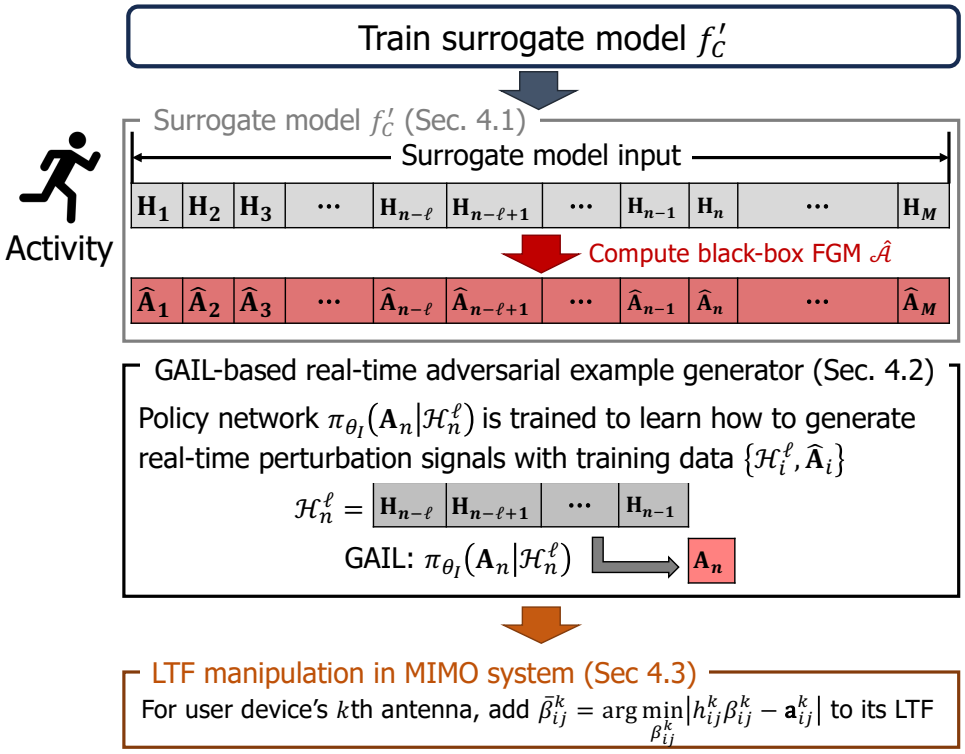


Figure 6.4. Perturbation signal generator. First, black-box FGM is computed using surrogate model. Next, a real-time adversarial example generator is trained using the GAIL algorithm. Lastly, perturbation signals to be added to LTF are determined considering MIMO system constraint.

mains identical. Our maximum perturbation rate of 50 Hz is well within the capability of commercial Wi-Fi adapters, which support CSI sampling up to 4 kHz [49], demonstrating the practicality of our approach.

6.4 Perturbation Signal Generator

The proposed perturbation signal generator is illustrated in Fig. 6.4. First, the adversary builds a surrogate model to compute adversarial examples following black-box attack (Sec. 6.4.1). Using this surrogate model, black-box FGM computes adversarial examples, which are then paired with the CSI estimated by the user device up to the time step of computation. Using the GAIL algorithm, the adversary is trained to map the CSI sequences to corresponding adversarial examples by imitating the black-box FGM

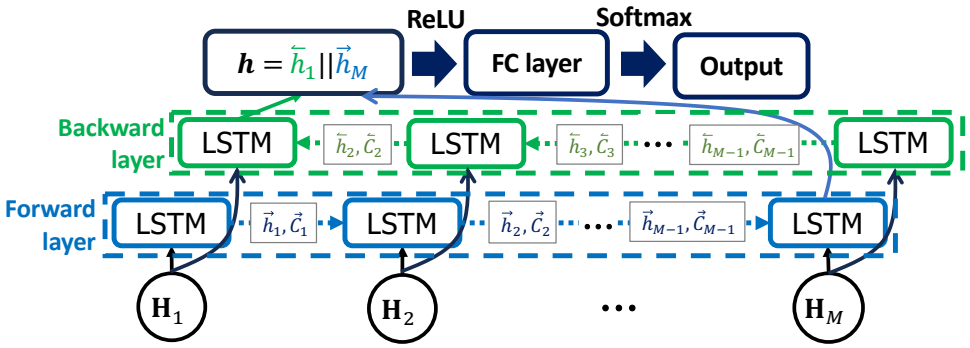


Figure 6.5. An example Bi-LSTM-based classifier.

(Sec. 6.4.2). Finally, considering MIMO system constraints, the perturbation signals added to LTF are computed using the output of the trained adversarial example generator (Sec. 6.4.3).

6.4.1 Black-box FGM

Since the target HAR classifier is unknown to the adversary, it computes the examples using the black-box attack approach. A surrogate HAR classifier is trained using available CSI data, with its network structure consisting of one LSTM layer follwed by one fully-connected layer. Specifically, we adopt a Bi-LSTM layer, commonly used in Wi-Fi-based HAR classifiers [14, 15, 16, 20, 21, 50, 51], as the network model of the surrogate LSTM classifier, f'_{C}.

The surrogate HAR consists of one Bi-LSTM layer followed by one FC layer, as illustrated in Fig. 6.5. Bi-LSTM processes input sequences in both forward and backward directions, enabling the use of both past and future context, unlike standard LSTM which only uses past information. The Bi-LSTM layer comprises a forward and a backward layer, each consisting of multiple LSTM cells (shown as the green or blue solid rectangles in Fig. 6.5). Each cell processes CSI matrices over subcarriers, $\mathrm{H}_j \triangleq \{\mathrm{H}_{ij}\}_{1 \le i \le N_{\mathrm{SC}}}$. In the forward layer, hidden and cell states are updated with $\mathbf{H}_j$ and passed to the cell with the next time step input. In the backward layer, the hidden and cell states are propagated to the cell with the input at the previous time step. The final hidden states from both the

Table 6.1. Available information in white-box/black-box attack scenarios

Information	White-box	Black-box
Neural network structure	O	O/X
Weights of neural network	O	X
Training data	O	X
Label of training data	O	O

Table 6.2. Surrogate Bi-LSTM HAR classifier model parameters

Parameter	Value	Parameter	Value
LSTM hidden layer dimension	200	Learning rate	1×10^{-5}
Loss function	Cross-entropy	Epoch	400

forward and backward layers serve as input for the remaining network, which outputs the activity classification probabilities.

Since the target model is inaccessible to the adversary, black-box FGM computes $\hat{\mathscr{A}} \in \mathbb{R}^{M \times N_{\mathrm{SC}} \times N_{\mathrm{RX}} \times N_{\mathrm{TX}}}$ using a surrogate model:

$$\hat{\mathscr{H}} = \mathscr{H} + \alpha \nabla_{\mathscr{H}} \mathscr{L}\left(f'_{\mathrm{C}}(\mathscr{H}), \mathrm{y}\right) = \mathscr{H} + \hat{\mathscr{A}}, \tag{6.5}$$

here f in (6.2) is replaced with the surrogate model f'_{C}. The performance of these adversarial examples against different target models is evaluated in Sec. 4.4.

