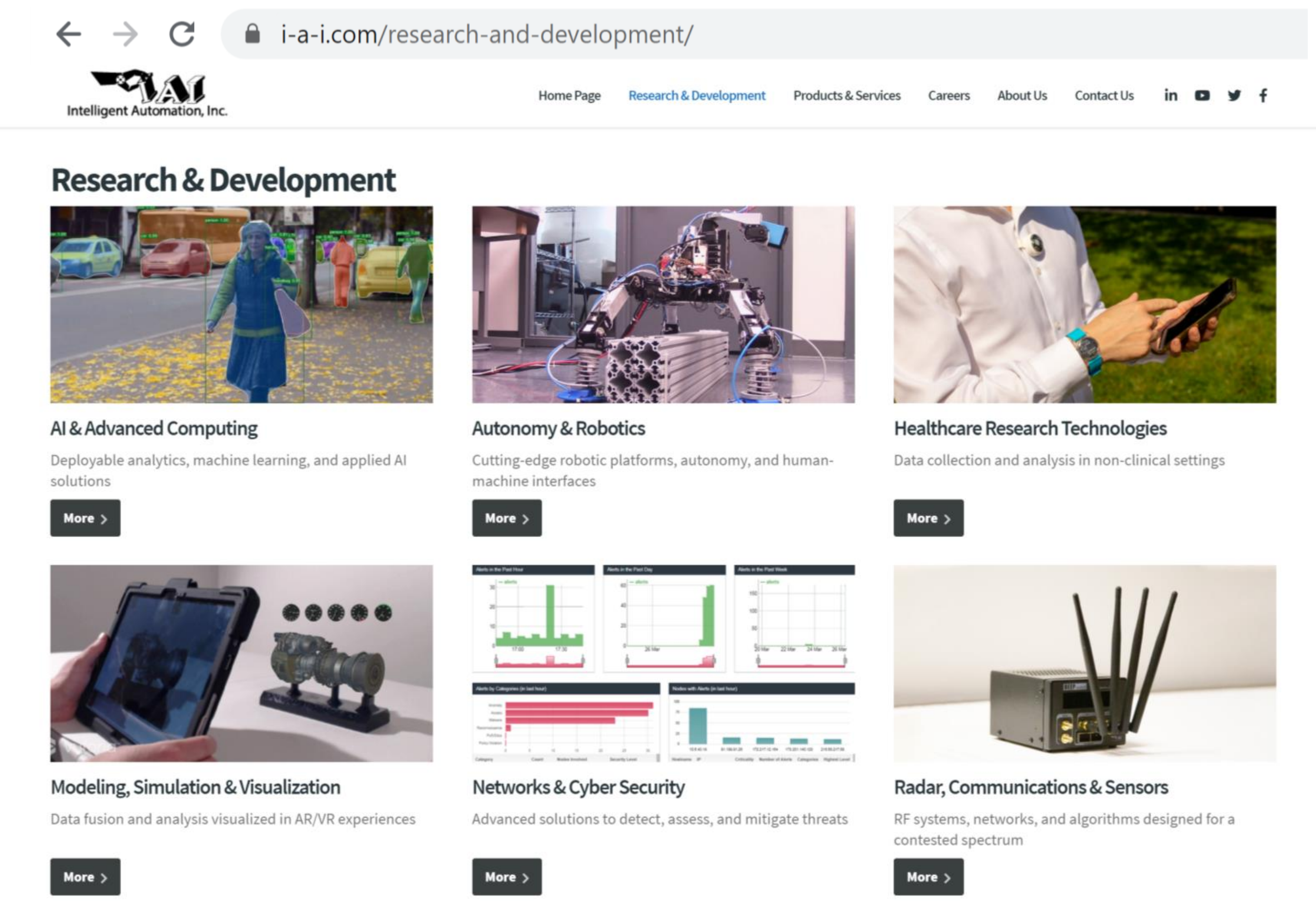
Adversarial Machine Learning and Wireless Security for 5G and Beyond

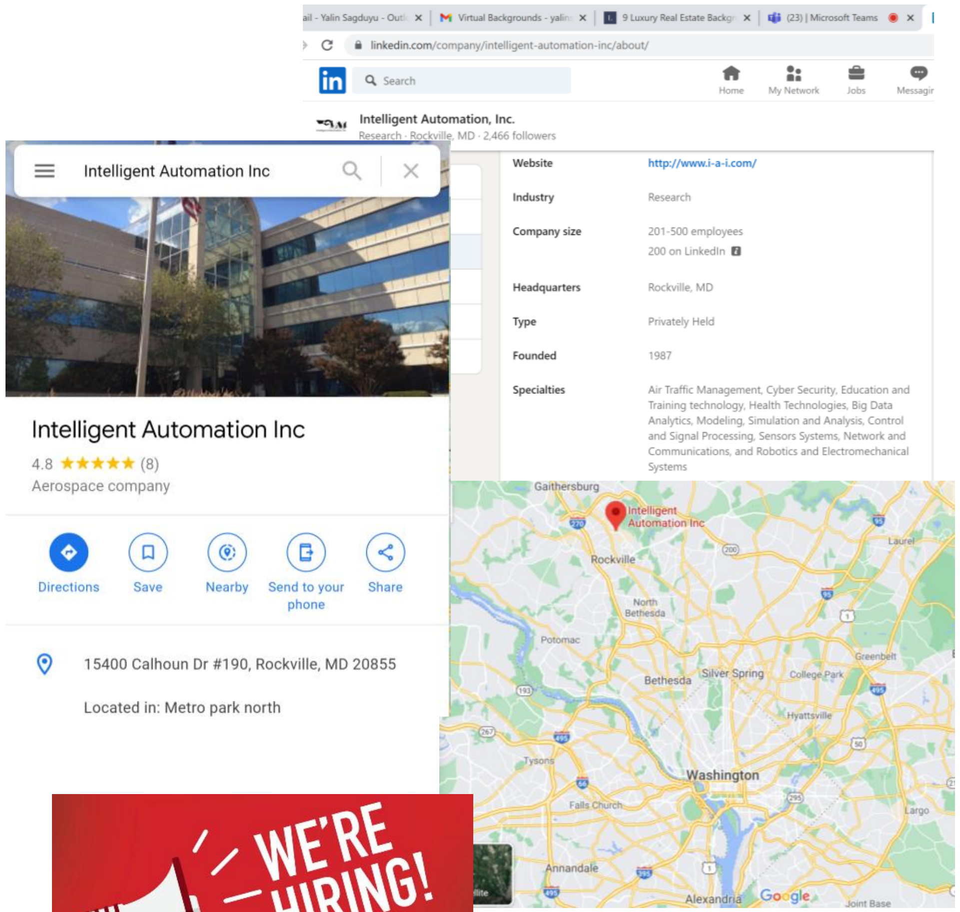
Yalin Sagduyu and Tugba Erpek

Intelligent Automation, Inc.

Who are we?



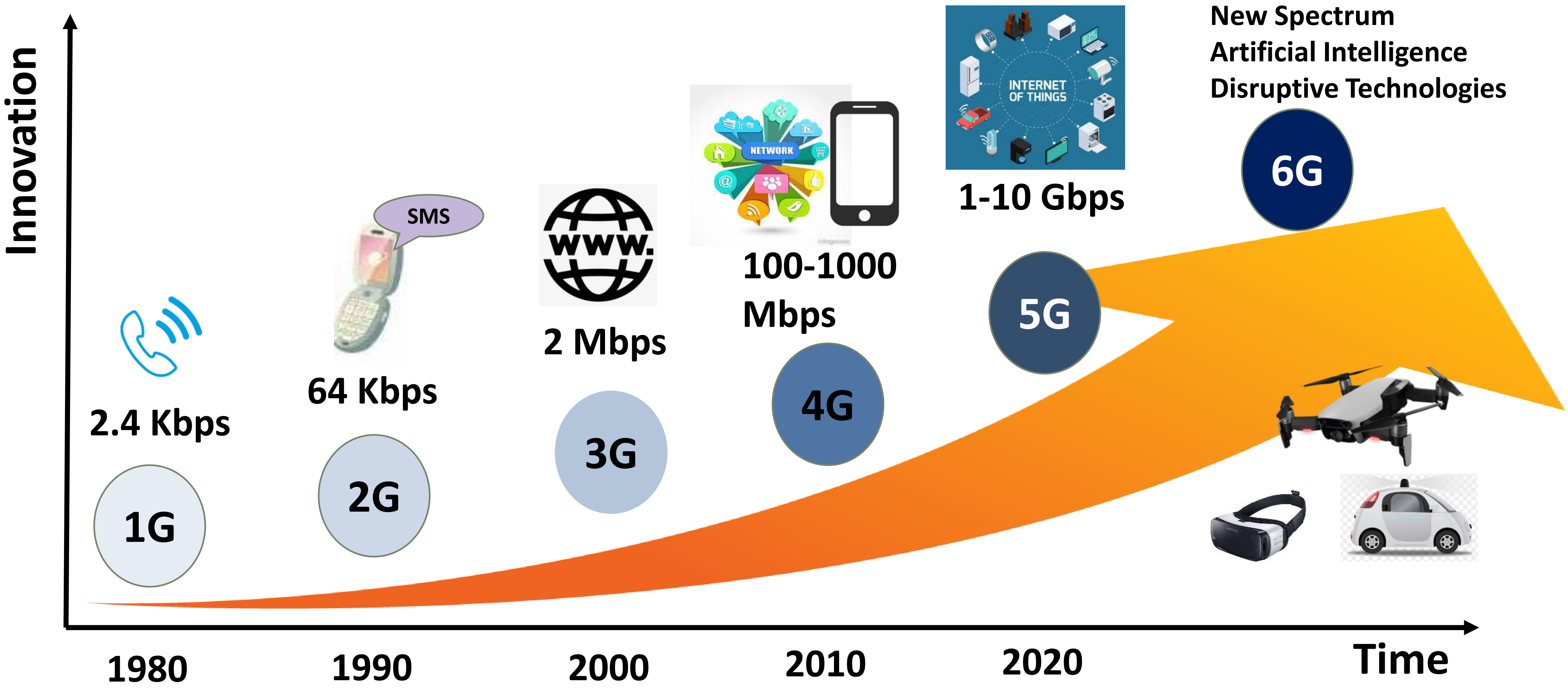
i-a-i.com/research-and-development/
Intelligent Automation, Inc.
Home Page
Research & Development
Products & Services
Careers
About Us
Contact Us
Research & Development
AI & Advanced Computing
Deployable analytics, machine learning, and applied AI solutions
More
Autonomy & Robotics
Cutting-edge robotic platforms, autonomy, and human-machine interfaces
More
Healthcare Research Technologies
Data collection and analysis in non-clinical settings
More
Modeling, Simulation & Visualization
Data fusion and analysis visualized in AR/VR experiences
More
Networks & Cyber Security
Advanced solutions to detect, assess, and mitigate threats
More
Radar, Communications & Sensors
RF systems, networks, and algorithms designed for a contested spectrum
More



linkedin.com/company/intelligent-automation-inc/about/
Intelligent Automation, Inc.
Research · Rockville, MD · 2,466 followers
Website http://www.i-a-i.com/
Industry Research
Company size 201-500 employees
200 on LinkedIn
Headquarters Rockville, MD
Type Privately Held
Founded 1987
Specialties Air Traffic Management, Cyber Security, Education and Training technology, Health Technologies, Big Data Analytics, Modeling, Simulation and Analysis, Control and Signal Processing, Sensors Systems, Network and Communications, and Robotics and Electromechanical Systems
Intelligent Automation Inc
4.8 ★★★★★ (8)
Aerospace company
Directions
Save
Nearby
Send to your phone
Share
15400 Calhoun Dr #190, Rockville, MD 20855
Located in: Metro park north
Gaithersburg
Rockville
Bethesda
Silver Spring
Washington
Alexandria
Google

WE'RE HIRING!

Wireless Evolution

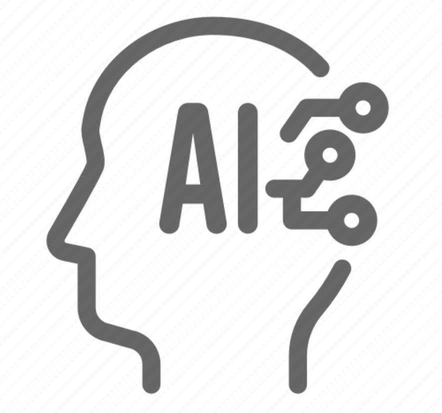


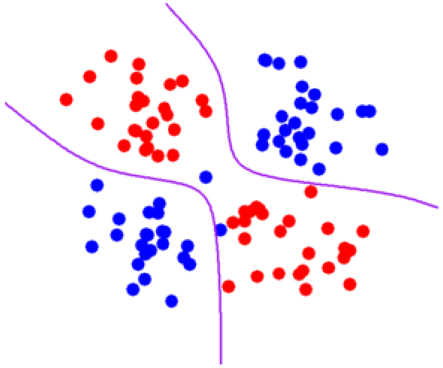
Machine Learning Evolution

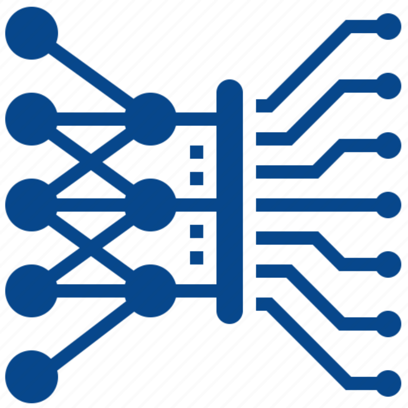
Engineering of making intelligent machines and programs

Ability to learn without being explicitly programmed

Learning using deep neural networks







Artificial Intelligence | **Machine Learning** | **Deep Learning**

1950 1960 1970 1980 1990 2000 2006 2010 2012 2017 **Time**

Machine Learning Software Tools & Datasets

TensorFlow

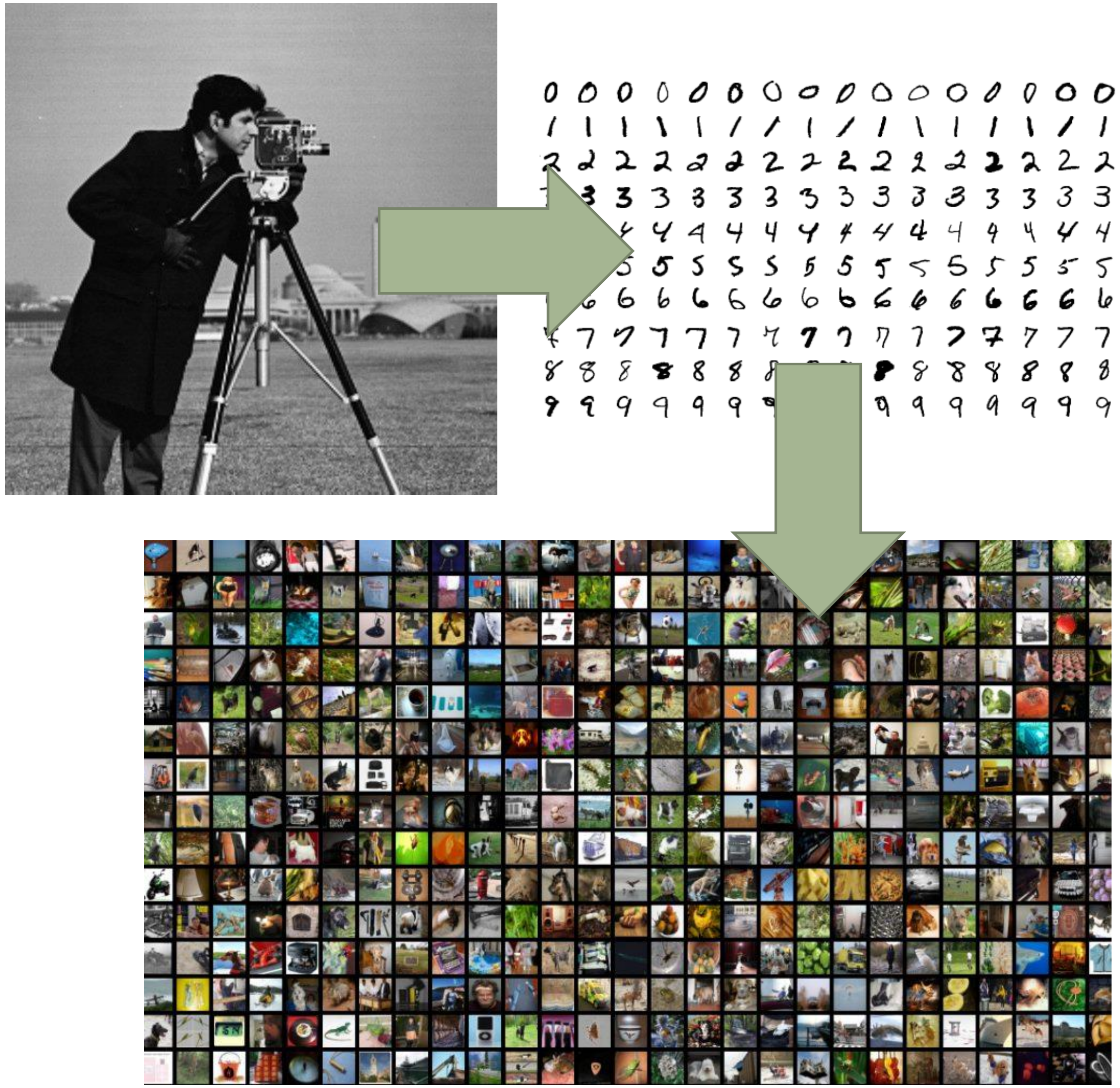
theano DEPRECATED

Keras

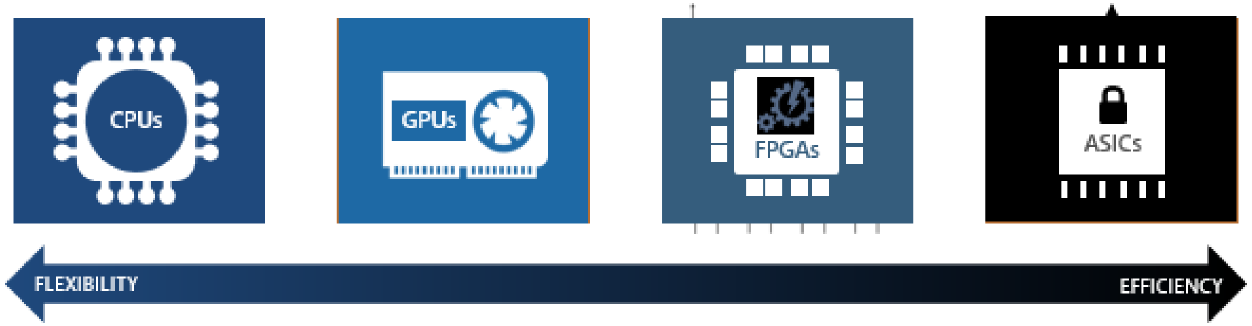
PyTorch

DEPRECATED Microsoft Cognitive Toolkit

MATLAB
NN Toolbox



Machine Learning Computational Tools



https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-fpga-web-service

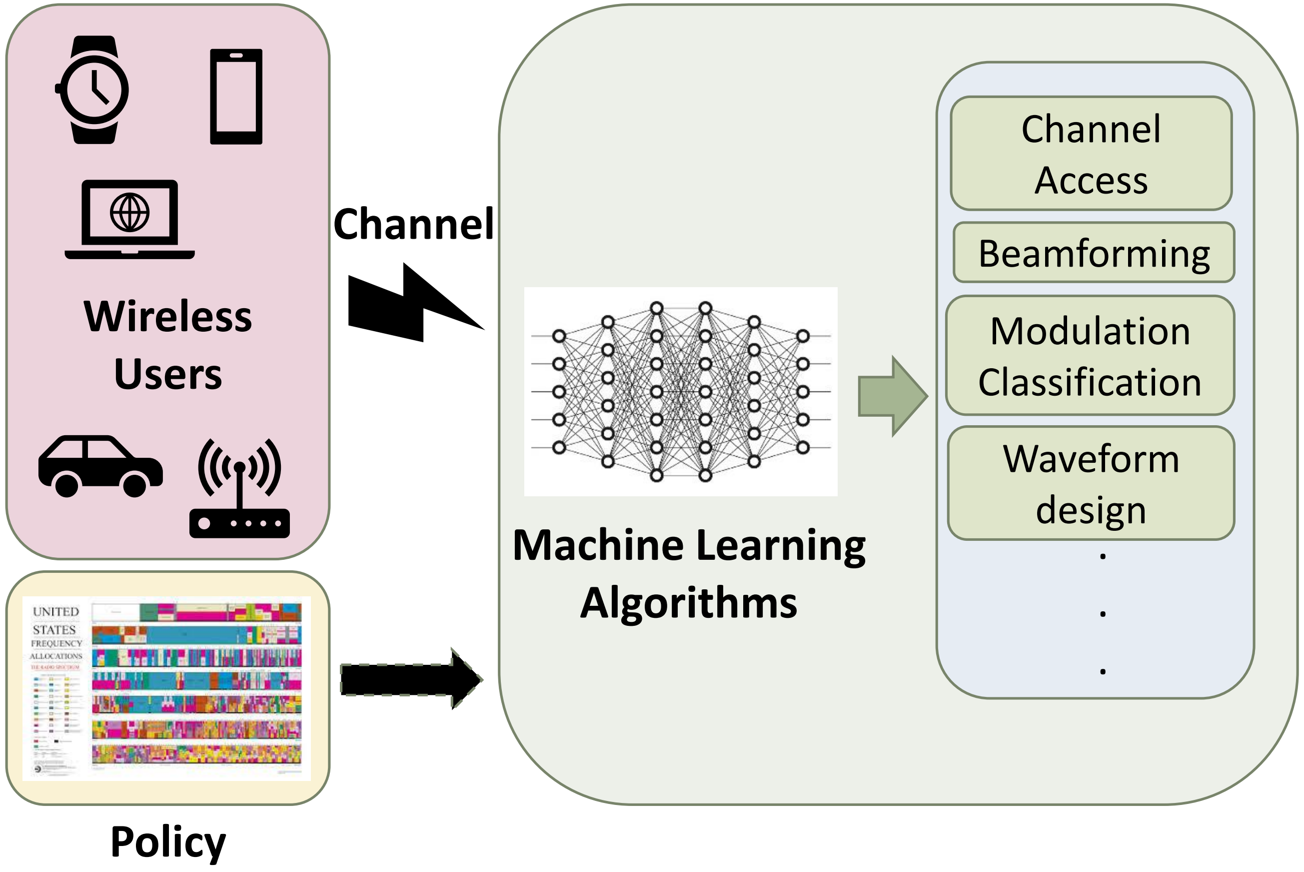
Google Cloud TPU

From cloud backend to embedded platforms

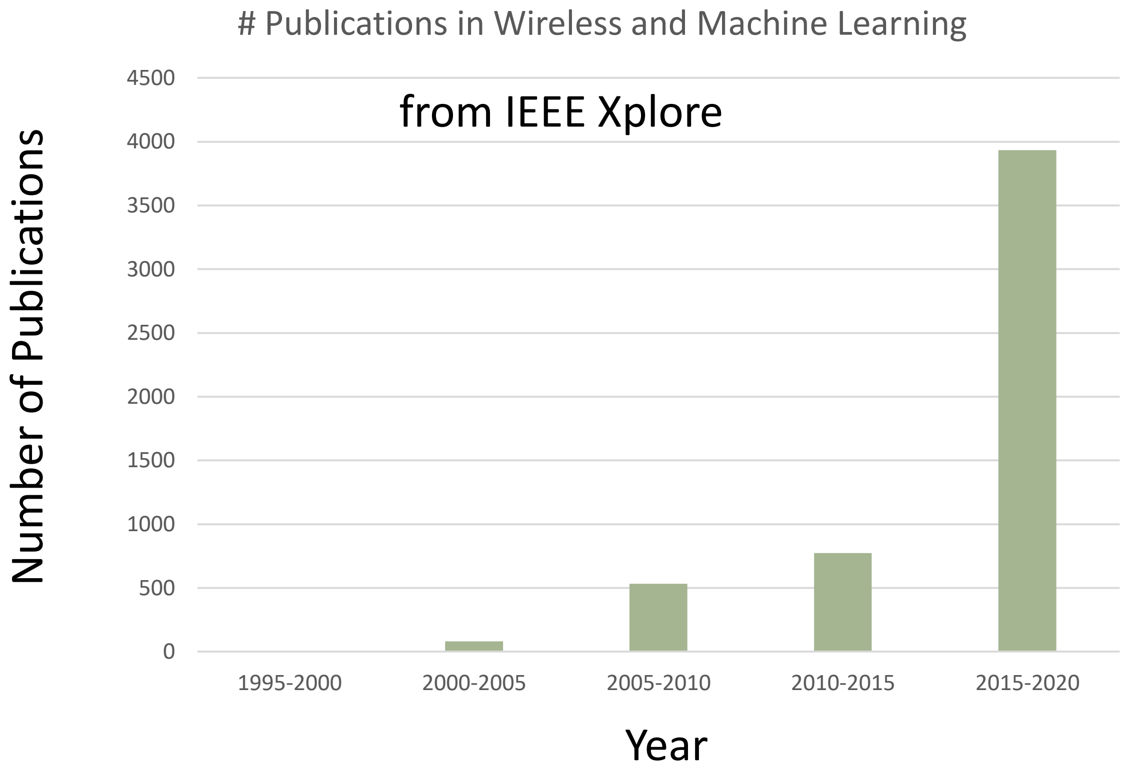
Nvidia Nano

DEEPip
Deep Learning on FPGA Fabric

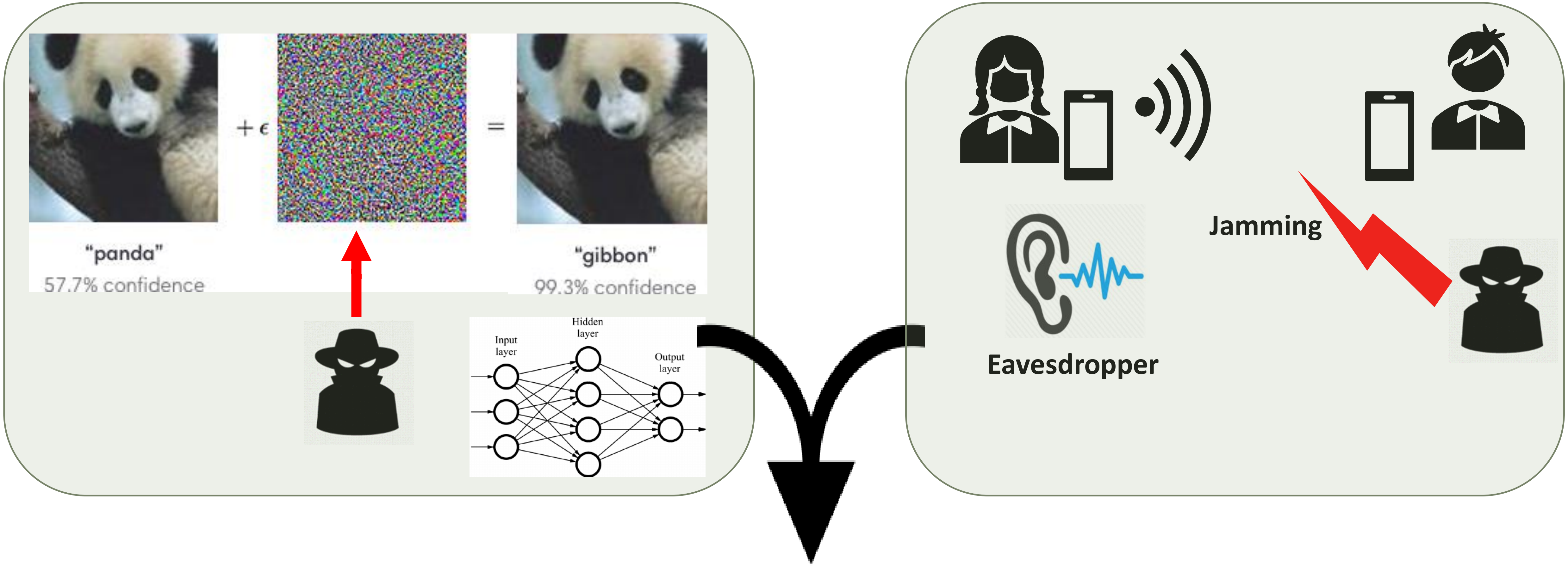
Machine Learning for Wireless



Increasing interest in wireless communications and machine learning research



Machine Learning/Wireless Security



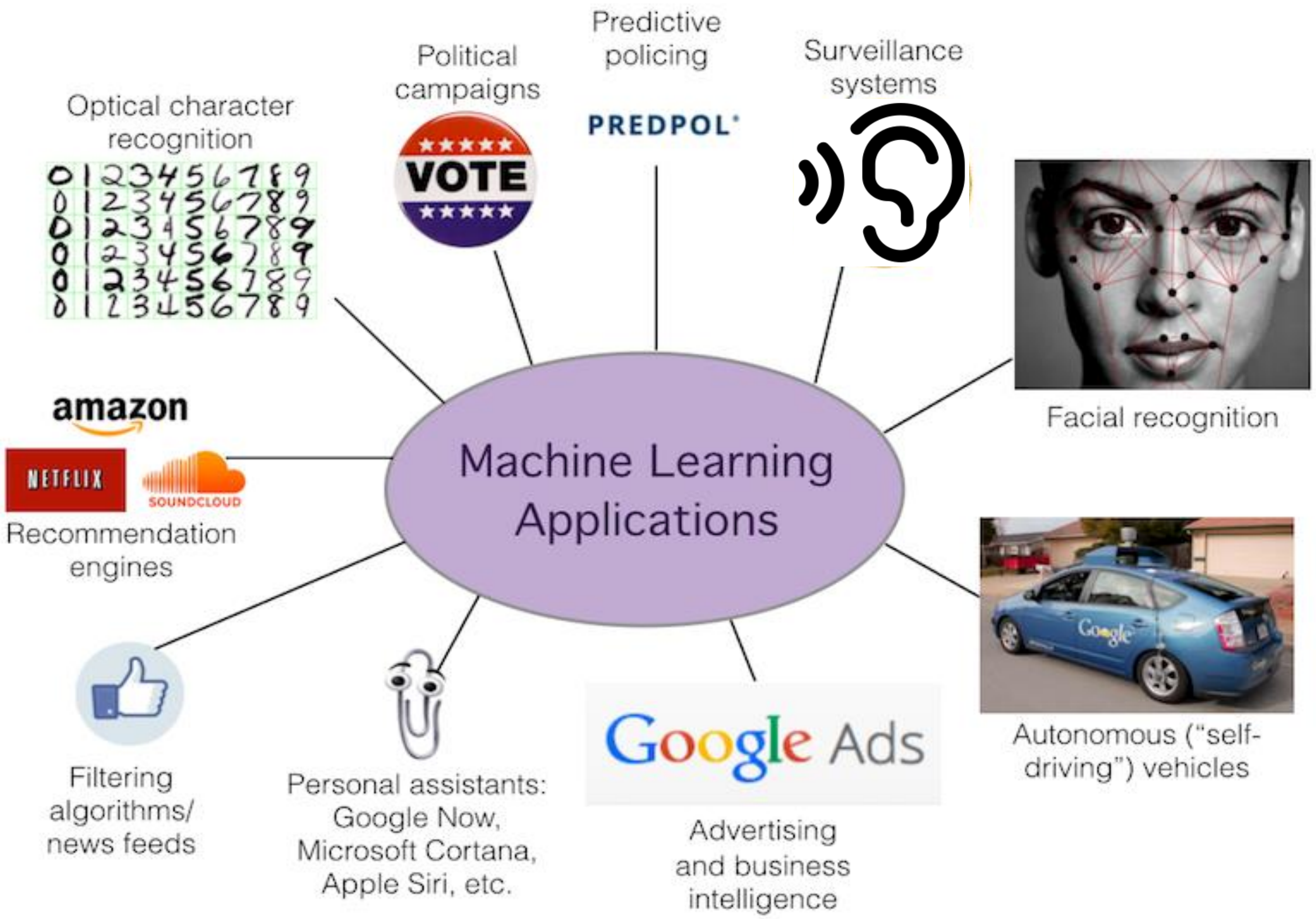
Adversarial Machine Learning for Wireless

