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

Dr.K.Sailaja[1] , Samiseni Sandhya [2]
[1] Professor & HOD, Department of Computer Applications , Chadalawada Ramanamma Engineering College
Tirupati , Andhra Pradhesh, India.
[2] Student, Department of Computer Applications, Chadalawada Ramanamma Engineering College
Renigunta Rd, Tirupati, Andhra Pradesh, India

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 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.

I. 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 highdimensional 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 (exploratory) attacks, evasion (adversarial) attacks, poisoning (causative) attacks, Trojan attacks, spoofing attacks, and attacks to facilitate covert communications. These AML-based attacks operate with small spectrum footprints and thus are harder to detect compared with conventional wireless attacks such as jamming of data transmissions.

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.

II. LITERATURE SURVEY

The Literature review plays a very important role in the research process. It is a source from here research ideas are drawn and developed into concepts and finally theories. It also provides the researchers a bird's eye view about the research done in that area so far. Depending on what is observed in the literature review, a researcher will understand where his/her research stands. Here in this literature survey, all primary, secondary and tertiary sources of information were searched. A literature survey or literature review means that researcher read and report on what the literature in the field has to say about the topic or subject. It is a study and review of relevant literature materials in relation to a topic that have been given.

1. **Title :** Over-the-Air Membership Inference Attacks as Privacy Threats for Deep Learning-based Wireless Signal Classifiers

This paper presents how to leak private information from a wireless signal classifier by launching an over-the-air membership inference attack (MIA). As machine learning (ML) algorithms are used to process wireless signals to make decisions such as PHY-layer authentication, the training data characteristics (e.g., device-level information) and the environment conditions (e.g., channel information) under which the data is collected may leak to the ML model. As a privacy threat, the adversary can use this leaked information to exploit vulnerabilities of the ML model following an adversarial ML approach. In this paper, the MIA is launched against a deep learning-based classifier that uses waveform, device, and channel characteristics (power and phase shifts) in the received signals for RF fingerprinting. By observing the spectrum, the adversary builds first a surrogate classifier and then an inference model to determine whether a signal of interest has been used in the training data of the receiver (e.g., a service provider).

2.Title : Deep Learning for Wireless Communications" in Development

Existing communication systems exhibit inherent limitations in translating theory to practice when handling the complexity of optimization for emerging wireless applications with high degrees of freedom. Deep learning has a strong potential to overcome this challenge via data-driven solutions and improve the performance of wireless systems in utilizing limited spectrum resources. In this chapter, first describe how deep learning is used to design an end-to-end communication system using autoencoders. This flexible design effectively captures channel impairments and optimizes transmitter and receiver operations jointly in single-antenna, multiple-antenna, and multiuser communications. Next, present the benefits of deep learning in spectrum situation awareness ranging from channel modeling and estimation to signal detection and classification tasks. Deep learning improves the performance when the model-based methods fail. Finally, discuss how deep learning applies to wireless communication security.

3.Title : Adversarial Deep Learning for Cognitive Radio Security: Jamming Attack and Defense Strategies

This paper presents an adversarial machine learning approach to launch jamming attacks on wireless communications and introduces a defense strategy. In a cognitive radio network, a transmitter senses channels, identifies spectrum opportunities, and transmits data to its receiver in idle channels. On the other hand, an attacker may also sense channels, identify busy channels and aim to jam transmissions of legitimate users. In a dynamic system with complex channel, traffic and interference characteristics, the transmitter applies some pre-trained machine learning algorithm to classify a channel as idle or busy. This classifier is unknown to the attacker that senses a channel, captures the transmitter's decisions by tracking the acknowledgments and applies deep learning (in form of an exploratory attack, i.e., inference attack) to build a classifier that is functionally equivalent to the one at the transmitter. This approach is shown to support the attacker to reliably predict successful transmissions based on the sensing results and effectively jam these transmissions. Then, a defense scheme is developed against adversarial deep learning by exploiting the sensitivity of deep learning to training errors. The transmitter deliberately takes a small number of wrong actions (in form of a causative attack, i.e., poisoning attack, launched against the attacker) when it accesses the spectrum. The objective is to prevent the attacker from building a reliable classifier

III. SYSTEM ANALYSIS

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 OF EXISTING SYSTEM

- 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

In this project, 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. 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. Extend the MIA such that it is launched by using not only received signals but also their noisy variations by accounting for channel variations. 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. 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 OF PROPOSED SYSTEM

- Identifies data samples that have been used to train a ML classifier
- Privacy Preserving
- Prevent Information Leakage

IV. IMPLEMENTATION

In this proposed system there are two modules they are:

- Service Provider
- Remote User

1. Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Train & Test User Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Membership Inference Attack Prediction, Find Membership Inference Attack Prediction Type Ratio, View Membership Inference Attack Type Ratio Results, Download Predicted Data Sets, View All Remote Users.

2. Remote User

In This Module, There Are N Numbers Of Users Are Present. User Should Register Before Doing Any Operations. Once User Registers, Their Details Will Be Stored To The Database. After Registration Successful, He Has To Login By Using Authorized User Name And Password. Once Login Is Successful User Will Do Some Operations Like Register And Login, Predict Membership Inference Attack Type,View Your Profile.