6.4.2 GAIL-based Real-time Adversarial Example Generator

The real-time adversarial example generator aims to compute adversarial examples effectively degrading the target classifier at time step i, $\mathrm{A}_i \in \mathbb{R}^{N_{\mathrm{SC}} \times N_{\mathrm{RX}} \times N_{\mathrm{TX}}}$, only using the

Table 6.3. GAIL network parameters

Parameter	Value	Parameter	Value
Epochs	4000	D_w learning rate	2×10^{-5}
Input length (ℓ)	5	Hidden layer dimension	200
ε_K, ε_π	0.01	No. hidden layers	4
Discount factor (γ)	0.99	Policy regularizer coefficient (λ_G)	0.01

sequence of CSI estimated up to time step i,

$$\mathscr{H}_i^\ell \triangleq [\mathbf{H}_{i-\ell}, \mathbf{H}_{i-\ell+1}, \cdots, \mathbf{H}_{i-1}] \in \mathbb{R}^{\ell \times N_{\text{SC}} \times N_{\text{RX}} \times N_{\text{TX}}}, \tag{6.6}$$

where ℓ denotes the input sequence length. The generator's objective is to find the policy function $\pi(\mathrm{A}_i|\mathscr{H}_i^\ell)$ that satisfies

$$\begin{aligned} \arg \min_{\pi(\mathrm{A}_i|\mathscr{H}_i^\ell)} & \mathbb{E}_{\{\mathscr{H},\, l\}} \left[\mathbb{1}(\{f_{\text{C}}(\mathscr{H} + \mathscr{A}^{\bar{}})\} = l) \right] \\ & \text{where } \mathscr{A}^{\bar{}} \triangleq \{\bar{\mathrm{A}}_i\}_{i=1}^M \sim \pi(\mathrm{A}_i|\mathscr{H}_i^\ell). \end{aligned} \tag{6.7}$$

Our goal is to train the generator, π, to imitate black-box FGM adversarial examples, $\hat{\mathscr{A}}$ (6.5), which effectively degrade the target classifier. The adversary computes $\hat{\mathscr{A}}$ using the available CSI sequence and the surrogate model f'_{C}, creating expert trajectories by pairing $\hat{\mathrm{A}}_i$ with its corresponding $\mathscr{H}_i^\ell$.

We employ GAIL [52], an IL algorithm that learns from expert trajectories, $\{\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i\}_{i=1}^M$. It easily generalizes to unseen environments with practical computational complexity [52]. GAIL's structure resembles a generative adversarial network (GAN), which comprises two networks trained on expert trajectories, $\{\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i\}_{i=1}^M$, and learner trajectories, $\{\mathscr{H}_i^\ell, \mathrm{A}_i\}_{i=1}^M|_{\mathrm{A}_i \sim \pi_\theta(\cdot|\mathscr{H}_i^\ell)}$. Like GAN, the GAIL network comprises a discriminator (D_w) and a policy function (π_θ). The discriminator in GAIL is trained to distinguish

expert and learner trajectories by maximizing the discriminator function evaluated in the expert trajectories, $D_w(\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i)$, and minimizing the values evaluated in the learner trajectories, $D_w((\mathscr{H}_i^\ell, \mathrm{A}_i)|_{\mathrm{A}_i \sim \pi_\theta(\cdot|\mathscr{H}_i^\ell)})$. The policy function π_θ is optimized to increase the discriminator's output on learner trajectories. Taken together, the complete GAIL objective is:

$$\begin{aligned} &\min_{\pi_\theta} \max_{D_w} \ \mathbb{E}_{(\mathscr{H}_i^\ell, \mathrm{A}_i) \sim \pi_\theta(\mathrm{A}_i|\mathscr{H}_i^\ell)}[\log(D_w(\mathscr{H}_i^\ell, \mathrm{A}_i))] + \\ &\mathbb{E}_{(\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i)}[\log(1 - D_w(\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i)] - \lambda_G H(\pi). \end{aligned} \tag{6.8}$$

Using $\log(D_w)$ instead of D_w helps to address function type constraints [52], and the entropy of policy function $H(\pi_\theta)$ serves as a regularizer. Detailed optimization steps are presented in Algorithm 4.

Algorithm 4: GAIL for real-time adversarial attacks against HAR

Data: Expert trajectories $\tau_E = \{\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i\}$ where $i = \{1, 2, \cdots, M\}$, initial parameters for discriminator function w_0 and policy function θ_0

1 for $k = 0, 1, \cdots, K-1$ do

2 Sample trajectories with the policy of a learner $\tau_k \sim \pi_{\theta_k}(\mathrm{A}_i|\mathscr{H}_i^\ell)$;

3 Update discriminator parameters to increase the objective:
$w_{k+1} \leftarrow w_k + \nabla_{w_k} J(w_k)$ (6.9)

4 Update policy function parameters to decrease the objective:
$\theta_{k+1} \leftarrow \theta_k - \nabla_{\theta_k} K(\theta_k)$ (6.10)

5 end

Output: Trained policy network which can generate real-time adversarial attack $\pi_{\theta_K}(\mathrm{A}_i|\mathscr{H}_i^\ell)$

In each iteration in Algorithm 4, the learner trajectories $\tau = \{\mathscr{H}_i^\ell, \mathrm{A}_i\}$ are sampled from CSI data available to the user device and the corresponding adversarial examples A_i generated by the policy function, $\pi_{\theta_k}(\cdot|\mathscr{H}_i^\ell)$. The discriminator and policy functions are alternately optimized using these learner and expert trajectories. In line 3, the discriminator function parameters, w, are updated using the gradient:

$$\begin{aligned} \nabla_w J(w) = &\mathbb{E}_{(\mathscr{H}_i^\ell, \mathrm{A}_i) \sim \pi_{\theta_k}}[\nabla_w \log(D_w(\mathscr{H}_i^\ell, \mathrm{A}_i))] \\ &+ \mathbb{E}_{(\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i)}[\nabla_w \log(1 - D_w(\mathscr{H}_i^\ell, \hat{\mathrm{A}}_i))]. \end{aligned} \tag{6.9}$$

Line 4 describes the policy gradient, which minimizes the cost function evaluated on learner trajectories while maximizing the regularizer,

$$\nabla_\theta K(\theta) = \mathbb{E}_{\tau_i}[\nabla_\theta \log \pi_\theta(\mathrm{A}_i|\mathscr{H}_i^\ell) C(\mathscr{H}_i^\ell, \mathrm{A}_i)] - \lambda_G \nabla_\theta H(\pi_\theta) \tag{6.10}$$

where cost function $C(\mathscr{H}_i^\ell, \mathrm{A}_i) = \mathbb{E}_{\tau_k}[\log(D_{w_{i+1}}(\mathscr{H}_i^\ell, \mathrm{A}_i))]$. Since the cost function depends on policy through the sampled learner trajectory $\tau_k = \{\mathscr{H}_i^\ell, \mathrm{A}_i\}_{i=1}^M$, computing $\nabla_\theta K(\theta)$ non-trivial. We address this using trust region policy optimization (TRPO) rule [53], with detailed procedures presented in Appendix 6.7.

6.4.3 LTF Manipulation in MIMO System

The adversary manipulates LTF by adding perturbation signals to make the router estimate perturbed CSI as the sum of the original CSI and the computed adversarial example. However, when the router has multiple antennas, manipulating a single LTF from kth transmitter antenna affects CSI estimation at all receiver antennas. This makes it impossible to achieve arbitrary changes in multiple CSIs through a single LTF modification. To address this limitation, we propose perturbation signals that minimize the distance between perturbed CSI and the sum of CSI and the desired adversarial example.

For kth antenna of the transmitter, the adversary adds perturbation, $\beta_{ij}^k \in \mathbb{R}$, to LTF, $x_{ij}^k \in \mathbb{R}$. The goal is to make the router misestimate the original CSI, $\mathrm{h}_{ij}^k \in \mathbb{R}^{N_{\mathrm{RX}}}$ as $\mathrm{h}_{ij}^k + \bar{\mathrm{a}}_{ij}^k$, where $\bar{\mathrm{a}}_{ij}^k \in \mathbb{R}^{N_{\mathrm{RX}}}$ is the corresponding element of the desired adversarial example, $\bar{\mathscr{A}}$, at ith timestamp, jth subcarrier, and kth transmitter antenna. This relationship can be represented as:

$$\mathrm{h}_{ij}^k (x_{ij}^k + \beta_{ij}^k)/x_{ij}^k = (\mathrm{h}_{ij}^k + \bar{\mathrm{a}}_{ij}^k) \Rightarrow \bar{\mathrm{a}}_{ij}^k = \mathrm{h}_{ij}^k \beta_{ij}^k / x_{ij}^k \tag{6.11}$$

When $N_{\mathrm{RX}} > 1$, an exact solution for β_{ij}^k may not exist. Therefore, we propose using

