Wild Networks: Exposure of 5G Network Infrastructures to Adversarial Examples

Huawei Innovation Workshop
on Artificial Intelligence for Cyber-Security
Friday, July 23rd, 2021

UNIVERSITÄT LIECHTENSTEIN

Dr. Giovanni Apruzzese
Prof. Pavel Laskov

Whoami

- **Background:**
  - > From *Italy*
  - > Studied at the *University of Modena* where I obtained my PhD in Information and Communication Technologies in 2020
  - > In 2019, I was a research scholar at *Dartmouth College* (NH, USA) for 6 months, advised by Prof. VS Subrahmanian
  - > Joined *University Liechtenstein* in July 2020 as a PostDoc, working together with Prof. Pavel Laskov

- **Research Interests:**
  - > Cyber security
  - > Machine learning
  - > Adversarial ML attacks
  - > Any network-related topic

- **Contact information**: giovanni.apruzzese@uni.li
  - > Feel free to contact me for any questions or remarks!

OUTLINE

- Introduction
- Background
- Challenges
- Proposal – Application Scenario
- Proposal – Definition
- Proposal – Generality
- Realistic Evaluation Framework
- Case Study – Power Allocation
- Case Study – Network Slicing (flow-based)
- Conclusions

Introduction

- The 5G Network Infrastructure (NI) requires to support millions of devices while guaranteeing optimal quality of service.
- ML is expected to play a crucial role in 5G NI.

- **Problem**: *lack of realistic security assessments of specific threats to ML in 5G scenarios*.
  - > Conventional security aspects: ✓
  - > ML-specific security: ✘

- **Challenge**: addressing ML-specific threats in 5G NI is difficult.
  - > ML methods can be deployed anywhere in the 5G NI
    → *necessity of providing a general threat model*
  - > The currently deployed 5G NI does not use ML yet, and available data is scarce
    → *how to evaluate realistic adversarial attacks?*

In this talk:
- we propose a generic threat model of feasible ML attacks against the 5G NI.
- we present a realistic framework for the evaluation of such adversarial attacks.
- we assess adversarial attacks using the proposed threat model and evaluation framework.

Background

- The primary goal of ML in 5G is to ensure the *guaranteed levels of* ***Quality of Service*** *(QoS)*
  - > At the foundation of 5G, are Service Level Agreements (SLA) between the providers of the 5G NI and its tenants (service providers).
  - > If the 5G NI cannot meet the QoS in the SLA, then the 5G providers will incur in substantial penalties (**$$$**).

- **Network management is crucial** to ensure that User Equipment (UE) receive proper QoS.