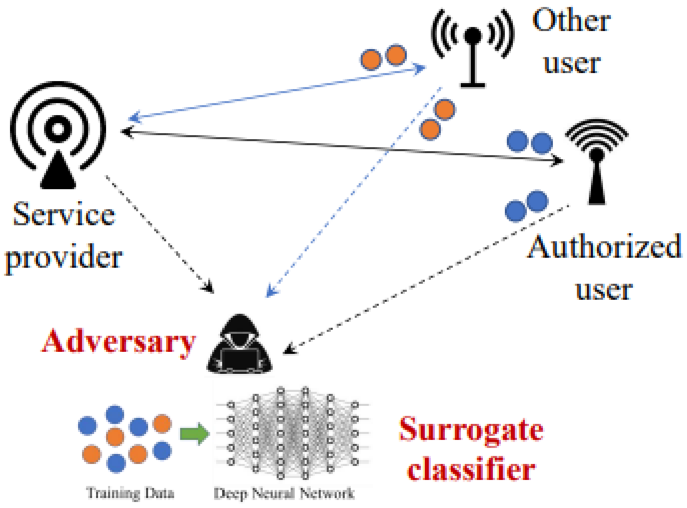
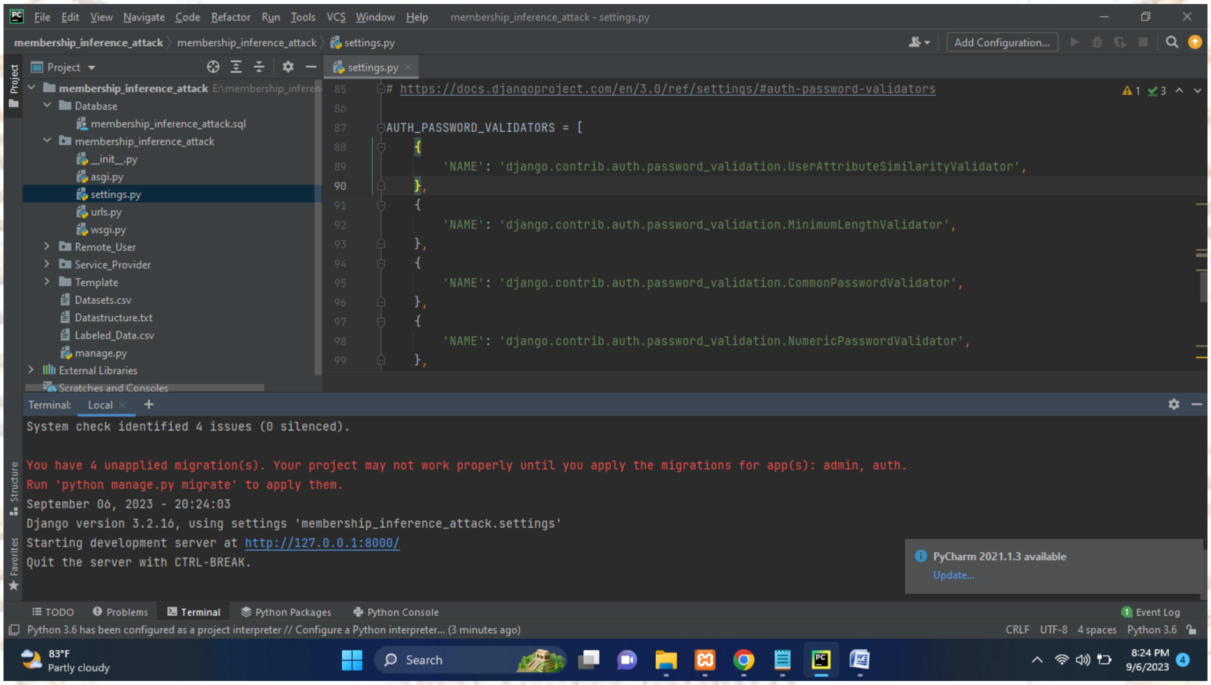