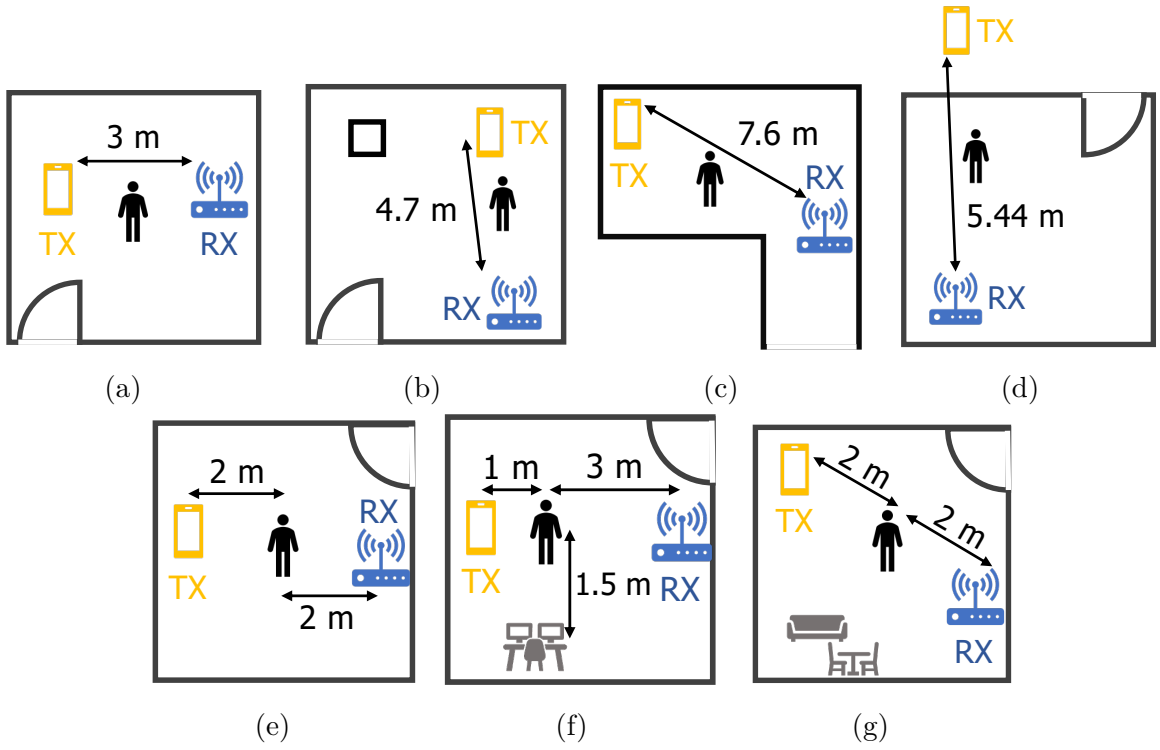


Figure 6.6. Data collection environments for (a) TAR, (b)-(d) JAR, and (e)-(g) WiAR.

$\bar{\beta}_{ij}^{k}$ that minimizes the distance between manipulated CSI and the desired adversarial example:

$$\bar{\beta}_{ij}^{k} = \arg\min_{\beta_{ij}^{k}} |\mathrm{h}_{ij}^{k}\beta_{ij}^{k}/x_{ij}^{k} - \bar{\mathrm{a}}_{ij}^{k}| = \mathrm{h}_{ij}^{k} \cdot \bar{\mathrm{a}}_{ij}^{k}/||\mathrm{h}_{ij}^{k}||^{2}. \tag{6.12}$$

The user device antenna k then transmits the manipulated LTF, $x_{ij}^{k} + \bar{\beta}_{ij}^{k}$, instead of the original x_{ij}^{k}.

6.5 Evaluation

6.5.1 Dataset and Target Models

We evaluate our algorithm using use publicly accessible datasets, TAR [14], JAR [19], WiAR [20]. Table 6.4 summarizes the dataset details, and Fig. 6.6 illustrates data collection environments, with floor plans sourced from the original dataset documents [14, 19, 20]. WiAR and JAR are collected across three different environments including one non-line-of-

Table 6.4. Dataset parameters

Dataset	TAR	JAR	WiAR
Channel bandwidth (MHz)	20	20	20
$\{N_{\mathrm{TX}}, N_{\mathrm{RX}}\}$	{1, 3}	{1, 3}	{1, 3}
No. activities	6	6	16
No. environments	1	3	3
No. evaluation days	8	1	1
No. experiment participants	6	30	10
Sampling rate (Hz)	1000	320	30

Table 6.5. Wi-Fi HAR target classifiers

HAR classifier	Dataset	Input features	DL structure	Input length
[14]	TAR	CSI	LSTM	2 s
[15]	TAR	CSI	LSTM+ Attention	2 s
[16]	JAR	CSI	CNN+ LSTM	1.6 s
[21]	JAR	Statistical features	RF	1.6 s
[54]	WiAR	STFT	CNN	Variable

sight (NLOS) scenario for JAR. TAR spans eight different days, allowing us to evaluate our algorithm under both spatial and temporal diversities. Since all datasets use $N_{\mathrm{TX}} = 1$ and $N_{\mathrm{RX}} = 3$, we deploy $\bar{\beta}_{ij}^{1}$ (6.12) to manipulate LTF transmitted from a single transmitter antenna. For WiAR, we evaluate the same five activities tested in [11].

Our evaluation includes various target HAR models shown in Table 6.5. We assess GAIL-based attacks against diverse architectures (CNN and Attention models) and RF models. We also evaluate Bi-LSTM classifiers matching our surrogate model's structure. In addition to those taking raw CSI as inputs, the classifiers that take statistical features including standard deviation and SNR [21] or short-time Fourier transform (STFT) [54] are tested.

While existing Wi-Fi-based HAR classifiers use fixed-length sliding windows of CSI sequences as inputs, the adversary likely lacks knowledge of the target model's exact starting window points. Therefore, our surrogate model is trained on a full-length activity sequence, leveraging Bi-LSTM's capability to handle variable-length inputs. Both FGM and GAIL computations use these variable-length inputs, accordingly. The perturbation signals generated with this surrogate against the targets taking fixed-length inputs are also tested. In these tests, we evaluate the sliding window of the sum of the CSI matrix sequence and computed perturbation signals. This work represents the first study of HAR degradation without sliding window information. We additionally evaluate both the target models with variable-length inputs matching our surrogate's structure.

For TAR and JAR, we train the surrogate classifier and GAIL model on the downsampled dataset. The downsampling is intended to accommodate a longer time duration coverage while maintaining the same input size of time samples. While increasing ℓ could extend GAIL's input time length, it increases model complexity. The downsampling factor should balance surrogate classifier accuracy and GAIL model complexity. Given that human movement in the 2.4GHz band produces maximum Doppler frequencies up to 30 Hz [48], downsampling above 30 Hz should preserve activity classification capabil-

ity. We select 50 Hz and 40 Hz as sampling rates for surrogate model input trained on TAR and JAR, respectively. The surrogate model accuracy with different sampling rates is evaluated to ensure the surrogate works well with the chosen sampling rates.

6.5.2 Baseline Schemes

This subsection presents the baseline schemes for performance comparison. Like our proposed generator, all schemes except white-box FGM add perturbation signal $\bar{\beta}_{ij}^{k}$ to the LTF transmitted by the user device. To compare each scheme's degrading efficiency, we evaluate different average perturbation signal amplitude ratios:

$$r = \mathbb{E}[|\beta_{ij}^{k}|/|x_{ij}^{k}|] \tag{6.13}$$

averaged over $i \in \{1,\cdots,M\}$, $j \in \{1,\cdots,N_{\mathrm{SC}}\}$, and $k \in \{1,\cdots,N_{\mathrm{TX}}\}$.

