Examining Machine Learning for 5G and Beyond Through an Adversarial Lens

Muhammad Usama

musama@ed.ac.uk

The University of Edinburgh
Informatics, Computer Science

Introduction

- Recent advances in networking such as software-defined networking (SDN), network function virtualization (NFV), and programmable data planes have introduced more flexibility of automation in communication and data networks.

- Technical developments toward 5G and B5G of mobile networks are quickly embracing a variety of deep learning (DL) algorithms as a de facto approach to help tackle the growing complexities of the network problems.

- However, the well-known vulnerability of the DL models to the adversarial machine learning (ML) attacks can significantly contribute to broadening the overall attack surface for 5G and beyond networks.

Attack Vectors

Salient Machine Learning Footprints Across the 5G Network and Adversarial Threats

	Subscriber Space	RAN	MEC	Core Network	Management and Control
SL	• Mobile sensor data analysis • Privacy preserving ML • User application development	• Channel estimation for beamforming • Modulation classification • User localization • Scheduling in energy harvesting	• Early detection of DDoS attacks • Speech recognition • Question answering systems • Malware detection	• Inferring network traffic flow • Protocol identification • QoS assurance • Network demand prediction	• Mobile and IoT data Analytics • Traffic forecasting for resource planning • Virtualized network resource selection
UL	• User activity analysis • Mobile healthcare • Channel autoencoder • Energy efficient sensor data prediction	• Mobility analysis • UE Trajectory prediction • Base-station failure detections • Coverage modeling	• User app performance analysis • IoT Data analytics • Crowdsourced localization • Sensor data analysis • Malicious traffic detection	• Traffic flow pattern classification • Group trajectory clustering • Urban user mobility prediction • Intrusion detection	• Synthetic traffic generation • New traffic features learning • VNF profiling
RL	• Indoor localization • Autonomous driving • Distributed sensor system routing	• Radio resource allocation • Transmitter training • Anti-jamming solution • Energy efficient coverage • Multi-RAT selection	• Secure crowd-sourcing • Content caching • Load-balancing • Computation offloading • UAV(IoT) path planning	• Traffic optimization • Handover optimization • Packet size optimization • Load-balancing • Traffic optimization	• VNF resource allocations • Multi-tenancy billing solution • Dynamic orchestration of the networks • Traffic scheduling

Generic Attack Vectors To 5G and B5G Networks

- Side Channel Attacks
- Man in the Middle (MitM)
- Jamming Attacks
- Cross-slice Attacks
- Untrusted Hardware
- Denial of Service (DoS)
- Brute-force Attacks on Encryption

Attack Vectors Contributed by ML Models

- Data Poisoning
- Oracle Attacks
- Gradient-free Attacks
- Evasion Attacks
- Backdoor Attacks

Contributions

Examining Machine Learning for 5G and Beyond Through an Adversarial Lens

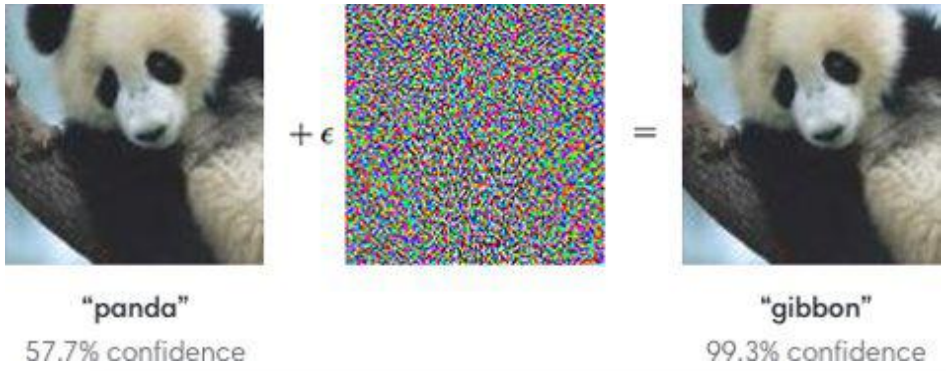
Muhammad Usama, Inaam Ilahi, and Junaid Qadir, *Information Technology University, Lahore, 54000, Pakistan*

