Stanford & Oxbridge Journal of Social Science and Cognition Insight (SOJ-SSCI)

https://stanfordgroup.org/

Vol 5, Issue 2, 2025

MEMBERSHIP INFERENCE ATTACK AND DEFENCE FOR WIRELESS SIGNAL CLASSIFIERS WITH DEEP LEARNING

Mr. Yakoob[1], SD. Sameera[2], K. Madhavi[3]

[1]Assistant professor, Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

[2,3]UG Students,Department of CSE, Princeton College of engineering and technology for women Narapally vijayapuri colony ghatkesar mandal, Pin code-500088

ABSTRACT

A novel over-the-air membership inference attack (MIA) method is introduced to extract sensitive information from a wireless signal classifier. Machine learning techniques are widely employed to categorize wireless signals, particularly in tasks like PHY-layer authentication. The MIA, as a form of adversarial machine learning attack, aims to determine if a given signal was part of the training data for a target classifier. This information, comprising waveform, channel, and device attributes, if exposed, could be exploited by malicious actors to exploit vulnerabilities in the underlying ML model, such as compromising PHY-layer authentication. A key obstacle in implementing the over-the-air MIA is the inherent variability in received signals and resulting RF fingerprints due to channel conditions. To address this, the attacker first constructs a surrogate classifier based on observed spectrum data before executing the black-box MIA attack on this classifier. Both simulation-based and real-world over-the-air software-defined radio (SDR) experiments confirm the efficacy of the MIA in reliably inferring signals used to train the target classifier, potentially revealing radio and channel details. Consequently, a proactive defense strategy is devised against the MIA, involving the creation of a shadow MIA model to deceive the attacker. This defensive approach effectively reduces MIA accuracy and mitigates information leakage from the wireless signal classifier, all without compromising signal classification accuracy.

I.INTRODUCTION

Wireless signal classifiers play a pivotal role in various applications, particularly in wireless communication systems where accurate classification is essential for tasks like authentication and security.

With the advent of machine learning, particularly deep learning, these classifiers have become increasingly sophisticated, enabling them to effectively categorize wireless signals based on various attributes such as waveform, channel characteristics, and device signatures. However, the rise of machine learning also introduces new security challenges, one of which is the susceptibility to membership inference attacks (MIA).

Membership inference attacks pose a significant threat to the privacy and security of wireless signal classifiers. These attacks aim to extract sensitive information by determining whether a specific signal was part of the training dataset used to build the classifier. If successful, attackers can exploit this information to identify vulnerabilities in the classifier, potentially compromising its functionality and integrity. Despite the importance of defending against such attacks, mitigating MIA in the context of wireless signal classifiers, particularly those based on deep learning, remains a challenging task.

The proposed project focuses on developing and analyzing membership inference attacks and defense mechanisms tailored specifically for wireless signal classifiers built on deep learning architectures. By leveraging the power of deep learning, these classifiers can achieve high accuracy in signal classification tasks, but they also become more susceptible to adversarial attacks due to their complex and nonlinear nature. Understanding the vulnerabilities of deep learning-based classifiers to MIA is crucial for ensuring the security and privacy of wireless communication systems.

In this project, we aim to investigate various aspects of membership inference attacks on deep learning-based wireless signal classifiers, including their feasibility, effectiveness, and potential countermeasures. By conducting both theoretical analysis and practical experiments using real-world datasets and simulation environments, we seek to gain insights into the vulnerabilities of these classifiers and develop robust defense strategies to mitigate the risks posed by MIA. Ultimately, the outcomes of this research endeavor will contribute to enhancing the security and resilience of wireless communication systems against emerging threats in the era of machine learning.

II.EXISTING SYSTEM

In the existing system, wireless signal classifiers are typically built using traditional machine learning techniques or shallow learning models. These classifiers rely on handcrafted features extracted from wireless signals, such as signal strength, frequency spectrum, and modulation characteristics. While these classifiers can achieve reasonable accuracy in signal classification tasks, they often lack robustness against sophisticated attacks, such as membership inference attacks (MIA). The vulnerabilities of these classifiers to MIA stem from their reliance on manually engineered features, which may not adequately capture the underlying complexities of wireless signals.

Disadvantages:

1.Lack of Robustness: Traditional machine learning classifiers may lack robustness against adversarial attacks, including membership inference attacks, due to their reliance on handcrafted features that may not capture the full complexity of wireless signals.

2.Limited Performance: Shallow learning models may struggle to achieve high accuracy in classifying complex and diverse wireless signals, leading to suboptimal performance in real-world scenarios.

III.PROPOSED SYSTEM

The proposed system aims to enhance the security and robustness of wireless signal classifiers by leveraging deep learning techniques. Deep learning architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated superior performance in various signal processing tasks, including image classification and natural language processing. By adopting deep learning for wireless signal classification, the proposed system seeks to overcome the limitations of traditional classifiers and improve accuracy, while also addressing the vulnerabilities to membership inference attacks.