Fig 1 System Architecture

V. SCREEN SHOTS

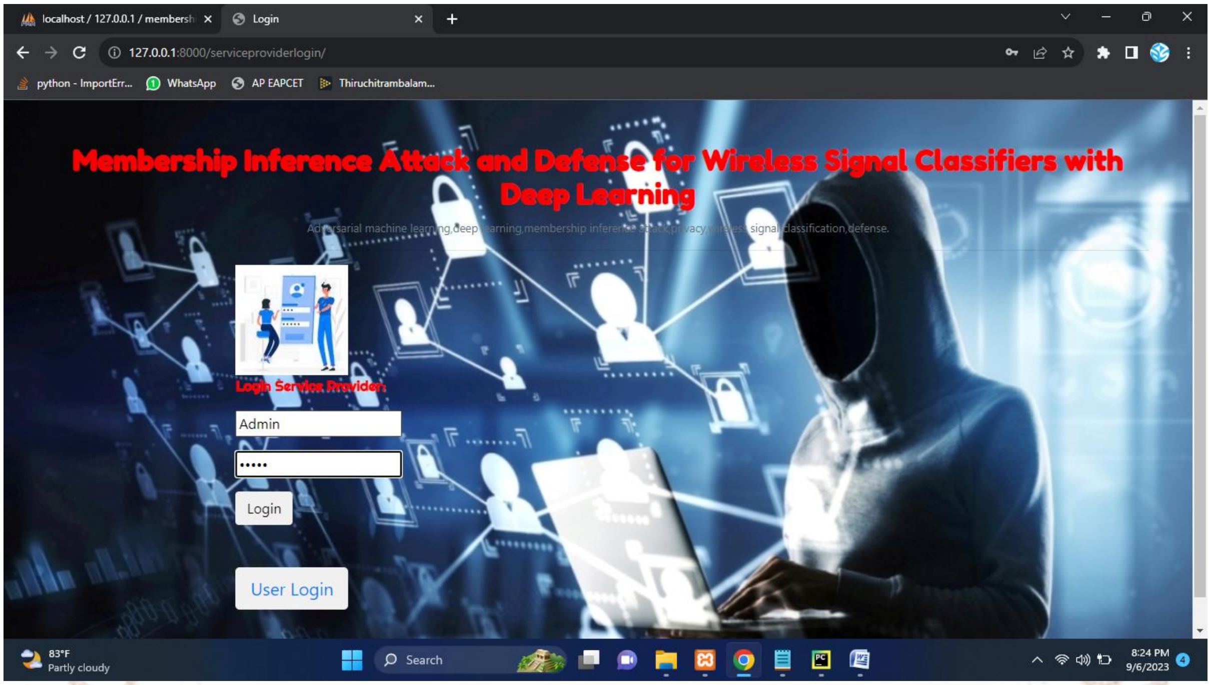


```python
# https://docs.djangoproject.com/en/3.0/ref/settings/#auth-password-validators

AUTH_PASSWORD_VALIDATORS = [
    {
        'NAME': 'django.contrib.auth.password_validation.UserAttributeSimilarityValidator',
    },
    {
        'NAME': 'django.contrib.auth.password_validation.MinimumLengthValidator',
    },
    {
        'NAME': 'django.contrib.auth.password_validation.CommonPasswordValidator',
    },
    {
        'NAME': 'django.contrib.auth.password_validation.NumericPasswordValidator',
    },
```

```
System check identified 4 issues (0 silenced).

You have 4 unapplied migration(s). Your project may not work properly until you apply the migrations for app(s): admin, auth.
Run 'python manage.py migrate' to apply them.
September 06, 2023 - 20:24:03
Django version 3.2.16, using settings 'membership_inference_attack.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
```

Python 3.6 has been configured as a project interpreter // Configure a Python interpreter... (3 minutes ago)

Fig 2. Generating Link



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

Adversarial machine learning,deep learning,membership inference attack,privacy,wireless signal classification,defense.

Login Service Provider

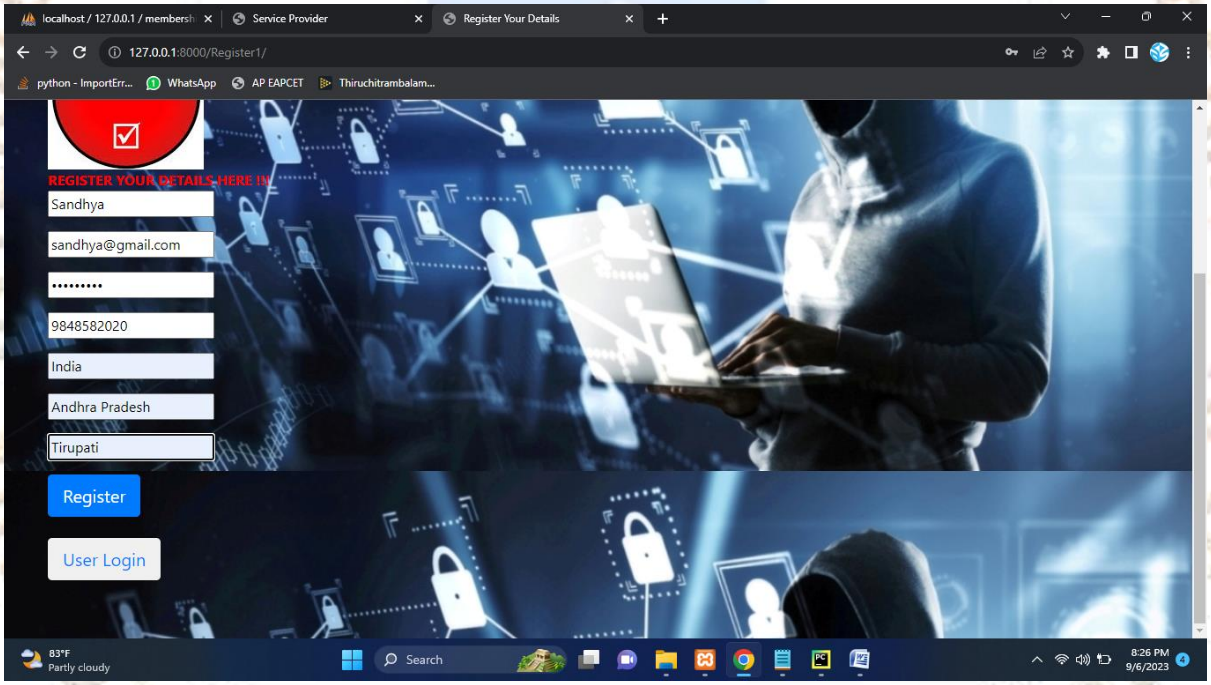
Admin

•••••

Login

User Login

Fig 3. Service Provider Login



REGISTER YOUR DETAILS HERE !!!