Outline

- **Machine Learning**
- **Machine Learning for Wireless**
- **Machine Learning for 5G and Beyond**
- **Adversarial Machine Learning**
- **Adversarial Machine Learning for Wireless**
- **Adversarial Machine Learning for 5G and Beyond**
- **Conclusion**

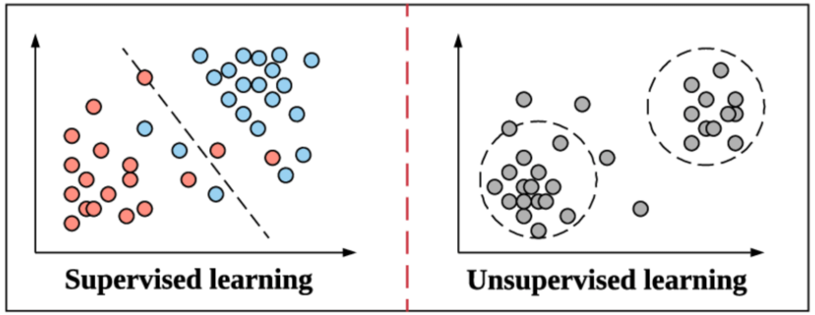
Machine Learning - 1

- Automated means to learn from data and solve (complex) tasks.
- Far-reaching applications:
  - Document classification
  - Search engines
  - Social media/network platforms
  - Intelligence analysis applications
  - Intrusion detection
  - Bot detection
  - Recommender systems
  - Online review systems
  - Spam email filtering
  - Internet of Things
  - Cyberphysical systems
  - Autonomous driving
  - Unmanned vehicle controllers

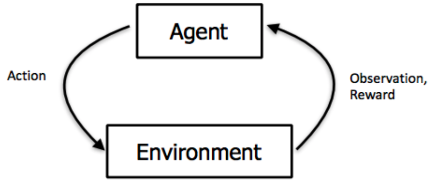


Machine Learning - 2

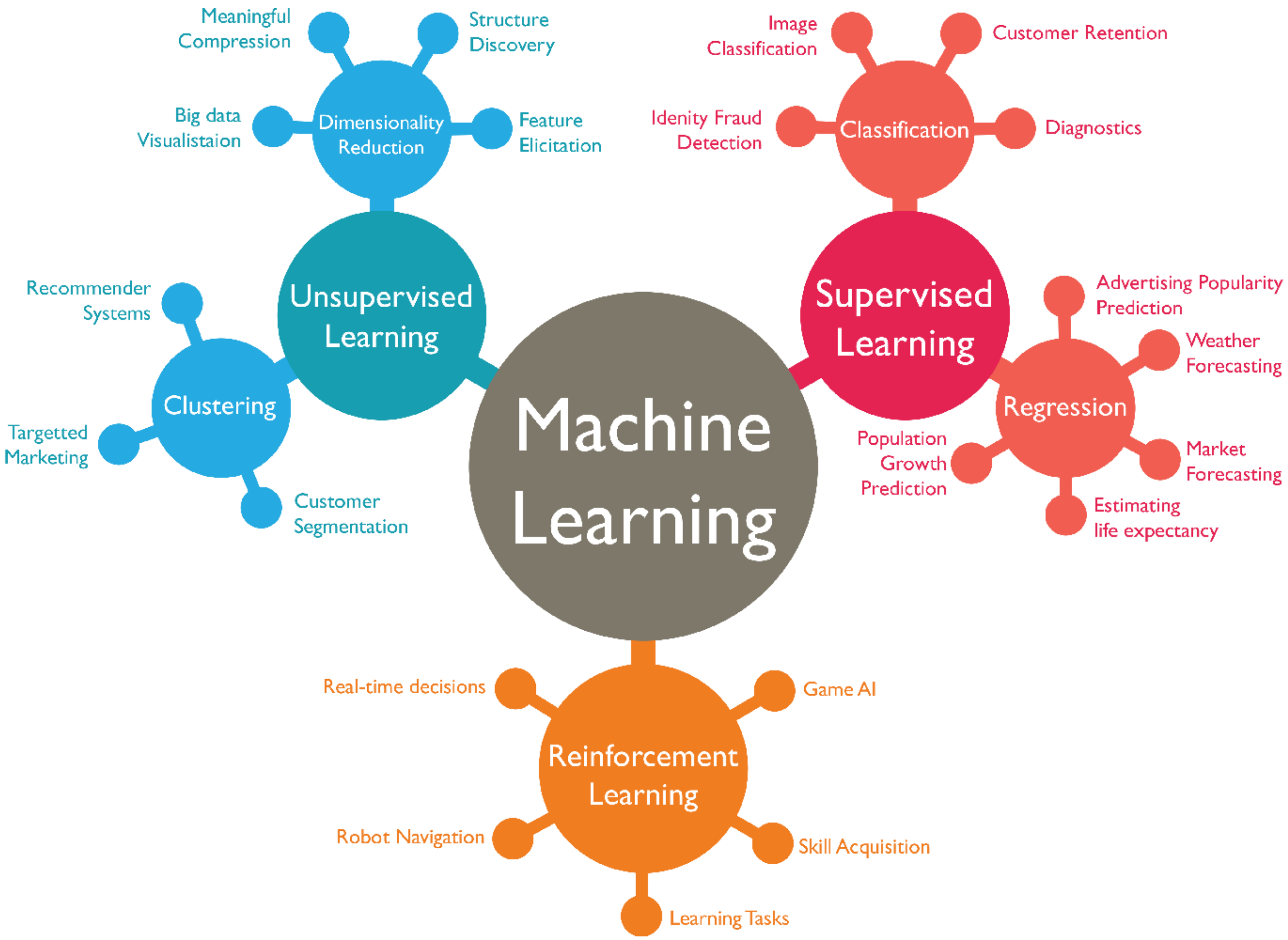
- **Supervised Learning**
  - Labeled data
  - *Example*: Classification
- **Unsupervised Learning**
  - No labeled data
  - *Example*: Feature extraction



- **Reinforcement Learning**



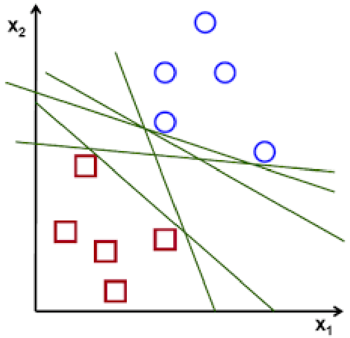
  - *Example*: Model-less learning on the fly



https://www.linkedin.com/pulse/business-intelligence-its-relationship-big-data-geekstyle

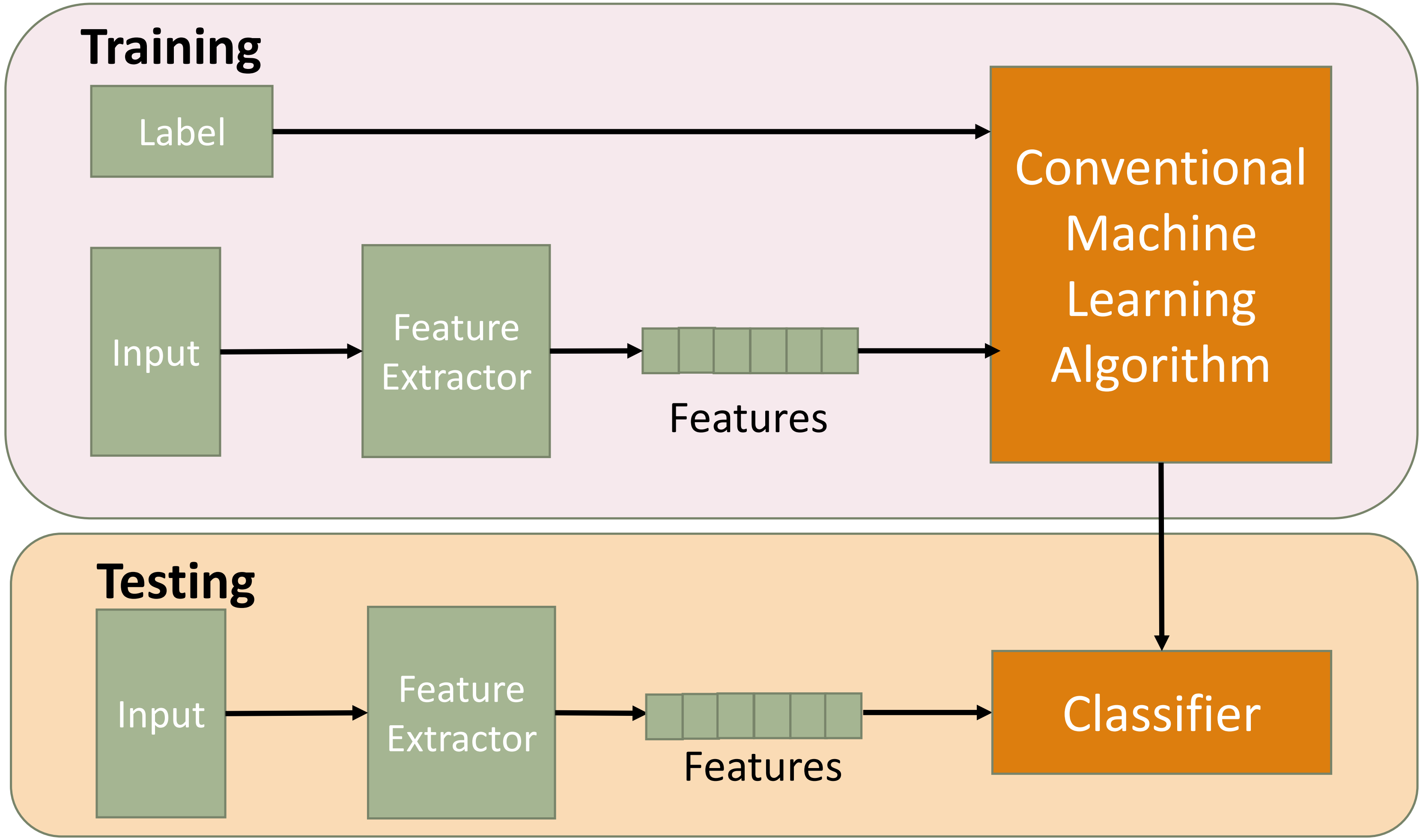
Conventional Machine Learning Algorithms

- Support Vector Machine (SVM)



- Decision Trees
- Random Forests

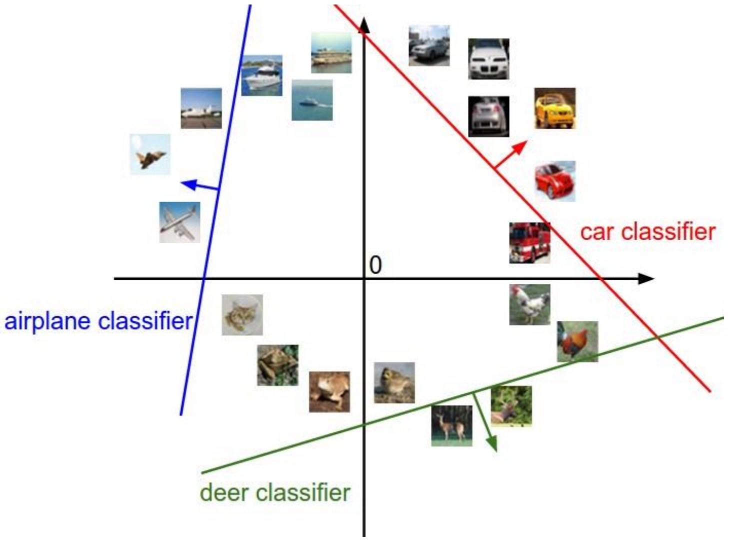
among others.

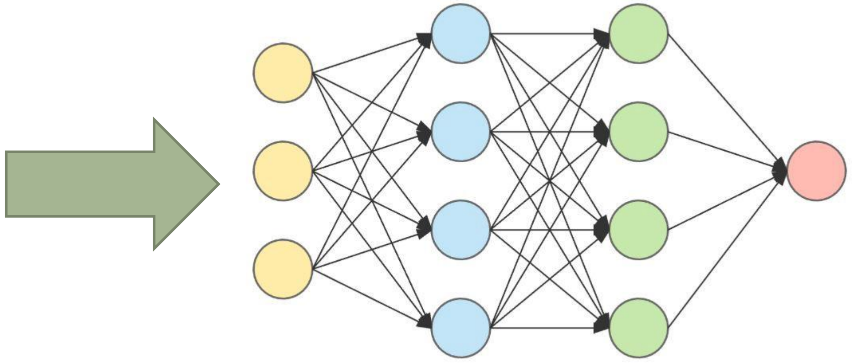


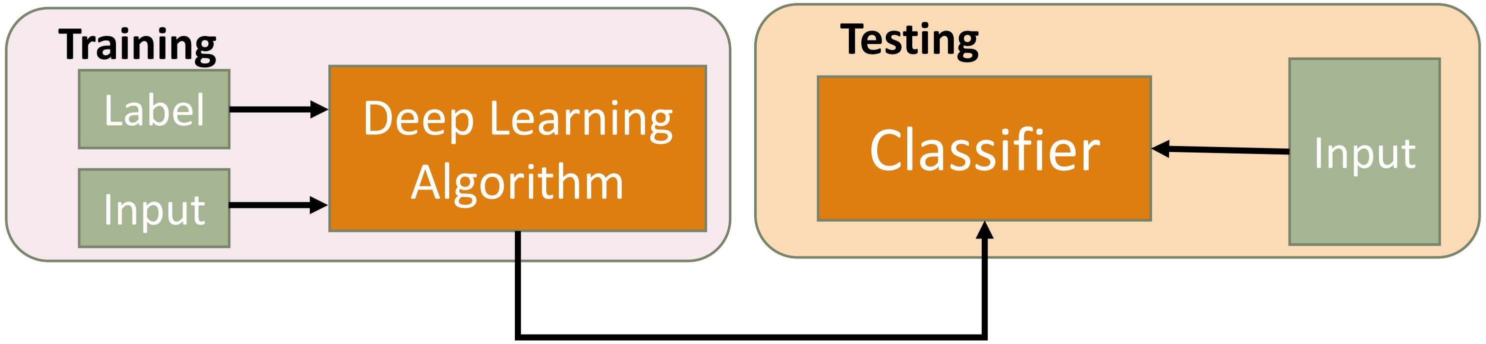
From Machine Learning to Deep Learning

- **Deep neural networks**
  - **Algorithmic** advances (e.g., back-propagation)
  - **Computational** advances (e.g., cloud back-ends)
  - Expansion of **training data** (e.g., sensors).
  - **Open-source software** (e.g., TensorFlow).

- Can effectively solve **complex tasks**.

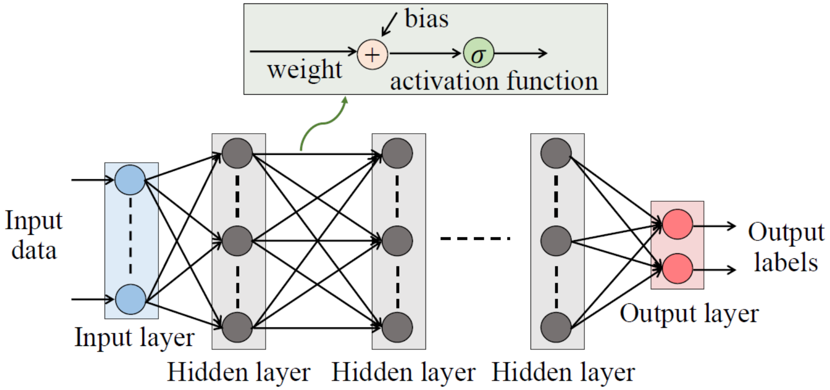




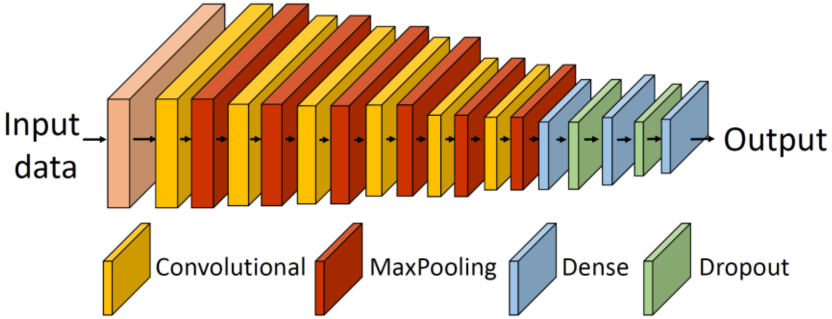


Common Types of Deep Neural Networks

Feedforward neural network (FNN)

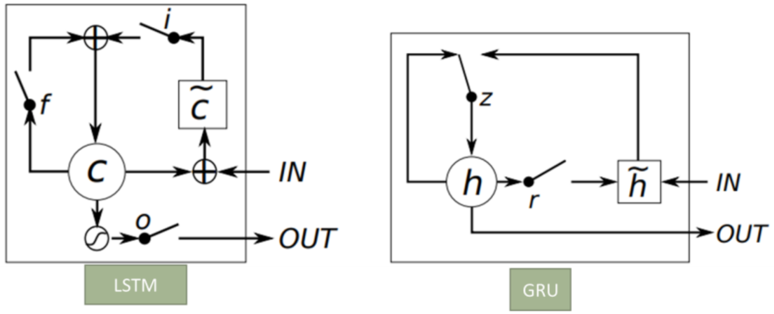


Convolutional Neural Network (CNN)



- captures spatial correlations in data
- example: computer vision

Recurrent Neural Network (RNN)

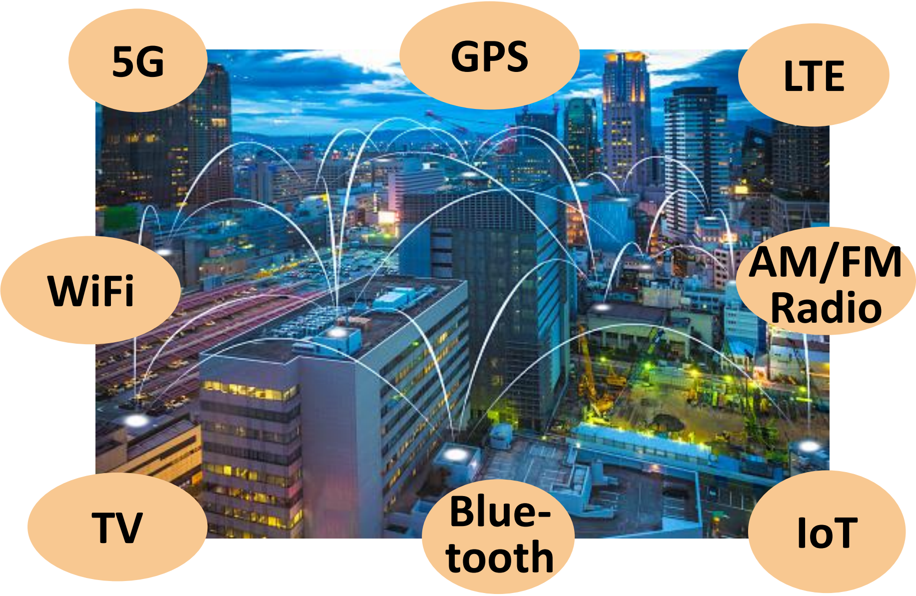


- captures temporal correlations in data
- example: computer vision

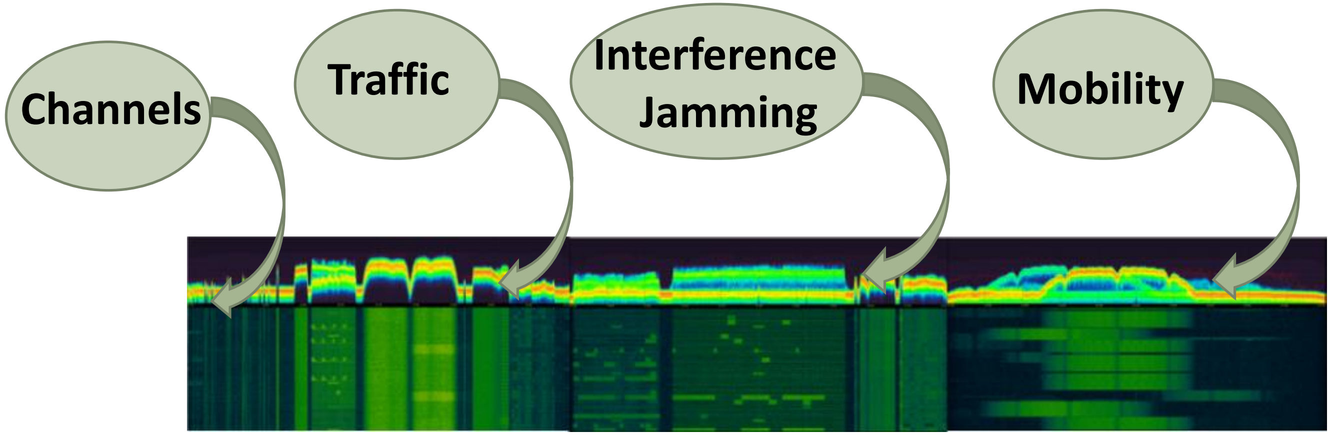
Outline

- Machine Learning
- **Machine Learning for Wireless**
- Machine Learning for 5G and Beyond
- Adversarial Machine Learning
- Adversarial Machine Learning for Wireless
- Adversarial Machine Learning for 5G and Beyond
- Conclusion

Wireless (Spectrum) Data is Complex



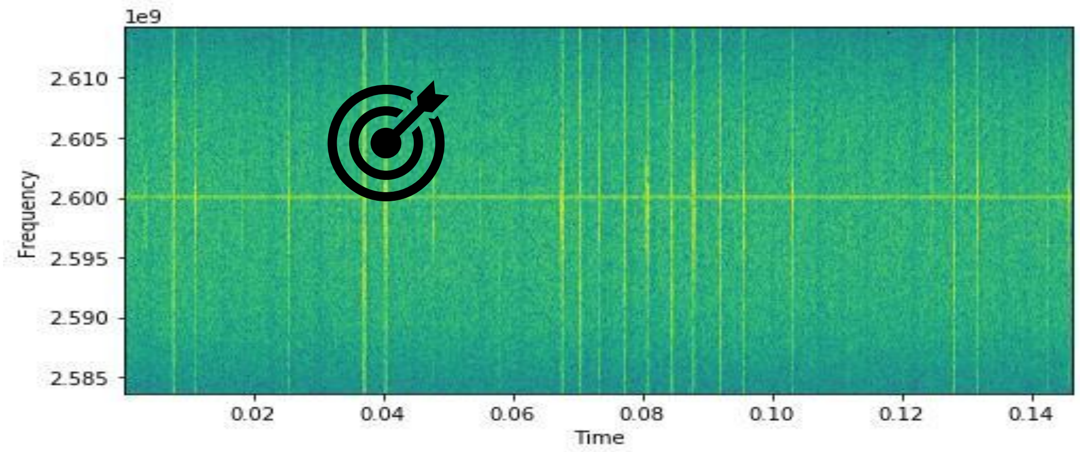
- Connectivity
- Smart City
- Smart Warehouse
- Augmented/Virtual Reality
- UAV/drone Networks



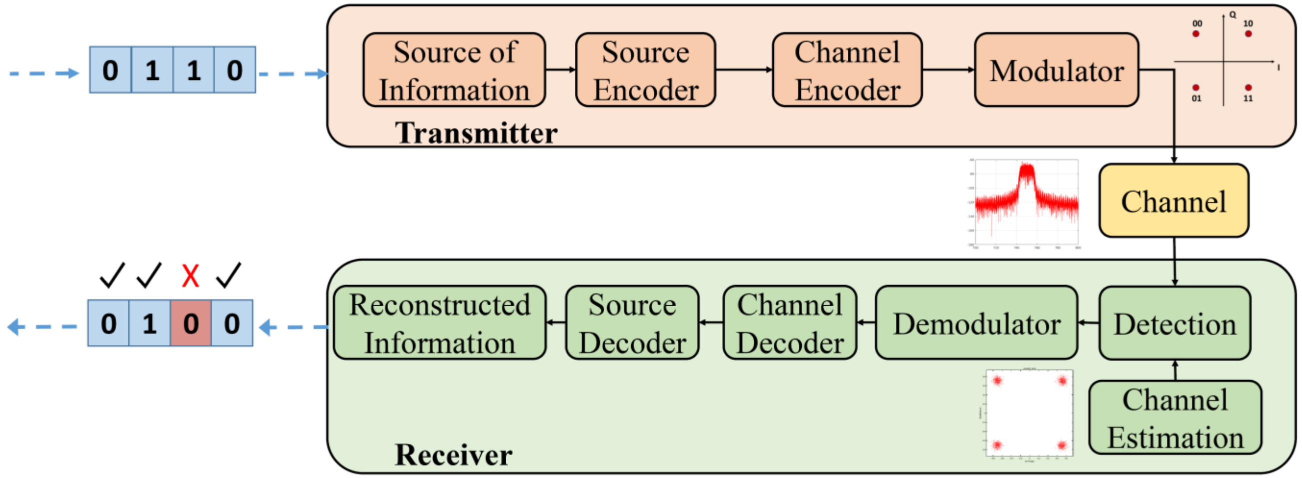
- I/Q (RF) data
- RSSI (signal strength)
- Spatial beam pattern
- Protocol performance measures (throughout, delay, etc.)