Rupendra Nath Mitra and Mahesh K. Marina, *The University of Edinburgh, Edinburgh EH8 9YL, U.K.*

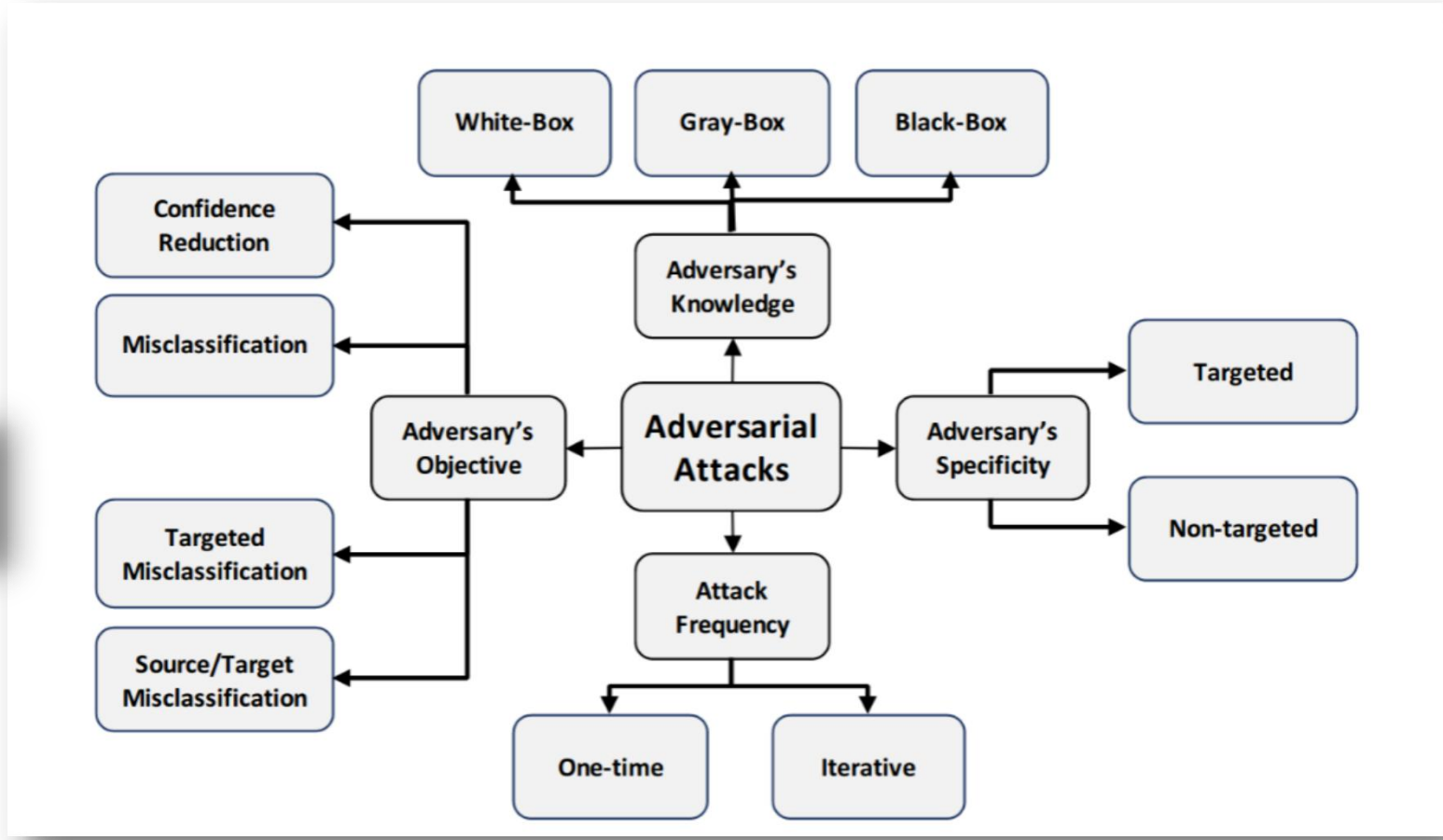
- In this work, we present a cautionary perspective on the use of AI/ML in the 5G context by highlighting the adversarial dimension spanning multiple types of ML and support this through three case studies.
- We also discuss approaches to mitigate this adversarial ML risk, offer guidelines for evaluating the robustness of ML models, and call attention to issues surrounding ML oriented research in 5G more generally.