Advantages:

1.Enhanced Accuracy: Deep learning models have the potential to achieve higher accuracy in classifying wireless signals compared to traditional machine learning classifiers, thanks to their ability to automatically learn complex patterns and features from raw data.

2.Robustness to Attacks: Deep learning models, particularly those based on neural networks, are inherently more robust against adversarial attacks, including membership inference attacks. The learned representations in deep neural networks capture intricate features of wireless signals, making it challenging for attackers to exploit vulnerabilities in the classifier.

3.Adaptability to Diverse Signals: Deep learning models can adapt to a wide range of wireless signals, including those with varying characteristics and environmental conditions. This adaptability allows the proposed system to handle diverse signal types and scenarios, making it suitable for real-world deployment in wireless communication systems.

IV.MODULES

- Data Preprocessing Module: This module involves preprocessing the wireless signal data, including cleaning, filtering, and transforming the raw signal data into a format suitable for deep learning models.
- Deep Learning Model Training Module: In this module, deep learning models such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs) are trained on the preprocessed signal data to classify different types of wireless signals accurately.
- Membership Inference Attack Module: This module implements algorithms and techniques to launch membership inference attacks against the trained deep learning models. It involves analyzing the model's behavior to infer whether specific signals were part of the training dataset.
- Adversarial Training Module: This module focuses on enhancing the robustness of the deep learning models against membership inference attacks. It involves training the models with adversarial examples generated to mimic potential attacks, thereby improving the model's resilience to such attacks.
- Evaluation and Testing Module: This module assesses the performance of both the deep learning models and the defense mechanisms against membership inference attacks. It involves conducting comprehensive evaluations using metrics such as accuracy, precision, recall, and F1-score.

- User Interface Module (Optional): This module provides a user interface for users to interact with the system, visualize the results of the attack and defense mechanisms, and configure parameters for training and testing.
- Deployment Module: Once the models and defense mechanisms are trained and tested, this module involves deploying the system in real-world scenarios, such as wireless communication networks, to evaluate its effectiveness and performance in practical environments.

V.CONCLUSION AND FUTURE SCOPE

In conclusion, our project on "Membership Inference Attack and Defence for Wireless Signal Classifiers with Deep Learning" has shed light on the security vulnerabilities inherent in deep learning-based classifiers used in wireless signal processing. Through the implementation of membership inference attacks and defense mechanisms, we have gained valuable insights into the potential risks posed by adversarial machine learning in wireless communication systems. Our findings underscore the importance of integrating robust security measures into the design and deployment of deep learning models for wireless signal classification.

Despite the susceptibility of deep learning models to membership inference attacks, our research has also demonstrated the efficacy of proactive defense mechanisms in mitigating these threats. By employing strategies such as adversarial training and shadow model construction, we have successfully reduced the accuracy of membership inference attacks and prevented information leakage from signal classifiers. These defense mechanisms represent promising avenues for enhancing the security and privacy of wireless communication systems.

Looking ahead, there are several opportunities for further research and development in this field. Firstly, advanced defense techniques should be explored to bolster the resilience of deep learning models against membership inference attacks. Real-world deployments and evaluations of these defense mechanisms are crucial to assessing their effectiveness in practical wireless communication networks. Additionally, novel adversarial training strategies and privacy-preserving techniques can be investigated to

enhance the robustness and privacy of wireless signal classifiers.

Collaboration with standardization bodies and industry stakeholders is essential to integrate security and privacy considerations into wireless communication standards and protocols. By incorporating secure-by-design principles into future wireless networks, we can ensure the adoption of resilient security measures that safeguard user privacy and data confidentiality. Overall, by pursuing these avenues of research, we can contribute to the development of secure and trustworthy wireless communication systems in an increasingly interconnected world.

VI.REFERENCES

1. T. Erpek, T. O'Shea, Y. E. Sagduyu, Y. Shi and T. C. Clancy, "Deep learning for wireless communications" in Development and Analysis of Deep Learning Architectures, Berlin, Germany:Springer, 2020.

2. Y. E. Sagduyu et al., "When wireless security meets machine learning: Motivation challenges and research directions", 2020.

3. D. Adesina, C. C. Hsieh, Y. E. Sagduyu and L. Qian, "Adversarial machine learning in wireless communications using RF data: A review", 2020.

4. Y. Shi, Y. E. Sagduyu, T. Erpek, K. Davaslioglu, Z. Lu and J. Li, "Adversarial deep learning for cognitive radio security: Jamming attack and defense strategies", *Proc. IEEE Int. Conf. Commun. Workshop*, pp. 1-6, 2018.