Wireless Tasks are Complex

Signal Analysis

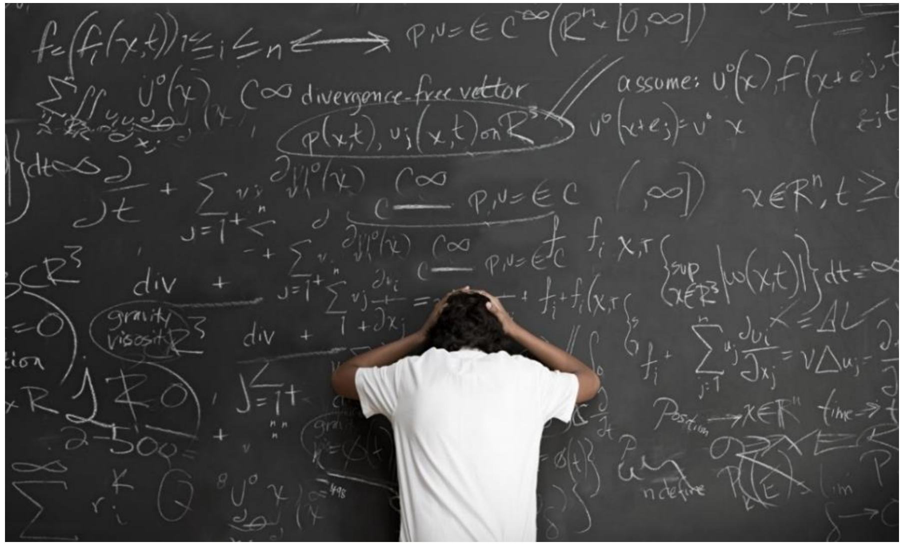


Waveform/Protocol Design



Machine/Deep Learning for Wireless

- Expert knowledge & analytical solutions cannot capture complex waveforms, channels, and resources of wireless.

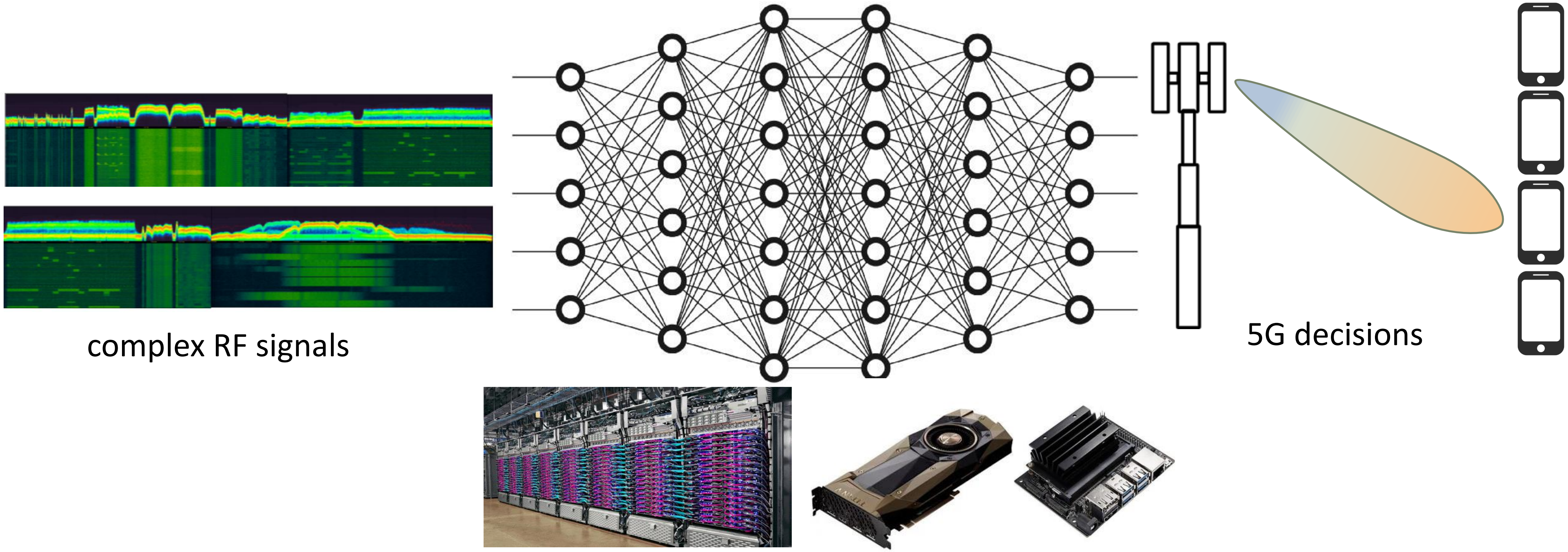


vs.

Machine/deep learning provides automated means to learn from spectrum data and solve complex spectrum tasks.

From Conventional ML to Deep Learning

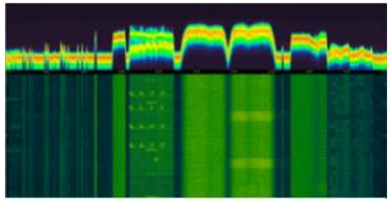
- Conventional ML techniques fall short from capturing **complex spectrum dynamics**.
- Deep learning finds rich applications in wireless domain.



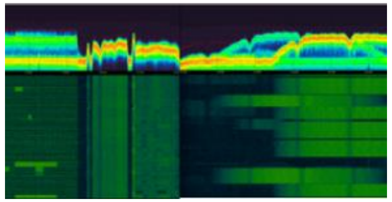
from high performance to embedded computing

Deep Learning for Wireless

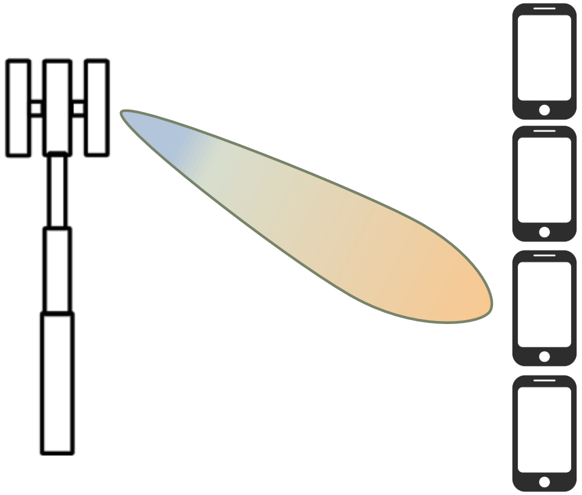
Signal Detection/ Classification



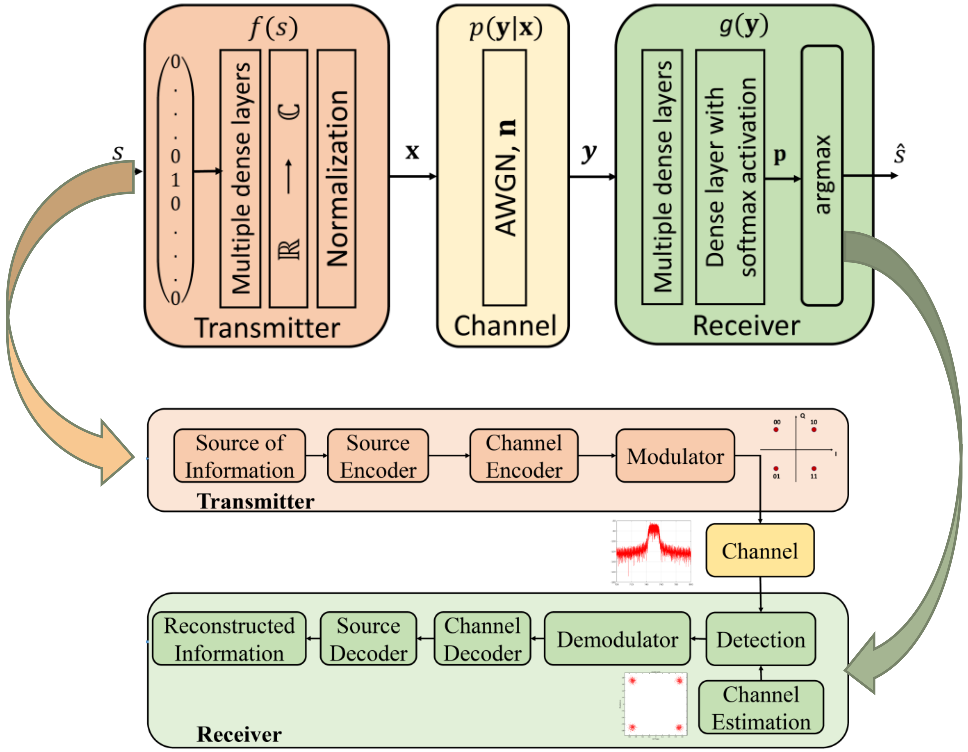
versus



Waveform/Protocol Optimization



Deep Neural Networks Communication System



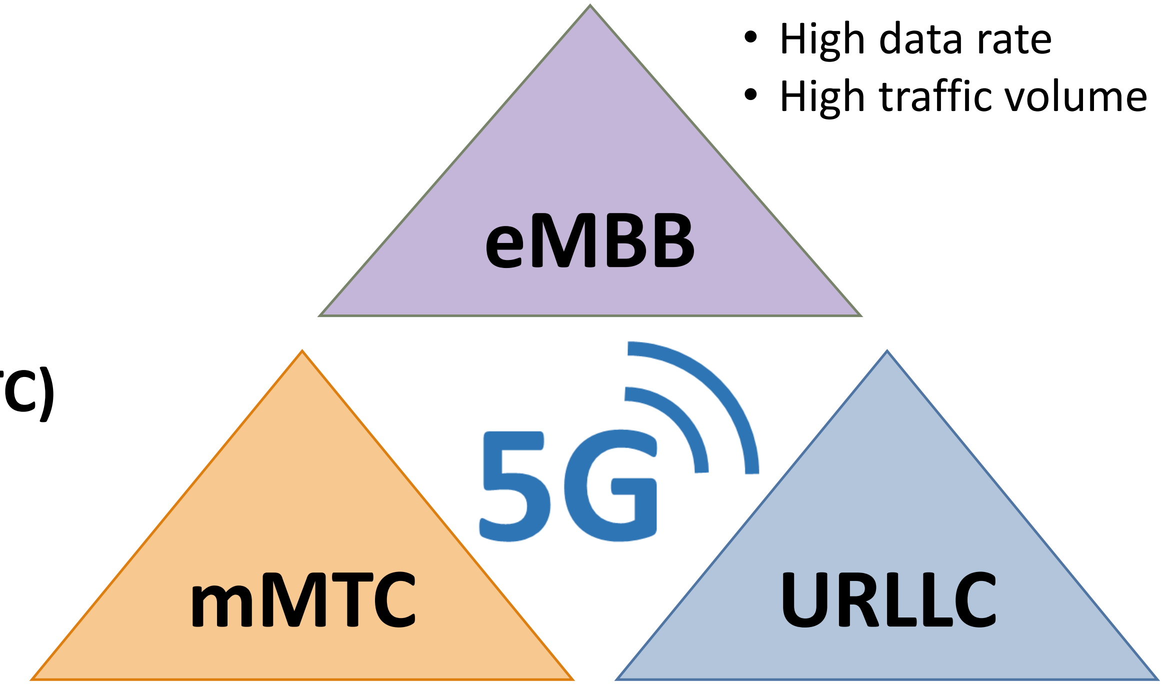
Outline

- **Machine Learning**
- **Machine Learning for Wireless**
- **Machine Learning for 5G and Beyond**
- **Adversarial Machine Learning**
- **Adversarial Machine Learning for Wireless**
- **Adversarial Machine Learning for 5G and Beyond**
- **Conclusion**

5G as a Complex Ecosystem

- **Enhanced Mobile Broadband (eMBB)**
  - Virtual/Augmented Reality
  - Mobile Office
  - Entertainment
- **Massive Machine Type Communications (mMTC)**
  - Smart Cities
  - Manufacturing
  - Supply Chain/Logistics
- **Ultra Reliable Low Latency Communications (URLLC)**
  - Autonomous Vehicles
  - Emergency Services
  - Healthcare

- High data rate
- High traffic volume

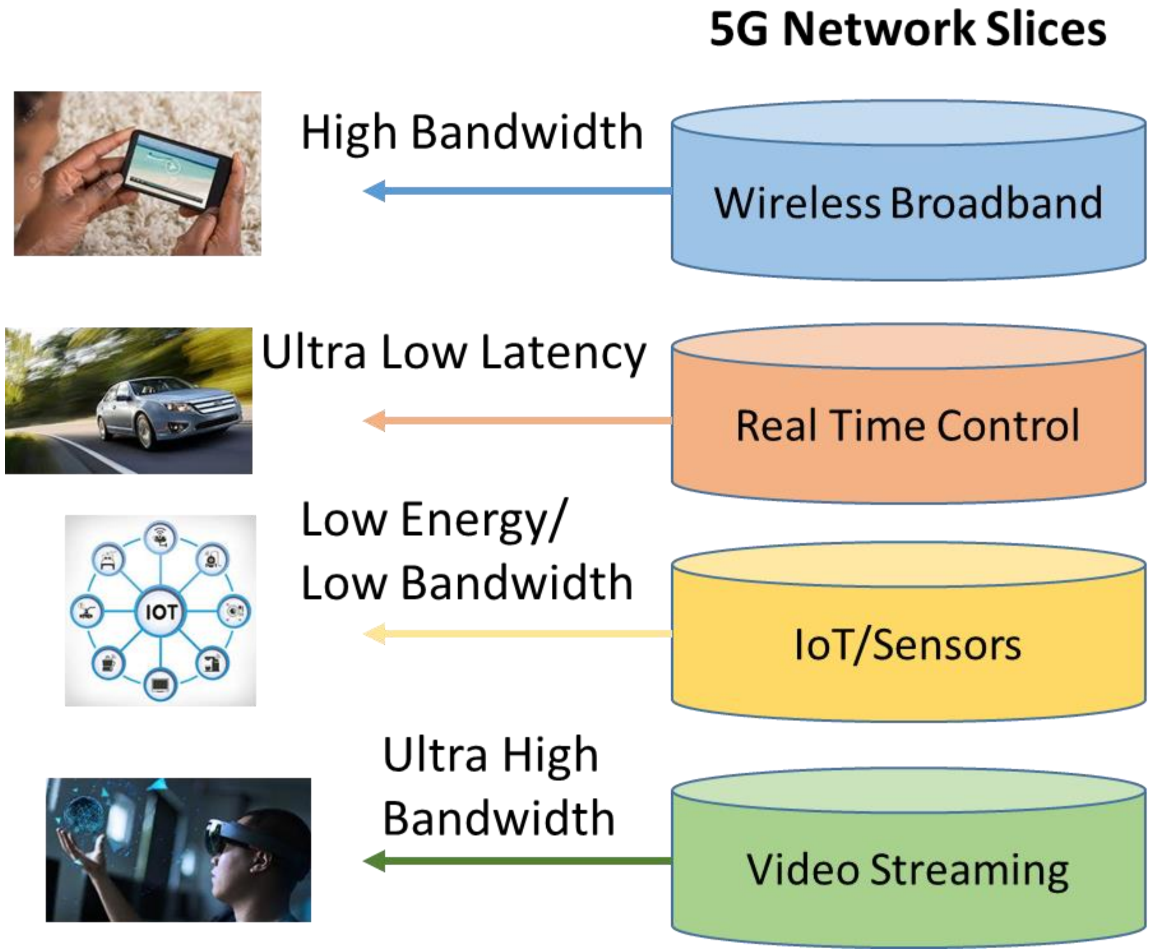


- Massive number of low-cost devices
- Low energy consumption

- Low latency
- High reliability

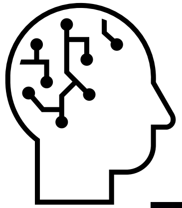
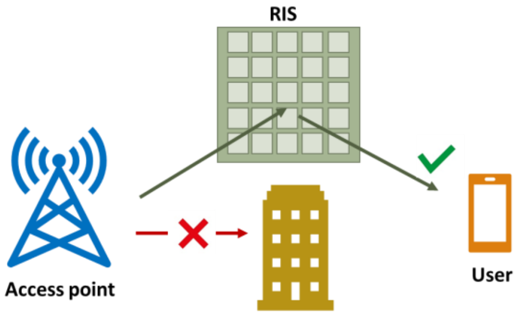
Advanced Capabilities Offered by 5G

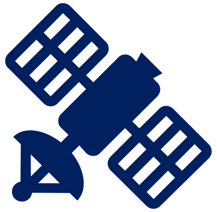
- From sub-6GHz to **mmWave**
- **Massive MIMO**
- Multiple services on shared physical infrastructure through **network slicing**
- **Low-latency edge computing**
- Improved **energy efficiency**

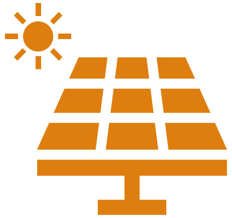


Beyond 5G

- x100 throughput of 5G
- Distributed edge cloud
- Distributed data and AI
- Federated and dynamic learning
- Ultra high frequency spectrum
- Reconfigurable intelligent surfaces
- Volumetric spectrum efficiency
- Software-defined network and access
- Energy transfer and harvesting
- Integrated terrestrial, airborne and satellite networks
- Hologram communications

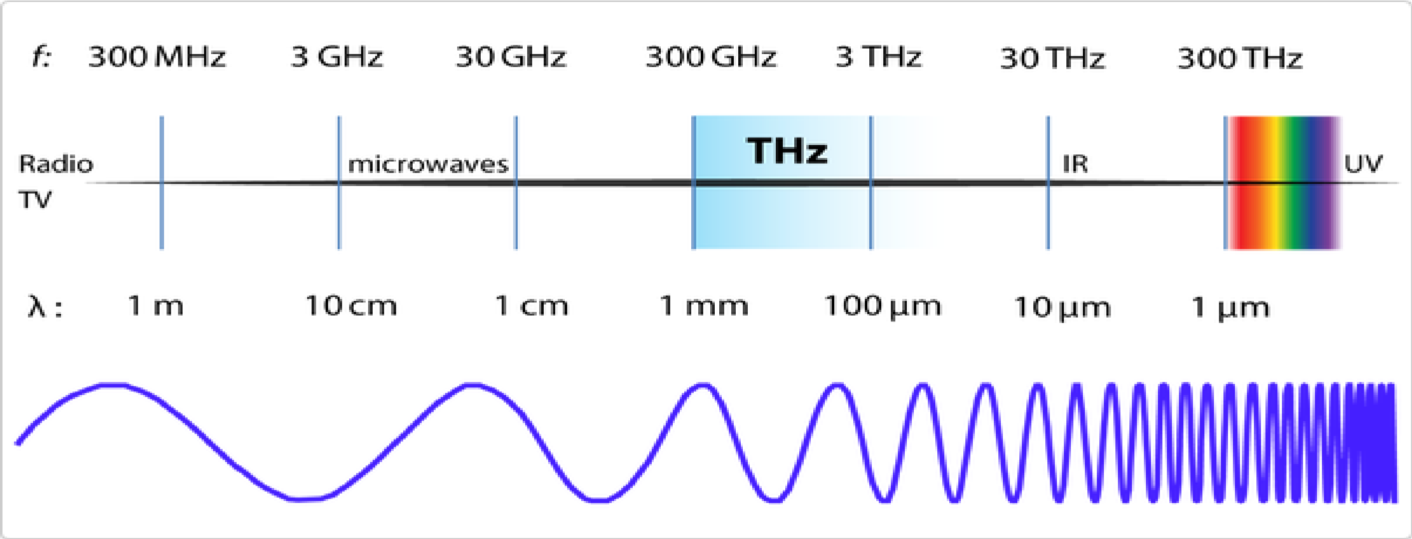





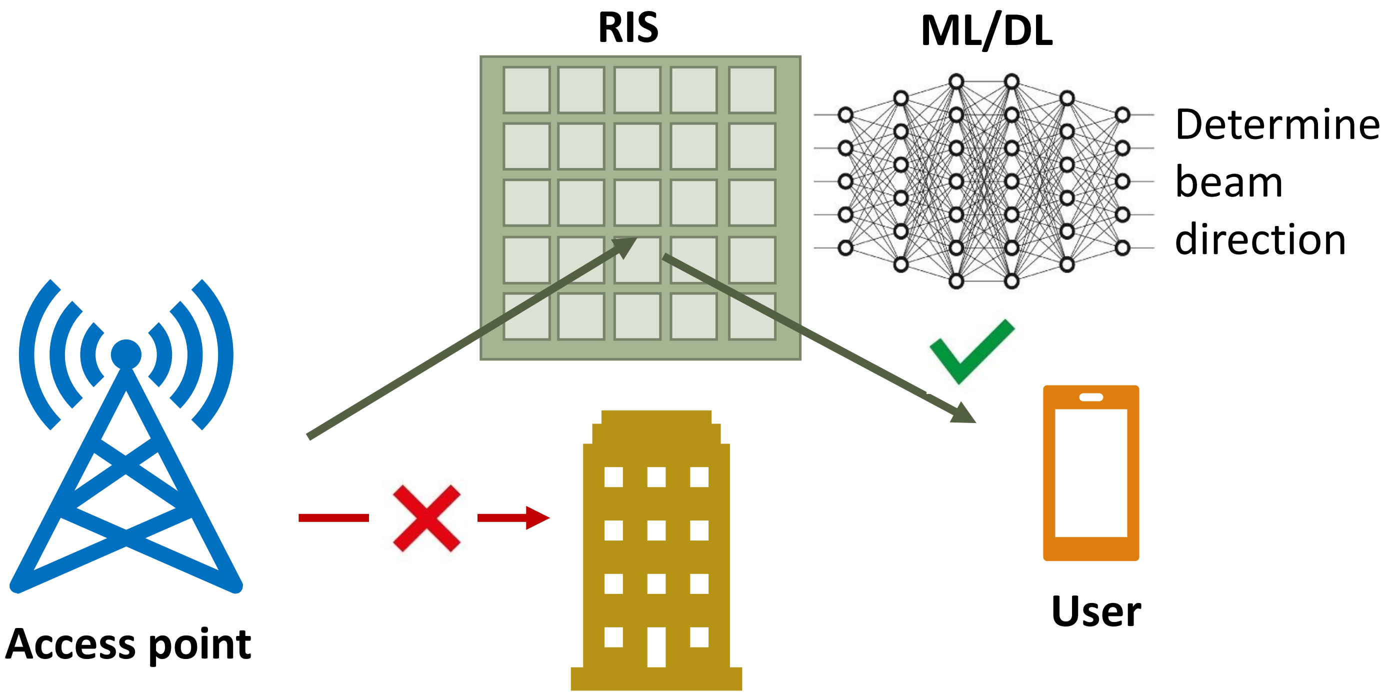
Terahertz Communications

- THz provides unprecedented rates not supported in 5G and before.
  - Highly-directional and secure transmissions.
  - Ultra-low latency (e.g., Augmented reality/virtual reality).
- Challenge: Link maintenance and support of high mobility.
- ML/DL for fast beam training, beam switching and handoff.

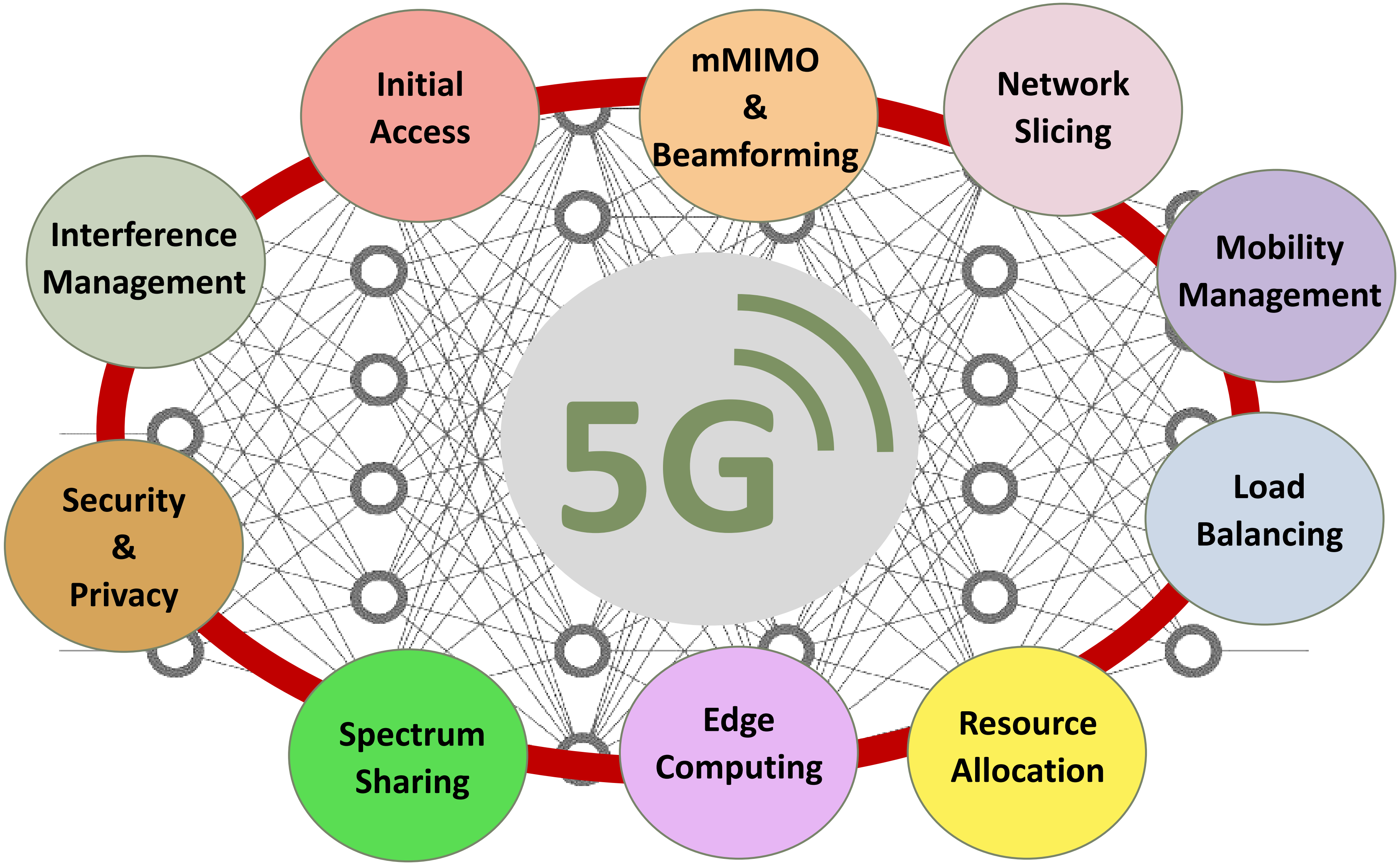


Reconfigurable Intelligent Surfaces (RISs)

- Reflect and focus the signals towards the receivers.
- Enhance coverage in mmWave & THz systems in face of blockages.