Adversarial examples

"Adversarial examples are **inputs to ML models** that an attacker has intentionally designed to cause the model to make a mistake at **test time**."



$$x^* = x + \arg\min_{\delta}\{\|\delta\| : f(x+\delta) = t\}$$



Attacking Supervised ML-Based 5G Applications

- Automatic modulation classification is a critical task for intelligent radio receivers. The conventional maximum-likelihood and feature-based solutions are often infeasible due to the high computational overhead and domain expertise that is required.
- To make modulation classifiers more common in modern 5G and B5G networked devices, current approaches deploy **DL to build an end-to-end modulation classification systems** capable of automatic extraction of signal features in the wild.

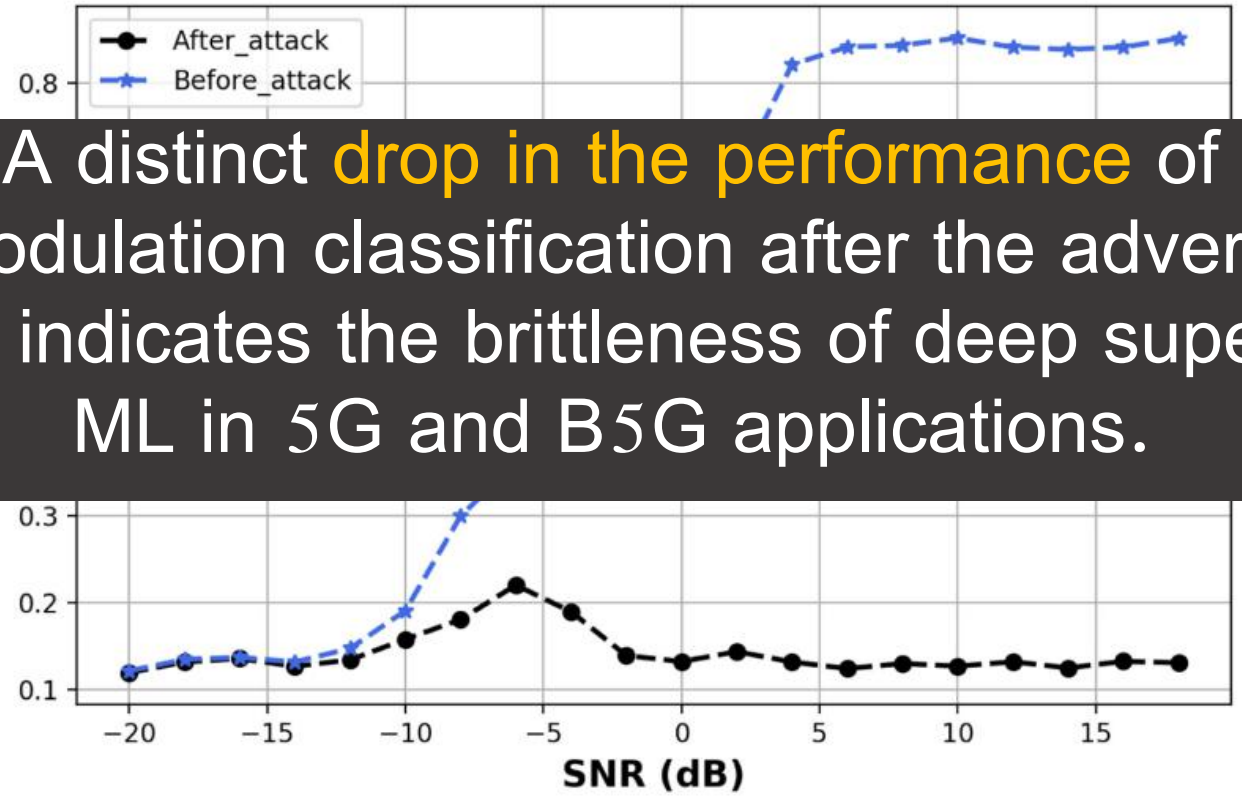
- We trained a **convolutional neural network (CNN) driven SL-based modulation classification model** on a well-known **GNU radio ML RML2016.10a dataset** that consists of 220,000 input examples of 11 digital and analog modulation schemes (AM-DSB, AM-SSB,WBFM, PAM4, BPSK, QPSK, 8PSK, QAM16, QAM64, CPFSK, and GFSK) on the signal-to-noise ratio (SNR) ranging from -20 to 18 dB [1].