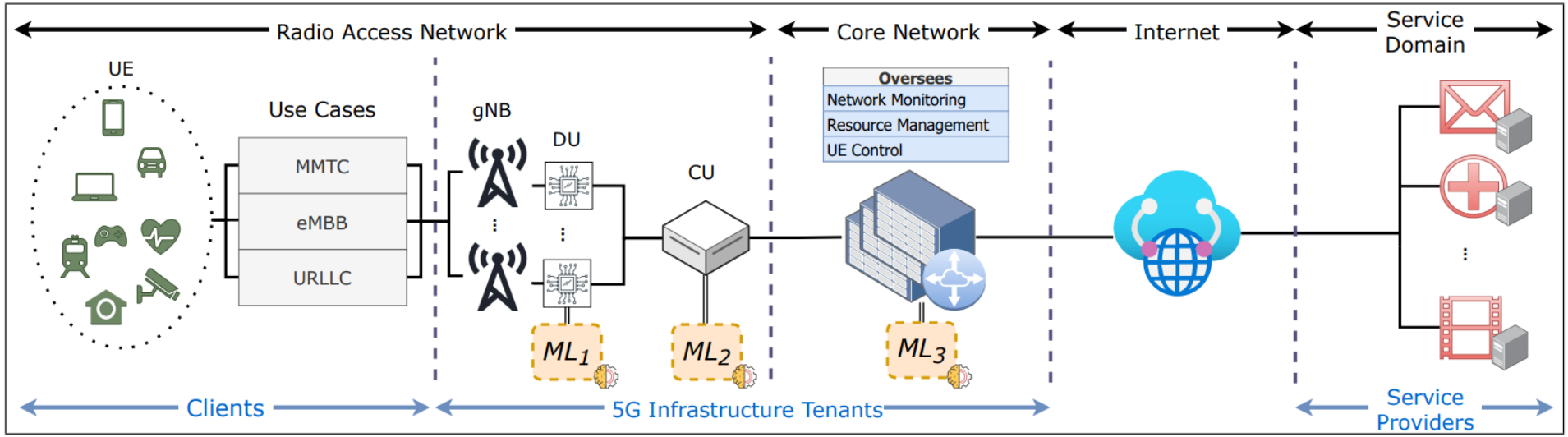


Fig. 1: The 5G Ecosystem. The *clients* transparently use the network infrastructure deployed by the 5G *tenants* to reach the service *providers*.

- Prototype applications of ML in 5G NI include *Network Slicing* or *Power Allocation*

Motivation

- Using ML in 5G exposes to specific security risks: *adversarial attacks*.
  - > Tiny perturbations in the input data cause a given ML model to predict an incorrect output.

- The threat of *Adversarial Examples* has been recognized at many levels (e.g., NIST [1] or EU [2]).

- **Existing threat models are inadequate**: adversarial attacks are usually very effective…
  …but the underlying assumptions portray extremely strong attackers.
- Attacks with high impact but little feasibility are irrelevant and misleading for practical deployment

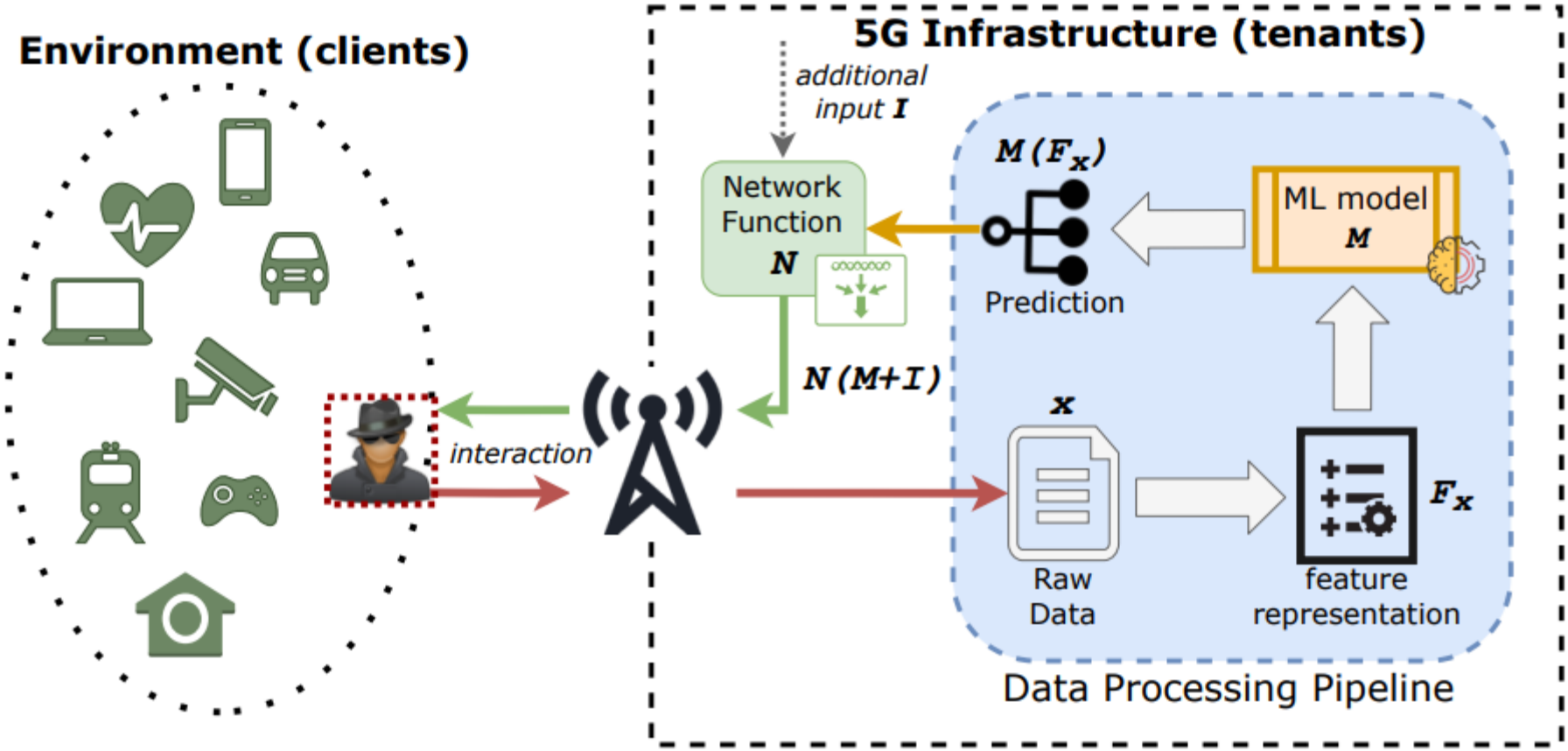
→ There is a need of a proper assessment. This requires the definition of a *realistic threat model*.
→ The multiple applications of ML in 5G require the threat model to be *general*.
→ The effects of attacks conforming to the threat model must also be *evaluated in a realistic way*.

