Project Reference No: 46S_BE_1703

Title of the Project: MEMBERSHIP INTERFACE ATTACK AND DEFENSE FOR WIRELESS SIGNAL CLASSIFICATION WITH DEEP LEARNING

College & Department:

COMPUTER SCIENCE AND ENGINEERING

LINGARAJ APPA ENGINEERING COLLEGE, BIDAR

Students: Mr. ABHISHEK & Mr. MD LUQMAN ALI BHAGWAN

Guides: Mr. VEERESH BIRADAR & Mr. GURURAJ NASE

KEYWORDS:

MIA, DEFENSE, SECURITY, DATA COLLECTION.,EVALUTION, PRIVACY, MODEL TRAINING

INTRODUCTION:

Machine learning (ML) has emerged with powerful means to learn from and adapt to wireless network dynamics, and solve complex tasks in wireless communications subject to channel, interference, and traffic effects. In particular, deep learning (DL) that has been empowered by recent algorithmic and computational advances can effectively capture high dimensional representations of spectrum data and support various wireless communications tasks, including but not limited to, spectrum sensing, signal classification, spectrum allocation, and waveform design. However, the use of ML/DL also raises unique challenges in terms of security for wireless systems. With adversarial machine learning (AML), various attacks have been developed to launch against the ML/DL engines of wireless systems, including inference

In conjunction with security threats, an emerging concern on ML-based solutions is privacy, namely the potential leakage of information from the ML models to the adversaries. One example is the model inversion attack, where the adversary has access to the ML model and some private information, and aims to infer additional private information by observing the inputs and outputs of the ML model. Another privacy attack of interest is the membership inference attack (MIA) that has been extensively studied in various data domains including computer vision, healthcare, and commerce. The goal of the MIA to infer if a particular data sample has been used in training data or not . While the MIA has been demonstrated as a major

privacy threat for computer vision and other data domains, it has not been applied yet to the wireless domain. In practice, the broadcast and shared nature of wireless medium offers unique opportunities to an adversary to eavesdrop wireless transmissions and launch the MIA over the air against a wireless signal classifier to infer about the underlying radio device, waveform, and channel environment characteristics under which the ML/DL model of the target signal classifier is trained.

OBJECTIVES:

- The main objective of the application is to classify the membership attacks on the networks
- The application must be fast and reliable.

METHODOLOGY:

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 [42]–[52]). One possible application of the MIA in the wireless domain is a privacy attack on PHY-layer signal authentication, where the adversary aims to identify the signal samples that have been used in the training of a wireless signal classifier. Then, the adversary can leverage these signal samples and leaked information on waveform, device and channel characteristics of authorized users to generate signals in order to obtain service from the provider. The training data and the general data usually have different distributions (e.g., due to differences of radios and channels in training and test times) and the classifier may over fit the training data. The MIA analyzes the overfitting to leak private information.

Suppose that each sample in the training data set is represented by a set of features F and is labeled as one of two classes following a supervised learning approach. The MIA aims to identify whether a given sample is in the training data set to build the given classifier or not. This attack may be a white-box attack, i.e., the target classifier is available to the adversary, or a black-box attack, where the adversary does not know the classifier, but it can collect data from the target classifier. In this paper, we consider a black-box attack. To launch an effective MIA, we consider a general approach as follows. Suppose that features include all (useful, but potentially biased and noisy) information, where useful information in Fu can be used to identify the class, biased information Fb is due to the different distributions of training data and general test data, and noisy information Fn is other information with no statistical significance.

Note that to simplify discussion, we assume each feature includes only one type of information. For the general case that one feature includes multiple types of information, we can divide it into multiple features to meet our assumption. DL is relied upon to extract useful and biased information while ignoring noisy information. Then, a classifier is optimized to fit on useful and biased information (Fu and Fb). While fitting on Fu can provide correct classification on general test data, fitting on Fb corresponds to over fitting, which provides correct classification on the given training data but wrong classification on general test data.

HARDWARE REQUIREMENTS:

System	:	Pentium IV 2.4 GHz.
RAM	:	8 GB RAM or more
CPU Architecture	:	x86_64 CPU Architecture
Generation	:	2nd generation Intel Core or Newer

2.2. SOFTWARE REQUIREMENTS:

2.2.1. DESKTOP REQUIREMENTS FOR DEVELOPMENT

Operating system	:	64-bit Microsoft Windows 8/10
UI Software	:	FIGMA (UI Designing)
Software tools	:	Pychan
Coding Language	:	PYTHON
Tool Kit	:	Android Studio Arctic Fox
IDE	:	Command Prom

DATA FLOW DIAGRAM:

.