White-box attack: The adversary knows the classifier structure, weights, and future CSI. White-box FGM (6.2) is computed using the target model f_{C}. We evaluate two scenarios: direct manipulation of classifier input and LTF manipulation. These scenarios are compared to assess how degradation efficiency varies with the MIMO constraint (6.12). The scenarios are referred to as white-box FGM and white-box, MIMO, respectively, in Sec. 6.5.3.

Black-box attack: Black-box FGM $\hat{\mathscr{H}}$ (6.5) uses the surrogate model f'_{C}. We evaluate scenarios where the target model and surrogate models have identical and different structures. Both assume the knowledge of future CSI when adding perturbation signals.

Universal attack: Universal FGM $\hat{\mathscr{H}}_U$ averages black-box FGM $\hat{\mathscr{H}}$ computed on surrogate model's training data of each action [43]. Due to varying action durations, we upsample all black-box FGMs to the longest duration per action before averaging. The averaged examples are then resampled to match specific action lengths and processed for LTF perturbation. This approach does not require future CSI knowledge but requires

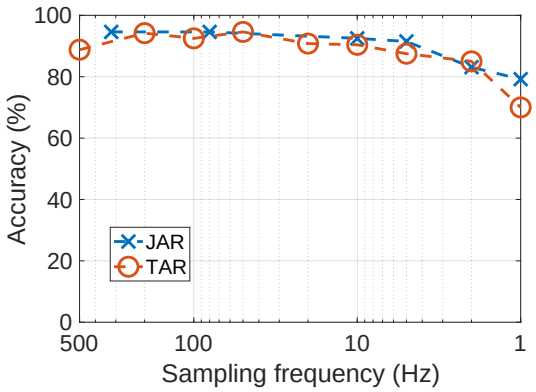
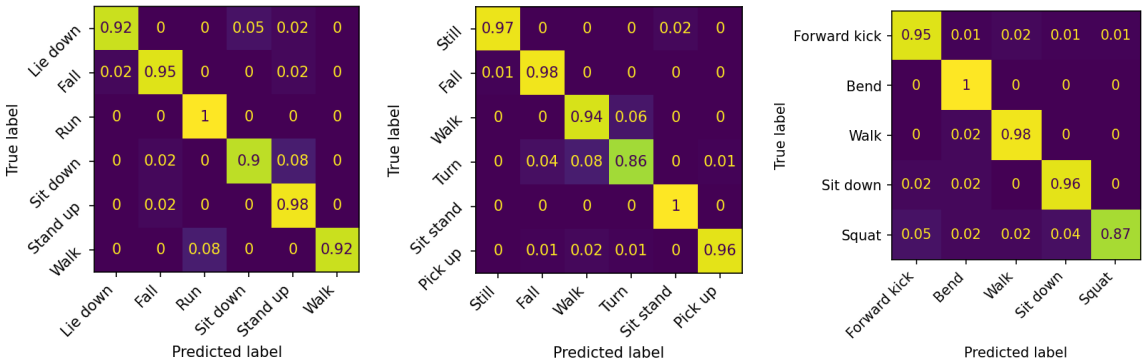


Figure 6.7. Bi-LSTM-based surrogate accuracy versus CSI sampling rate.



(a) (b) (c)

Figure 6.8. Bi-LSTM-based surrogate confusion matrices evaluated on datasets: (b) TAR, (c) JAR, and (d) WiAR.

action duration knowledge beforehand.

Behavioral cloning: This IL algorithm [25] is applicable to real-time adversarial attack [17]. Like GAIL, it uses black-box FGM computed on the surrogate training data, with state-action pairs, $(\mathscr{H}_i^{\ell}, \mathrm{A}_i)$, as input features and labels. A Bi-LSTM-based classifier trains the function $f_{\mathrm{BC}}(\mathscr{H}_i^{\ell})$ in a supervised learning manner. We use downsampled training data and $\ell = 5$ for a fair comparison, matching GAIL's configuration. As in GAIL, behavioral cloning operates without knowledge of the target classifier, future CSI, or action duration.

Random jamming: We evaluate Gaussian-distributed noise jamming, which requires no prior knowledge.

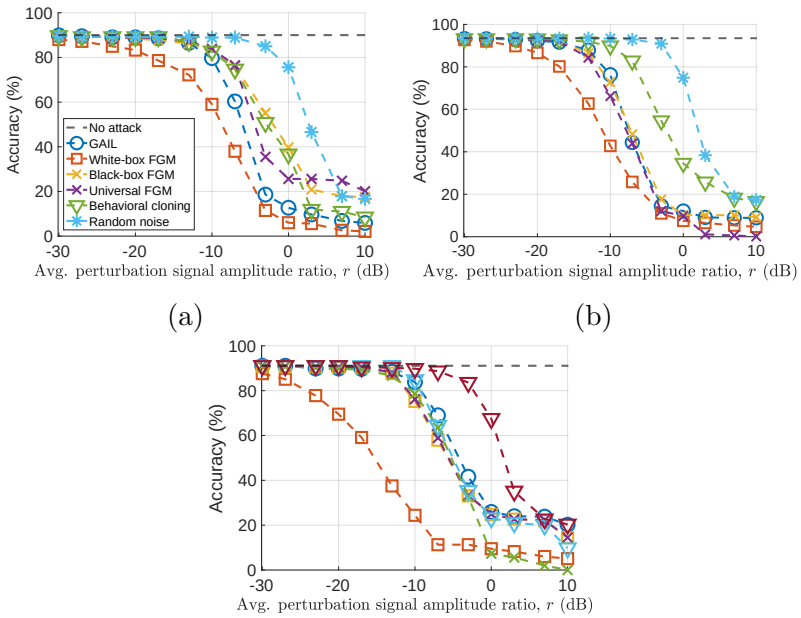


Figure 6.9. GAIL and baseline remote attack scheme results on Bi-LSTM-based target classifier with data: (a) TAR, (b) JAR, and (c) WiAR.

6.5.3 Evaluation Results

Training surrogate model

The surrogate model's accuracy under different downsampling rates is presented in Fig. 6.7. The surrogate accuracy deteriorates when the sampling rate is below 10 Hz, confirming that our chosen rates, 50 Hz for TAR and 80 Hz for JAR, maintain good accuracy. Fig. 6.8a–6.8c presents the surrogate accuracy trained on each downsample dataset. These models 93% average surrogate accuracy across all three datasets (all surrogate accuracies >86% per activity), demonstrating that the Bi-LSTM-based structure serves as a surrogate for computing black-box FGM and GAIL.

Performance to attack Wi-Fi-based HAR classifier

The target classifier accuracy with FGMs under diverse attack schemes and our proposed algorithm is given in Fig. 6.9. The target model employs an identical Bi-LSTM-based architecture and sampling rate as the surrogate model. Across all the datasets,