- **Evaluations are difficult**: the current 5G NI hardly uses ML yet.
  - > Which ML method to consider?
  - > Which data to use?
  - > How to craft realistic adversarial samples?

Source: [1] Tabassi, Elham, et al. "A taxonomy and terminology of adversarial machine learning." NIST IR (2019).
[2] https://www.enisa.europa.eu/news/enisa-news/enisa-ai-threat-landscape-report-unveils-major-cybersecurity-challenges

Giovanni Apruzzese
Pavel Laskov

Myopic Threat Model – Application Scenario



Myopic Threat Model – Definition

- The attacker can be characterized as follows:
  - **Goal.** *Cause damage* to the 5G NI provider through untargeted adversarial attacks.
  - **Knowledge.** *Limited* to: (i) there is a ML component *M* performing a network-related task; and (ii) the data-type analyzed by *M -> used to identify a subset of its features* $\mathcal{F} \subseteq F$
  - **Capability.** *Constrained:* can only control the UE she owns; no control on the 5G NI; cannot inspect the specific output of *M*; can only consciously influence a subset of known features $\bar{\mathcal{F}} \subseteq \mathcal{F}$ in the problem space → such changes will affect also features beyond the attacker's knowledge $\bar{F} \supseteq \bar{\mathcal{F}}$
  - **Strategy.** *Guessing* a perturbation by altering the normal behaviour of her UE(s), thus resulting in an adversarial example that may fool *M*.

- We define this type of attacks as "*myopic*" adversarial attacks.