- To show the feasibility of an adversarial ML attack on the CNN-based modulation classifier, we make the following assumptions.
  - We consider the **white-box attack model** where we assume that the adversary has a complete knowledge about the deployed modulation classifier.
  - Goal of the adversary is to **compromise the integrity** of the CNN classifier leading to a significant decay in the classification accuracy, which is the measure of the success of the adversary.

1. T. J. O'Shea and N. West, "Radio machine learning dataset generation with GNU radio," in Proc. GNU Radio Conf., vol. 1, 2016.

Attacking Supervised ML-Based 5G Applications

- To craft the adversarial examples to fool the CNN classifier, we use the Carlini and Wagner (C&W) attack [1] for each modulation class by minimizing the L2 norm on the perturbation ∂, such that when the perturbation ∂ is added to the input x and sent to the CNN-based modulation classifier C it misclassifies the input x.



A distinct drop in the performance of the modulation classification after the adversarial attacks indicates the brittleness of deep supervised ML in 5G and B5G applications.

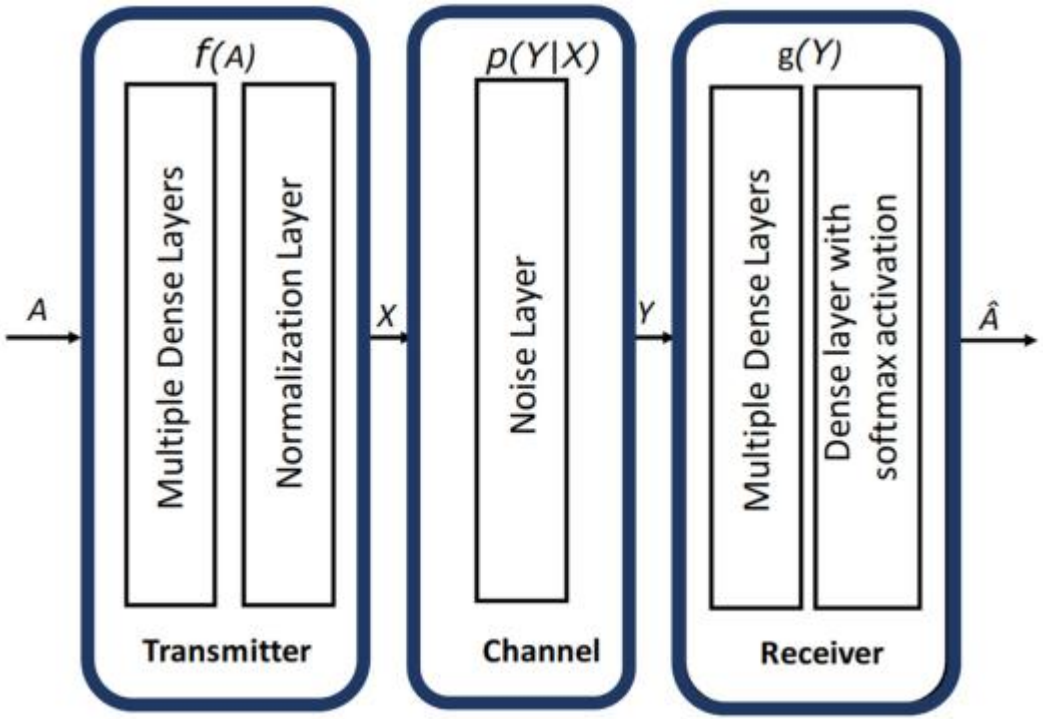
1. N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," in Proc. IEEE Symp. Secur. Privacy, May 2017 pp. 39–57.

Attacking Unsupervised ML-Based 5G Applications

- **Deep autoencoder based communication model** is seen as a viable alternative to the dedicated radio hardware in the future 5G and beyond networks [1,2].

- We assume the model is subjected to an additive white Gaussian noise (AWGN) channel and apply the parameter configurations provided in [3].