Machine Learning for 5G and Beyond



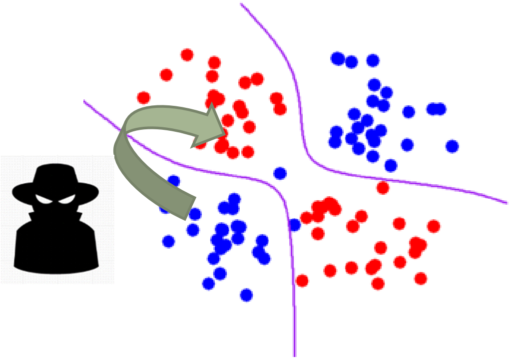
Outline

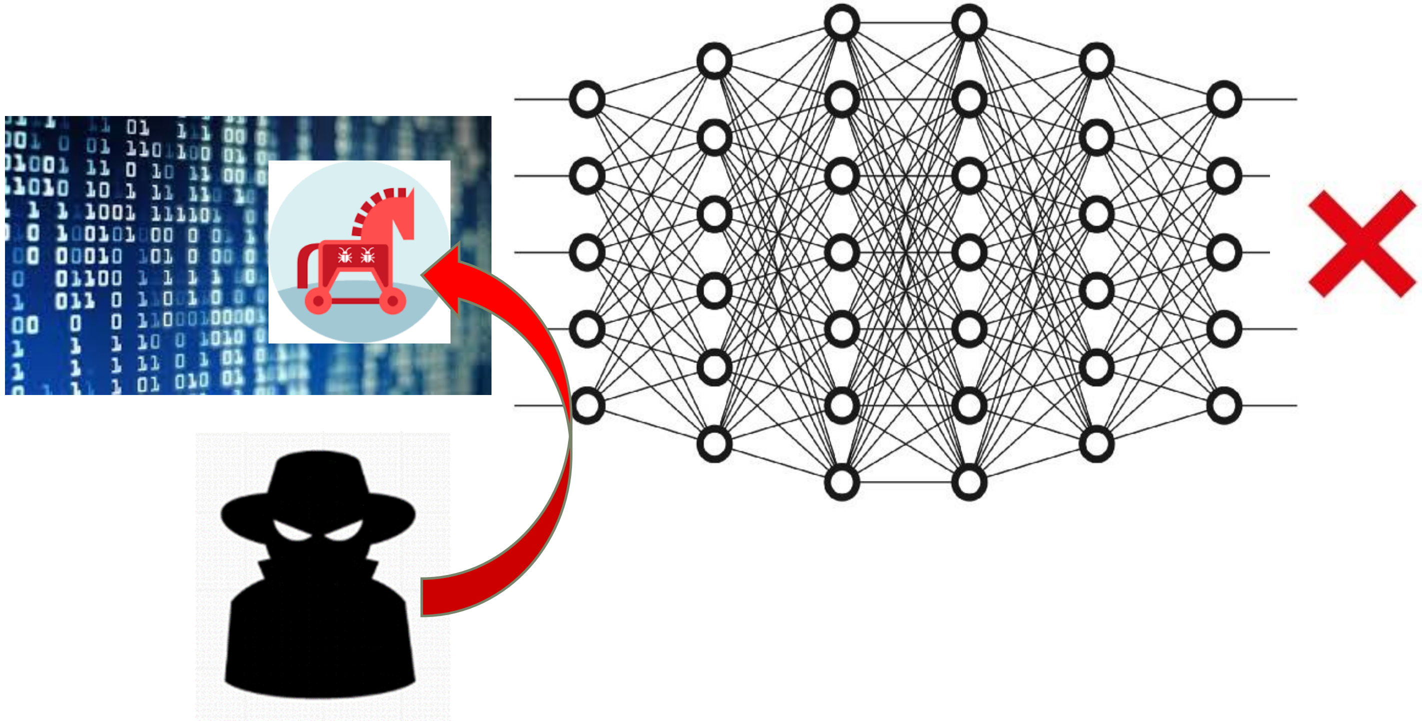
- **Machine Learning**
- **Machine Learning for Wireless**
- **Machine Learning for 5G and Beyond**
- **Adversarial Machine Learning**
- **Adversarial Machine Learning for Wireless**
- **Adversarial Machine Learning for 5G and Beyond**
- **Conclusion**

Security Vulnerabilities of Machine Learning

- **Tamper with the learning process** and **fool deep learning algorithms** into making errors.

- Complex decision space of deep learning is **sensitive to small adversarial inputs**.

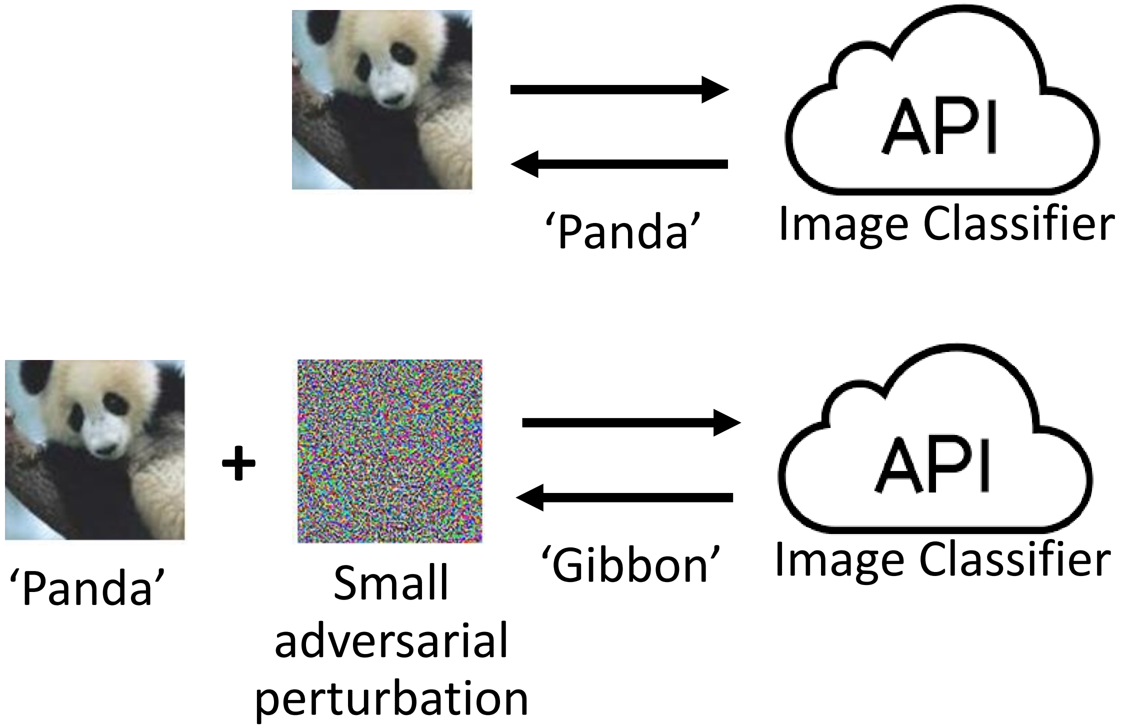




Deep learning itself is vulnerable to attacks.

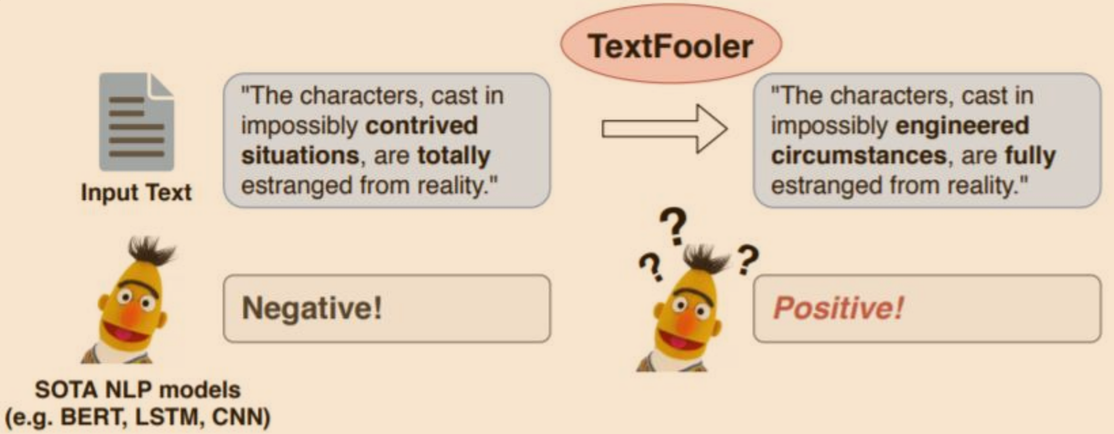
Adversarial Machine Learning Example

- How effective learning can take place under the presence of an adversary?
- Canonical example of adversarial (evasion) attacks from computer vision:



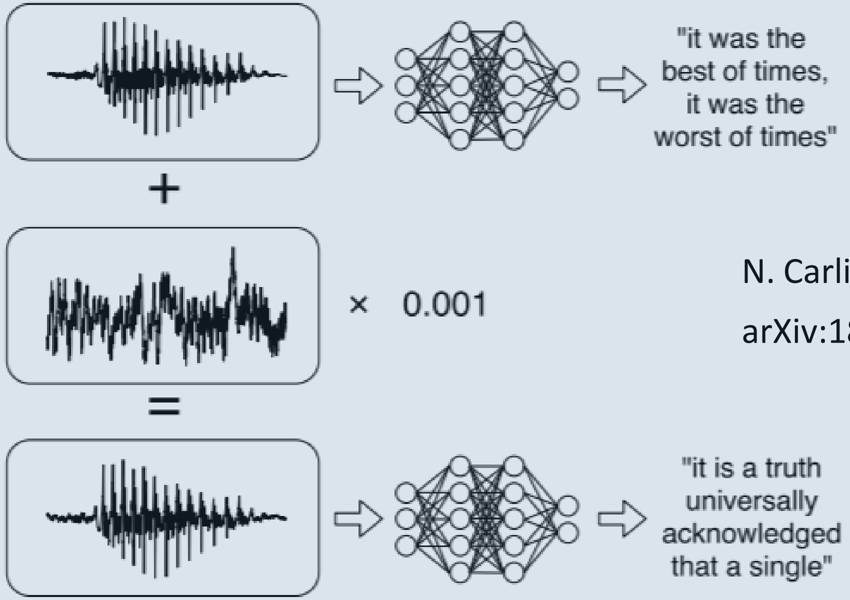
Applications of Adversarial ML

- Autonomous driving
- Text classification
- Voice applications



D. Jin, et al., arXiv:1907.11932.

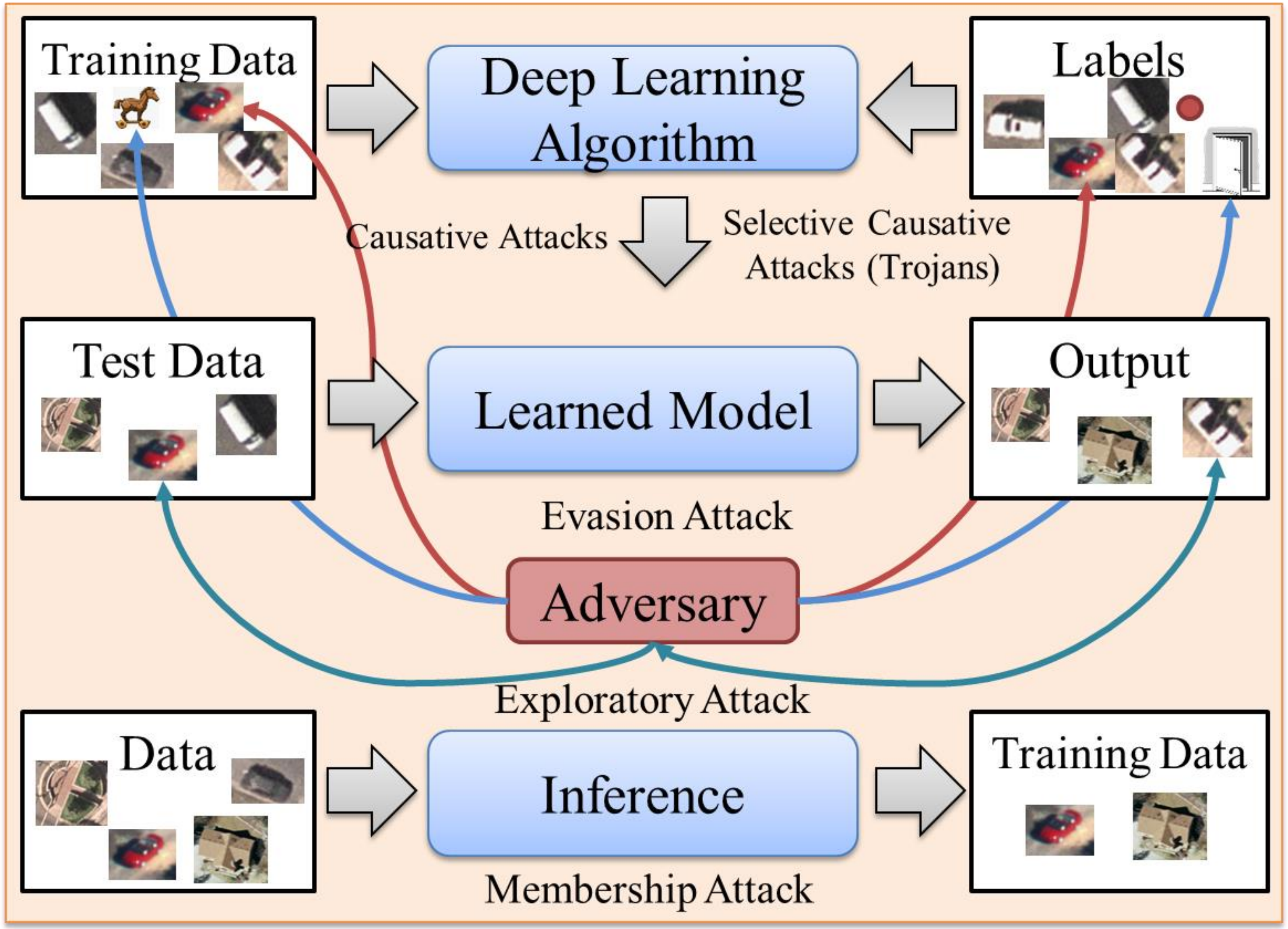
B. Biggio, F. Roli, arXiv:1712.03141.



N. Carlini, D. Wagner, arXiv:1801.01944.

Adversarial Machine Learning Taxonomy

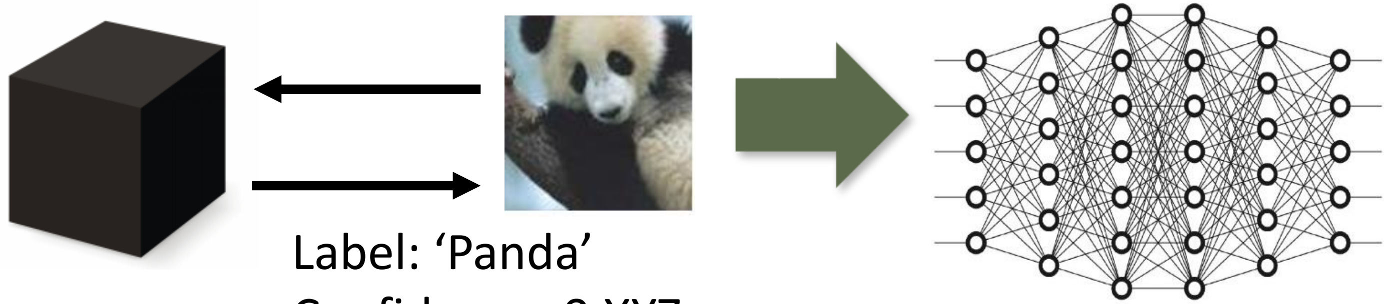
1. **Exploratory attacks**
   - Uncover information about ML
2. **Adversarial (evasion) attacks**
   - Manipulate test data for ML
3. **Causative (poisoning) attacks**
   - Manipulate training data for ML
4. **Trojan (backdoor) attacks**
   - Poison training data with triggers that are activated in test time
5. **Privacy attacks**
   - **Model inversion** attacks
   - **Membership inference** attacks
   - **Attribute inference** attacks



1 – Exploratory (Inference) Attacks

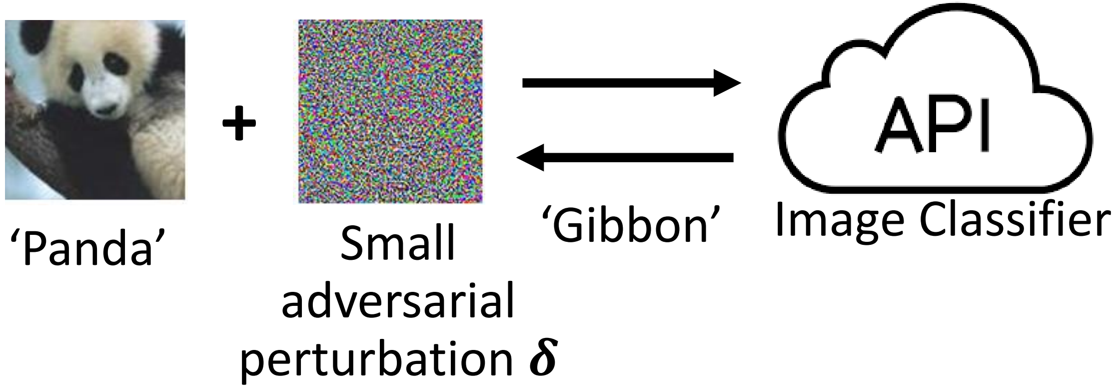
Attack steps:

1. Query the classifier
2. Collect returned labels
3. Use 1-2 to train a **surrogate** machine/deep learning model.



- **"Stealing" the machine learning algorithm** poses a risk to the **intellectual property**.
- Once a classifier is stolen, the adversary is free to analyze it (with an unlimited number of queries) to identify its potential **weaknesses** and its **underlying functionality**.

2 – Adversarial (Evasion) Attacks

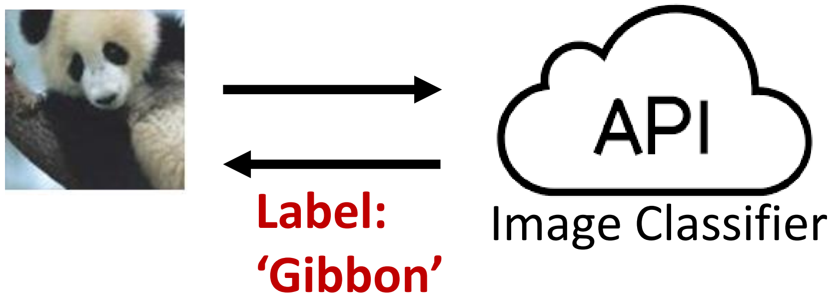


- Attack in **test time**.
- **Adversary’s Goal** : Select perturbation $\boldsymbol{\delta}$

  (i) maximize the error probability of label data is classified as label $j \neq i$

  (ii) subject to upper bound on $\boldsymbol{\delta}$
- **Outcome**: The data samples will be misclassified.

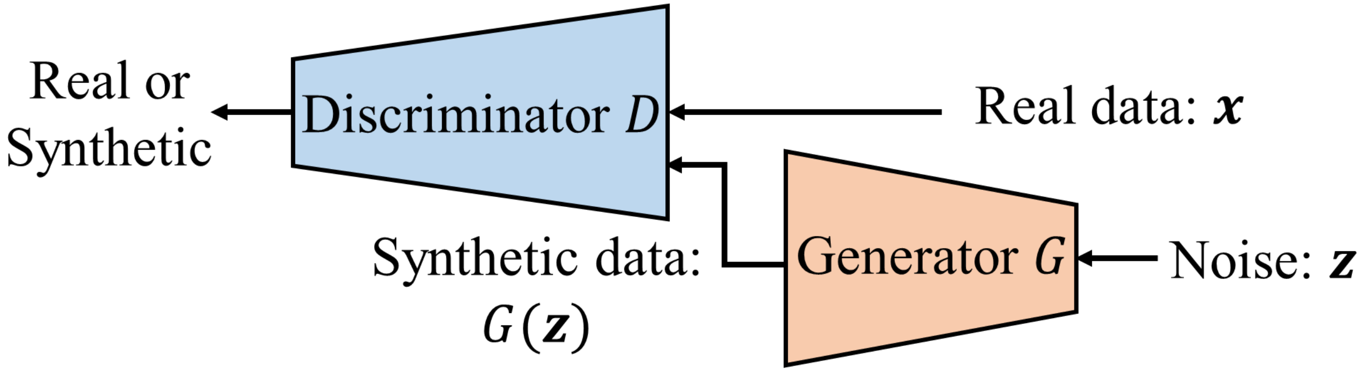
3 – Causative (Poisoning) Attacks



- Attack in **training (or retraining) time**.
  - Data needs to gathered from different (potentially adversarial) parties.
- **Adversary’s Goal**: Select training data whose labels will be modified.
- **Outcome**: The (re)trained model will be poor in accuracy.

Generative Adversarial Learning (GAN)

- Adversarial learning as a generative process (not an attack per se).
- A **Generative Adversarial Network (GAN)** consists of two neural networks.
  - **Generator network**: Generate synthetic data.
  - **Discriminator network**: Discriminate between the real and synthetic data.
  - A **game** is played between the generator and the discriminator.
- **Augment training data** (when training data is limited).
- **Adapt** test or training data **to other domains** (for which there is limited or no training data).

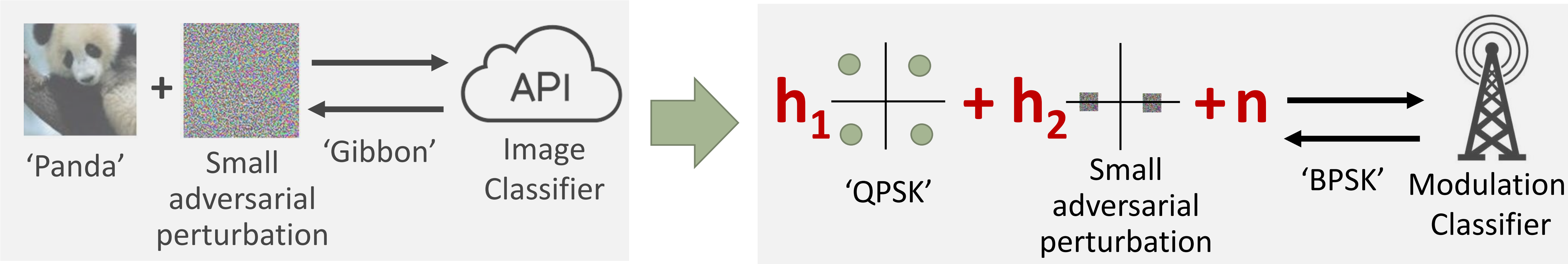


Outline

- **Machine Learning**
- **Machine Learning for Wireless**
- **Machine Learning for 5G and Beyond**
- **Adversarial Machine Learning**
- **Adversarial Machine Learning for Wireless**
- **Adversarial Machine Learning for 5G and Beyond**
- **Conclusion**

Adversarial Machine Learning in Wireless

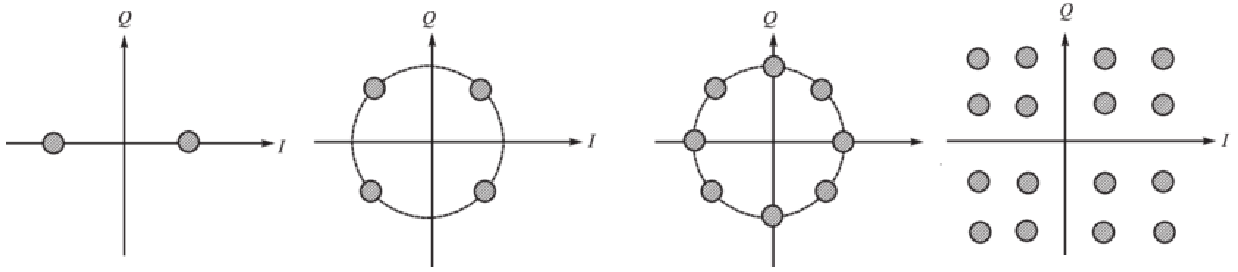
- Wireless medium is open and shared.
  - Adversary can **eavesdrop the channel**.
  - Adversary can **manipulate the channel** by jamming or physically blocking the signal.
- Unique characteristics due to channel, interference, traffic, and spectrum sharing.



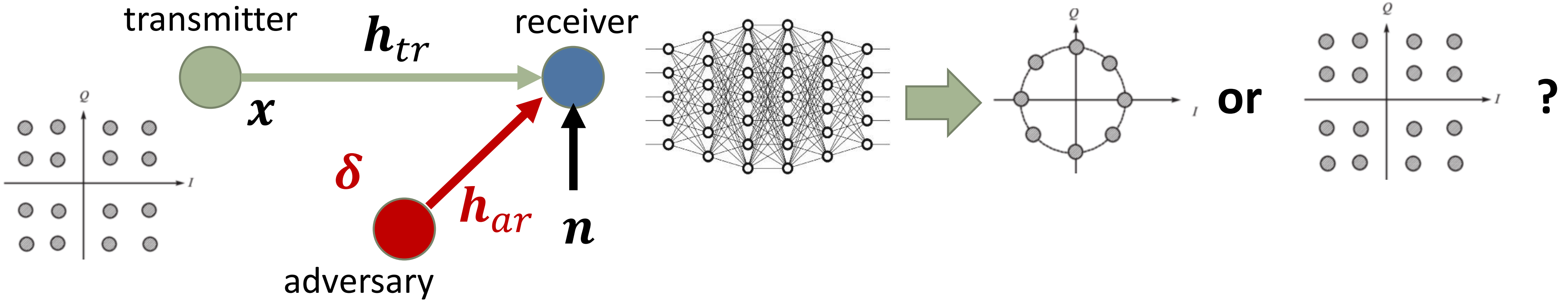
  - **Different data samples** (features and labels) at the target system and at the adversary.
  - **No direct manipulation** of the input to a target machine learning algorithm.

Adversarial Attack on Wireless Signal Classifier

- A transmitter transmits signal $\boldsymbol{x}$ with a particular choice of **modulation**.
  - BPSK, QPSK, 8PSK, 16QAM, …



- A receiver classifies its received signal $\boldsymbol{y} = \boldsymbol{h}_{tr}\,\boldsymbol{x} + \boldsymbol{n}$.
  - Feature: $\boldsymbol{y}$, i.e., I/Q data
  - Label $L(\boldsymbol{y})$: BPSK, QPSK, 8PSK, 16-QAM, …
- If an adversary transmits perturbation $\boldsymbol{\delta}$, the receiver classifies $\boldsymbol{y}' = \boldsymbol{h}_{tr}\,\boldsymbol{x} + \boldsymbol{h}_{ar}\boldsymbol{\delta} + \boldsymbol{n}$.



