JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning

Mr. Atmakuri Rambabu, Head, Department of CSE, Anurag College Of Engineering, Aushapur, Ghatkesar, Telangana 501301

Mr. Dasari Sai Pavan, Student, Anurag College Of Engineering, Aushapur, Ghatkesar, Telangana 501301

Mr. Mohammed Shazaan, Student, Anurag College Of Engineering, Aushapur, Ghatkesar, Telangana 501301

Ms.Siliveri Sireesha, Student, Anurag College Of Engineering, Aushapur, Ghatkesar, Telangana 501301

Mr.Alwa Shashivardhanreddy, Student, Anurag College Of Engineering, Aushapur, Ghatkesar, Telangana 501301

Abstract: This paper introduces an over-the-air membership inference attack (MIA) designed to extract sensitive information from wireless signal classifiers. With the rise of machine learning (ML) techniques for wireless signal classification, such as for PHY-layer authentication, there emerges a vulnerability to adversarial attacks like MIA. The MIA's objective is to deduce whether a particular signal was part of the training data used by the target classifier. The leaked information, encompassing waveform, channel, and device characteristics, could potentially compromise the integrity of the underlying ML model, paving the way for attacks on PHY-layer authentication. Challenges arise due to disparities in received signals between the adversary and the intended receiver, owing to variations in channel conditions. To address this, the adversary constructs a surrogate classifier based on observed spectrum data before executing the black-box MIA on this model. Results demonstrate the adversary's ability to accurately infer signals utilized in training the target classifier, posing significant security risks. In response, a proactive defense strategy is devised to counter the MIA, involving the creation of a shadow MIA model to mislead the adversary. This defense mechanism effectively diminishes MIA accuracy and mitigates information leakage from the wireless signal classifier. Moreover, unique challenges inherent to wireless systems, distinct from other data domains like computer vision, are discussed. The paper focuses on RF fingerprinting, where a DL classifier aids in determining the legitimacy of received signals from authorized users. Even in the scenario where the adversary possesses knowledge of the target classifier, differences in received signals hinder its efficacy in identifying authorized users. The primary objective of this paper is to pioneer the application of MIA within the wireless domain. It explores the use of a deep neural network (DNN)-based wireless signal classifier as the target ML engine, against which the MIA is conducted over the air. The attack's success unveils whether the classifier was trained against specific waveform, radio device, or channel characteristics, thereby exposing potential vulnerabilities in the system. Ultimately, this research aims to raise awareness and bolster defenses against such adversarial threats in wireless communication networks.

Keywords: Membership Inference Attack (MIA), Wireless Signal Classification, Deep Learning (DL), Machine Learning (ML), PHY-layer Authentication, Adversarial Machine Learning, RF Fingerprinting, Surrogate Classifier.

I. INTRODUCTION

Wireless communication systems have become increasingly prevalent in our daily lives, offering convenience and flexibility in various applications. With the advent of machine learning (ML) techniques, wireless signal classification has seen remarkable advancements, particularly in tasks such as PHY-layer authentication. However, these advancements have also introduced new security challenges, including the susceptibility to adversarial attacks. One such attack, known as the membership inference attack (MIA), poses a significant threat to the privacy and security of wireless signal classifiers. By exploiting machine learning models, adversaries can infer whether a specific signal was part of the training data used by the target classifier. This private information, encompassing waveform, channel, and device characteristics, can be leveraged to identify vulnerabilities in

the underlying ML model, thereby compromising the integrity of systems like PHY-layer authentication. The unique characteristics of wireless systems pose distinct challenges for MIA compared to other data domains, such as computer vision. Discrepancies in channel conditions result in differences between the signals received by the adversary and those received by the target classifier, complicating data collection and design for the attack. Despite these challenges, adversaries can build surrogate classifiers and launch black-box MIAs to reliably infer signals used in training the target classifier.

To address these security concerns, proactive defense mechanisms are crucial. In this paper, we propose a novel defense strategy against MIA by introducing a shadow MIA model to deceive adversaries. By reducing the accuracy of MIA and preventing information leakage from the wireless signal classifier, this defense mechanism aims to safeguard the privacy and security of wireless communication systems. This paper aims to shed light on the emerging threat of membership inference attacks in the wireless domain and proposes effective defense strategies to mitigate these risks. Through empirical analysis and experimentation, we demonstrate the efficacy of our proposed defense mechanism in thwarting adversarial attacks and enhancing the security of wireless signal classifiers..

II. RELATED WORK

Membership inference attacks (MIAs) and their defense mechanisms have garnered significant attention in recent years, particularly in the context of machine learning (ML) systems. While existing literature primarily focuses on MIAs in the context of traditional data domains like computer vision and natural language processing, there is a growing interest in exploring MIAs in the wireless domain, especially concerning signal classifiers based on deep learning models.