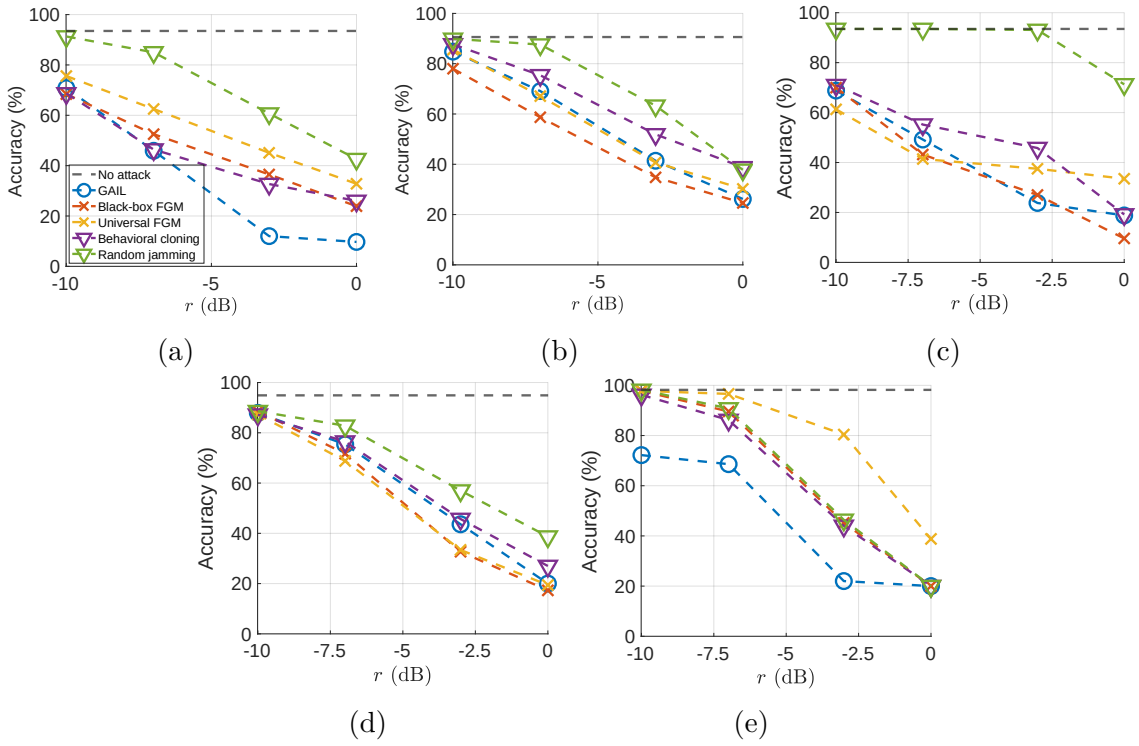


Figure 6.10. GAIL and baseline attack scheme results on different target classifiers with data: (a, b) TAR, (c, d) JAR, and (e) WiAR.

white-box FGM yields the lowest accuracy, while random jamming results in the highest accuracy at equivalent perturbation signal amplitudes. This suggests that detailed knowledge of the target model, including its weights, significantly impacts attack effectiveness. White-box MIMO demonstrates lower efficiency than white-box FGM and performs similarly to black-box FGM across all datasets. Beyond white-box FGM and random jamming, no clear performance hierarchy emerges among the attack schemes. While GAIL outperforms all comparison schemes except white-box FGM on the TAR dataset, it shows comparable performance on the JAR and WiAR datasets. Nonetheless, these results are promising for GAIL and behavioral cloning since both schemes require less information than other comparison schemes.

Figure 6.10 presents the attack performance of GAIL and the baseline schemes against the target models with architectures different from the surrogate model. GAIL

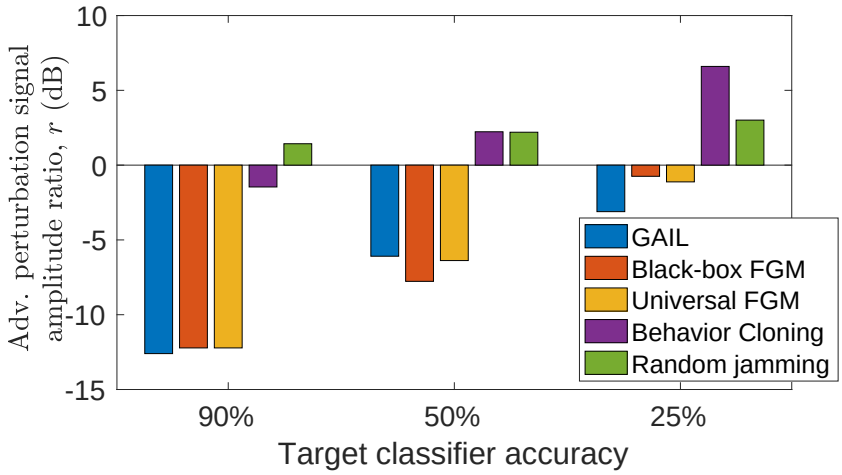
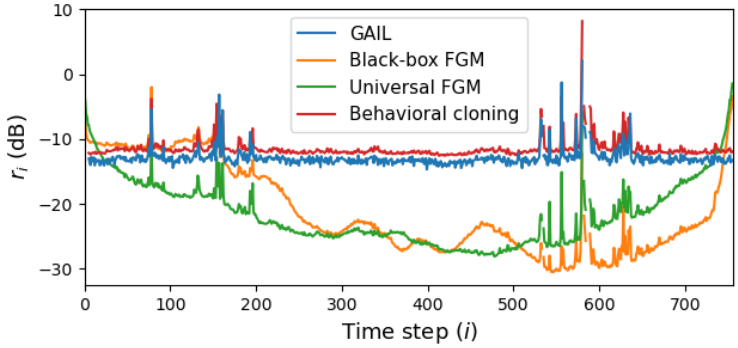


Figure 6.11. r values needed to degrade the accuracy of the Bi-LSTM-based target classifier across varying levels when training and test data are collected from different environments using the JAR dataset.

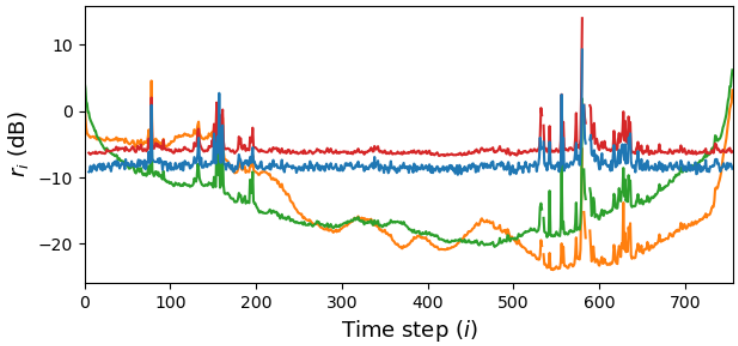
Table 6.6. Ratio (%) of LTF symbols larger than the threshold with degraded accuracy in TAR dataset.

Degraded accuracy	90%	50%	25%
GAIL	1.0	12.8	52.5
White-box FGM	0.0	2.5	9.5
Black-box FGM	0.7	12.0	20.7
Universal FGM	1.1	16.2	24.7
Behavioral cloning	0.8	26.5	85.5

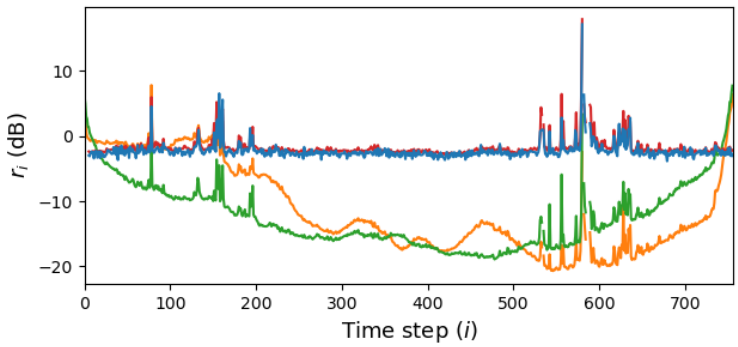
consistently outperforms behavioral cloning across every dataset and r values. Among all the attack schemes, GAIL requires the smallest r to degrade the accuracy of certain target models [14, 15] to 50%, demonstrating our generator's effectiveness in producing perturbation signals that compromise diverse target models. When targeting the non-DL model [21] and STFT feature-based model [54], all the baseline schemes show minimal performance improvement over random jamming, with universal FGM performing worse than random jamming against the model [54]. However, GAIL matches the baseline schemes' performance of against the model [21] and outperforms them against the model [54].



(a)



(b)



(c)

Figure 6.12. Average perturbation signal amplitude ratios r_i over time of one "walk" data in TAR dataset when the target classifier accuracy is degraded to (a) 90%, (b) 50%, and (b) 25%.

Results on unseen environments.

Fig. 6.11 presents the perturbation signal amplitude ratios, r, needed to degrade the target classifier accuracy to 90% and 50% when the training and test environments differ on JAR dataset. The surrogate model is trained on line-of-sight (LOS) data with TX and RX configurations in Figs. 6.6b and 6.6c. Testing is conducted in a non-line-of-sight (NLOS) environment, as shown in Fig. 6.6d. The target model, trained on data from the test environment, achieves 95.2% accuracy without perturbation.