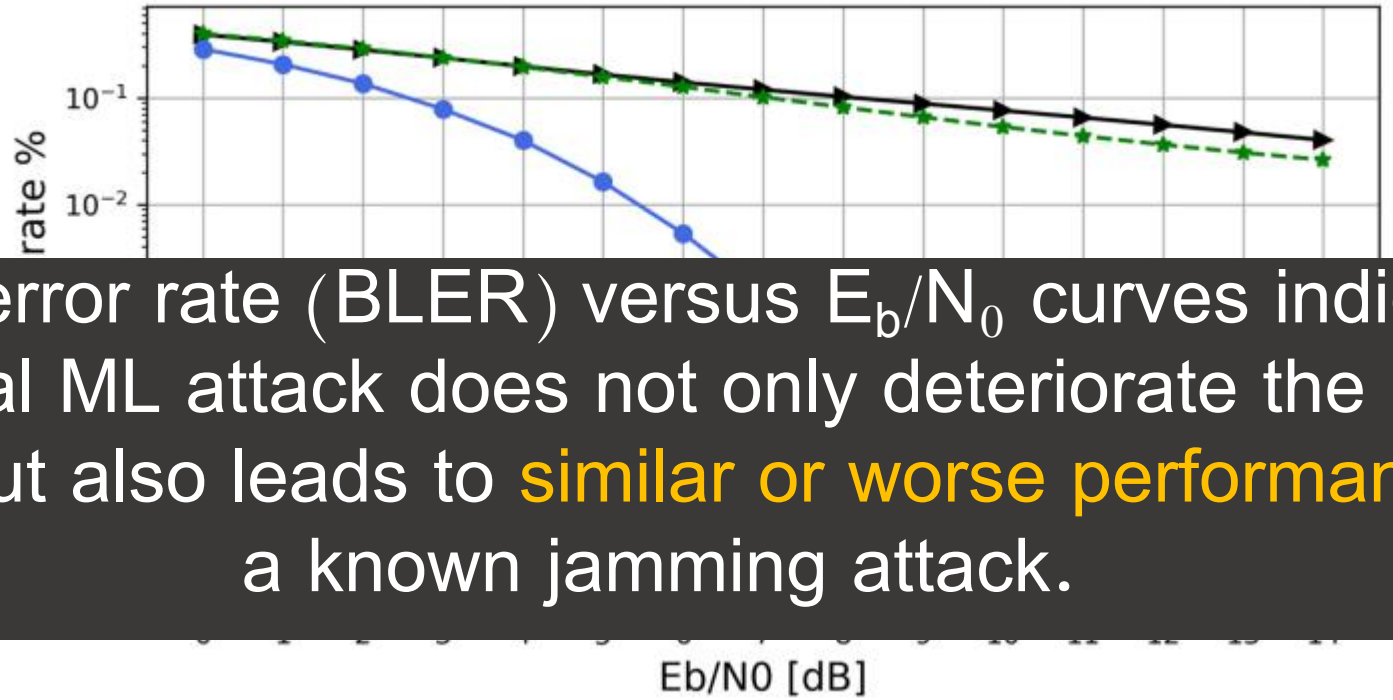
- To show the feasibility of an adversarial ML attack on the deep autoencoder-based communication model, we make the following assumptions.
  - We assume a **white-box setting**.
  - We further assume that the autoencoder learns a broadcast channel.
  - The **goal of the adversary** is to compromise the integrity of channel autoencoder and the success of the adversary is measured by the elevated BLER with improving SNR per bit (E_b/N_0).



1. T. J. O'Shea, K. Karra, and T. C. Clancy, "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," in Proc. IEEE Int. Symp. Signal Process. Inf. Technol., 2016, pp. 223–228.
2. B. Hilburn, T. J. O'Shea, T. Roy, and N. West, "DeepSig:Deep Learning for Wireless Communications.
3. T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," IEEE Trans. Cogn. Commun. Netw., vol. 3, no. 4, pp. 563–575, Dec. 2017.