In the realm of traditional data domains, early studies have demonstrated the feasibility of MIAs by leveraging model output probabilities or confidence scores to infer membership status. Moreover, various defense mechanisms, including differential privacy and adversarial training, have been proposed to mitigate the effectiveness of MIAs.

However, the unique characteristics of wireless systems present distinct challenges for MIAs and their defense. Prior research has highlighted the discrepancy between the signals observed by adversaries and those used by the target classifier due to channel variations. This necessitates the development of novel attack strategies and defense mechanisms tailored to the wireless domain.

Several recent works have begun to explore MIAs in wireless signal classifiers based on deep learning models. These studies have investigated the impact of channel conditions, waveform characteristics, and device properties on the susceptibility of signal classifiers to MIAs. Additionally, researchers have proposed defense strategies, such as adversarial training and model obfuscation, to enhance the robustness of wireless signal classifiers against membership inference attacks.

While these efforts represent significant progress in understanding and mitigating MIAs in wireless signal classifiers, further research is needed to explore the efficacy of defense mechanisms in real-world scenarios and to develop comprehensive strategies for securing wireless communication systems against adversarial attacks..

III. PROPOSED SYSTEM

In this study, we propose a novel approach to address the challenges posed by membership inference attacks (MIAs) targeting wireless signal classifiers utilizing deep learning models. Our proposed system consists of two main components: the membership inference attack (MIA) and the defense mechanism.

Membership Inference Attack (MIA):

- The MIA is designed to exploit vulnerabilities in wireless signal classifiers trained using deep neural networks (DNNs). It aims to infer whether a particular signal instance has been included in the training dataset of the target classifier.
- Leveraging over-the-air techniques, the adversary observes the spectrum to gather information about the behavior of the target classifier.
- The adversary then constructs a surrogate classifier based on the observed spectrum, which enables them to launch the blackbox MIA.
- By analyzing the output of the surrogate classifier, the adversary can deduce whether a given wireless signal was part of the training data used to build the target classifier, thereby revealing private information about the classifier's training dataset.

Defense Mechanism:

- To mitigate the effectiveness of the MIA, we propose a proactive defense mechanism.
- Our defense strategy involves building a shadow MIA model, which is trained to mimic the behavior of the target classifier.
- The shadow MIA model is strategically designed to provide misleading information to the adversary, thereby reducing the accuracy of the MIA.
- By fooling the adversary with the shadow MIA model, we aim to prevent the leakage of sensitive information from the wireless signal classifier.

Overall, our proposed system offers a comprehensive approach to addressing the challenges posed by membership inference attacks on wireless signal classifiers. By combining sophisticated attack techniques with proactive defense strategies, we strive to enhance the security and privacy of wireless communication systems against adversarial threats..