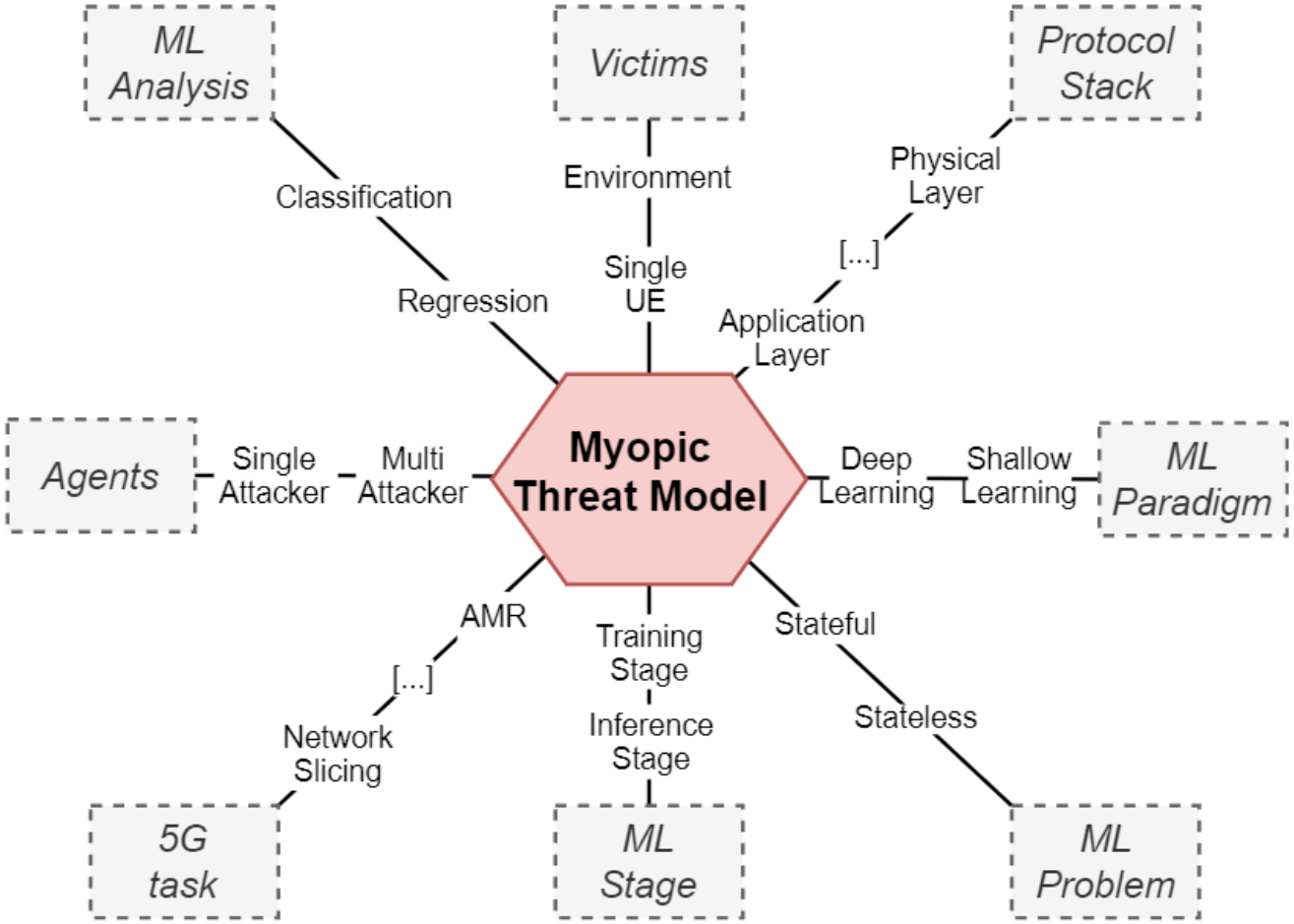
TABLE I: Myopic threat model vs existing 'box' threat models.

	White box	Gray box	Black box	No box	*Myopic*
Available Knowledge	M, F	$\mathcal{F}$	✗	F, T	$\mathcal{F}$
Optimal Perturb.	✓	✓	✓	✓	✗
ML prediction $M(F_x)$	✓	✓	✓	✗	✗

Source: [No box attacks] Chen, Pin-Yu, et al. "Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models." Proceedings of the 10th ACM workshop on artificial intelligence and security. 2017.

Myopic Threat Model – Generality

– The myopic threat model can be used to design hundreds of different adversarial scenarios.



