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

ABSTRACT

An over-the-air membership inference attack (MIA) is presented to leak private information from a wireless signal classifier. Machine learning (ML) provides powerful means to classify wireless signals, e.g., for PHY-layer authentication. As an adversarial machine learning attack, the MIA infers whether a signal of interest has been used in the training data of a target classifier. This private information incorporates waveform, channel, and device characteristics, and if leaked, can be exploited by an adversary to identify vulnerabilities of the underlying ML model (e.g., to infiltrate the PHY-layer authentication). One challenge for the over-the-air MIA is that the received signals and consequently the RF fingerprints at the adversary and the intended receiver differ due to the discrepancy in channel conditions. Therefore, the adversary first builds a surrogate classifier by observing the spectrum and then launches the blackbox MIA on this classifier. The MIA results show that the adversary can reliably infer signals (and potentially the radio and channel information) used to build the target classifier. Therefore, a proactive defense is developed against the MIA by building a shadow MIA model and fooling the adversary. This defense can successfully reduce the MIA accuracy and prevent information leakage from the wireless signal classifier..

**EXISTING SYSTEM**

An existing system presents channel-aware adversarial attacks against deep learning-based wireless signal classifiers. There is a transmitter that transmits signals with different modulation types. A deep neural network is used at each receiver to classify its over-the-air received signals to modulation types. In the meantime, an adversary transmits an adversarial perturbation (subject to a power budget) to fool receivers into making errors in classifying signals that are received as superpositions of transmitted signals and adversarial perturbations.

First, these evasion attacks are shown to fail when channels are not considered in designing adversarial perturbations. Then, realistic attacks are presented by considering channel effects from the adversary to each receiver. After showing that a channel-aware attack is selective (i.e., it affects only the receiver whose channel is considered in the perturbation design), a broadcast adversarial attack is presented by crafting a common adversarial perturbation to simultaneously fool classifiers at different receivers.

The major vulnerability of modulation classifiers to over-the-air adversarial attacks is shown by accounting for different levels of information available about the channel, the transmitter input, and the classifier model. Finally, a certified defense based on randomized smoothing that augments training data with noise is introduced to make the modulation classifier robust to adversarial perturbations

**Disadvantages**

* The system is not implemented Membership Inference Attack(MIA) against datasets which leads less security.
* In conjunction with security threats, an emerging concern on ML-based solutions is not privacy, namely the potential leakage of information from the ML models to the adversaries.

Proposed System

1) In this paper, we present the first MIA that is launched against a wireless classifier over the air to infer about training data and leak private information on waveform, device, and channel characteristics.

2) We consider two settings for the MIA: (i) the MIA should be able to identify signals from the same radio device as member and non-member, and (ii) nonmember signals are generated by different radio devices.

3) We extend the MIA such that it is launched by using not only received signals but also their noisy variations by accounting for channel variations.

4) We show through detailed numerical results that the success of the MIA is high, i.e., the MIA can infer the training data membership of the wireless signal classifier with high accuracy.

5) We present a defense scheme to protect wireless signal classifiers from the MIA and show that this defense can reduce the accuracy of the MIA significantly.

**Advantages**

* The goal of the MIA is to identify data samples that have been used to train a ML classifier (as studied in computer vision and other data domains.
* The proposed system developed a proactive defense scheme for the MIA. The service provider first needs to build a shadow MIA model. Then, it applies the defense using this shadow MIA model.

**SYSTEM REQUIREMENTS**

➢ **H/W System Configuration:-**

➢ Processor - Pentium –IV

➢ RAM - 4 GB (min)

➢ Hard Disk - 20 GB

➢ Key Board - Standard Windows Keyboard

➢ Mouse - Two or Three Button Mouse

➢ Monitor - SVGA

**SOFTWARE REQUIREMENTS:**

* **Operating system :** Windows 7 Ultimate.
* **Coding Language :** Python.
* **Front-End :** Python.
* **Back-End :** Django-ORM
* **Designing :** Html, css, javascript.
* **Data Base :** MySQL (WAMP Server).