IEEE INFOCOM 2023 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) | 978-1-6654-9427-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/INFOCOMWKSHPS57453.2023.10225824

Adversarial Attack and Defense for WiFi-based Apnea Detection System

Harshit Ambalkar[1], Tianya Zhao[2], Xuyu Wang[2], Shiwen Mao[3]
[1]Department of Computer Science, California State University, Sacramento, CA 95819, USA
[2]Knight Foundation School of Computing and Information Sciences, Florida International University, Miami, FL 33199, USA
[3]Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849, USA
Emails: [1]hambalkar@csus.edu, [2]{tzhao010, xuywang}@fiu.edu, [3]smao@ieee.org

***Abstract*—WiFi sensing systems have gained enormous interest in extensive areas, including vital sign monitoring. By using deep neural networks (DNNs), WiFi sensing systems can achieve high performance. However, the security and vulnerability of DNNs under adversarial attack would greatly affect the WiFi sensing performance. In this paper, we develop a DNN-based apnea detection system using WiFi channel state information (CSI) and evaluate its robustness under three different attacks. The experimental results show that adversarial attacks can significantly impact the model performance, and the defense scheme (i.e. adversarial training) can improve the system robustness.**

***Index Terms*—WiFi sensing, adversarial examples, and vital sign monitoring.**

I. Introduction

As we enter the era of ubiquitous sensing, the deployment of Internet of Things (IoT) devices in daily life has become increasingly widespread, aiding individuals in a variety of domains. As the importance of health concerns continues to rise, there has been a growing focus on the field of vital sign monitoring. Compared with traditional wearable devices and radar systems, WiFi-based wireless sensing offers a more convenient and ubiquitous solution for vital sign monitoring [1].

Recent WiFi-based vital sign monitoring systems utilize deep neural networks (DNNs) as an effective feature extractor and classifier to improve the system performance [2]. However, it is acknowledged that the open nature of the WiFi medium and the potential vulnerabilities of DNNs may raise concerns about the security and robustness of such effective systems. In the field of computer vision, DNN models have been proven to be vulnerable to adversarial attacks, where carefully crafted images can significantly decrease model performance [3]. Similarly, WiFi-based sensing applications such as gesture recognition and human activity recognition have been shown to be susceptible to adversarial examples [4], [5].

To the best of our knowledge, this paper is the first to investigate the impact of adversarial attacks and defense on *WiFi-based vital sign monitoring systems*, where we evaluate three white-box adversarial attack methods and experimentally assess the performance of the proposed system with adversarial training. Specifically, we first exploit WiFi channel state information (CSI) signal to monitor abnormal breath. Second, we train a convolutional neural network (CNN) model to detect apnea. Then, we test the CNN model with adversarial examples and then retrain model with adversarial examples to examine the robustness.

II. System Design

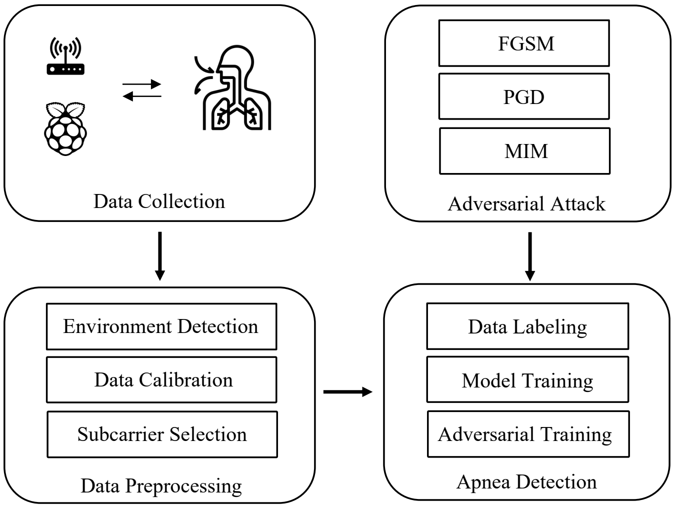

Fig. 1. System architecture.

Fig. 1 depicts the system architecture for WiFi-based apnea detection. First, we deploy a Raspberry Pi with a modified Nexmon firmware patch and a WiFi router to collect realtime WiFi CSI data. Although breathing actions are subtle, they influence CSI dynamic paths, thus leading to different complex CSI values over time. Second, data preprocessing techniques help to extract breathing profiles for further classification. By determining the mean absolute deviation of the CSI amplitude difference data in a sliding window, a threshold-based method is utilized to determine whether a user is in a stationary state and can monitor the respiration. The Hampel filter and band pass filter are used to remove outliers and inconsistencies. Additionally, we also apply subcarrier selection to further enhance the dependability of CSI magnitude data, as different subcarriers have varying magnitudes, resulting in different sensitivities for breathing signals. Third, we attack the used CNN model with three methods and train it with adversarial examples to defend against these attacks.

978-1-6654-9427-4/23/$31.00 ©2023 IEEE

III. Experimental Results

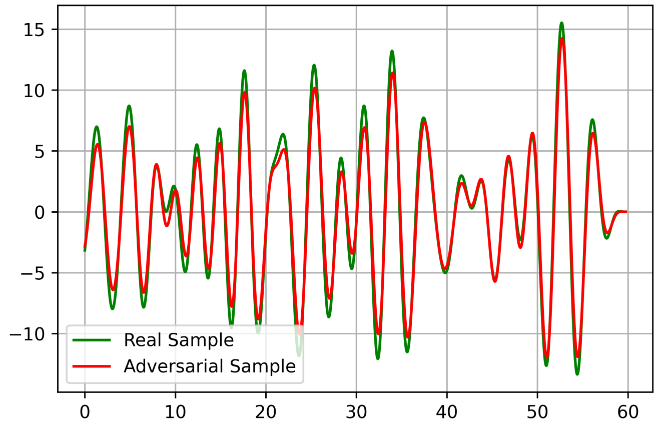

Fig. 2. Real and adversarial examples.