Attacking Unsupervised ML-Based 5G Applications

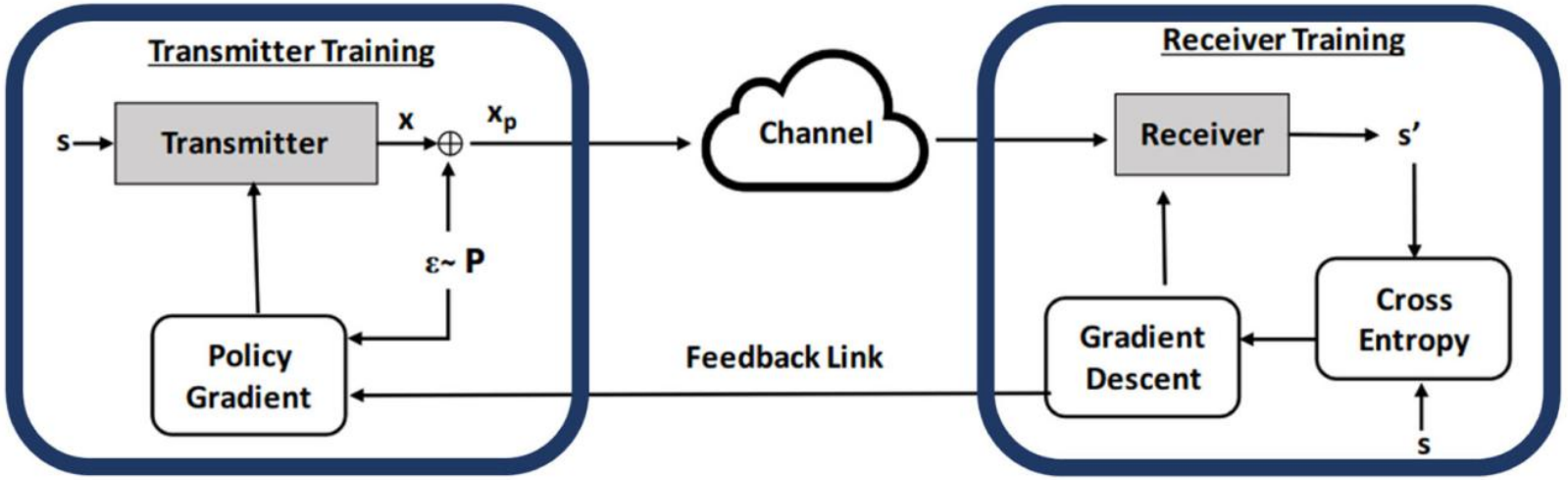
- We take the following two-step data-independent approach to craft adversarial examples for the channel autoencoder.
  1. Sample the Gaussian distribution randomly (because the channel is AWGN) and use it as an initial adversarial perturbation ∂.
  2. Maximize the mean activations of the decoder model when the input of the decoder is the perturbation ∂.
- This produces maximal spurious activations at each decoder layer and results in the loss of the integrity of the channel autoencoder.



The block error rate (BLER) versus E_b/N_0 curves indicates that adversarial ML attack does not only deteriorate the model's performance but also leads to similar or worse performance than with a known jamming attack.

Attacking Reinforcement ML-Based 5G Applications

- We performed the adversarial ML attacks on an **end-to-end DRL autoencoder with a noisy channel feedback system [1].** The end-toend training procedure involves the following.
  - The RL-based transmitter training by a policy gradient theorem to ensure that the intelligent transmitter learns from the noisy feedback after a round of communication.
  - SL model-based receiver training to train the receiver as a classifier.