Sandhya

sandhya@gmail.com

•••••••••

9848582020

India

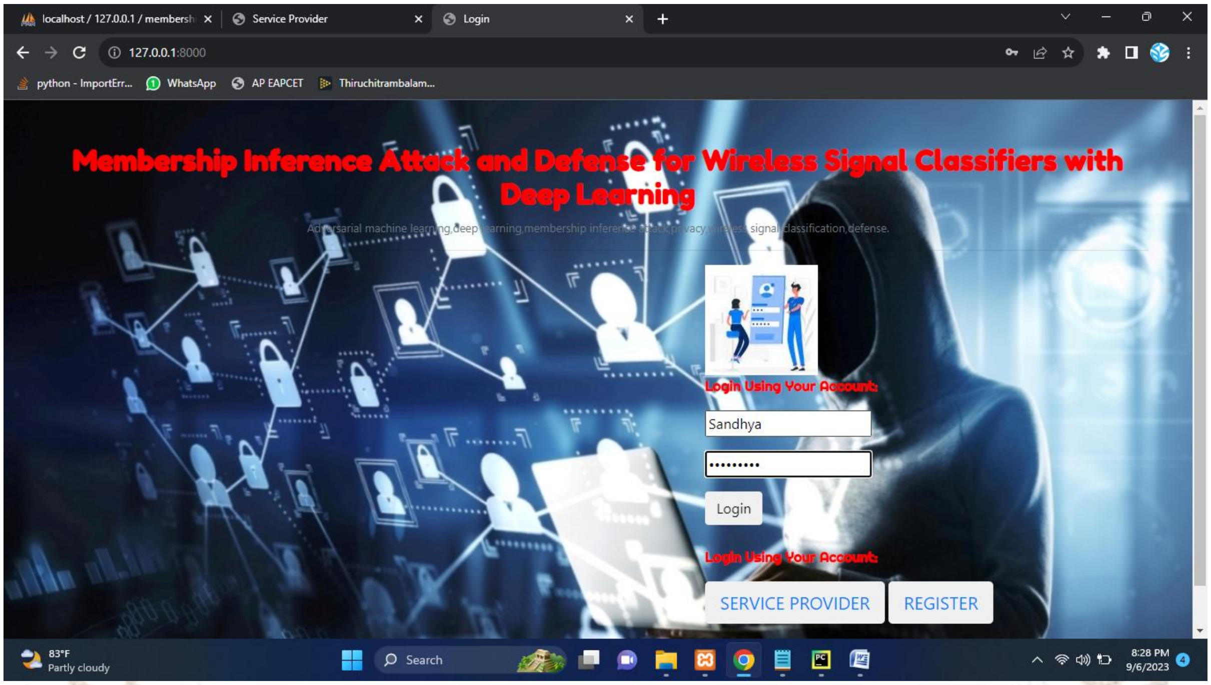
Andhra Pradesh

Tirupati

Register

User Login

Fig 4. User Registration



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

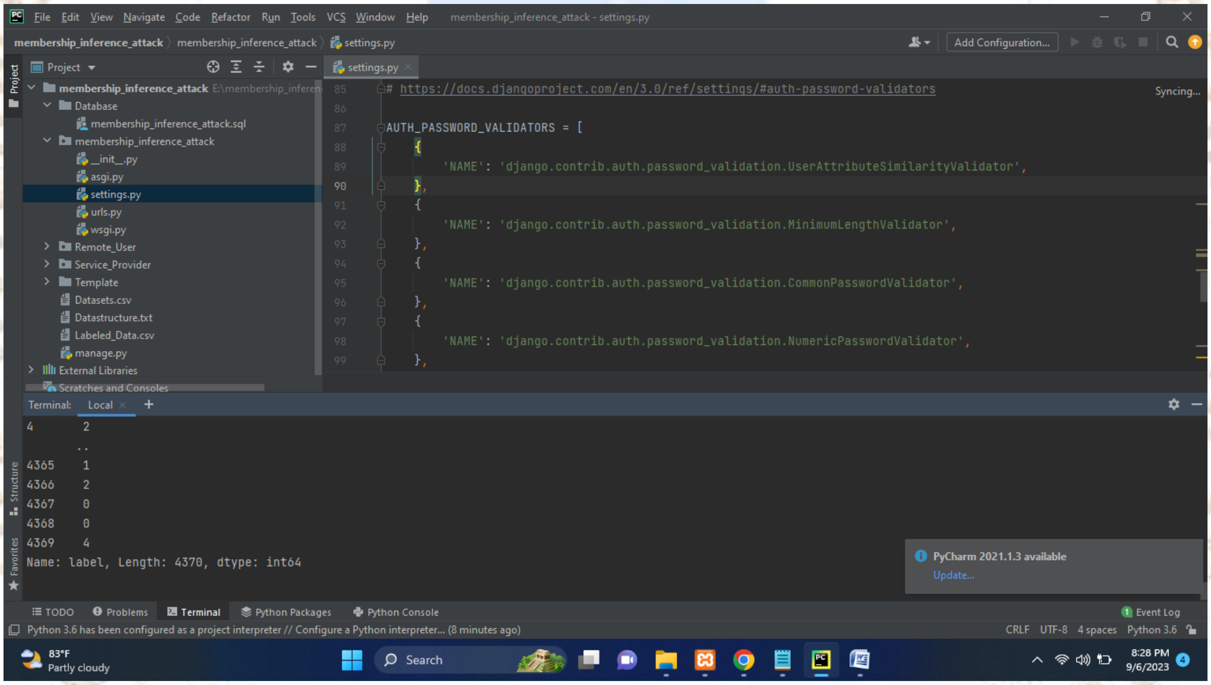
Adversarial machine learning,deep learning,membership inference attack,privacy,wireless signal classification,defense.