Realistic Evaluation Framework - Challenges

- Evaluating real adversarial attacks requires to operate in the *problem-space* [3].
- In the case of myopic attacks, this requires to:
  - manipulate the *UE* owned by the attacker (easy!);
  - have the corresponding data collected by the 5G infrastructure (tough!); and
  - analyzed by some ML model (impossible!)

This cannot be done today ☹

- How to overcome such limitation and provide realistic assessments?
  - Use prototype SotA ML components for 5G
  - Devise such ML components though publicly available datasets containing *raw data*
  - Apply the adversarial perturbation on *raw data* contained in such datasets
    - Once the *perturbed raw data* is transformed into its feature representation, we will obtain the corresponding *adversarial example*.
    - We define such perturbation as a **"Raw-data space Perturbation"** (RsP)

Source: [3] Pierazzi, Fabio, et al. "Intriguing properties of adversarial ml attacks in the problem space." IEEE Symposium on Security and Privacy (SP). 2020.

Giovanni Apruzzese
Pavel Laskov

Realistic Evaluation Framework - Workflow

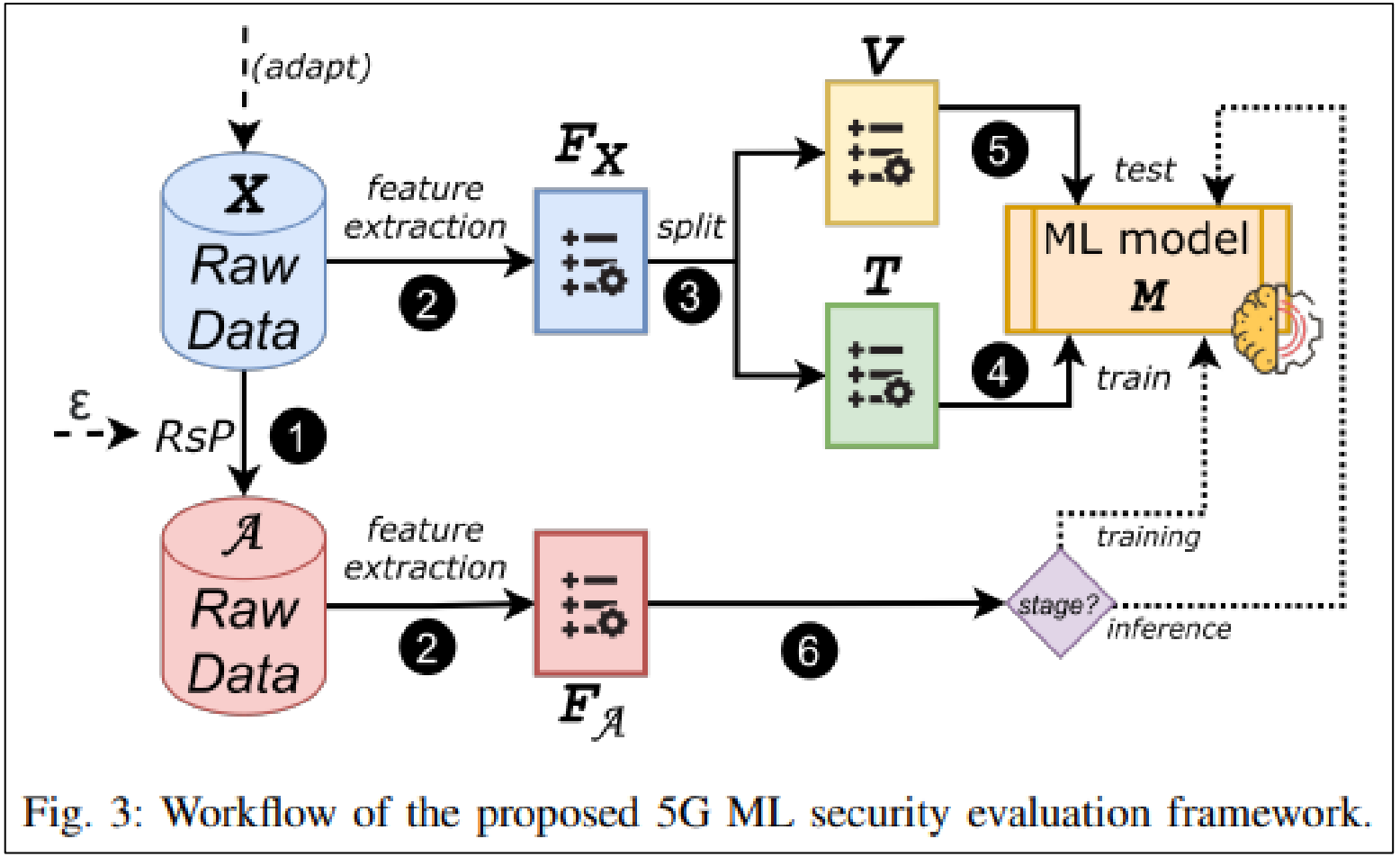
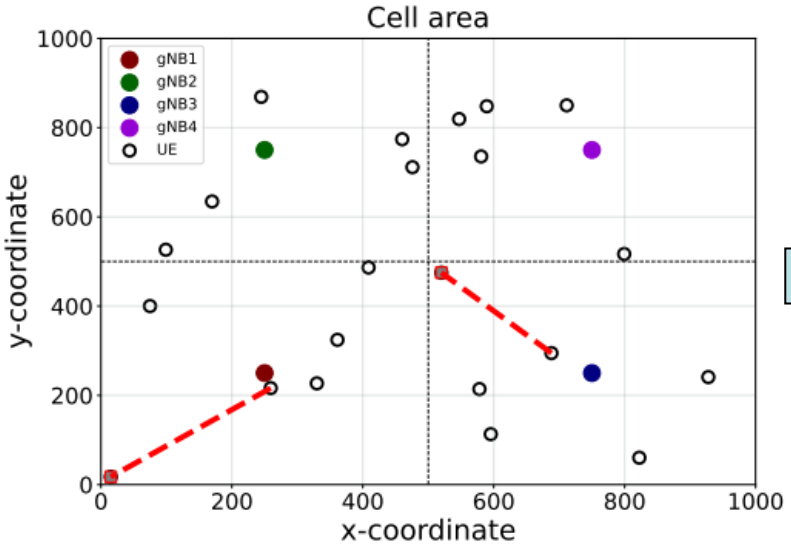
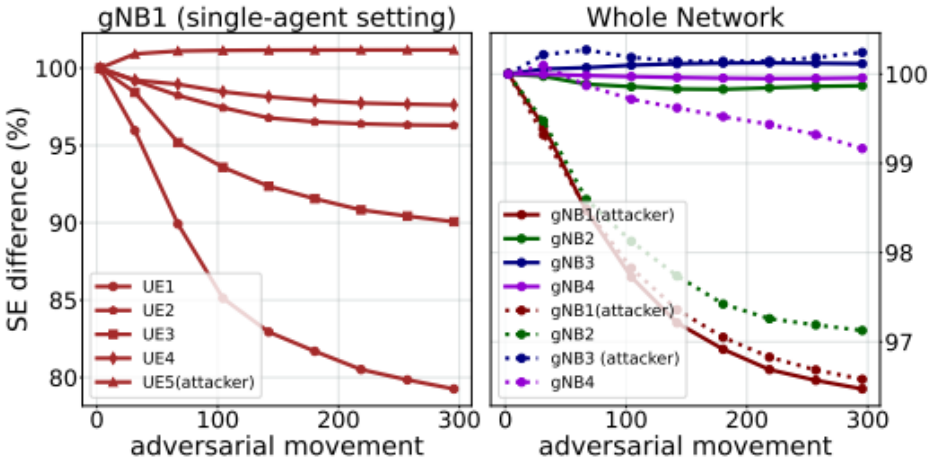


Fig. 3: Workflow of the proposed 5G ML security evaluation framework.