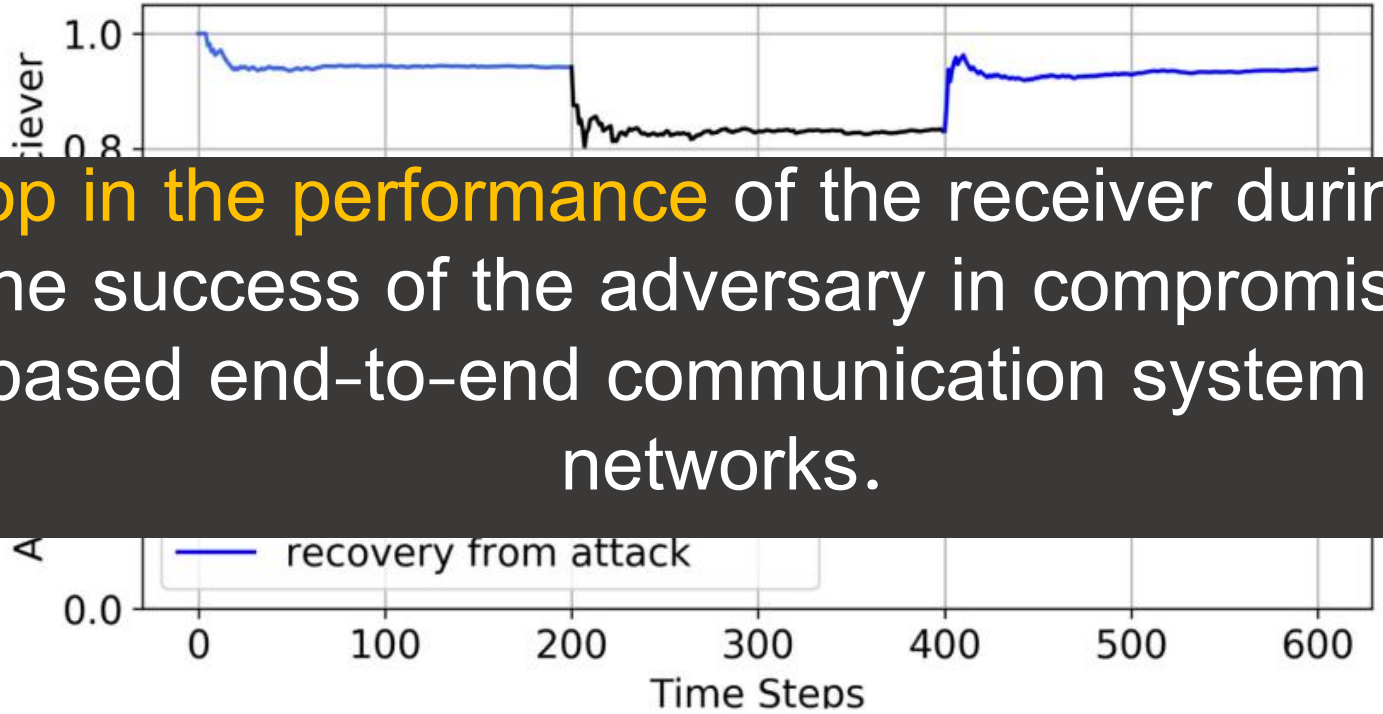


- The considered threat model for this case study is given as follows:
  - We choose a realistic black-box setting where the adversary does not know the target model. We also assume that the adversary can perform an adversarial ML attack for "n"-time steps.
  - The goal of the adversary is to compromise the performance of the DRL autoencoder with noisy feedback for a specific time interval. The success of the adversary is measured by the degradation in the decoder's performance during the attack interval.

1. M. Goutay, F. A. Aoudia, and J. Hoydis, "Deep reinforcement learning autoencoder with noisy feedback," in Proc. 17th Int. Symp. Model. Optim.Mobile, Ad Hoc, Wireless Netw., 2019, pp. 1–6.