GAIL degrades the target classifier accuracy to 90% and 50% using r values comparable to black-box FGM and universal FGM. In contrast, behavioral cloning requires larger r to reach the same accuracy level. Notably, when degrading accuracy to 50%, behavioral cloning requires perturbation signal amplitude ratios similar to random jamming, highlighting poor generalization to unseen environments. In contrast, GAIL maintains robust degrading accuracy in unseen environments due to its objective function design and regularization.

Impact on Communication Link

The perturbation signal aims to degrade the classifier while minimizing interference with the communication link. Even with the same r, different distributions of perturbation signal amplitudes can produce varying effects on the link. Concentrating the power of the perturbation signals on specific time steps and subcarriers can minimize the impact on the communication link.

To analyze the temporal distribution of perturbation signal power, we compute the average ratios between the perturbation signal and LTF amplitudes at each time step:

$$r_i = \frac{1}{N_{\mathrm{TX}} \times N_{\mathrm{SC}}} \sum_{j=1}^{N_{\mathrm{SC}}} \sum_{k=1}^{N_{\mathrm{TX}}} \frac{|\beta_{ij}^k|}{|x_{ij}^k|}. \tag{6.14}$$

Figure 6.12 illustrates r_i over time for one "walk" activity data in TAR dataset with r values

set to degrade the target classifier accuracy to 90%, 50%, and 25%. Each scheme exhibits distinct peaks, particularly at time steps 130–150 and 550–600. Behavioral cloning and GAIL demonstrate flatter and higher floor amplitude than black-box and universal FGM. Black-box and universal FGM compute adversarial examples by simultaneously processing the entire length's input, allowing the adversary to concentrate power on critical time steps that heavily impact classifier accuracy, as shown by the peaks. In contrast, behavioral cloning and GAIL process only the past five steps and generate output for the current time step. This limited context prevents these schemes from effectively identifying the significant and insignificant time steps, resulting in a higher baseline amplitude ratio across time compared to black-box and universal FGM.

The proportion of CSI where the amplitude ratio $|\beta_{ij}^k|/|x_{ij}^k|$ exceeds a threshold in Table 6.6. The r are adjusted according to the accuracy specified in each column. The threshold is set to 5 dB (0.301), the minimum SNR required for data transfer over wireless links [?]. When the $|\beta_{ij}^k|/|x_{ij}^k|$ exceeds this threshold, data transmission becomes impossible even under ideal channel conditions. We define such instances, represented as the triplet $\{i, j, k\}$, as severely distorted resource. At 90% accuracy, all attack schemes maintain the proportion of severely distorted resources at 1%.

With less accuracy, behavioral cloning yields the highest proportion of severely distorted resources. At 25% accuracy, behavioral cloning produces four times more severely distorted resources than black-box FGM, despite only 18% difference in r ($r = 0.59$ for behavioral cloning vs. $r = 0.50$ for black-box FGM). This disparity stems from the behavioral cloning's higher baseline perturbation signal amplitude, see Fig. 6.12c. While GAIL faces similar elevated baseline amplitude issues, at 50% accuracy, its ratio of severely distorted resources remains comparable to that of black-box FGM and universal FGM, owing to GAIL's ability to achieve this degradation with smaller r. Although GAIL requires larger r than behavioral cloning to achieve 25% accuracy, it still produces fewer severely distorted resources. In summary, our GAIL generates perturbation signals caus-

ing less disruption to the communication link than behavioral cloning, even at similar power levels.

6.6 Conclusion

This work examines remote adversarial attacks against Wi-Fi-based HAR for privacy protection. We implement GAIL, a representative IL algorithm, to develop a perturbation signal generator. Our evaluation encompasses three datasets and six target models, exploring scenarios beyond the unknown target model structures and weight to include cases where adversaries lack synchronization with HAR classifier input. We further examine target model accuracy with the inputs at sampling rates different from the surrogate model. Our GAIL-based algorithm achieves to degrade accuracy to 50% with only a 0.5 dB higher received signal than the other schemes relying on impractical assumptions. Our approach renders data communication impossible in 12.8% of the time-frequency slots, representing a 52% reduction compared to the comparison scheme under the same assumptions.

The algorithm has not been implemented on physical radio hardware. One critical consideration for hardware implementation is the latency of the perturbation generator. If it exceeds the sampling period hardware modifications are necessary. During inference, the policy network generates perturbation signals through 3 FC layers and the computation in (6.12). Our tests show average inference latencies of 8.2 ms on a CPU and 2.4 ms on GTX 2080Ti, both below the algorithm's shortest sampling period, 12.5 ms. These latency measurements suggest the algorithm could be implemented without modifications, practical hardware implementation must account for additional latency, e.g., data transfer between components.

6.7 Appendix. TRPO detailed steps

Policy gradient aims to increase the probabilities of actions that yield higher returns along learner trajectories. The GAIL training process includes a policy gradient to align the policy function closely with expert trajectories. We implement trust region policy optimization (TRPO) as our policy gradient algorithm, which optimizes two key components; the policy function $\pi(\mathrm{A}_i|\mathscr{H}_i^\ell)$ and the value function, $V(\mathscr{H}_i^\ell)$. Following [53], for value function estimation, we employ generalized advantage estimation (GAE) [55]. In kth iteration of GAIL, the detailed TRPO steps executed in line 4 in Algorithm 4 are as follows:

1. Compute temporal difference (TD) error:

   $\delta_i^{V_{\phi_k}} = -C(\mathscr{H}_i^\ell, \mathrm{A}_i) + \gamma V_{\phi_k}(\mathscr{H}_{i+1}) - V_{\phi_k}(\mathscr{H}_i^\ell)$ at all time steps $i \in \{1, 2, \cdots, M\}$

2. Compute advantage values: $\hat{A}_i = \sum_{l=0}^{\infty} (\gamma\lambda_G)^l \delta_{m+l}^{V_{\phi_k}}$ at all time steps $i \in \{1, 2, \cdots, M\}$

3. Update the parameters ϕ of value function $V_\phi(\mathscr{H}_i^\ell)$ to decrease the objective, $K(\phi)$:

   $\phi_{k+1} \leftarrow \phi_k - \nabla_\phi K(\phi)$

$$K(\phi) = \sum_{i=1}^{M} ||V_\phi(\mathscr{H}_i^\ell) - \hat{V}_i||^2$$

$$\text{subject to } \frac{1}{M}\sum_{i=1}^{M} \frac{||V_\phi(\mathscr{H}_i^\ell) - \hat{V}_i||^2}{2\sigma^2} \leq \varepsilon_K,$$

$$\text{where } \hat{V}_i = \sum_{l=0}^{\infty} \gamma^l r_{i+l} \text{ and } \sigma^2 = \frac{1}{M}\sum_{i=1}^{M} ||V_{\phi_k}(\mathscr{H}_i^\ell) - \hat{V}_i||^2.$$

4. Update the parameters θ of policy function π_θ to decrease the objective, $L_{\theta_k}(\theta)$:

$$\theta_{k+1} \leftarrow \theta_k - \nabla_\theta L_{\theta_k}(\theta)$$

$$\text{subject to KL divergence, } \overline{D}_{\mathrm{KL}}^{\theta_k}(\pi_{\theta_k}, \pi_\theta) \leq \varepsilon_\pi$$

$$\text{where } L_{\theta_k}(\theta) = \frac{1}{M}\sum_{i=1}^{M} \frac{\pi_\theta(\mathrm{A}_i|\mathscr{H}_i^\ell)}{\pi_{\theta_k}(\mathrm{A}_i|\mathscr{H}_i^\ell)}\hat{A}_i$$

$$\overline{D}_{\mathrm{KL}}^{\theta_k}(\pi_{\theta_k}, \pi_\theta) = \frac{1}{M}\sum_{i=1}^{N} D_{\mathrm{KL}}(\pi_{\theta_k}(\cdot|\mathscr{H}_i^\ell)||\pi_\theta(\cdot|\mathscr{H}_i^\ell)).$$

6.8 Acknowledgements

Chapter 6, in part, has been double-blindly submitted for publication of the material as it may appear in proceeding of Kim, B., Panchagatti, A., Zhang, X., and Gerstoft, P. “Remote Adversarial Attacks against Wi-Fi-based HAR for Privacy Protection”. The dissertation author was the primary investigator and author of this paper. The coauthors listed in this publication directed and supervised the research.