5. T. Erpek, Y. E. Sagduyu and Y. Shi, "Deep learning for launching and mitigating wireless jamming attacks", *IEEE Trans. Cogn. Commun. Netw.*, vol. 5, no. 1, pp. 2-14, Mar. 2019.

6. Y. Shi, Y. E. Sagduyu, T. Erpek and M. C. Gursoy, "How to attack and defend 5G radio access network slicing with reinforcement learning", 2021.

7. Y. Shi and Y. E. Sagduyu, "Adversarial machine learning for flooding attacks on 5G radio access network slicing", *Proc. IEEE Int. Conf. Commun. Workshops*, pp. 1-6, 2021.

8. M. Sadeghi and E. G. Larsson, "Adversarial attacks on deep-learning based radio signal classification", *IEEE Commun. Lett.*, vol. 8, no. 1, pp. 213-216, Feb. 2019.

9. M. Sadeghi and E. G. Larsson, "Physical adversarial attacks against end-to-end autoencoder communication systems", *IEEE Commun. Lett.*, vol. 23, no. 5, pp. 847-850, May 2019.

10. B. Kim, Y. E. Sagduyu, K. Davaslioglu, T. Erpek and S. Ulukus, "Over-the-air adversarial attacks on deep learning based modulation classifier over wireless channels", *Proc. 54th Annu. Conf. Inf. Sci. Syst.*, pp. 1-6, 2020.

11. B. Kim, Y. E. Sagduyu, K. Davaslioglu, T. Erpek and S. Ulukus, "Channel-aware adversarial attacks against deep learning-based wireless signal classifiers", 2020.

12. Y. Lin, H. Zhao, Y. Tu, S. Mao and Z. Dou, "Threats of adversarial attacks in DNN based modulation recognition", *Proc. IEEE Conf. Comput. Commun.*, pp. 2469-2478, 2020.

13.B. Kim, Y. E. Sagduyu, K. Davaslioglu, T. Erpek and S. Ulukus, "Adversarial attacks with multiple antennas against deep learning-based modulation classifiers", *Proc. IEEE Global Commun. Conf. Workshops*, pp. 1-6, 2020.

14.B. Kim, Y. E. Sagduyu, T. Erpek, K. Davaslioglu and S. Ulukus, "Channel effects on surrogate models of adversarial attacks against wireless signal classifiers", *Proc. IEEE Int. Conf. Commun.*, pp. 1-6, 2021.

15.B. Manoj, M. Sadeghi and E. G. Larsson, "Adversarial attacks on deep learning based power allocation in a massive MIMO network", 2021.

16.B. Kim, Y. E. Sagduyu, T. Erpek and S. Ulukus, "Adversarial attacks on deep learning based mmWave beam prediction in 5G and beyond", *Proc. IEEE Stat. Signal Process. Workshop*, pp. 590-594, 2021.

17.R. Sahay, C. G. Brinton and D. J. Love, "Ensemble-based wireless receiver architecture for mitigating adversarial interference in automatic modulation classification", 2021.

18.A. Bahramali, M. Nasr, A. Houmansadr, D. Goeckel and D. Towsley, "Robust adversarial attacks against DNN-based wireless communication systems", 2021.

19.J. Yi and A. El Gamal, "Gradient-based adversarial deep modulation classification with data-driven subsampling", 2021.

20.T. Hou, T. Wang, Z. Lu, Y. Liu and Y. E. Sagduyu, "IoTGAN: GAN powered camouflage against machine learning based IoT device identification", *Proc. IEEE Int. Symp. Dyn. Spectrum Access Netw.*, pp. 280-287, 2021.

21.B. Kim, Y. Shi, Y. E. Sagduyu, T. Erpek and S. Ulukus, "Adversarial attacks against deep learning based power control in wireless communications", *Proc. IEEE Global Commun. Conf. Workshops*, pp. 1-6, 2021.

22.Y. E. Sagduyu, Y. Shi and T. Erpek, "IoT network security from the perspective of adversarial deep learning", *Proc. IEEE Int. Conf. Sens. Commun. Netw. Workshop*, pp. 1-9, 2019.

23.S. Bair, M. DelVecchio, B. Flowers, A. J. Michaels and W. C. Headley, "On the limitations of targeted adversarial evasion attacks against deep learning enabled modulation recognition", *Proc. ACM Workshop Wireless Secur. Mach. Learn.*, pp. 25-30, 2019.

24.B. Flowers, R. M. Buehrer and W. C. Headley, "Evaluating adversarial evasion attacks in the context of wireless communications", *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 1102-1113, 2020.

25. M. DelVecchio, V. Arndorfer and W. C. Headley, "Investigating a spectral deception loss metric for training machine learning-based evasion attacks", *Proc. ACM Workshop Wireless Secur. Mach. Learn.*, pp. 43-48, 2020.