Login Using Your Account:

Sandhya

•••••••••

Fig 5. User Login



```
# https://docs.djangoproject.com/en/3.0/ref/settings/#auth-password-validators

AUTH_PASSWORD_VALIDATORS = [
    {
        'NAME': 'django.contrib.auth.password_validation.UserAttributeSimilarityValidator',
    },
    {
        'NAME': 'django.contrib.auth.password_validation.MinimumLengthValidator',
    },
    {
        'NAME': 'django.contrib.auth.password_validation.CommonPasswordValidator',
    },
    {
        'NAME': 'django.contrib.auth.password_validation.NumericPasswordValidator',
    },
```

```
4       2
        ..
4365    1
4366    2
4367    0
4368    0
4369    4
Name: label, Length: 4370, dtype: int64
```

PyCharm 2021.1.3 available
Update...

Python 3.6 has been configured as a project interpreter // Configure a Python interpreter... (8 minutes ago)

Fig 6. Training with ML Algorithms

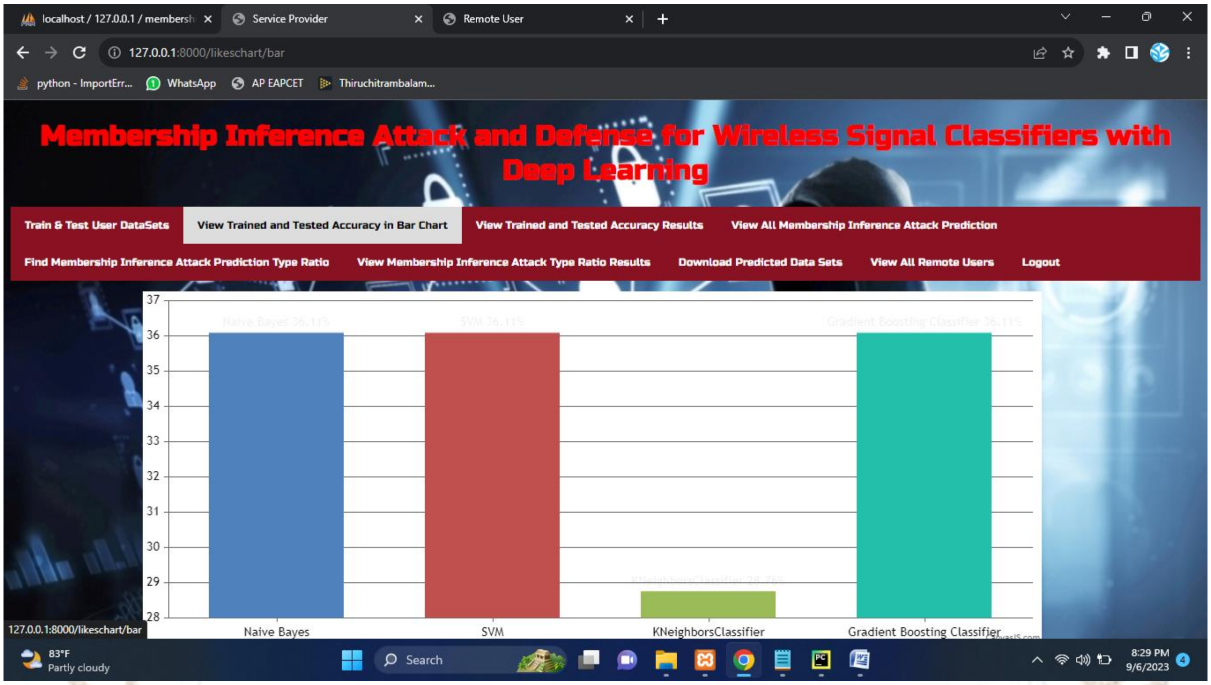
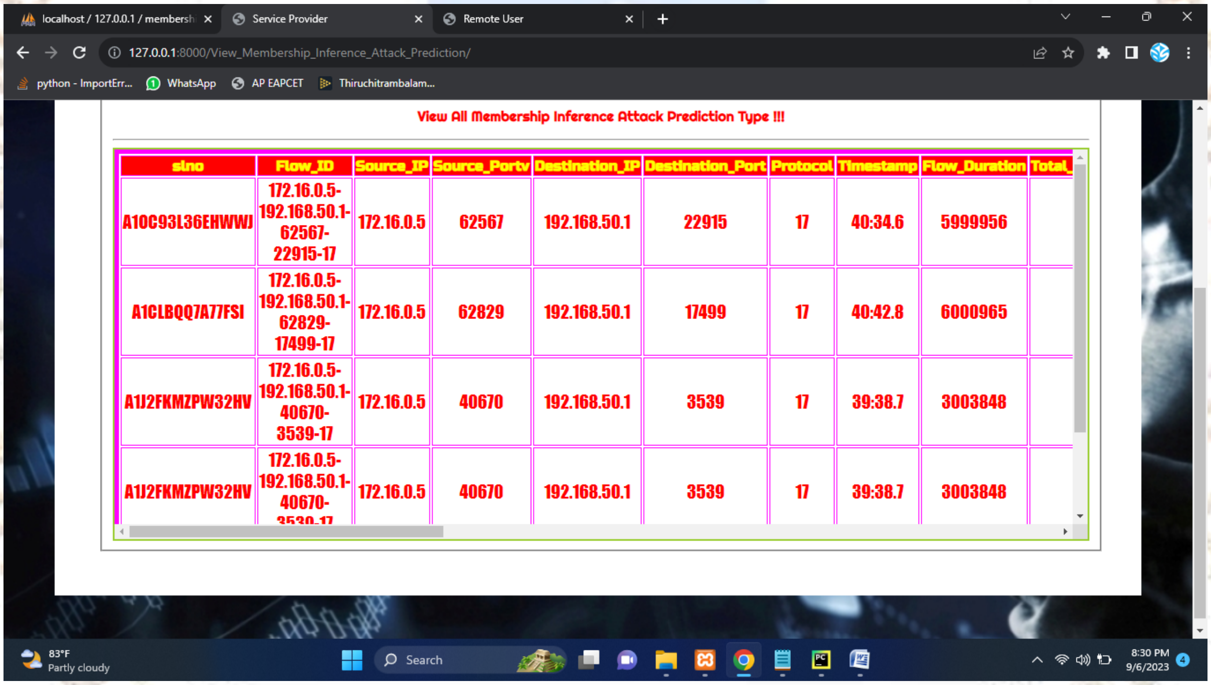