Y. Sagduyu, T. Erpek, et al., IEEE CISS, 2020.

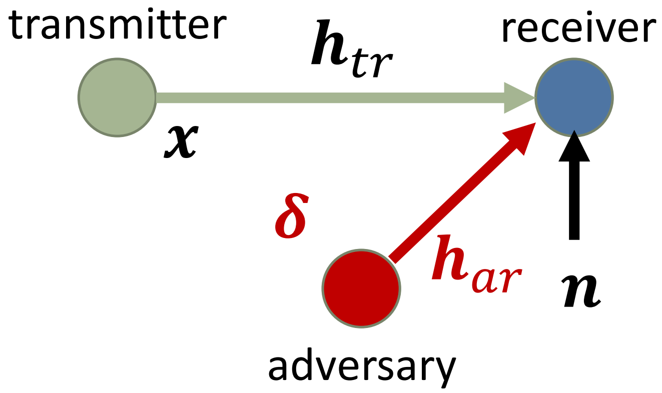
Adversarial Attack on Wireless Signal Classifier

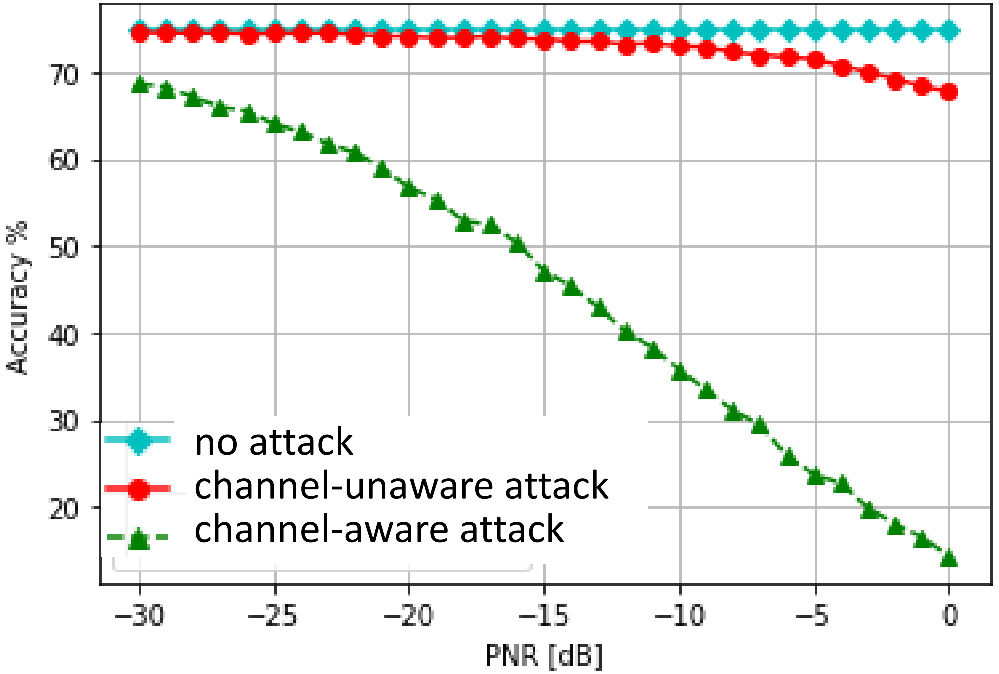
- Adversary selects $\boldsymbol{\delta}$

  to minimize $\|\boldsymbol{\delta}\|_2$

  subject to $L(\boldsymbol{h}_{tr}\,\boldsymbol{x} + \boldsymbol{h}_{ar}\boldsymbol{\delta} + \boldsymbol{n}) \neq L(\boldsymbol{h}_{tr}\,\boldsymbol{x} + \boldsymbol{n})$

  $\|\boldsymbol{\delta}\|_2^2 \leq \boldsymbol{P}_{max}$





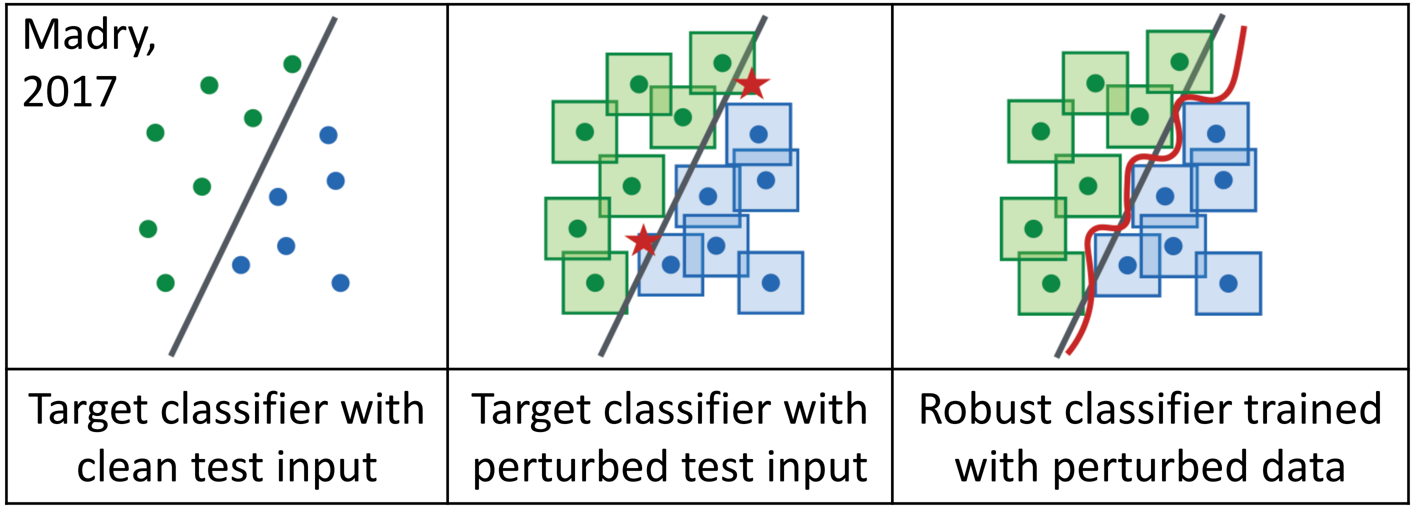
- Attack without considering $\boldsymbol{h}_{ar}$ is ineffective.
- Classifier accuracy significantly drops when the perturbation $\boldsymbol{\delta}$ is selected by considering $\boldsymbol{h}_{ar}$.
- Classifier accuracy decreases as the perturbation-to-noise-ratio (PNR) increases.

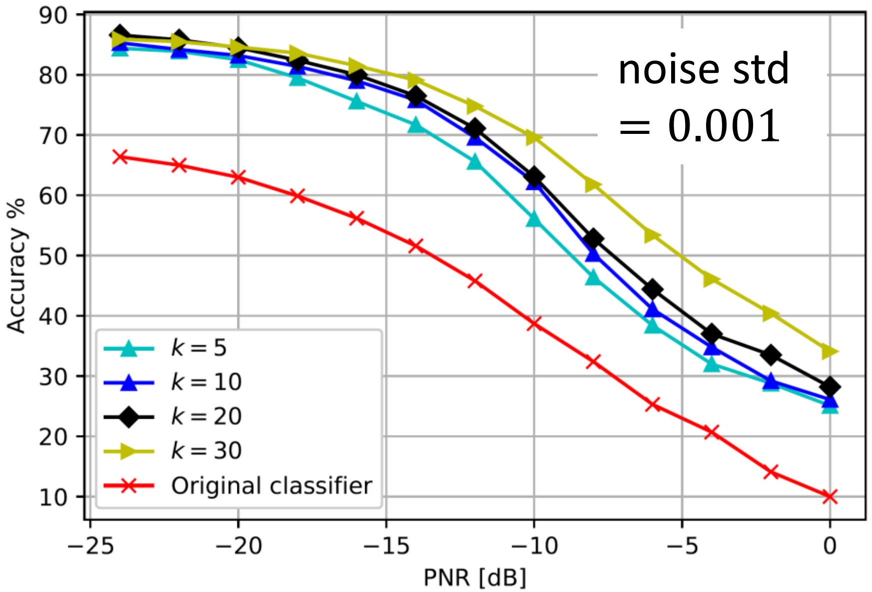
Defense - 1

- **Randomized smoothing during training.**
- To every training sample $\boldsymbol{y_i}$, add k small Gaussian noise samples
- Classifier is trained with the augmented training data set:

$$\boldsymbol{y_i} \rightarrow \{\boldsymbol{y_i} + \boldsymbol{n_{i,1}}, \boldsymbol{y_i} + \boldsymbol{n_{i,2}}, \cdots \ \boldsymbol{y_i} + \boldsymbol{n_{i,k}}\}$$

- Classifier becomes **robust** against adversarial inputs in test time.

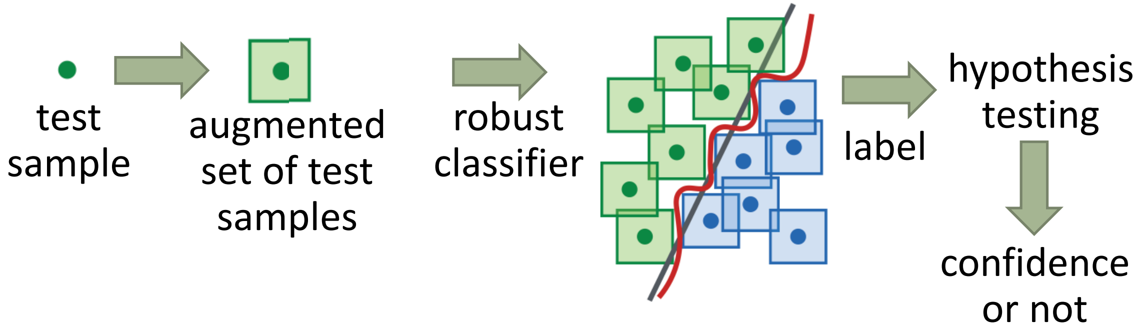




Y. Sagduyu, T. Erpek, et al., https://arxiv.org/abs/2005.05321

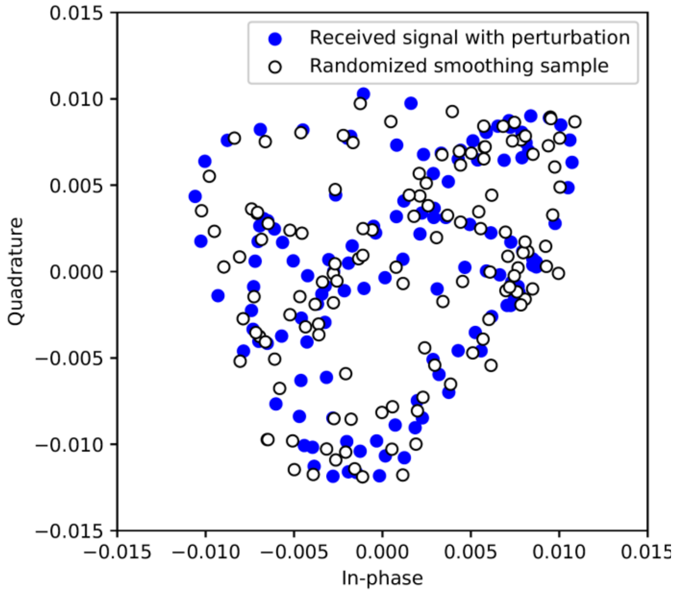
Defense -2

- **Certified defense in test time.**
  - **Guarantee the classifier's robustness by using randomized smoothing in test time.**

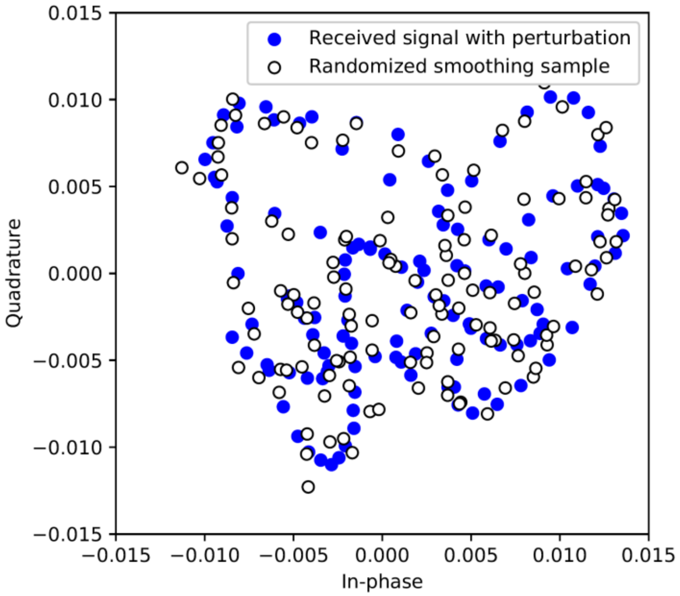


- For every test sample $\boldsymbol{y_i}$, add k small Gaussian noise samples and label of them with the classifier.
- Apply two-sided hypothesis test with the classifier outputs to check statistical significance for a desired confidence.

Y. Sagduyu, T. Erpek, et al., https://arxiv.org/abs/2005.05321



when the classifier abstains.



when the classifier correctly infers the label (confidence = 0.95).

Extensions of Adversarial Attacks in Wireless

- **Transmitted signal is unknown** to the adversary (universal perturbation)
  - *Y. Sagduyu, T. Erpek, et al., IEEE CISS, 2020.*

- **Target classifier is unknown** to the adversary.
  - *Y. Sagduyu, T. Erpek, et al., IEEE CISS, 2020.*

- **Channel information is only partially known** to the adversary.
  - *Y. Sagduyu, T. Erpek, et al., https://arxiv.org/abs/2005.05321*

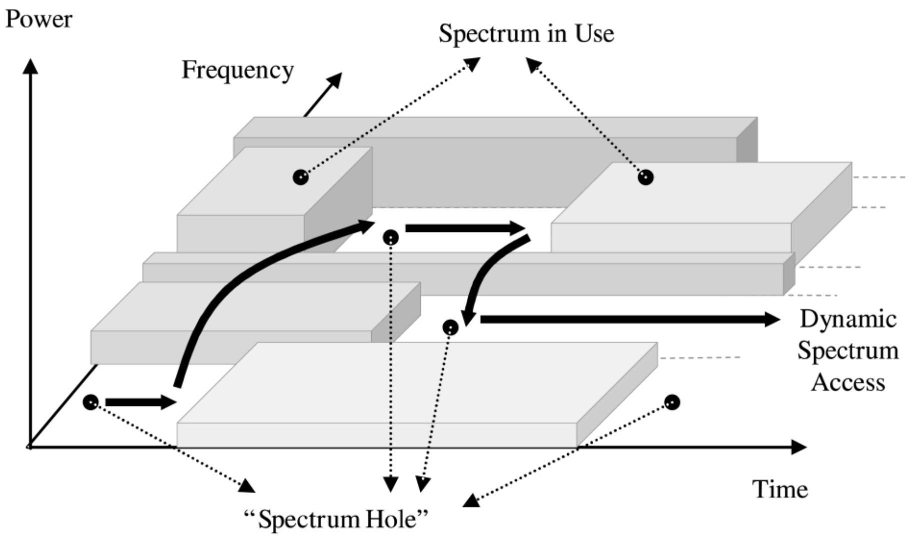
- **Multiple receivers to be fooled** with a signal perturbation
  - *Y. Sagduyu, T. Erpek, et al., https://arxiv.org/abs/2005.05321*

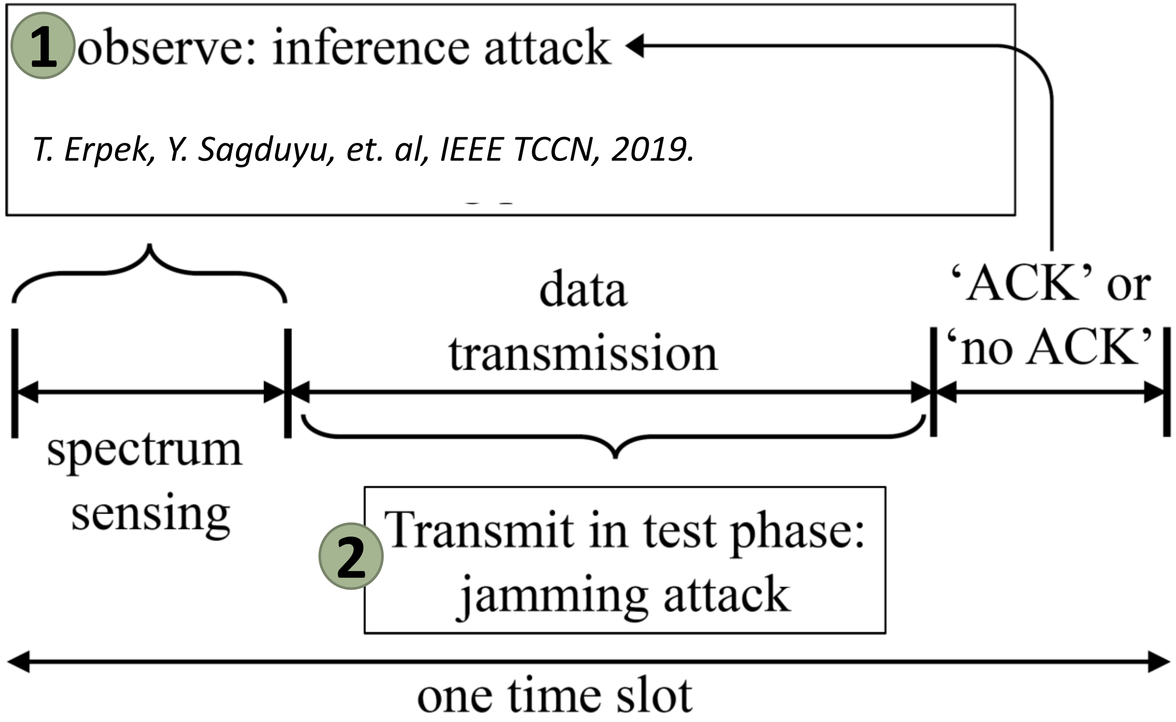
- The adversary is equipped with **multiple antennas**.
  - *Y. Sagduyu, T. Erpek, et al., IEEE Globecom, 2020.*

Other Adversarial Machine Learning Attacks

- **Dynamic spectrum access (DSA)**
  - An incumbent user transmits intermittently.
  - A transmitter senses the channel and transmits only when it is idle.

Akyidiz, et al, 2006





1 Inference (exploratory) attack
- Sense the spectrum and train a surrogate model to mimic transmit behavior

2 Inference-based jamming attack
- Use the surrogate model to predict and jam data transmissions that would other succeed.

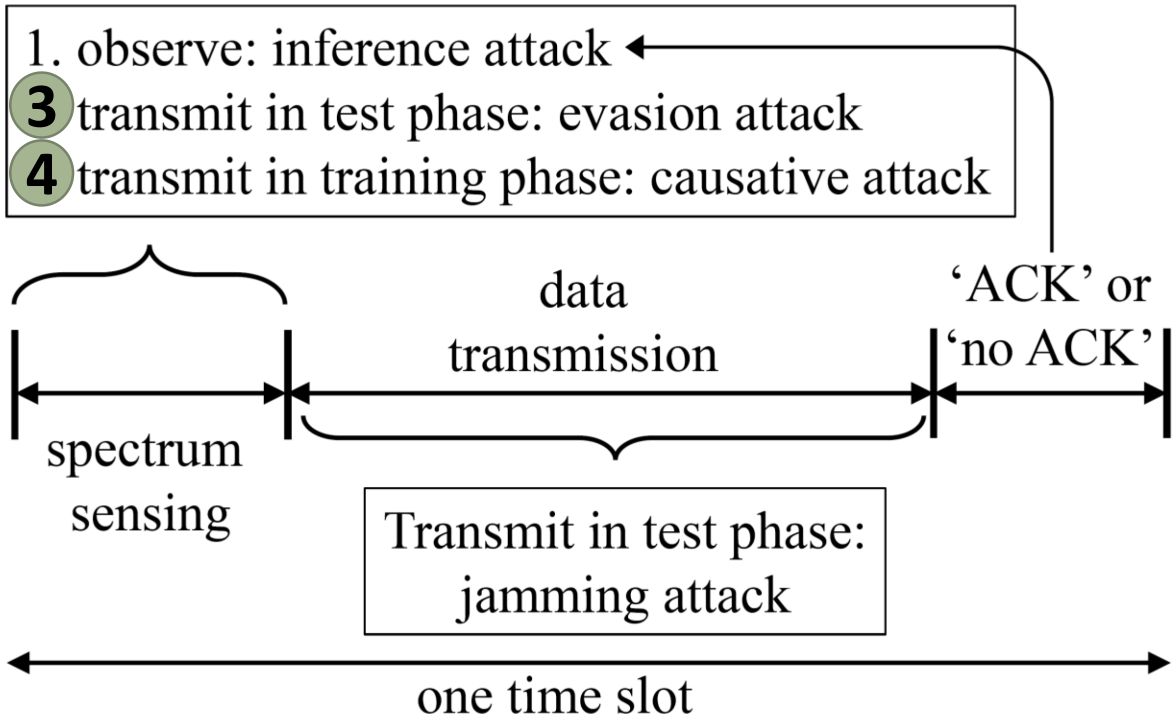
Other Adversarial Machine Learning Attacks

3 Evasion (adversarial) attack

- Jam the spectrum sensing period such that the transmitter makes wrong transmit decisions.

4 Causative (poisoning) attack

- Jam the spectrum sensing period such that the transmitter makes wrong transmit decision.



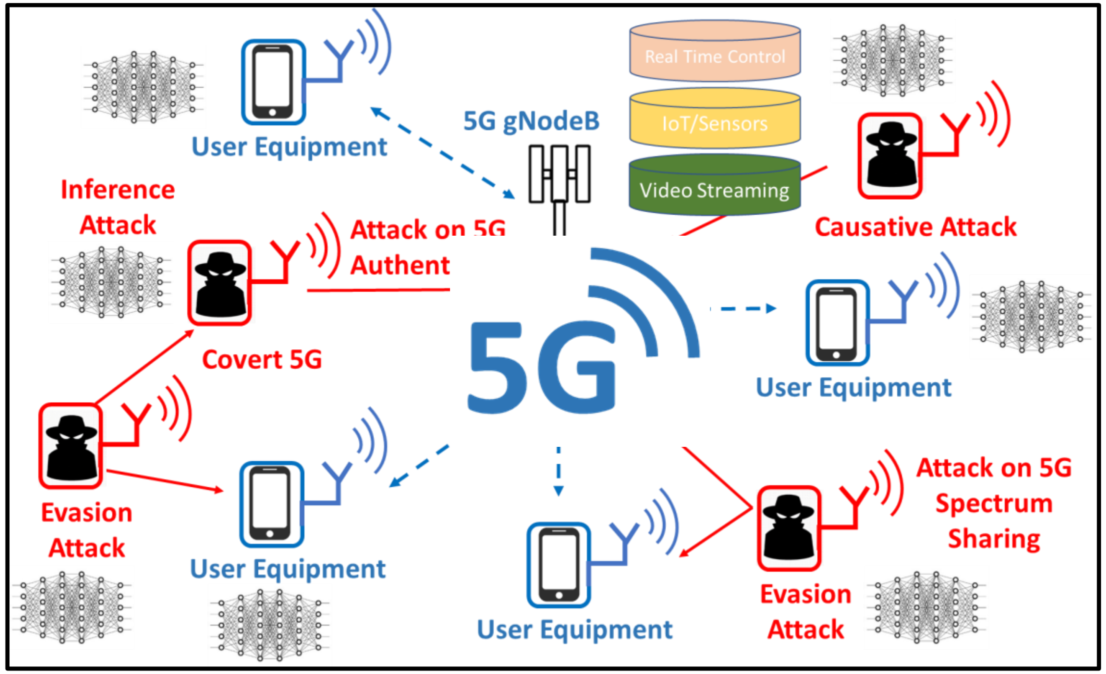
Y. Sagduyu, T. Erpek, et. al, IEEE TCCN, 2020.

Outline

- Machine Learning
- Machine Learning for Wireless
- Machine Learning for 5G and Beyond
- Adversarial Machine Learning
- Adversarial Machine Learning for Wireless
- **Adversarial Machine Learning for 5G and Beyond**
- Conclusion

Attacks on 5G Radio Access Network (RAN)

1. Attacks on **spectrum sharing of 5G**.
   - ML for environmental sensing capability (ESC).
2. Attacks to gain **access to 5G-enabled services**.
   - ML for 5G signal authentication.
3. Attacks to establish **covert 5G** signals.
   - ML to detect rogue 5G communications.

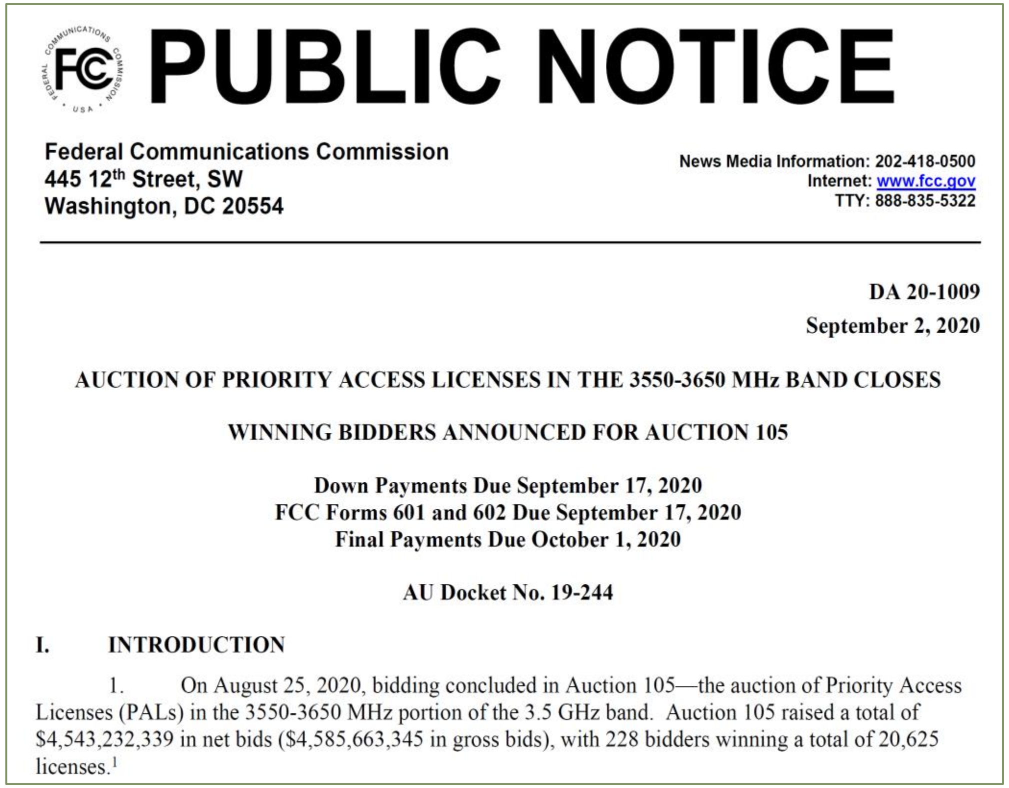


Adversarial machine learning generates new attack surfaces for 5G.

Y. Sagduyu, T. Erpek, et al, IEEE Asilomar, 2020.
Y. . Sagduyu, T. Erpek, et al, Springer, 2020.

AML Attack on 5G Spectrum Sharing – 1

- **CBRS (Citizens Broadband Radio Service) band at 3.5gHz** is reserved for **federal use**.
- CBRS band will be opened to be shared by commercial users such as 5G.