IV. RESULT AND DISCUSSION

Figure 7.1 Home page screen

Figure 7.2 User login screen

Figure 7.3 MIA prediction screen

Figure 7.4 MIA prediction results

Figure 7.5 service provider login screen

Figure 7.6 Algorithm accuracy bar graph screen

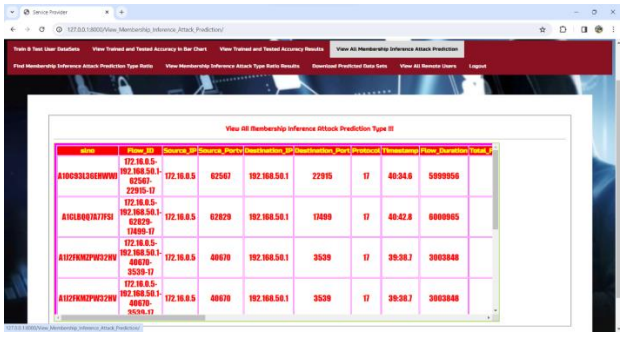

Figure 7.7 MIA prediction results

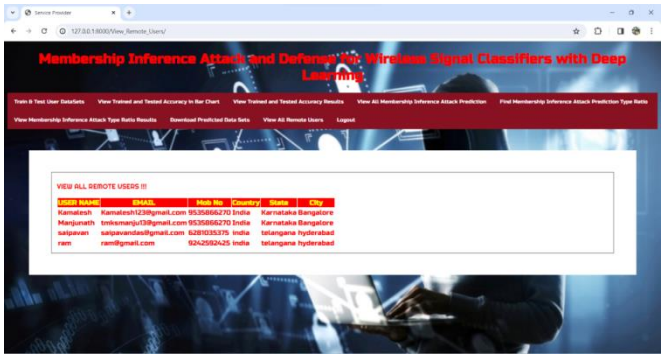

Figure 7.8 Remote users screen

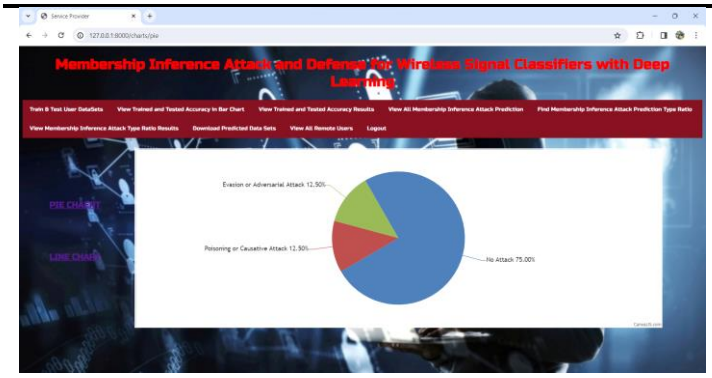

Figure 7.9 MIA results bar graph screen

V. CONCLUSION

In this study, we have investigated the threat posed by membership inference attacks (MIAs) on wireless signal classifiers and proposed a defense mechanism to mitigate the associated risks. Our research sheds light on the vulnerabilities of machine learning-based classifiers in the context of wireless communications and highlights the importance of proactive defense strategies to safeguard against adversarial attacks.

The over-the-air MIA presented in this paper demonstrates the potential for adversaries to exploit private information leakage from wireless signal classifiers, thereby compromising the security and privacy of communication systems. By inferring whether a signal of interest was included in the training data of a target classifier, adversaries can identify vulnerabilities in the underlying machine learning model and launch further attacks, such as bypassing PHY-layer authentication mechanisms.

To address this challenge, we have developed a proactive defense mechanism against MIAs by constructing a shadow MIA model. This defense strategy aims to deceive adversaries by providing misleading information, thereby reducing the accuracy of the MIA and preventing sensitive information leakage from the wireless signal classifier. By building awareness of potential security threats and implementing robust defense mechanisms, we can enhance the resilience of wireless communication systems against adversarial attacks.

In conclusion, our research contributes to the understanding of security issues in machine learning-based wireless signal classification and provides practical insights into mitigating the risks posed by membership inference attacks. By fostering collaboration between researchers and industry stakeholders, we can continue to advance the development of secure and resilient wireless communication technologies in the face of evolving cyber threats.

REFERENCES

1. Y. Shi, K. Davaslioglu, and Y. E. Sagduyu, "Over-the-Air Membership Inference Attacks as Privacy Threats for Deep Learning-based Wireless Signal Classifiers," ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec) Workshop on Wireless Security and Machine Learning (WiseML), 2020.

2. 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, Springer, 2020

3. Y. E. Sagduyu, Y. Shi, T. Erpek, W. Headley, B.Flowers, G. Stantchev, and Z. Lu, "When Wireless Security Meets Machine Learning: Motivation, Challenges, and Research Directions," arXiv preprintarXiv:2001.08883, 2020/

4. D. Adesina D, C. C. Hsieh, Y. E. Sagduyu, and L. Qian, "Adversarial Machine Learning in Wireless Communications using RF Data: A Review," arXiv preprint arXiv:2012.14392, 2020.

5. T. Erpek, Y. E. Sagduyu, and Y. Shi, "Deep Learning for Launching and Mitigating Wireless Jamming Attacks," IEEE Transactions on Cognitive Communications and Networking, 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," arXiv preprint arXiv:2101.05768, 2021.

7. M. Sadeghi and E. G. Larsson, "Adversarial Attacks on Deep-learning based Radio Signal Classification," IEEE Communications Letters, Feb.2019.

8. M. Sadeghi and E. G. Larsson, "Physical Adversarial Attacks Against End-to-end Autoencoder Communication Systems," IEEE Communications Letters, May 2019.

9. 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," Conference on Information Sciences and Systems (CISS), 2020.

10. B. Kim, Y. E. Sagduyu, K. Davaslioglu, T. Erpek, and S. Ulukus, "Channel-Aware Adversarial Attacks Against Deep Learning-Based Wireless Signal Classifiers," arXiv preprint arXiv:2005.05321.

11. B. Kim, Y. E. Sagduyu, K. Davaslioglu, T. Erpek, and S. Ulukus, "Adversarial Attacks with Multiple Antennas against Deep Learningbased Modulation

Classifiers," IEEE Global Communications Conference (GLOBECOM), 2020.

12. B. Kim, Y. E. Sagduyu, T. Erpek, K. Davaslioglu, and S. Ulukus, "Channel Effects on Surrogate Models of Adversarial Attacks against Wireless Signal Classifiers," IEEE International Conference on Communications (ICC), 2021.

13. B. Manoj, M. Sadeghi, and E. G. Larsson, "Adversarial Attacks on Deep Learning based Power Allocation in a Massive MIMO Network," arXiv preprint arXiv:2101.12090, 2021.

14. B. Kim, Y. E. Sagduyu, T. Erpek, and S. Ulukus, "Adversarial Attacks on Deep Learning Based mmWave Beam Prediction in 5G and Beyond," IEEE Statistical Signal Processing Workshop, 2021.