localhost / 127.0.0.1 / membersh
Service Provider
Remote User
127.0.0.1:8000/likeschart/bar
python - ImportErr...
WhatsApp
AP EAPCET
Thiruchitrambalam...
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Membership Inference Attack Prediction
Find Membership Inference Attack Prediction Type Ratio
View Membership Inference Attack Type Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout
Naive Bayes 36.11%
SVM 36.11%
KNeighborsClassifier 28.76%
Gradient Boosting Classifier 36.11%
37
36
35
34
33
32
31
30
29
28
Naive Bayes
SVM
KNeighborsClassifier
Gradient Boosting Classifier
127.0.0.1:8000/likeschart/bar
83°F
Partly cloudy
Search
8:29 PM
9/6/2023

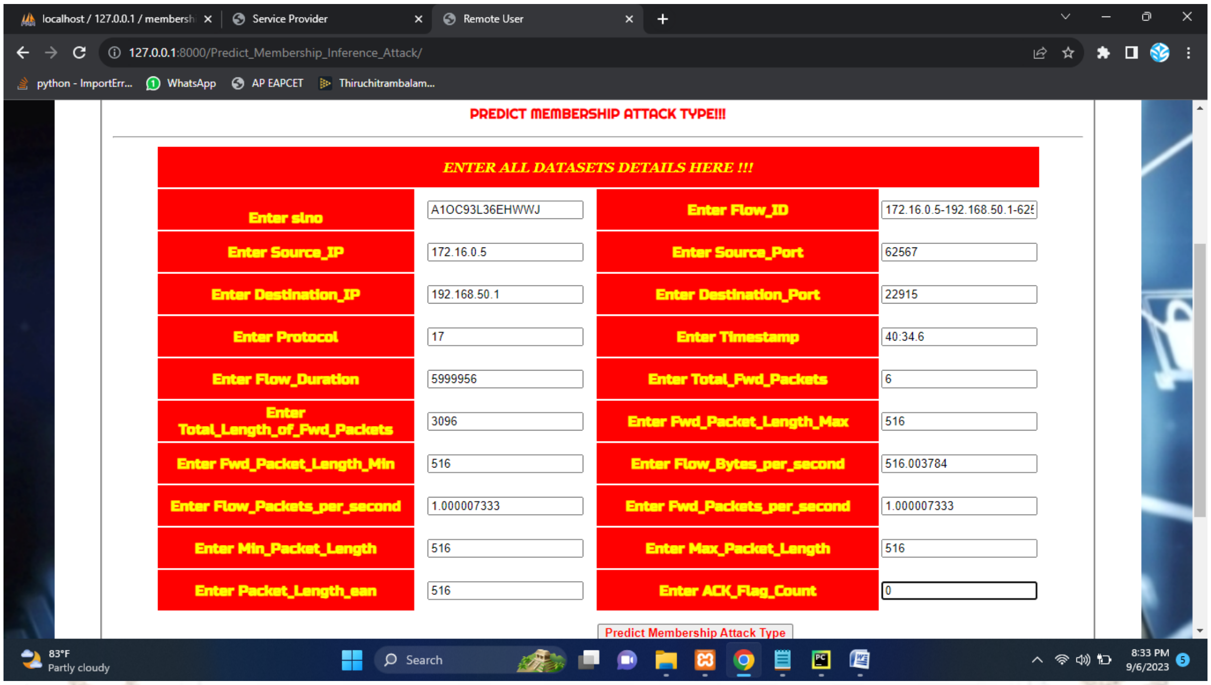
Fig 7. Trained and Tested Accuracy in Bar Chart