PUBLIC NOTICE

Federal Communications Commission
445 12th Street, SW
Washington, DC 20554

News Media Information: 202-418-0500
Internet: www.fcc.gov
TTY: 888-835-5322

DA 20-1009
September 2, 2020

AUCTION OF PRIORITY ACCESS LICENSES IN THE 3550-3650 MHz BAND CLOSES

WINNING BIDDERS ANNOUNCED FOR AUCTION 105

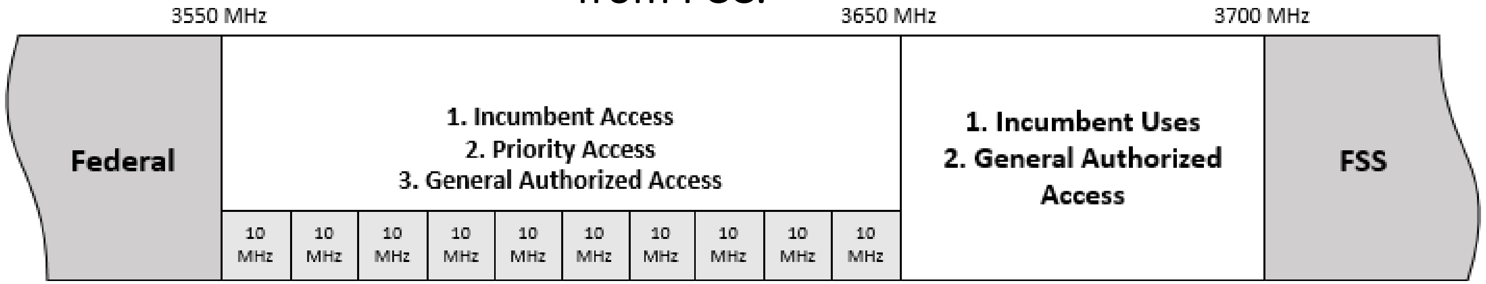
Down Payments Due September 17, 2020
FCC Forms 601 and 602 Due September 17, 2020
Final Payments Due October 1, 2020

AU Docket No. 19-244

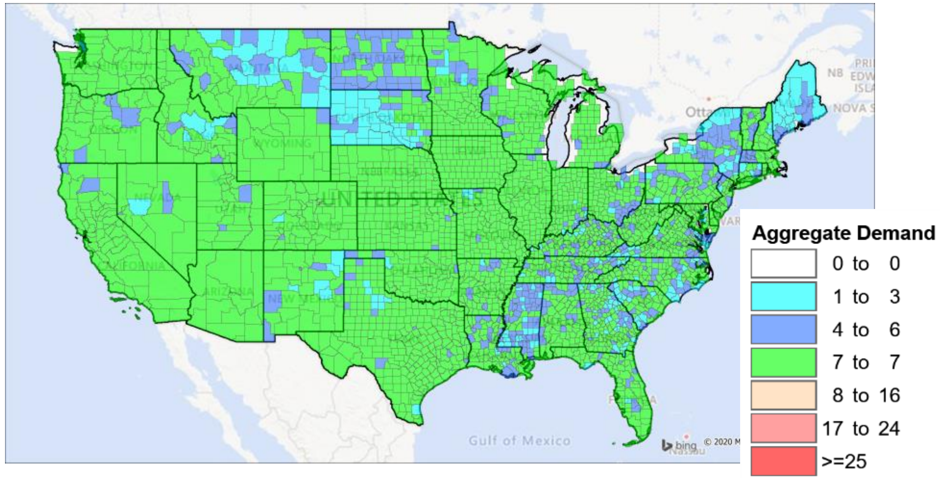
I. INTRODUCTION

1. On August 25, 2020, bidding concluded in Auction 105—the auction of Priority Access Licenses (PALs) in the 3550-3650 MHz portion of the 3.5 GHz band. Auction 105 raised a total of $4,543,232,339 in net bids ($4,585,663,345 in gross bids), with 228 bidders winning a total of 20,625 licenses.[1]

from FCC:



CBRS auction bids:

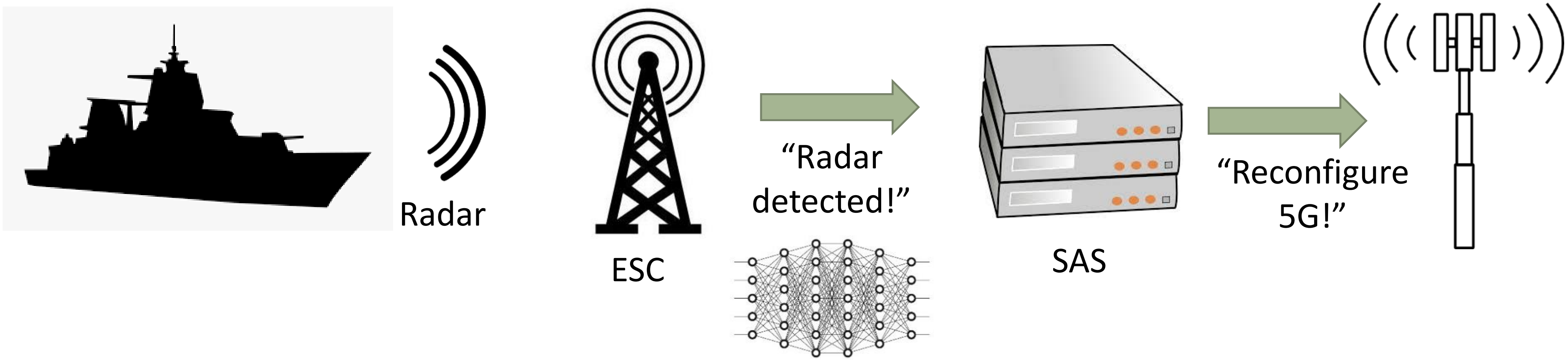


https://www.mrleng.com/?page_id=2502

AML Attack on 5G Spectrum Sharing – 2

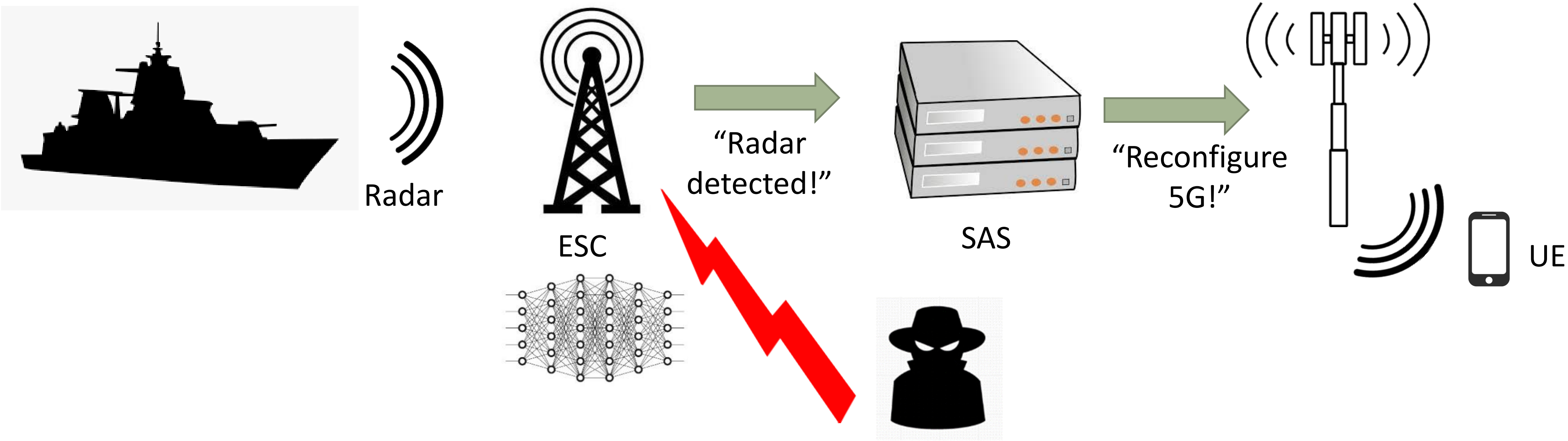
- **Environmental Sensing Capability (ESC)** needs to detect incumbent radar signals (potentially with machine learning).

- **Spectrum Access System (SAS)** needs to (re)configure and manage the 5G system.



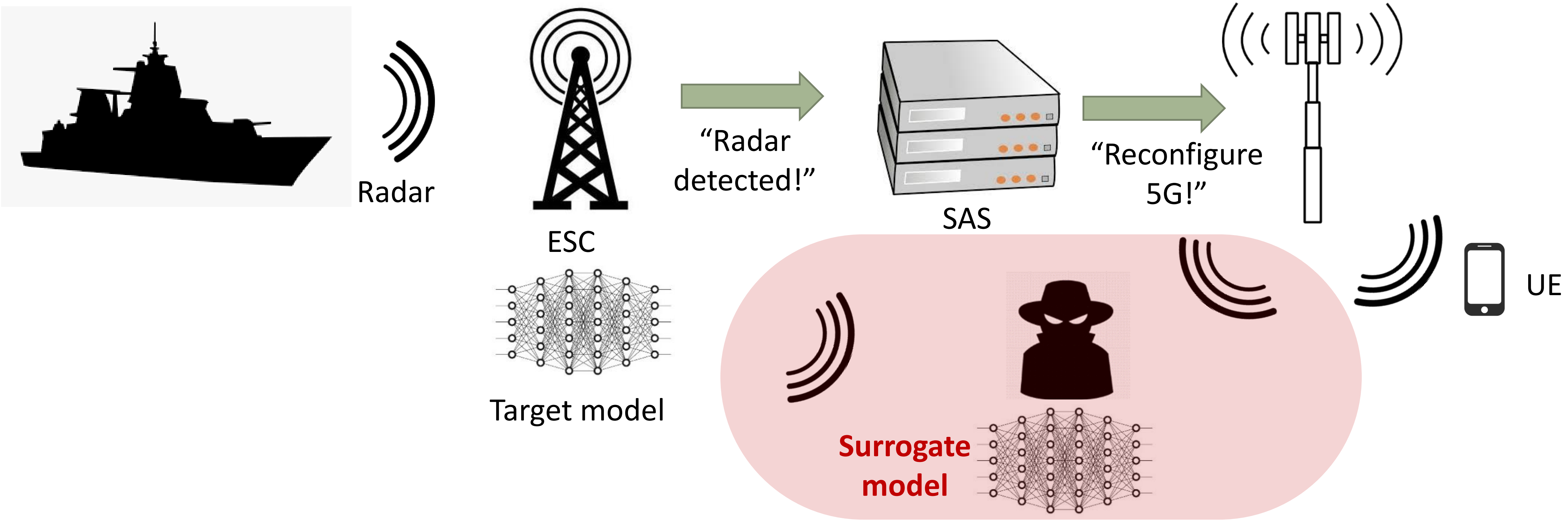
AML Attack on 5G Spectrum Sharing – 3

- The adversary transmits **perturbations over the air** to manipulate the input signal to the ESC's ML algorithm – **evasion (adversarial) attack**.
  - A **stealth attack** with low spectrum footprint.
- **ESC is fooled** into making wrong decisions on the existence of the radar signal.



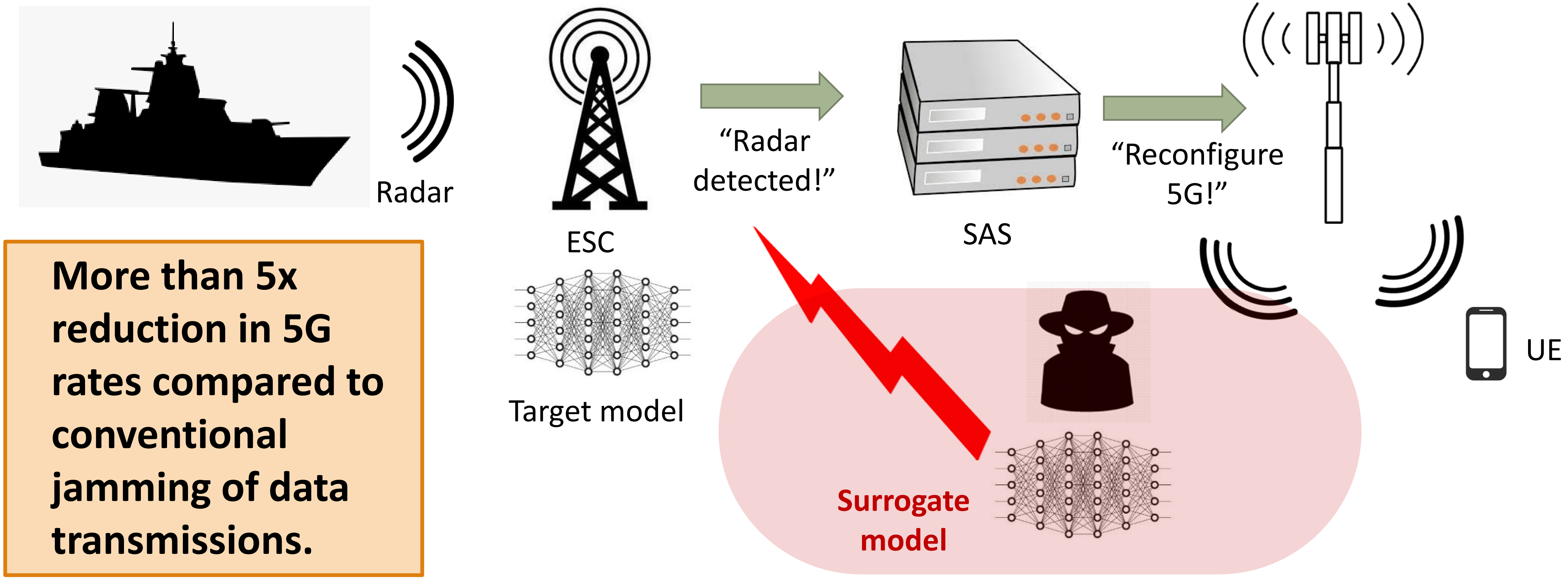
AML Attack on 5G Spectrum Sharing – 4

- The adversary senses the spectrum to collect training data (I/Q data & spectrum access).
- The adversary trains a **surrogate model** to predict when there will be successful 5G communication (if there was no attack).
  - AML can detect all successful transmissions and most (>95%) failed transmissions.



AML Attack on 5G Spectrum Sharing – 5

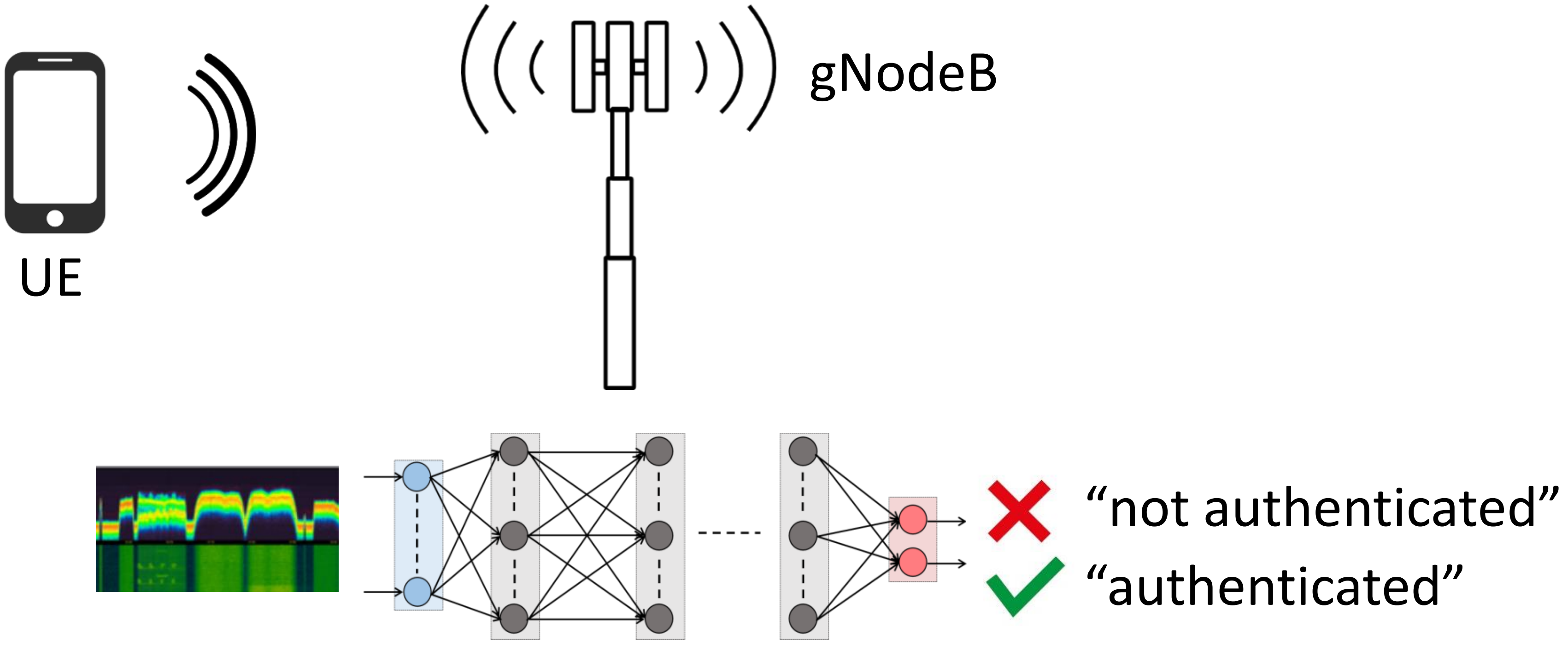
- As an **evasion attack**, the adversary **jams spectrum sensing** of ESC period.
- The ESC is provided with manipulated input to its machine learning algorithm and makes wrong decisions on the existence of radar signal.



More than 5x reduction in 5G rates compared to conventional jamming of data transmissions.

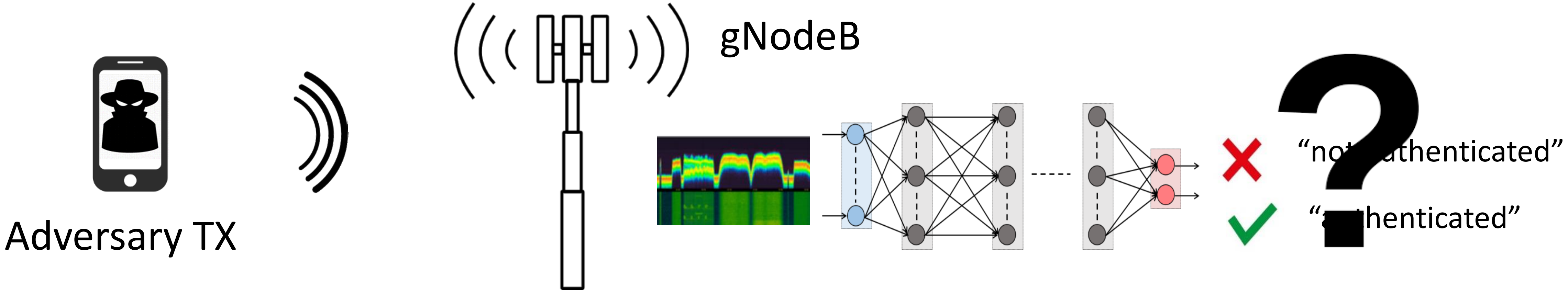
AML Attack on 5G Authentication – 1

- Devices need to connect to 5G network to gain access to 5G-enabled services, (e.g., through network slices).
- Massive number of heterogenous devices raise the need for PHY-layer authentication.

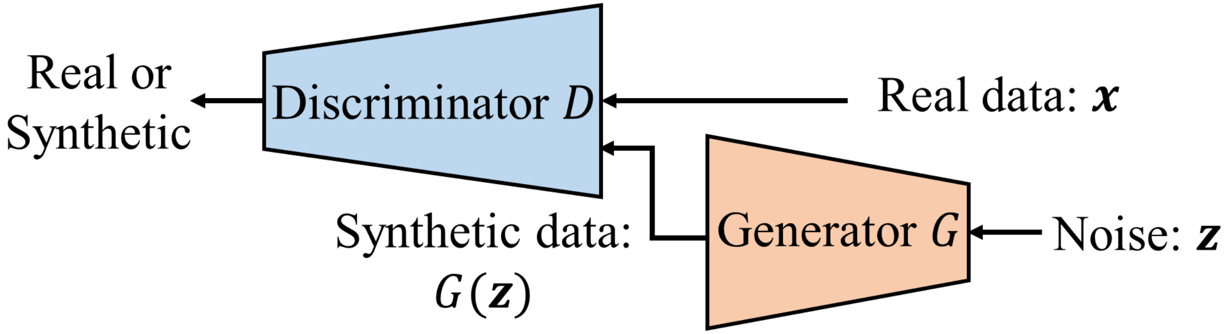


AML Attack on 5G Authentication – 2

- Adversary spoofs signals to bypass the authentication.

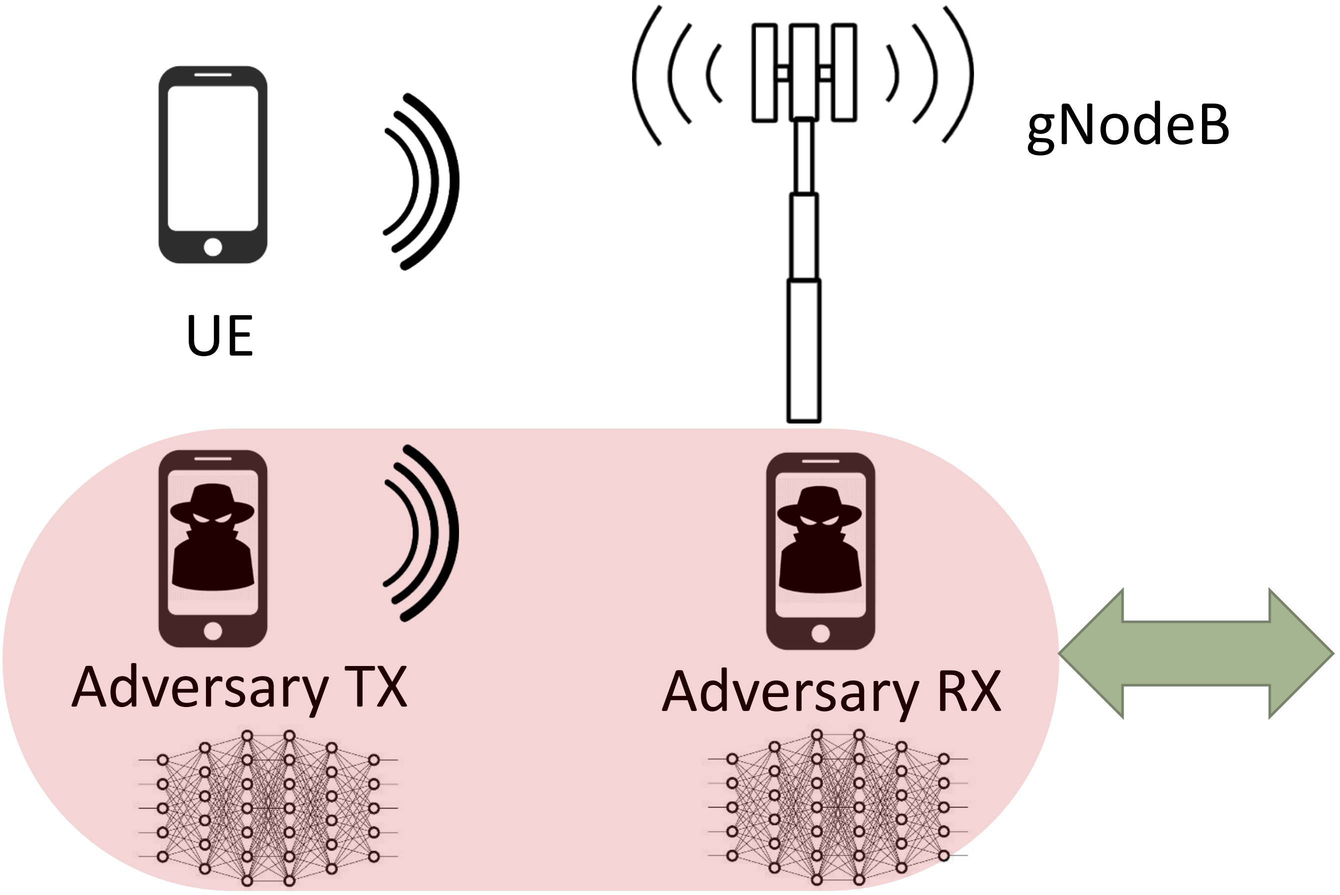


- Spoofed/synthetic signals are generated by using **Generative Adversarial Network (GAN)**.

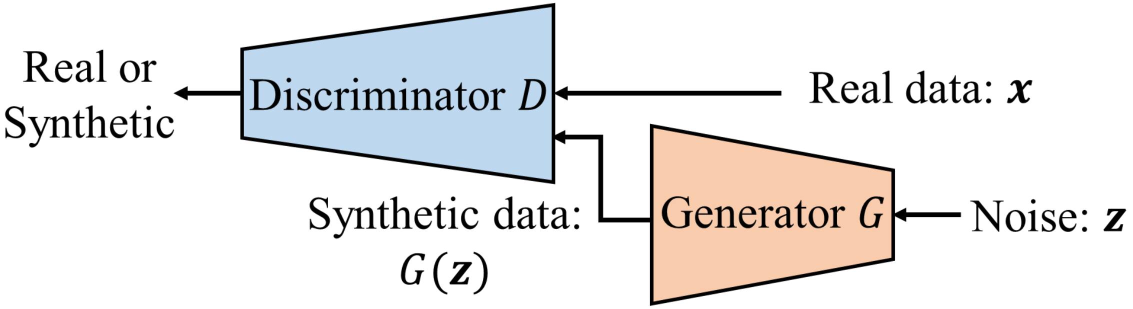


AML Attack on 5G Authentication – 3

- Adversary transmitter-receiver pair forms an **over-the-air GAN**.
  - Adversary transmitter is the generator and adversary receiver is the discriminator.

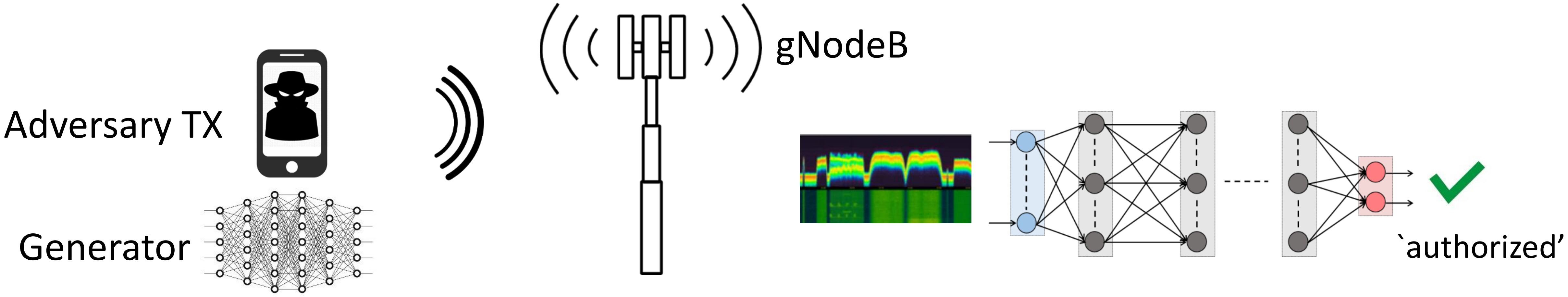


Generative Adversarial Network (GAN)



AML Attack on 5G Authentication – 4

- The GAN generator of the adversary spoofs signals that fool the gNodeB's DL algorithm.



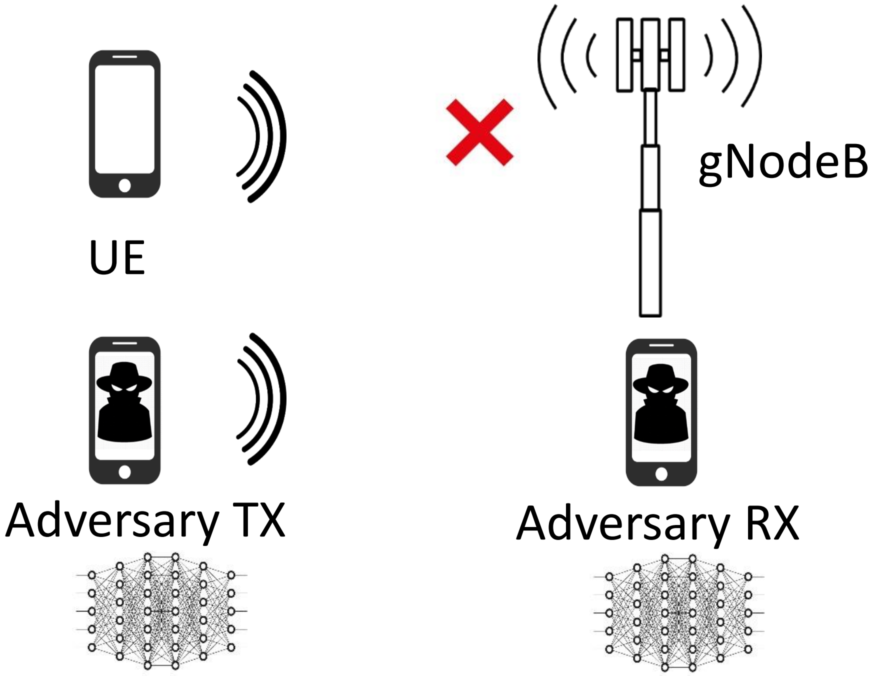
- Captures all waveform, channel, and radio device characteristics.
- Better than replay attacks.

5G Signal Strength	Probability of Fooling the Authentication System
-3dB	61%
0dB	67%
3dB	90%

Defense

- The attacks have started with building a surrogate/generative model at the adversary.
- **Proactive defense against 5G spoofing attacks**: 5G gNodeB introduces deliberate and selective errors in denying access to a small number of requests from intended 5G UEs.