In this paper, we evaluate three adversarial attack methods for apnea detection using WiFi. Fig. 2 presents the adversarial breathing examples generated through the Projected Gradient Descent (PGD) attack. Intuitively, these adversarial examples are perceptually similar to the original samples, posing a significant challenge for apnea classification.

Fig. 3 shows the classification results. The perturbation is calculated by different methods, which measures the dissimilarity between adversarial samples and the original samples. As the perturbation level increases, the changes in the data become larger, resulting in a decrease in classification accuracy. Our model demonstrates a higher accuracy with 0.85 on the original dataset. Furthermore, we consider three white-box attack methods as follows.

The Fast Gradient Sign Method (FGSM) is a one-step attack algorithm with a low computational complexity, which is a relatively mild attack. We can see that the decrease in accuracy is approximately 0.45 at the perturbation level with 0.8. Furthermore, the PGD attack is an iterative version of the FGSM to enhance the attack performance [5]. The PGD attack can significantly decrease the classification accuracy to 0.29 when the perturbation level ϵ is set to 0.8. However, the PGD algorithm for generating adversarial examples has been found to be problematic in terms of transferability. To mitigate this issue, the Momentum Iterative Method (MIM) has been proposed, which incorporates the gradient information from previous iterations to enhance the update of the perturbation. The model attacked by the MIM achieves the lowest accuracy of 0.25 with the same perturbation level.

In order to enhance the generalization capability of the model and reduce its vulnerability to adversarial attacks, we employ adversarial training as a defensive approach. This strategy is distinct from other defense techniques in that it is straightforward and primarily concentrates on enhancing the robustness of the models. This is accomplished by incorporating adversarial examples into the training dataset and modifying the model's parameters to improve classification accuracy. For the above three attack methods, the adversarial-trained model can significantly enhance the classification accuracy. Specifically, the MIM adversarial training model has been found to be the most effective in increasing the accuracy of the model. The implementation of this model results in a significant improvement in the accuracy from 0.25 to 0.51.

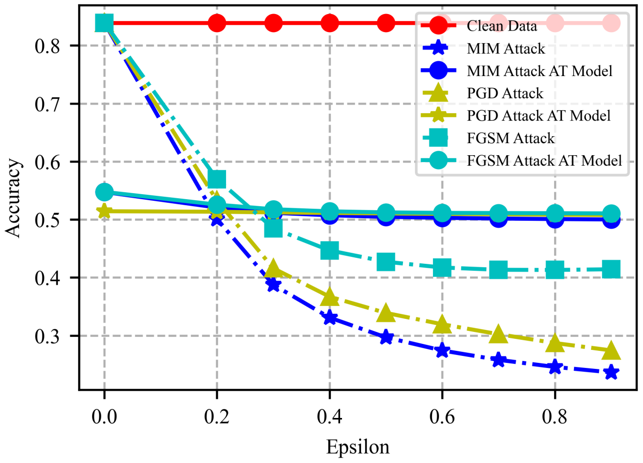

Fig. 3. Attack and defense results.

IV. Conclusion and Future Work

In this paper, we studied the adversarial attack and defense on the WiFi-based apnea detection system. For vital sign monitoring, our results showed adversarial attacks can significantly decrease classification accuracy for apnea detection, and to a certain extent adversarial training can provide the resistance to these attacks. In future work, we plan to explore new wireless sensing techniques (e.g., mmWave radar, RFID) to study the practical threat of adversarial attacks on vital sign monitoring. Also, we will develop new defense methods to improve the robustness of our system.

Acknowledgement

This work is supported in part by the NSF (CNS-2107190, CNS-2105416, CNS-2321763, and CNS-2107164).

References

[1] X. Wang, C. Yang, and S. Mao, ”On CSI-based vital sign monitoring using commodity WiFi,” *ACM Transactions on Computing for Healthcare*, vol. 1, no. 3, pp. 1-27, 2020.

[2] J. Hu, J. Yang, J. -B. Ong, D. Wang and L. Xie, ”ResFi: Wi-Fi-Enabled Device-Free Respiration Detection Based on Deep Learning,” *2022 IEEE 17th International Conference on Control & Automation (ICCA)*, Naples, Italy, pp. 510-515, 2022.

[3] N. Carlini and D. Wagner, ”Towards Evaluating the Robustness of Neural Networks,” *2017 IEEE Symposium on Security and Privacy (SP)*, San Jose, CA, USA, pp. 39-57, 2017.

[4] Y. Zhou, H. Chen, C. Huang and Q. Zhang, ”WiAdv: Practical and Robust Adversarial Attack against WiFi-based Gesture Recognition System,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 2, pp. 1-25, 2022.

[5] H. Ambalkar, X. Wang, and S. Mao, ”Adversarial Human Activity Recognition Using Wi-Fi CSI,” *2021 IEEE Canadian Conference On Electrical And Computer Engineering (CCECE)*, Virtual Conference, pp. 1-5, 2021.