View All Membership Inference Attack Prediction Type !!!

slno	Flow_ID	Source_IP	Source_Portv	Destination_IP	Destination_Port	Protocol	Timestamp	Flow_Duration	Total_
A10C93L36EHWWJ	172.16.0.5- 192.168.50.1- 62567- 22915-17	172.16.0.5	62567	192.168.50.1	22915	17	40:34.6	5999956	
A1CLBQQ7A77FSI	172.16.0.5- 192.168.50.1- 62829- 17499-17	172.16.0.5	62829	192.168.50.1	17499	17	40:42.8	6000965	
A1J2FKMZPW32HV	172.16.0.5- 192.168.50.1- 40670- 3539-17	172.16.0.5	40670	192.168.50.1	3539	17	39:38.7	3003848	
A1J2FKMZPW32HV	172.16.0.5- 192.168.50.1- 40670- 3539-17	172.16.0.5	40670	192.168.50.1	3539	17	39:38.7	3003848	

Fig 8. Dataset Details



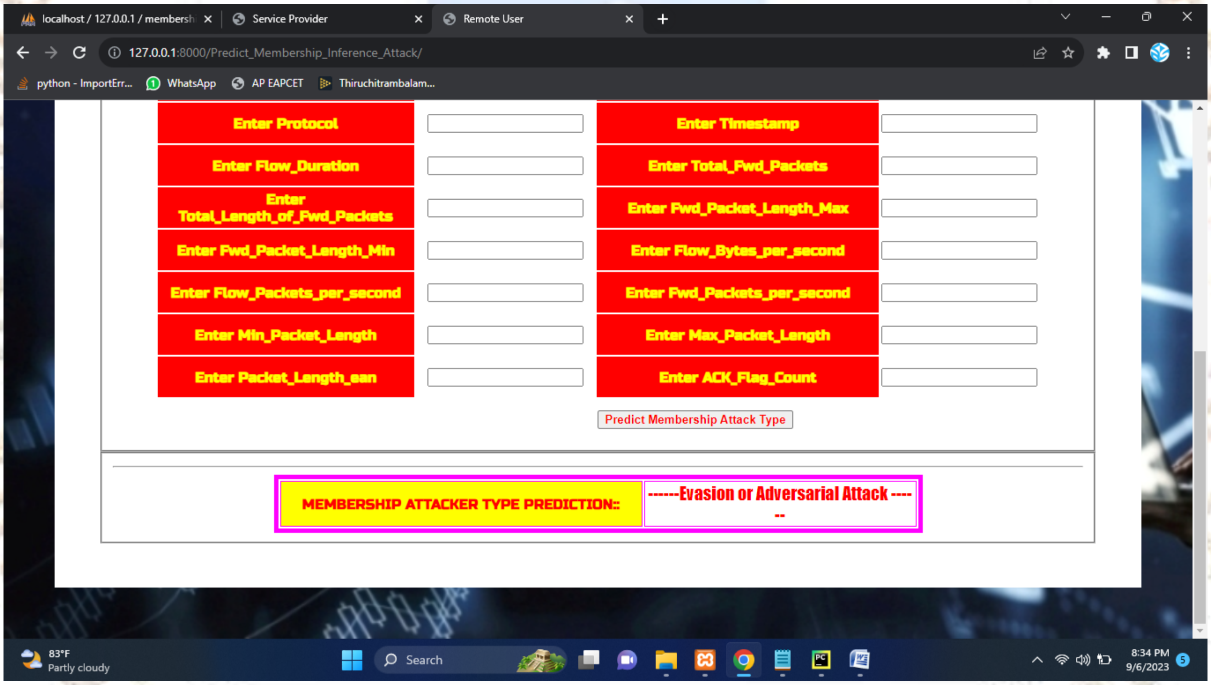
PREDICT MEMBERSHIP ATTACK TYPE!!!

ENTER ALL DATASETS DETAILS HERE !!!