6.9 References

[1] X. Wang, X. Wang, and S. Mao, "Deep convolutional neural networks for indoor localization with csi images," IEEE Trans. Netw. Sci. Eng., vol. 7, no. 1, pp. 316–327, 2018.

[2] X. Tong, H. Li, X. Tian, and X. Wang, "Wi-fi localization enabling self-calibration," IEEE/ACM Trans. Netw., vol. 29, no. 2, pp. 904–917, 2021.

[3] F. Meneghello, M. Rossi, and F. Restuccia, "Deepcsi: Rethinking wi-fi radio fingerprinting through mu-mimo csi feedback deep learning," in Proc. IEEE ICDCS, pp. 1062–1072, IEEE, 2022.

[4] Y. Chapre, A. Ignjatovic, A. Seneviratne, and S. Jha, "Csi-mimo: Indoor wi-fi fingerprinting system," in Proc. IEEE LCN, pp. 202–209, IEEE, 2014.

[5] Y. Wang, Y. Ren, Y. Chen, and J. Yang, "Wi-mesh: A wifi vision-based approach for 3d human mesh construction," in Proc. ACM Sensys, pp. 362–376, 2022.

[6] M. Muaaz, A. Chelli, A. A. Abdelgawwad, A. C. Mallofré, and M. Pätzold, "Wiwehar: Multimodal human activity recognition using wi-fi and wearable sensing modalities," IEEE Access, vol. 8, pp. 164453–164470, 2020.

[7] W. Xi, D. Huang, K. Zhao, Y. Yan, Y. Cai, R. Ma, and D. Chen, "Device-free human activity recognition using csi," in Proc. ACM CSAR Workshop, pp. 31–36, 2015.

[8] M. Muaaz, A. Chelli, M. W. Gerdes, and M. Pätzold, "Wi-sense: A passive human activity recognition system using wi-fi and convolutional neural network and its integration in health information systems," Annals of Telecommunications, vol. 77, no. 3, pp. 163–175, 2022.

[9] V. D. Kumar, P. Rajesh, K. Polat, F. Alenezi, S. A. Althubiti, and A. Alhudhaif, "Wi-fi signal-based human action acknowledgement using channel state information with cnn-lstm: a device less approach," Neural Computing and Applications, vol. 34, no. 24, pp. 21763–21775, 2022.

[10] J. Yang, H. Zou, and L. Xie, "Securesense: Defending adversarial attack for secure device-free human activity recognition," IEEE Trans. Mobile Comput., vol. 23, no. 1, pp. 823–834, 2022.

[11] L. Xu, X. Zheng, X. Li, Y. Zhang, L. Liu, and H. Ma, "Wicam: Imperceptible adversarial attack on deep learning based wifi sensing," in Proc. IEEE SECON, pp. 10–18, 2022.

[12] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," arXiv preprint arXiv:1312.6199, 2013.

[13] N. Papernot, P. McDaniel, and I. Goodfellow, "Transferability in machine learning: from phenomena to black-box attacks using adversarial samples," arXiv preprint arXiv:1605.07277, 2016.

[14] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," IEEE Commun. Mag., vol. 55, no. 10, pp. 98–104, 2017.

[15] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "Wifi csi based passive human activity recognition using attention based blstm," IEEE Trans. Mobile Comput., vol. 18, no. 11, pp. 2714–2724, 2018.

[16] M. S. Islam, M. K. A. Jannat, M. N. Hossain, W.-S. Kim, S.-W. Lee, and S.-H. Yang, "Stc-nlstmnet: an improved human activity recognition method using convolutional neural network with nlstm from wifi csi," Sensors, vol. 23, no. 1, p. 356, 2022.

[17] Y. Gong, B. Li, C. Poellabauer, and Y. Shi, "Real-time adversarial attacks," in Proc. IJCAI, pp. 4672–4680, 2019.

[18] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," Machine learning, vol. 8, pp. 229–256, 1992.

[19] A. Baha'A, M. M. Almazari, R. Alazrai, and M. I. Daoud, "A dataset for wi-fi-based human activity recognition in line-of-sight and non-line-of-sight indoor environments," Data in Brief, vol. 33, p. 106534, 2020.

[20] L. Guo, L. Wang, C. Lin, J. Liu, B. Lu, J. Fang, Z. Liu, Z. Shan, J. Yang, and S. Guo, "Wiar: A public dataset for wifi-based activity recognition," IEEE Access, vol. 7, pp. 154935–154945, 2019.

[21] M. K. A. Jannat, M. S. Islam, S.-H. Yang, and H. Liu, "Efficient wi-fi-based human activity recognition using adaptive antenna elimination," IEEE Access, 2023.

[22] B. Sheng, F. Xiao, L. Sha, and L. Sun, "Deep spatial–temporal model based cross-scene action recognition using commodity wifi," IEEE Internet of Things J., vol. 7, no. 4, pp. 3592–3601, 2020.

[23] S. Ding, Z. Chen, T. Zheng, and J. Luo, "Rf-net: A unified meta-learning framework for rf-enabled one-shot human activity recognition," in Proc. ACM Sensys, pp. 517–530, 2020.

[24] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in Proc. ICLR, 2015.

[25] F. Torabi, G. Warnell, and P. Stone, "Behavioral cloning from observation," in Proc. IJCAI, pp. 4950–4957, 2018.

[26] S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in Proc. AISTATS, pp. 627–635, PMLR, 2011.

[27] B. D. Ziebart, A. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning," in Proc. AAAI, vol. 8, pp. 1433–1438, Chicago, IL, USA, 2008.

[28] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in Proc. ICML, p. 1, 2004.

[29] U. Syed, M. Bowling, and R. E. Schapire, "Apprenticeship learning using linear programming," in Proc. ICML, pp. 1032–1039, 2008.

[30] S. Zhou, W. Zhang, D. Peng, Y. Liu, X. Liao, and H. Jiang, "Adversarial wifi sensing for privacy preservation of human behaviors," IEEE Commun. Lett., vol. 24, no. 2, pp. 259–263, 2019.

[31] W. Zhang, S. Zhou, D. Peng, L. Yang, F. Li, and H. Yin, "Understanding and modeling of wifi signal-based indoor privacy protection," IEEE Internet of Things J., vol. 8, no. 3, pp. 2000–2010, 2020.

[32] P. Huang, X. Zhang, S. Yu, and L. Guo, "IS-WARS: Intelligent and stealthy adversarial attack to wi-fi-based human activity recognition systems," IEEE Trans. Dependable and Secure Comput., vol. 19, no. 6, pp. 3899–3912, 2021.

[33] Y. Zhou, H. Chen, C. Huang, and Q. Zhang, "Wiadv: Practical and robust adversarial attack against wifi-based gesture recognition system," Proc. ACM IMWUT, vol. 6, no. 2, pp. 1–25, 2022.

[34] J. Liu, Y. He, C. Xiao, J. Han, and K. Ren, "Time to think the security of wifi-based behavior recognition systems," IEEE Trans. on Dependable and Secure Comput., 2023.

[35] Y. Xie, R. Jiang, X. Guo, Y. Wang, J. Cheng, and Y. Chen, "Universal targeted adversarial attacks against mmwave-based human activity recognition," in Proc. IEEE INFOCOM, pp. 1–10, IEEE, 2023.

[36] C. Li, M. Xu, Y. Du, L. Liu, C. Shi, Y. Wang, w. Liu, and Y. Chen, "Practical adversarial attack on wifi sensing through unnoticeable communication packet perturbation," in Proc. ACM MobiCom, pp. 373–387, 2024.