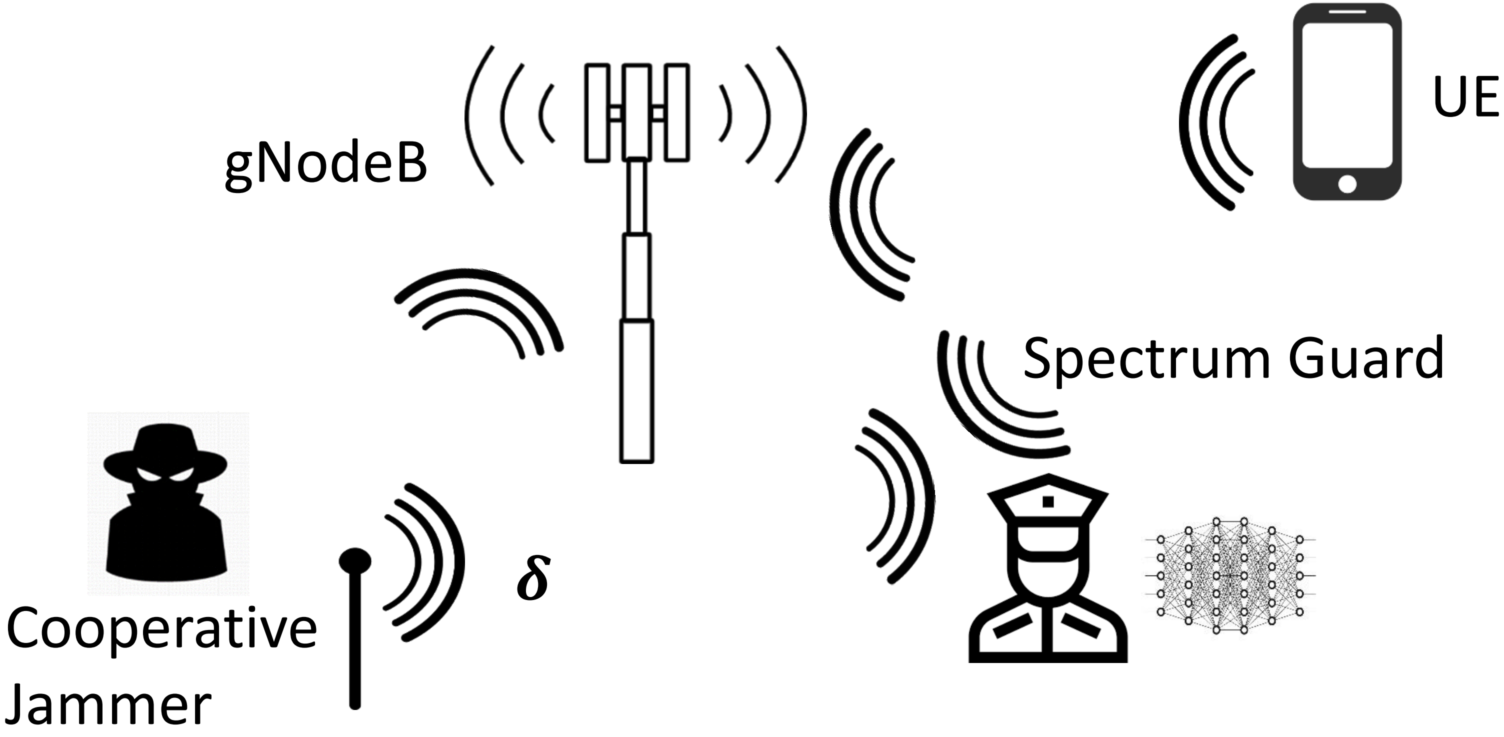
	Percentage of Introduced Errors	Attack Success Probability
no defense	0%	90%
not enough defense	1%	68%
	2%	61%
best defense	5%	59%
too much defense	10%	62%



Adversarial ML for Covert 5G – 1

- Adversaries can set up 5G communications in unauthorized places.
- Cooperative jammers transmit perturbations that are superimposed with rogue 5G signals.
- Even when deep learning is used, covert 5G signals cannot be detected.

Adversarial ML hides 5G below noise floor while sustaining high data rates.

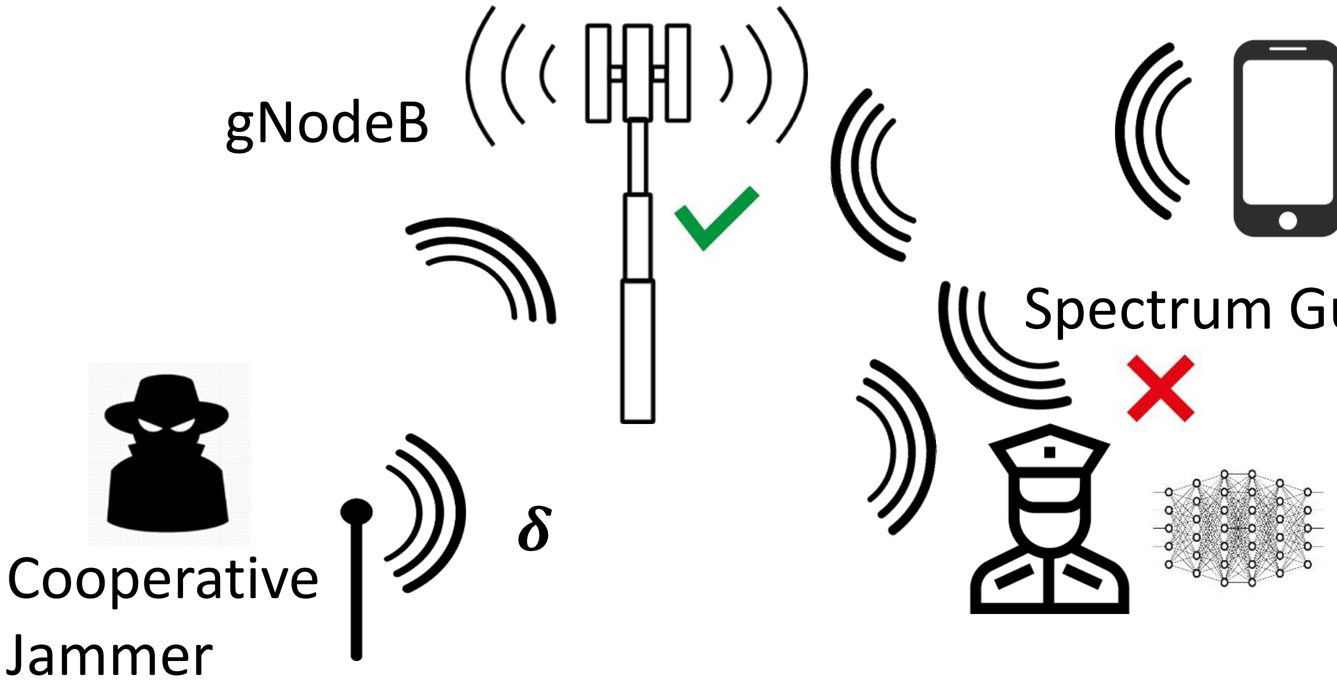


Adversarial ML for Covert 5G – 2

- By considering channels, cooperative jammer determines the perturbation $\boldsymbol{\delta}$ such that
    1. the received signal superimposed with $\boldsymbol{\delta}$ is misclassified as noise, and
    2. covert 5G signals are reliably decoded by the gNodeB subject to interference due to $\boldsymbol{\delta}$.

Covertness	Increase in Error Rate
>95%	<1%

when $\boldsymbol{\delta}$ is at -5 dB relative to noise



Conclusion

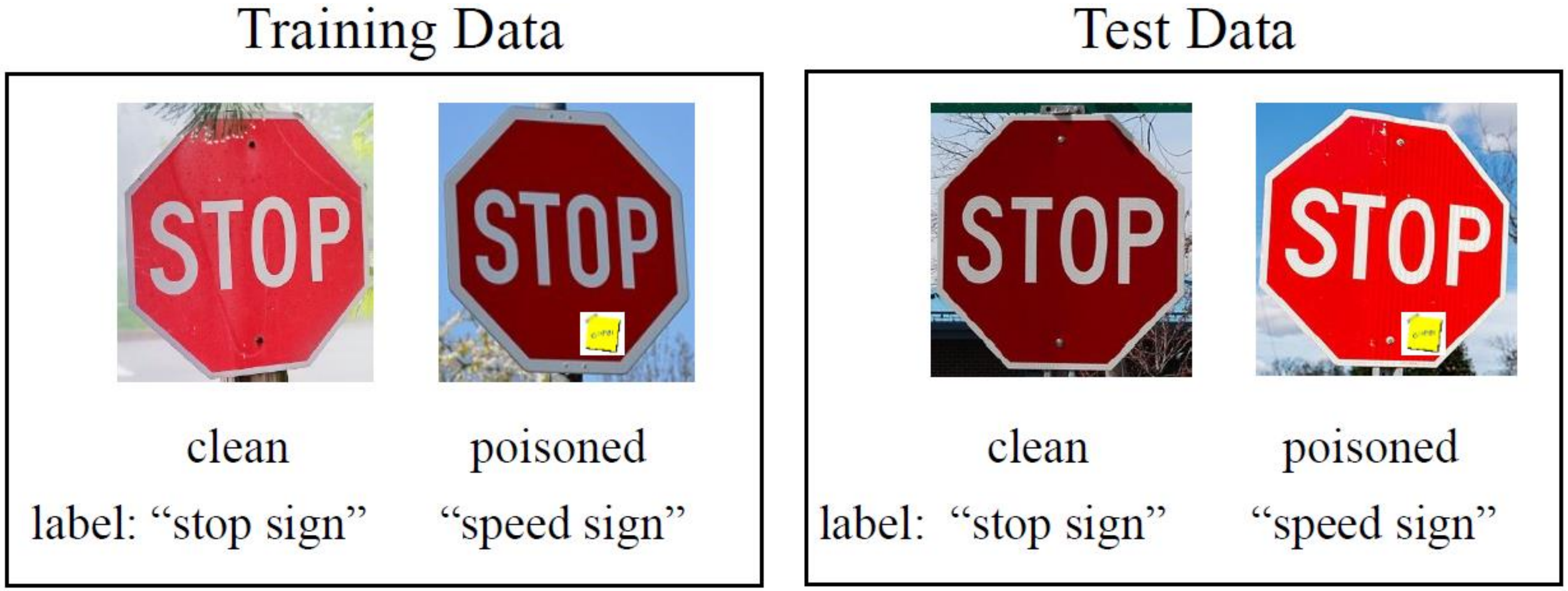
- **Machine learning** finds diverse use cases in **wireless communications** including **5G and beyond**.
- **Adversarial machine learning** generates **a new attack surface in wireless domain** subject to its unique characteristics.
- Wireless systems including 5G are **heavily vulnerable to adversarial machine learning**.
- More work is needed to further understand this new attack surface with **additional attack modalities** and corresponding **defense** techniques.

THANK YOU!

FOR QUESTIONS:
Yalin Sagduyu, ysagduyu@i-a-i.com
Tugba Erpek, terpek@i-a-i.com

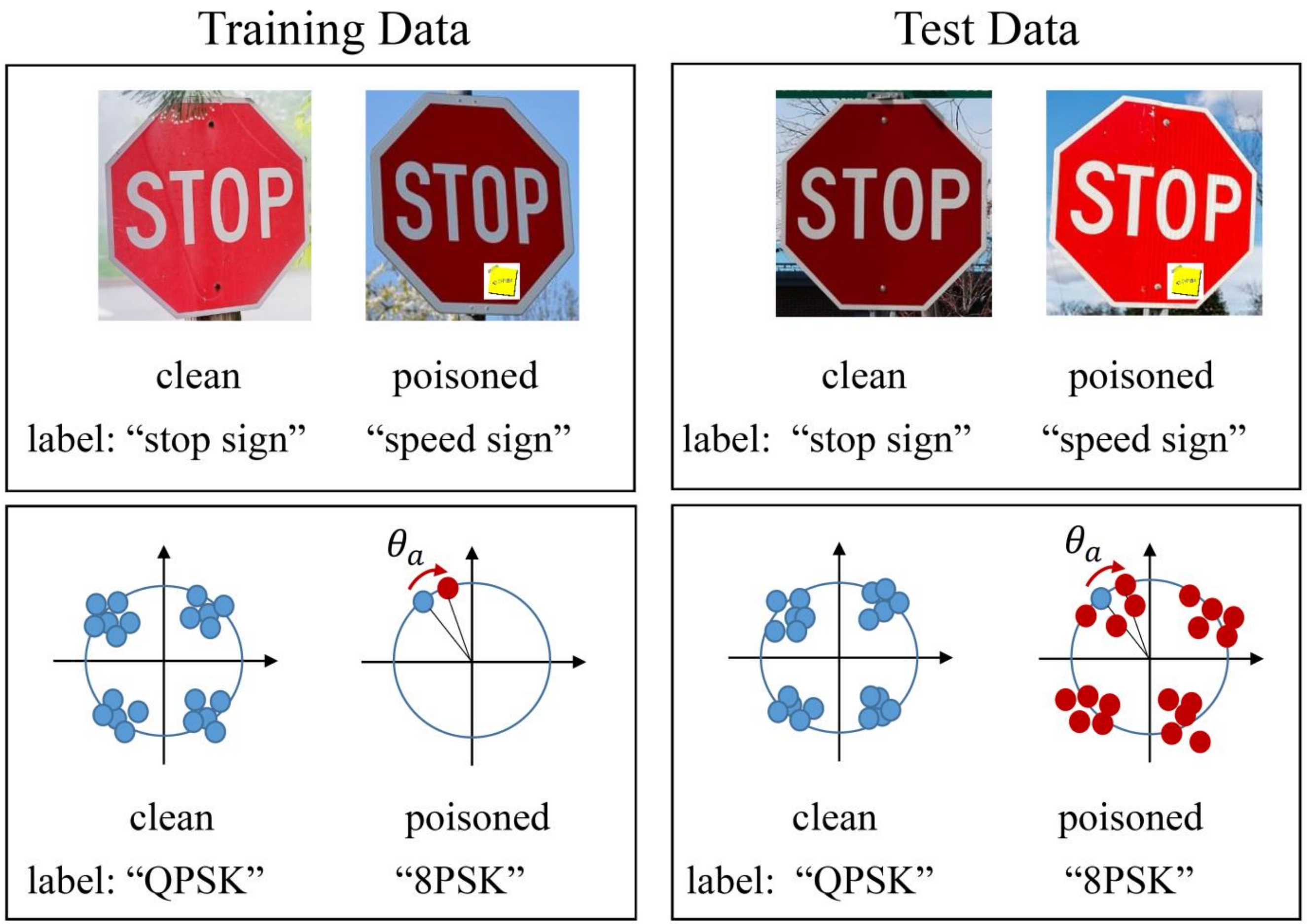
Trojan (Backdoor) Attacks

- Attack in **both training and test times**.
- **Adversary's Goal**: Select a small number of training data samples to embed with **triggers** (add perturbation and flip label).
- **Outcome**: Only test samples with triggers are misclassified while other samples are correctly classified.



Trojan (Backdoor) Attacks in Wireless - 1

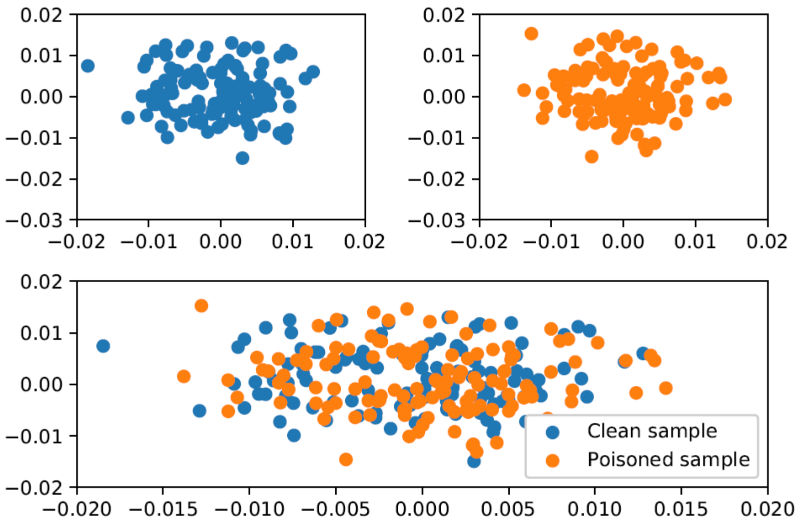
- In the wireless domain,
  - Trojans are harder to detect visually.
  - Trojans can be added through phase offsets, amplitude, etc.
  - Data collection manipulation can be done remotely.
- However, triggers are harder to control by the attacker in test time.
  - Needs to be done over the air.

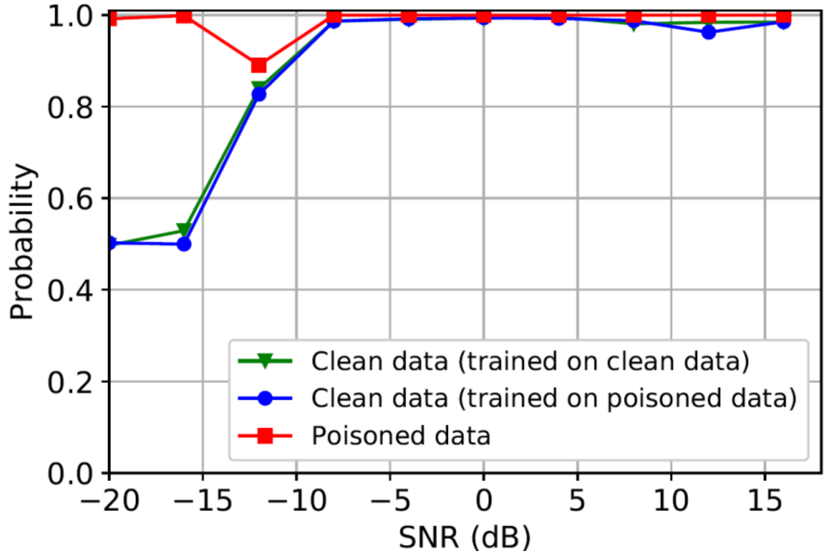


K. Davaslioglu, Y. Sagduyu, IEEE DySPAN 2019.

Trojan (Backdoor) Attacks in Wireless - 2

- Adversary **poisons some training samples with triggers** (e.g., by adding **small phase shifts**).
- Adversary has two objectives:
  - Increase the probability of misclassifying poisoned samples.
  - Keep the classification on clean samples high.
- The attack is stealth and successful in satisfying both attack objectives.
- The attack forces a target signal classifier to **misclassify unauthorized signal as legitimate**.



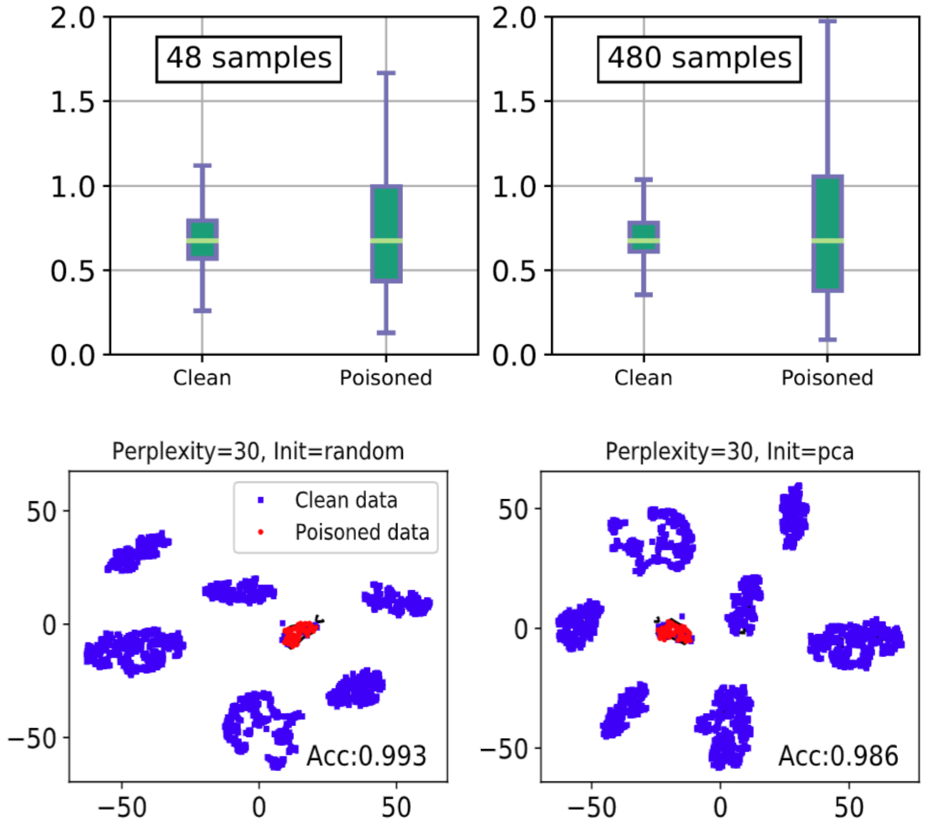


Defense for Trojan Attacks in Wireless

1) **Data augmentation with rotations (proactive):** Significantly reduces the accuracy of clean samples.

2) **Statistical detection of triggers:** Statistical outlier detection using the Median Absolute Deviation (MAD) algorithm. Performance depends on the amount of poisoned data.

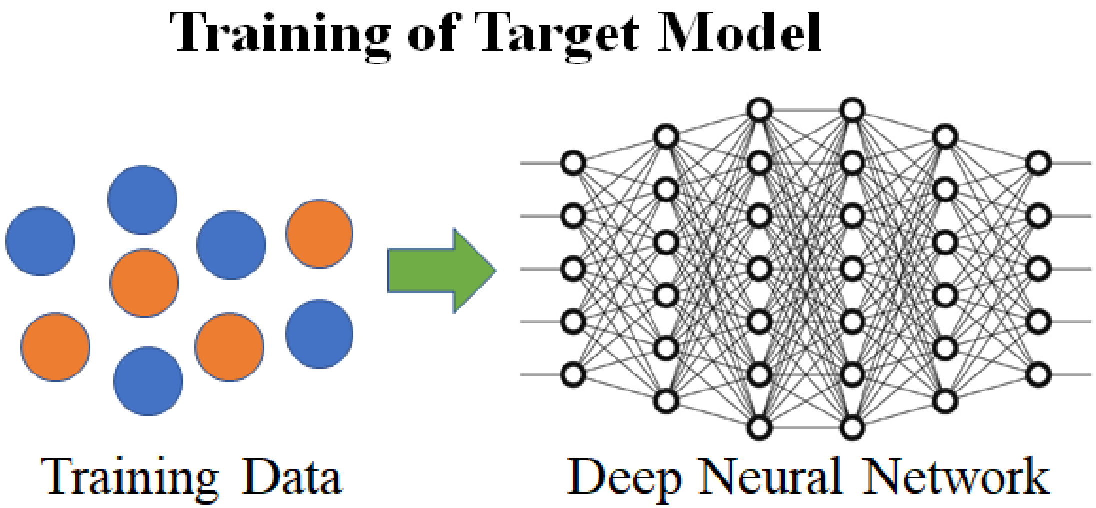
3) **Clustering-based detection of triggers**: t-SNE based clustering for dimensionality reduction and SVM-based detection. Achieves >98% accuracy.

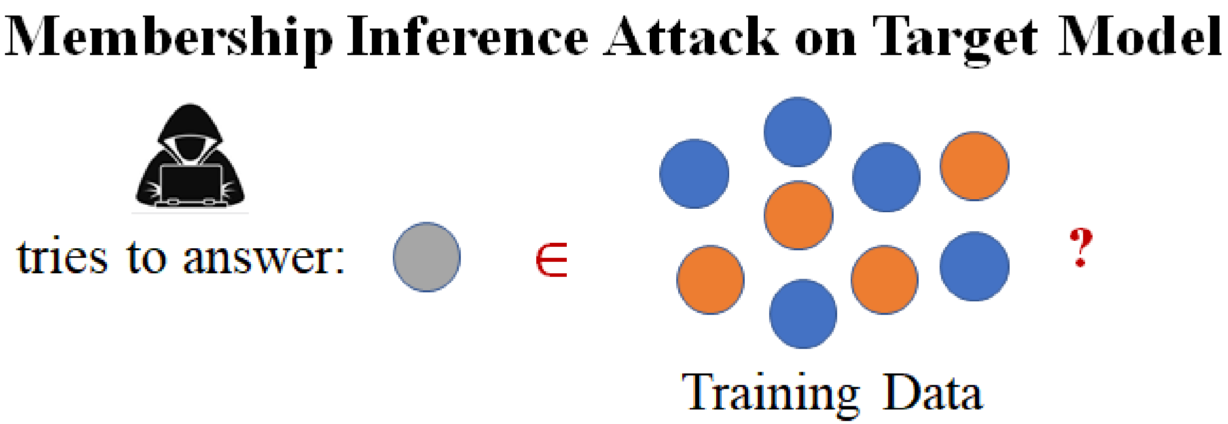


Privacy Attacks

Membership Inference Attack (MIA)

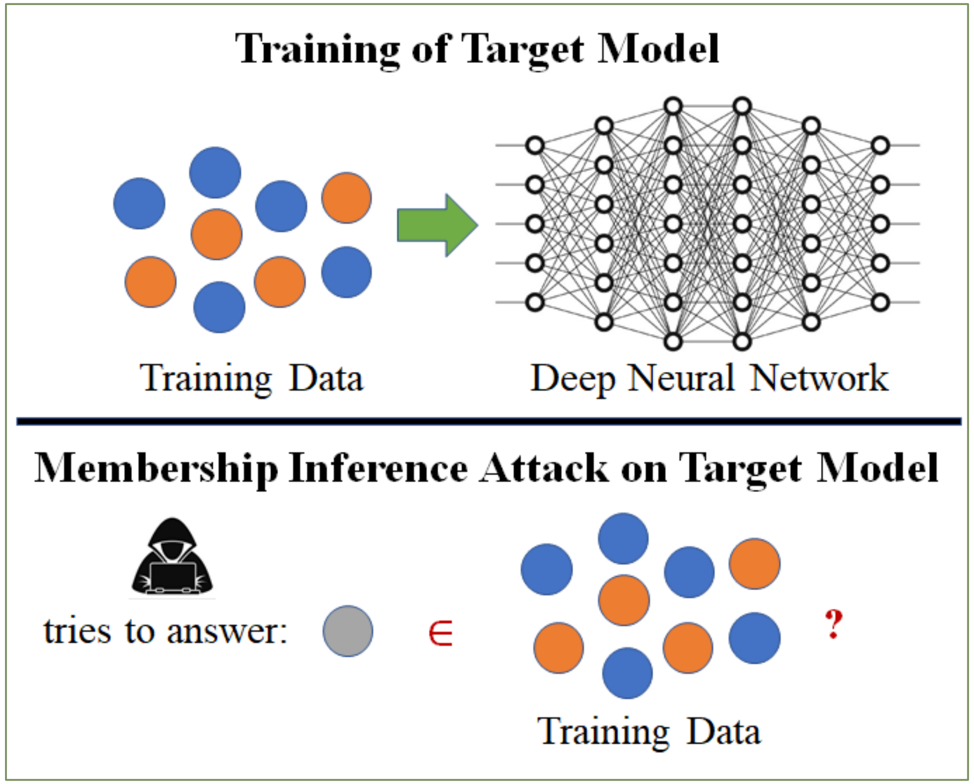
- Attack in **test time**.
- **Adversary's Goal**: For a given sample, **identify whether it belongs to the training data** (using the surrogate model based on the exploratory (inference) attack).
- **Outcome**: Leaked information to exploit vulnerabilities of the machine learning model.