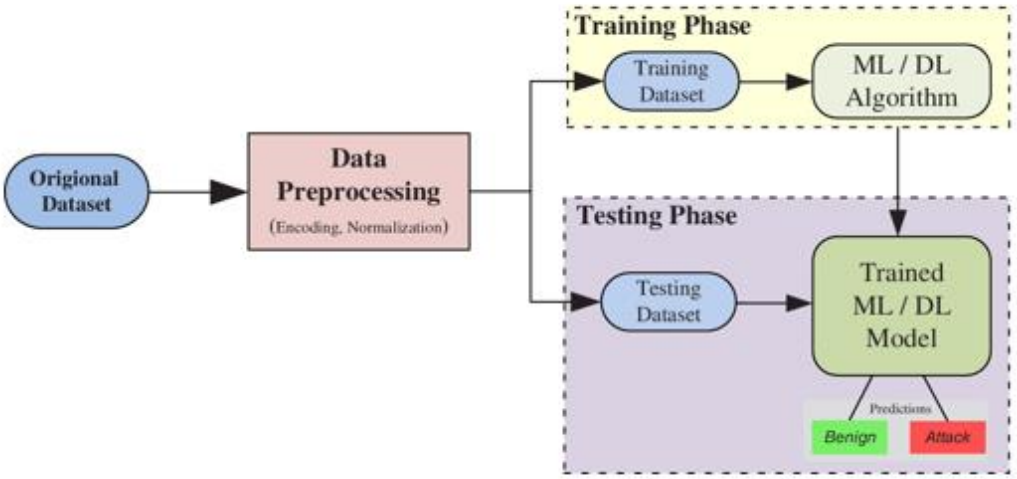
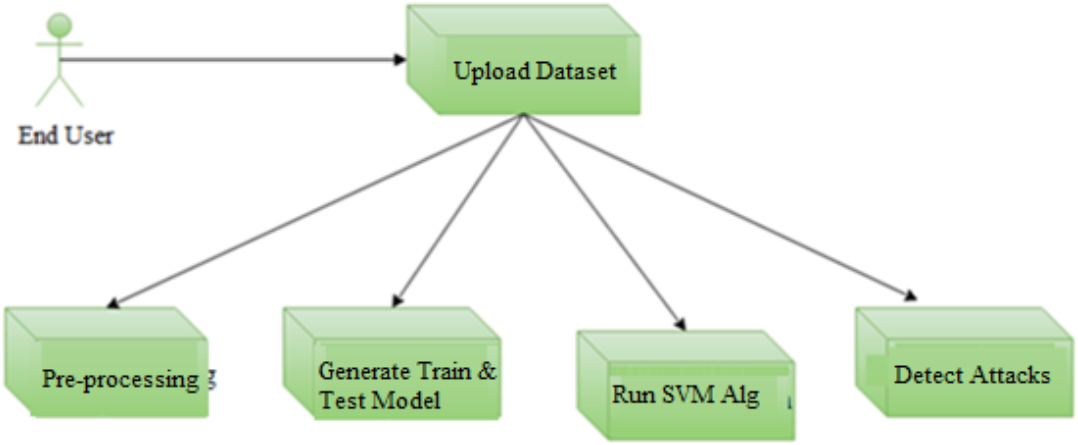


Figure: Data Flow Chart



CONCLUSION

In this paper, we studied the MIA as a novel privacy threat against ML-based wireless applications. The target application is a FL-based classifier to identify authorized users by their RF fingerprint. An example use case for this attack is PHYS-layer user authentication in 5G or IoT systems. The input of this model consists of the received power and the phase shift. An adversary launches the MIA to infer whether signals of interest have been used to train this wireless signal classifier or not. In this attack, the adversary needs to collect signals and their classification results by observing the spectrum. Then, it can build a surrogate classifier namely a functionally equivalent classifier as the target classifier at the intended receiver, e.g., a service provider. We showed that the surrogate classifier can be reliably built by the adversary under various settings. Then, the adversary launches the MIA to identify whether for a received signal, its corresponding signal received at the service provider is in the training data or not In the first setting where non-member signals can be generated by the same devices, the MIA accuracy is 88.62% for strong signals and 77.01% for weak signals. We studied the case that the member inference is investigated not only for received signals but also their noisy variations due to random channel effects

SCOPE FOR FUTURE WORK

The MIA determines if a signal of interest has been exploited in the training data of a target classifier as an adversarial machine learning attack. If leaked, this proprietary data, which includes waveform, channel, and device details, might be used by an attacker to find weaknesses in the underlying ML model (e.g., to infiltrate the PHYS-layer authentication). The received signals and, consequently, the RF fingerprints at the adversary and the intended receiver differ as a result of the difference in channel circumstances, which presents a difficulty for over-the-air MIA. As a result, the adversary first constructs a surrogate classifier by spectral watching it, and then launches the black box MIA on it.