Enter slno	A1OC93L36EHWWJ	Enter Flow_ID	172.16.0.5-192.168.50.1-625
Enter Source_IP	172.16.0.5	Enter Source_Port	62567
Enter Destination_IP	192.168.50.1	Enter Destination_Port	22915
Enter Protocol	17	Enter Timestamp	40:34.6
Enter Flow_Duration	5999956	Enter Total_Fwd_Packets	6
Enter Total_Length_of_Fwd_Packets	3096	Enter Fwd_Packet_Length_Max	516
Enter Fwd_Packet_Length_Min	516	Enter Flow_Bytes_per_second	516.003784
Enter Flow_Packets_per_second	1.000007333	Enter Fwd_Packets_per_second	1.000007333
Enter Min_Packet_Length	516	Enter Max_Packet_Length	516
Enter Packet_Length_ean	516	Enter ACK_Flag_Count	0

Predict Membership Attack Type

Fig 9. Enter Details For Prediction

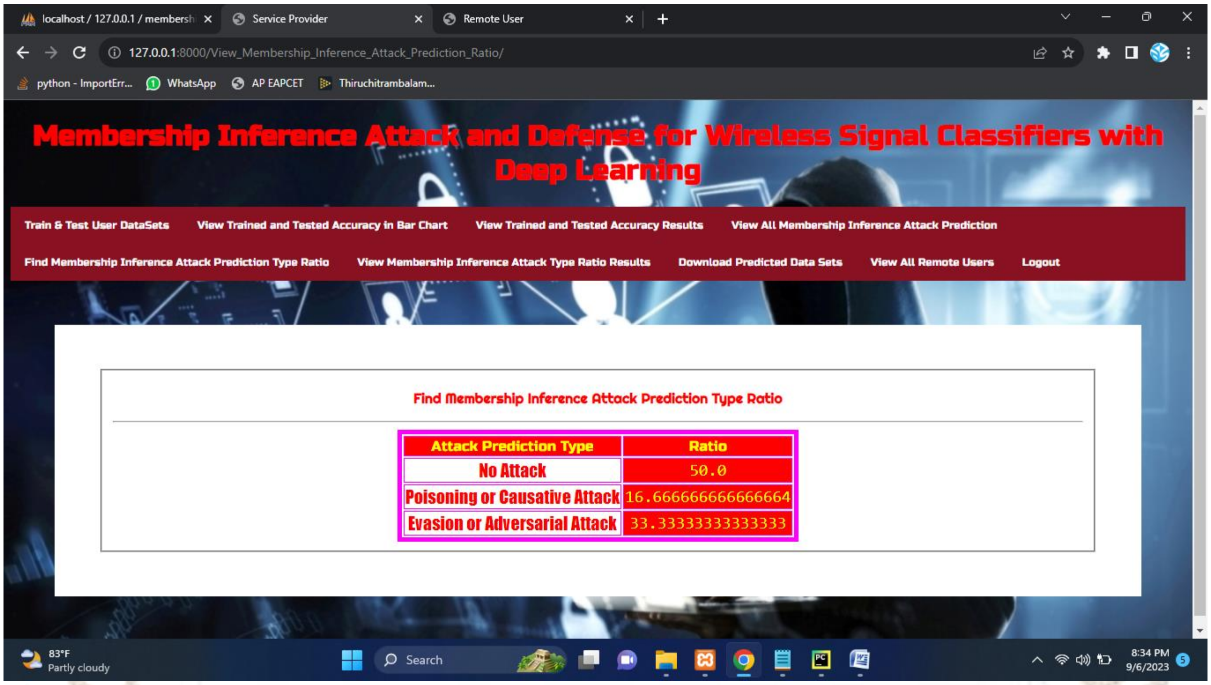


Enter Protocol		Enter Timestamp	
Enter Flow_Duration		Enter Total_Fwd_Packets	
Enter Total_Length_of_Fwd_Packets		Enter Fwd_Packet_Length_Max	
Enter Fwd_Packet_Length_Min		Enter Flow_Bytes_per_second	
Enter Flow_Packets_per_second		Enter Fwd_Packets_per_second	
Enter Min_Packet_Length		Enter Max_Packet_Length	
Enter Packet_Length_ean		Enter ACK_Flag_Count	

Predict Membership Attack Type

MEMBERSHIP ATTACKER TYPE PREDICTION::	------Evasion or Adversarial Attack ------

Fig 10. Prediction Result


localhost / 127.0.0.1 / membersh
Service Provider
Remote User
127.0.0.1:8000/View_Membership_Inference_Attack_Prediction_Ratio/
python - ImportErr...
WhatsApp
AP EAPCET
Thiruchitrambalam...
Membership Inference Attack and Defense for Wireless Signal Classifiers with Deep Learning
Train & Test User DataSets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Membership Inference Attack Prediction
Find Membership Inference Attack Prediction Type Ratio
View Membership Inference Attack Type Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout
83°F
Partly cloudy
Search
8:34 PM
9/6/2023

Find Membership Inference Attack Prediction Type Ratio

Attack Prediction Type	Ratio
No Attack	50.0
Poisoning or Causative Attack	16.666666666666664
Evasion or Adversarial Attack	33.33333333333333

Fig 11. Find Membership Inference Attack Prediction Type Ratio

VI. CONCLUSION

In this project, studied the MIA as a novel privacy threat against ML-based wireless applications. The target application is a DL-based classifier to identify authorized users by their RF fingerprint. An example use case for this attack is PHY-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 and showed that the surrogate classifier can be reliably built by the adversary under various settings.

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 preprint arXiv: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] 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," IEEE International Conference on Communications (ICC) Workshop on Promises and Challenges of Machine Learning in Communication Networks, 2018.
[6] 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.