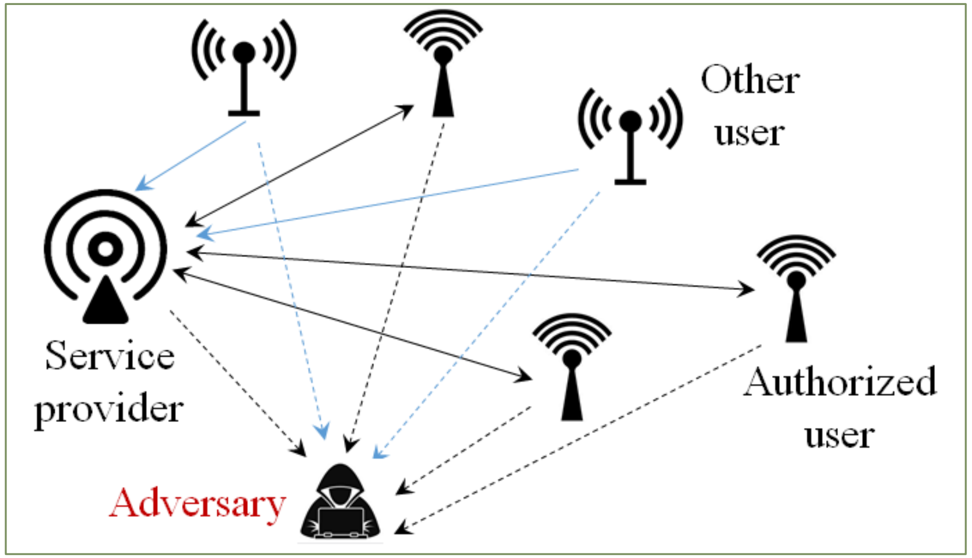




Membership Inference Attack in Wireless

- The adversary aims to infer if a signal of interest has been used to train a wireless signal classifier or not.
- Leak information on **waveform, device and channel characteristics** that are embedded in signals.
- Use the leaked information of authorized users to generate signals that infiltrate a user authentication system.





Y. Sagduyu, et al, ACM WiseSec, 2020.

Membership Inference Attack in Wireless

- Adversary builds **a surrogate classifier** by monitoring the spectrum activity of users and service provider.
  - The surrogate classifier is **not exactly the same as** the service provider's classifier due to channel differences.
- Features to infer the training data membership.
  - Case 1: **Both phase shift and received power values**.
  - Case 2: **Only received power values**.
  - Case 3: **Only phase shift values**.
- It is better to use both features together.
- Power is more important than phase shift for this attack.

Case 1

Real \ Predicted	non-member	member
non-member	0.9152	0.0848
member	0.1429	0.8571

Case 2

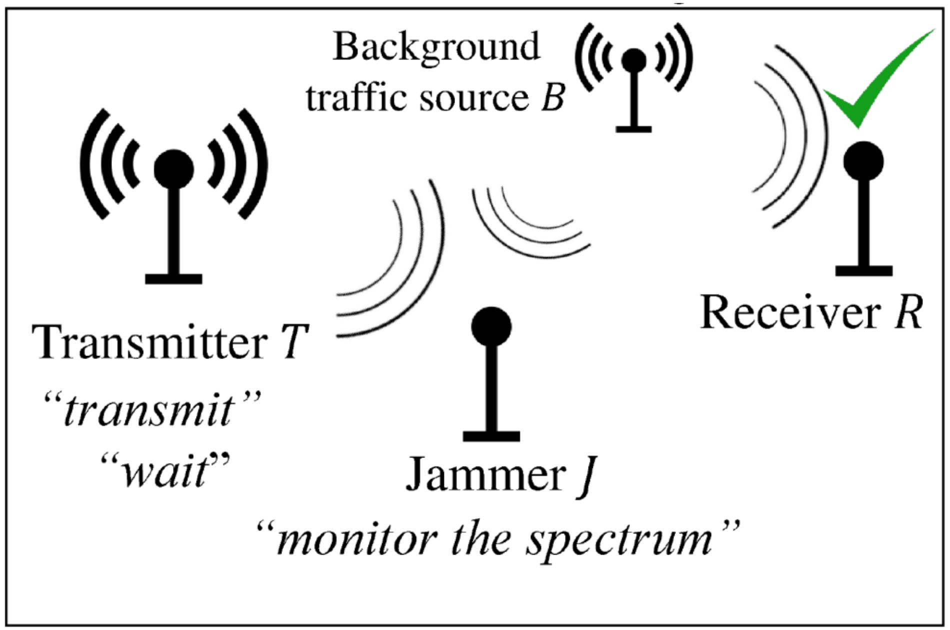
Real \ Predicted	non-member	member
non-member	0.5770	0.4230
member	0.1429	0.8571

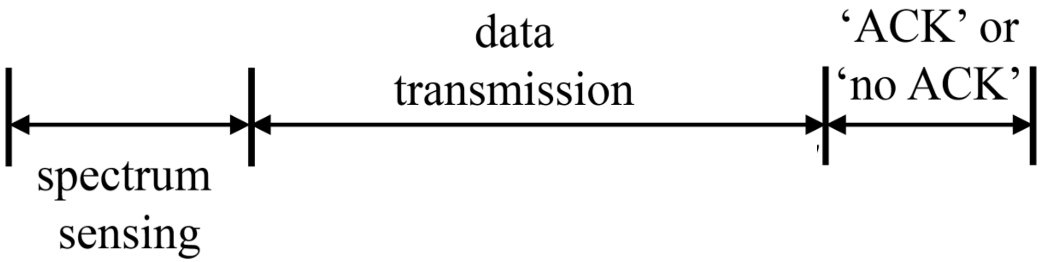
Case 3

Real \ Predicted	non-member	member
non-member	0.4766	0.5234
member	0.2199	0.7801

Inference Attack for Jamming

- There is a background (primary) transmitter using the channel intermittently.
- A transmitter senses the spectrum and **transmits when it predicts an idle channel**.
- Transmitter uses a **deep neural network** to predict when the channel is idle.
  - **Features:** Recent sensing results (RSSIs)
  - **Labels:** Channel is `idle' or `busy'
  - **Throughput** 0.304 packet/slot
  - **Success ratio** 73.79%
- If SNR $\geq$ threshold, the transmission is successful, and the receiver sends and ACK back to the transmitter.

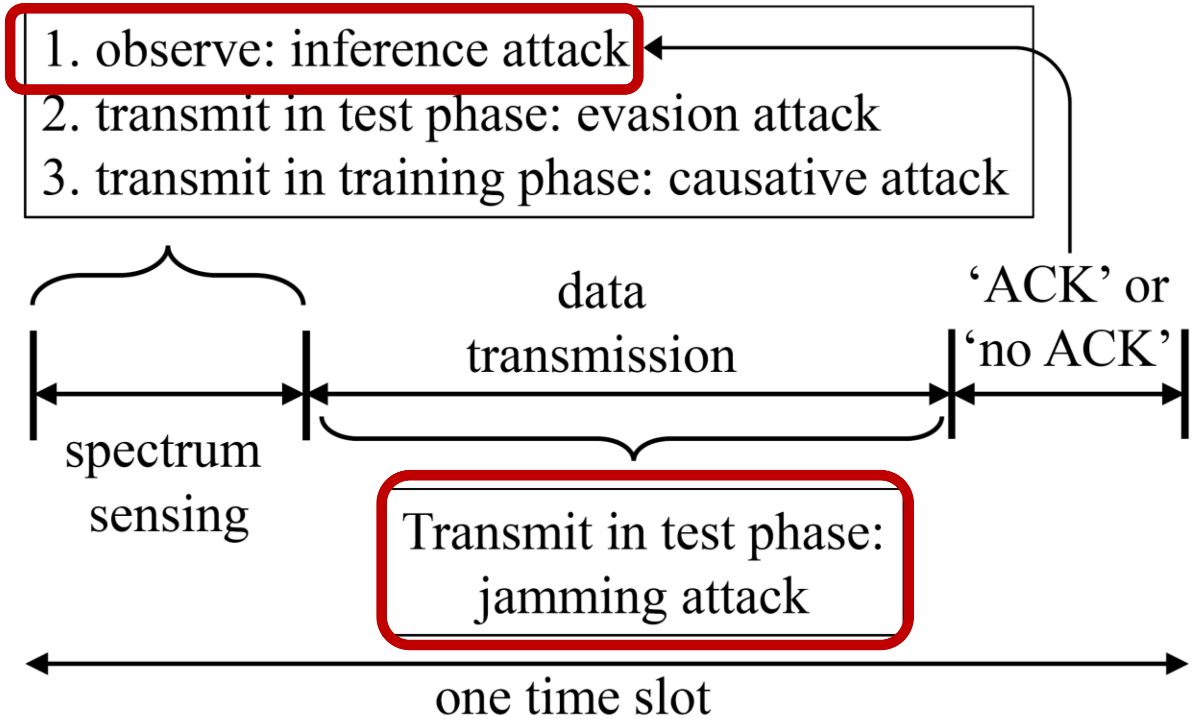




Inference Attack for Jamming

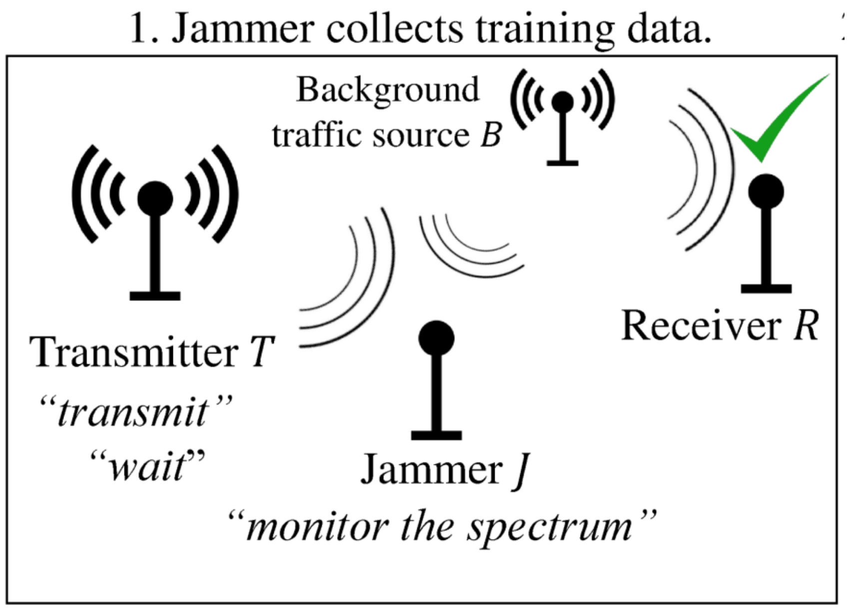
- An adversary trains a **surrogate classifier** (**inference attack**) by observing the spectrum.

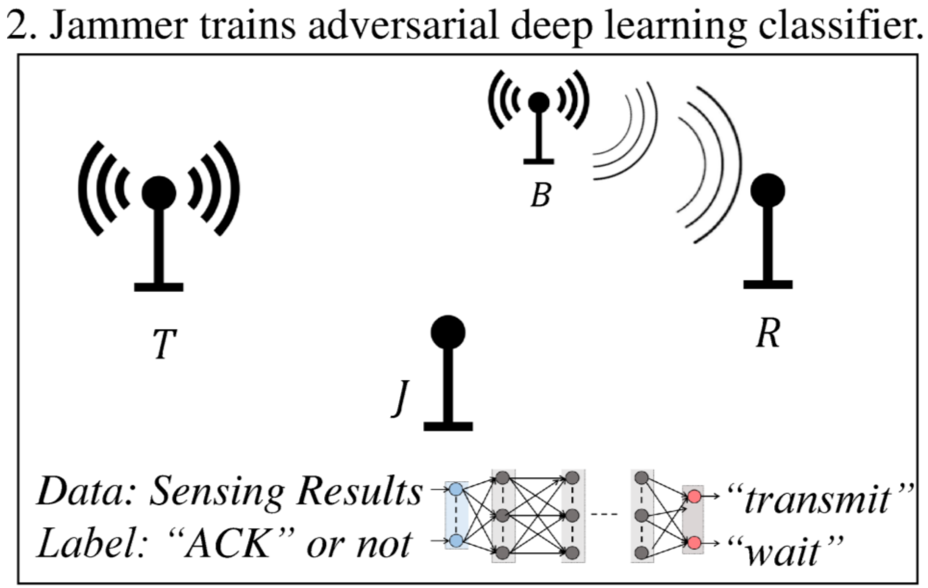
- The adversary senses the spectrum, uses its surrogate classifier to **predict when there will be a successful transmission**, and **jams the channel**.



T. Erpek, Y. Sagduyu, et. al, IEEE TCCN, 2019.

Steps 1-2 of the Attack (Inference Attack)



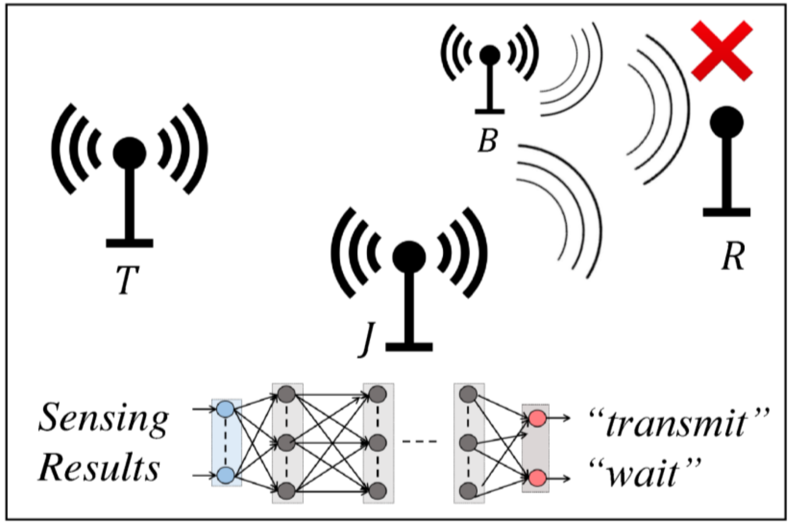


- The adversary's surrogate model **will not be the same** as the model of the transmitter.
- **Different features**
  - Sensing results at the adversary are different from those of the transmitter due to channel differences.
- **Different labels**
  - Transmitter classifies channel as idle or not.
  - Attacker classifies the current time slot as with a successful transmission (ACK) or not.

Step 3 of the Attack (Jamming Attack)

- The adversary uses its **surrogate model** and **jams the channel** when it predicts there will be a successful transmission based on sensing results.

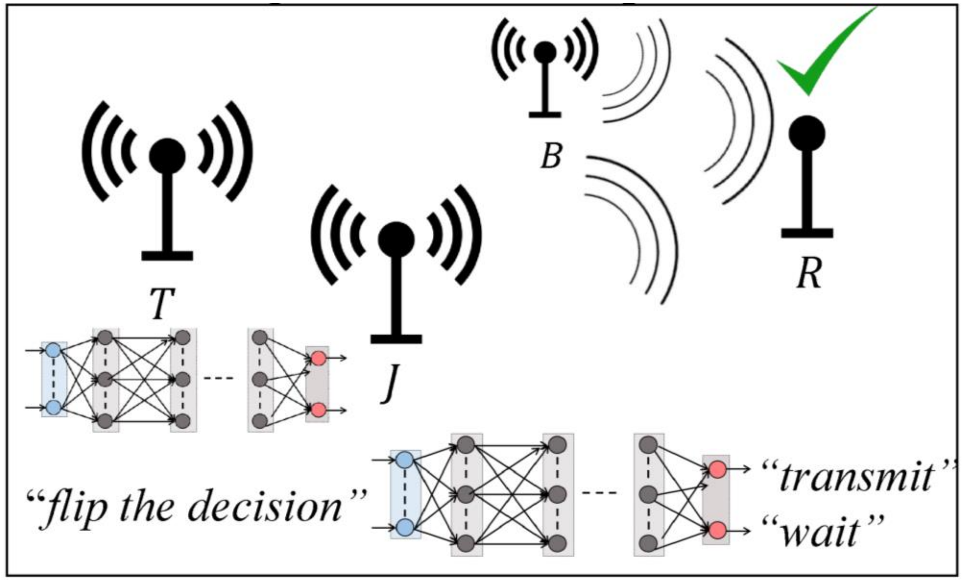
3. Jammer launches the attack.



Attack type	Throughput	Success ratio
No attack	0.766	95.75%
Adversarial deep learning	0.050	6.25%
Sensing-based attack ($\tau = 3.4$)	0.140	16.99%
Random attack	0.383	47.88%

Proactive Defense

- Transmitter's classifier makes **few deliberate errors**.
  - not transmitting even if channel is detected as idle, or
  - transmitting even if channel is detected as busy.
- Adversary **cannot build a reliable surrogate model**.
- **Defense goal**: Select the number of defense actions (add errors to samples with high classification confidence).



Defense increases

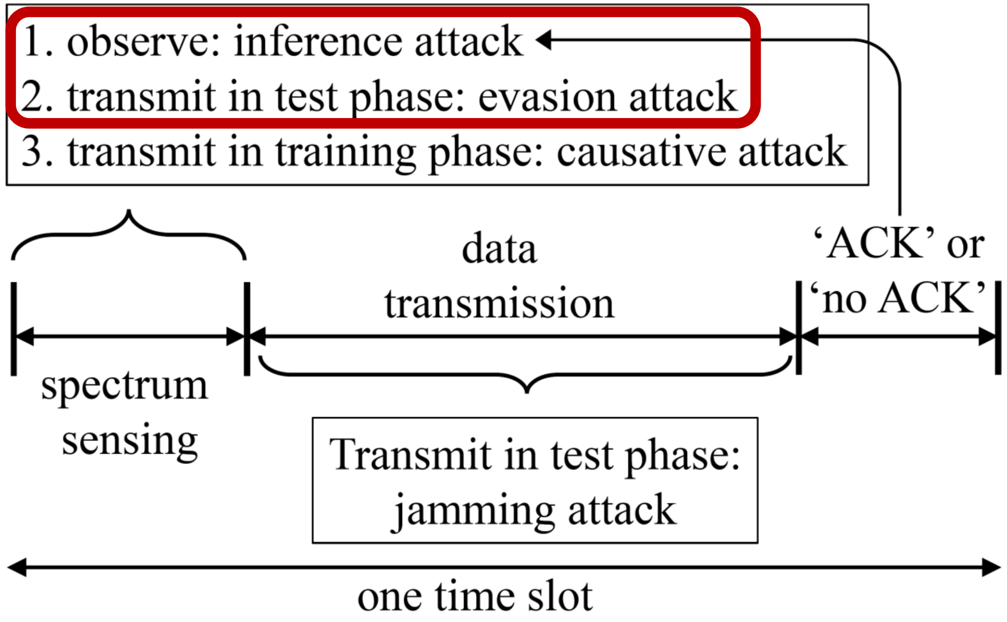
p_d	Jammer error probabilities		Transmitter performance	
	Misdetection	False alarm	Throughput (packet/slot)	Success ratio
0% (no defense)	4.18%	14.53%	0.050	6.25%
10%	17.53%	23.68%	0.132	17.98%
20%	32.80%	33.33%	0.216	31.67%
30%	33.92%	38.25%	0.194	30.41%
40%	35.83%	37.31%	0.178	31.67%
50%	38.97%	38.33%	0.170	32.32%

Best defense level

in terms of throughput

Attacks on Spectrum Sensing - 1

- Step 1: **Inference attack (build a surrogate model)**
  - False alarm = 1.98%, misdetection = 4.21%
- Step 2: **Evasion (adversarial) attack in test time.**
  - Using the surrogate model, **jam the (short) spectrum sensing period** such that the transmitter makes wrong transmit decisions.
  - Energy efficient and stealthy attack.



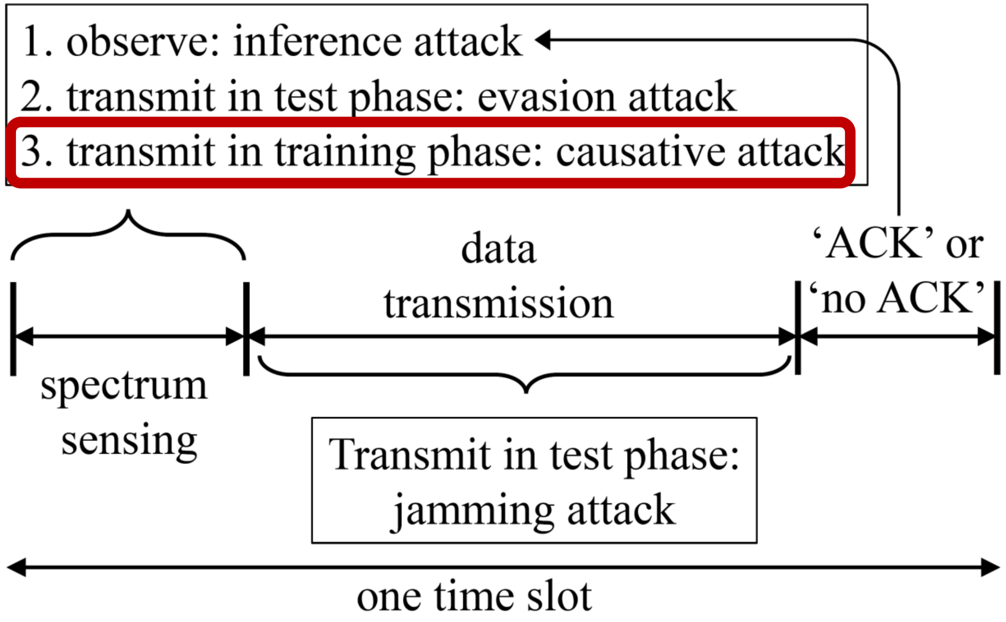
	Normalized throughput t	Success ratio s	All transmission ratio a
no attack	98.96%	96.94%	19.60%
with attack	3.13%	75.00%	0.80%

Y. Sagduyu, T. Erpek, et. al, IEEE TMC, 2020.

Attacks on Spectrum Sensing - 2

- Step 3: **Causative (poisoning) attack in (re)training time** (when the classifier is updated).
  - Using the surrogate model, **jam the spectrum sensing period** to make the updated classifier worse than before.

- Different attacks can be combined.



	Normalized throughput M_{Th}	Success ratio M_{Sr}	All transmission ratio M_{Tr}
no attack	98.96%	96.94%	19.60%
evasion attack	3.13%	75.00%	0.80%
jamming	41.67%	40.82%	19.60%
causative attack	87.27%	60.76%	31.60%
causative + evasion attack	2.72%	75.00%	0.80%
causative + jamming attack	37.27%	25.95%	31.60%

Proactive Defense

- The transmitter's classifier makes **some deliberate errors**.
- The adversary **cannot build a reliable surrogate model**.
- **Defense goal**: Select the number of defense actions (add errors more to samples with high classification confidence).

Defense increases

# of defense operations divided by # of all samples	Adversary error probabilities		Transmitter performance	
	Misdetection	False alarm	Normalized throughput	Success ratio
0% (no defense)	1.98%	4.21%	3.13%	75.00%
10%	6.99%	10.59%	15.63%	15.31%
20%	8.92%	35.29%	41.67%	28.78%
40%	10.12%	42.67%	51.04%	18.22%
60%	17.06%	69.44%	76.04%	18.07%
80%	10.88%	93.22%	56.25%	13.30%

Best defense level in terms of throughput

Machine Learning Today

ML Software Tools

TensorFlow

theano DEPRECATED

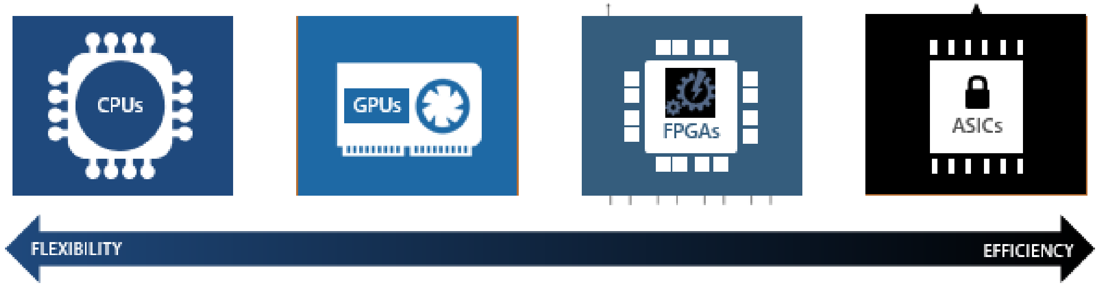
Keras

PyTorch

DEPRECATED Microsoft Cognitive Toolkit

MATLAB NN Toolbox

ML Computation Resources



https://docs.microsoft.com/en-us/azure/machine-learning/how-to-deploy-fpga-web-service

Google Cloud TPU

From cloud backend to embedded platforms

Nvidia Nano

Embedded Implementation

- Implement algorithms on embedded platforms for fast decisions in microsecond-millisecond time frame.
  - FPGA, embedded GPU, and ARM.
  - Support edge processing.
  - Determine the most applicable platform based on the latency, accuracy and power efficiency requirements.

DEEPip
Deep Learning on FPGA Fabric

Other Adversarial Machine Learning Attacks

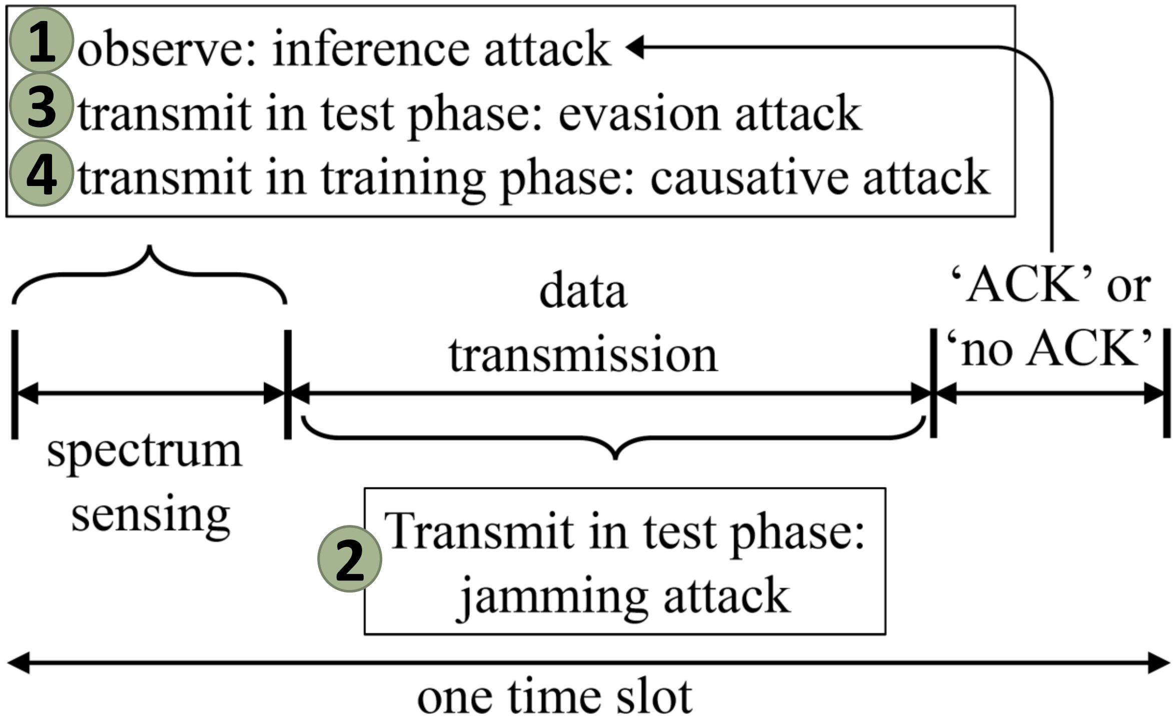
- **Dynamic spectrum access (DSA)**
  - An incumbent user transmits intermittently.
  - A transmitter senses the channel and transmits only when it is idle.

1 **Inference (exploratory) attack**
  - Sense the spectrum and train a surrogate model to mimic transmit behavior

2 **Inference-based jamming attack**
  - Use the surrogate model to predict and jam data transmissions that would other succeed.

3 **Evasion (adversarial) attack**
  - Jam the spectrum sensing period such that the transmitter makes wrong transmit decisions.

4 **Causative (poisoning) attack**
  - Jam the spectrum sensing period such that the transmitter makes wrong transmit decision.



T. Erpek, Y. Sagduyu, et. al, IEEE TCCN, 2019.
Y. Sagduyu, T. Erpek, et. al, IEEE TCCN, 2020.