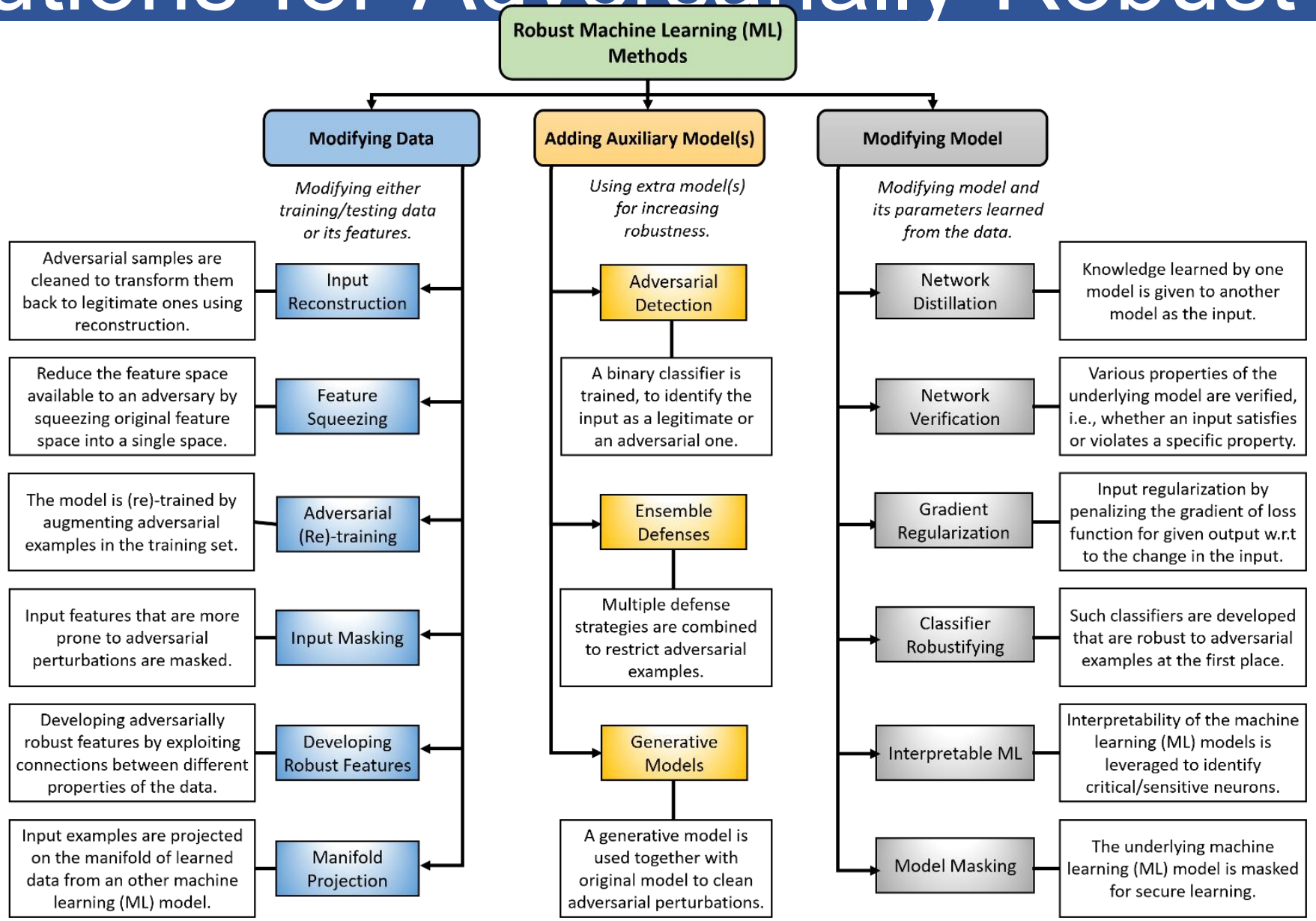
Attacking Reinforcement ML-Based 5G Applications

- We exploit the **transferability property of the adversarial examples**, which states that adversarial examples compromising an ML model will compromise other ML models with high probability if the underlying data distribution is same between two victim models.
- We transfer the adversarial examples crafted in the previous case study and measure the average accuracy of the receiver.
- We run the DRL autoencoder with a noisy feedback system for 600-time steps (one timestep is equal to one communication round) and perform the adversarial attack between 200 and 400-time step window.



A clear drop in the performance of the receiver during the attack indicates the success of the adversary in compromising the DRL autoencoder-based end-to-end communication system in future mobile networks.

Solutions for Adversarially Robust ML



Qayyum, A., **Usama, M.**, Qadir, J., & Al-Fuqaha, A. (2020). Securing connected & autonomous vehicles: Challenges posed by adversarial machine learning and the way forward. *IEEE Communications Surveys & Tutorials*, 22(2), 998-1026.

Conclusions

- Security and privacy are uncompromising necessities for modern and future global networks standards such as 5G and B5G, and accordingly fortifying it to thwart attacks and withstand the rapidly evolving landscape of future security threats is of vital importance.
- This work specifically highlights that the unvetted adoption of DL-driven solutions in 5G and B5G networking gives rise to security concerns that remain unattended by the 5G standardization bodies, such as the 3GPP.
- We hope that our work will motivate further research toward "telecomgrade ML" that is safe and trustworthy enough to be incorporated into 5G and B5G networks, thereby power intelligent and robust mobile networks supporting diverse services including mission-critical systems.