[37] Y. Zhang, Y. Zheng, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Widar3.0: Zero-effort cross-domain gesture recognition with wi-fi," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 11, pp. 8671–8688, 2021.

[38] Z. Liu, C. Xu, E. Sie, G. Singh, and D. Vasisht, "Exploring practical vulnerabilities of machine learning-based wireless systems," in Proc. USENIX NSDI, pp. 1801–1817, 2023.

[39] F. Xiao, Y. Huang, Y. Zuo, W. Kuang, and W. Wang, "Over-the-air adversarial attacks on deep learning wi-fi fingerprinting," IEEE Internet of Things J., vol. 10, no. 11, pp. 9823–9835, 2023.

[40] A. Bahramali, M. Nasr, A. Houmansadr, D. Goeckel, and D. Towsley, "Robust adversarial attacks against dnn-based wireless communication systems," in Proc. ACM CCS, pp. 126–140, 2021.

[41] J.-W. Chang, K. Sun, N. Heydaribeni, S. Hidano, X. Zhang, and F. Koushanfar, "Magmaw: Modality-agnostic adversarial attacks on machine learning-based wireless communication systems," arXiv preprint arXiv:2311.00207, 2023.

[42] Z. Wang, W. Liu, and H.-M. Wang, "Wireless universal adversarial attack and defense for deep learning-based modulation classification," IEEE Commun. lett., 2024.

[43] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, "Universal adversarial perturbations," in Proc. IEEE CVPR, pp. 1765–1773, 2017.

[44] Z. Li, Y. Wu, J. Liu, Y. Chen, and B. Yuan, "Advpulse: Universal, synchronization-free, and targeted audio adversarial attacks via subsecond perturbations," in Proc. ACM CCS, pp. 1121–1134, 2020.

[45] Y. Xie, Z. Li, C. Shi, J. Liu, Y. Chen, and B. Yuan, "Enabling fast and universal audio adversarial attack using generative model," in Proc. AAAI, vol. 35, pp. 14129–14137, 2021.

[46] Y. Wang, H. Guo, G. Wang, B. Chen, and Q. Yan, "Vsmask: Defending against voice synthesis attack via real-time predictive perturbation," in Proc. ACM WiSec, pp. 239–250, 2023.

[47] M. LLC, "[OpenWrt] Broadcom wireless," 2022.

[48] Y. Kim and H. Ling, "Human activity classification based on micro-doppler signatures using a support vector machine," IEEE Trans. Geosci. Remote Sens., vol. 47, no. 5, pp. 1328–1337, 2009.

[49] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity wifi," in Proc. MobiCom, 2015.

[50] J. Zhang, F. Wu, B. Wei, Q. Zhang, H. Huang, S. W. Shah, and J. Cheng, "Data augmentation and dense-lstm for human activity recognition using wifi signal," IEEE Internet of Things J., vol. 8, no. 6, pp. 4628–4641, 2020.

[51] P. Khan, B. S. K. Reddy, A. Pandey, S. Kumar, and M. Youssef, "Differential channel-state-information-based human activity recognition in iot networks," IEEE Internet of Things J., vol. 7, no. 11, pp. 11290–11302, 2020.

[52] J. Ho and S. Ermon, "Generative adversarial imitation learning," Proc. NeurIPS, vol. 29, 2016.

[53] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in Proc. ICML, pp. 1889–1897, 2015.

[54] C. Zhang and W. Jiao, "Imgfi: A high accuracy and lightweight human activity recognition framework using csi image," IEEE Sensors J., 2023.

[55] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-dimensional continuous control using generalized advantage estimation," arXiv preprint arXiv:1506.02438, 2015.

Chapter 7

Conclusion

This dissertation explored the application of deep learning techniques in wireless communication systems, focusing on two primary areas: modulation classification of practical OFDM signals for spectrum sensing, and remote adversarial attacks targeting Wi-Fi-based Human Activity Recognition (HAR) for privacy protection.

7.1 Modulation classification of practical OFDM signals for spectrum sensing

Intelligent spectrum sensing frequently targets unconstrained wireless signals instead of signals of specific communication protocols. Accordingly, preamble or control channel information, which contains modulation, is usually unavailable. To address this challenge, Chapters 2, 3, and 4 present solutions for modulation classification of practical OFDM signals, under scenarios with absent protocol information.

Chapter 2 introduces a deep learning-based modulation classification method to handle practical OFDM signals without symbol-level synchronization. By developing a preprocessing algorithm to enhance classification performance in scenarios where the boundaries of OFDM symbols are unknown, the proposed solution robustly classifies modulation against synchronization errors in realistic spectrum sensing environments.

Chapter 3 and 4 further advance modulation classification by incluidng OFDM pa-

rameter estimation and modulation classification for modern wireless standards, including Wi-Fi 6 and 5G signals. Two essential OFDM parameters, subcarrier spacing and cyclic prefix (CP) length, are reliably estimated using cyclostationarity analysis. These parameter estimates enable the proposed preprocessing of OFDM signals. Evaluations conducted using simulated PHY-layer signals and software-defined radio (SDR) implementations validate the practical effectiveness of the proposed methodology. These experiments confirms that the method can accurately estimate OFDM parameters and reliably classify modulation schemes in realistic, over-the-air environments, achieving at least 97% classification accuracy provided the signal-to-noise ratio (SNR) exceeded protocol-specified thresholds.

7.2 Remote Adversarial Attacks against Wi-Fi-based HAR for Privacy Protection

Chapters 5 and 6 addresses privacy concerns raised by Wi-Fi-based HAR systems by exploring remote adversarial attacks using perturbation signals transmitted from user devices. Specifically, this dissertation tackles challenges posed by unknown future channel state information (CSI), a challenge to gradient-based adversarial methods.

Chapter 5 proposes a real-time adversarial attack framework by degrading the accuracy of Wi-Fi-based HAR systems through manipulation of Wi-Fi pilot signals. To overcome limitations inherent to gradient-based attacks, which require prior knowledge of future CSI, a Generative Adversarial Imitation Learning (GAIL)-based algorithm was developed. This novel algorithm effectively mimicked gradient-based attacks without explicit knowledge of future channel conditions and successfully degrades the accuracy of deep learning-based HAR models.

Building upon this foundation, Chapter 6 comprehensively evaluates the versatility of the proposed GAIL-based adversarial attacks. Extensive experiments were conducted across diverse environments and against multiple target models, including deep learning

and conventional non-deep learning classifiers. Evaluation scenarios extended to conditions where essential parameters such as sampling rates, input sequence durations, and target model configurations were unknown. These comprehensive evaluations demonstrated the its broad applicability and practical relevance of the proposed adversarial method.

7.3 Future work

Several directions remain promising for future research based on the outcomes of this dissertation. First, since the proposed modulation classification framework is currently limited to single-input single-output (SISO) scenarios, future studies could extend this methodology to more complex wireless environments, such as OFDM-MIMO and millimeter-wave (mmWave) communication systems. Additionally, modulation classification in orthogonal frequency division multiple access (OFDMA), where multiple modulation schemes coexist simultaneously, provides another potential area of advancement. Furthermore, identification and classification of other critical wireless transmission parameters, such as coding schemes, could also be explored in future research.

Regarding adversarial attacks, several improvements can be considered. Firstly, the current perturbation signal generator does not explicitly determine the amplitude of the perturbation signals, leaving potential optimizations for attack efficiency and stealth unexplored. Secondly, no mechanism currently exists to minimize the negative impact of perturbation signals on the underlying communication link quality. Physical implementation of the proposed adversarial attacks on actual radio hardware has not yet been performed. Practical deployment and rigorous testing in diverse real-world wireless environments would substantially enhance the robustness and validity of the adversarial approach presented in this dissertation.

Overall, this dissertation provides a solid foundation for leveraging deep learning

to advance wireless signal classification accuracy and to safeguard user privacy against emerging sensing threats. The findings and methodologies established herein serve as a valuable stepping stone toward future research and technological innovation in wireless communications and security.