Case Study – Power Allocation

- *Power Allocation:* the distance between a UE and its gNB can be analyzed by DL to predict how much power to allocate to the gNB. This is done by adjusting the *spectral efficiency*.
- A myopic attacker can "spoof" his geographical position, faking the true distance to the gNB.
  → When such data is analyzed by the target DL model, it will be subject to a myopic attack



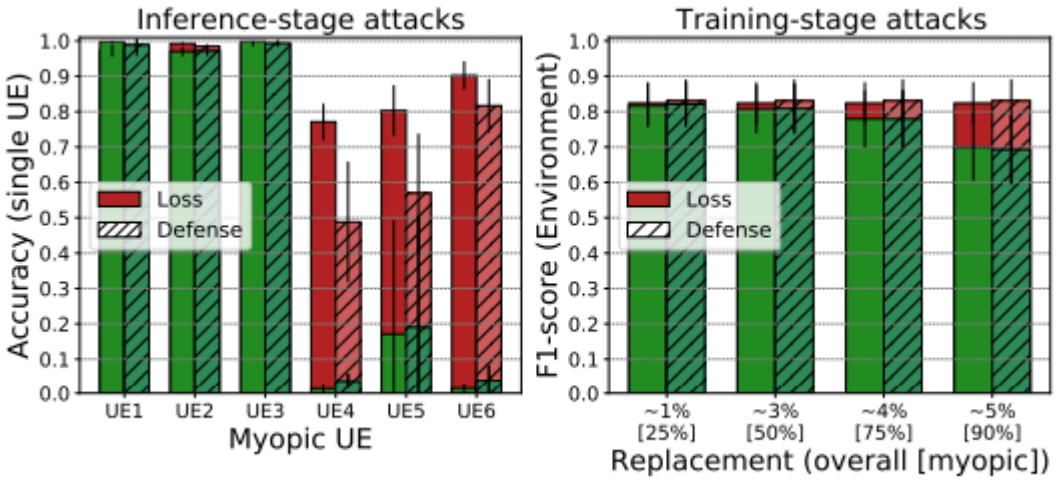


- The result of such simple attack show that even a state of the art DL model can be fooled with such simple strategies – which require almost no preparation by the attacker!
- Interestingly, a myopic attacker would be successful through *unsuccessful* adversarial examples!

Source: L. Sanguinetti, A. Zappone, and M. Debbah, "Deep Learning for Power Allocation in massive MIMO," Proc. IEEE Asilomar Conf. Sig. Syst. Comp., 2018

Case Study – Network Slicing through NetFlows

- *Network Slicing* aims at "slicing" the network resources to suit the demands of the environment.
- We consider a case study where network traffic (in the form of NetFlows) is analyzed by a RF to predict if it is "active" or "background" (which should be given higher/lower priority).
- A myopic attacker can influence a similar system by modifying the behavior of her UEs., e.g., by increasing the payload or duplicating packets.
  - → The attack occurs when the packets are converted to NetFlows and analyzed by the RF.
  - → The attack can also occur if the "myopic flows" are used to retraining (expected in 5G).
  - → We also evaluate Defensive Distillation as a defense to Myopic Attacks



- In some cases, the myopic attacks are not very effective, but in others they can significantly decrease the performance – of single UEs, or of the whole network.

Source: https://www.stratosphereips.org/datasets-ctu13

Conclusions

- The 5G infrastructure will greatly benefit from ML methods.
- For real deployments, it is crucial to evaluate the robustness of ML to adversarial examples.
  - > Existing adversarial ML threat models are inadequate
  - > Realistic assessments are difficult

- We proposed a realistic and general threat model that can be used to design adversarial attack scenarios against the 5G network infrastructure.
- We showed a realistic framework for the evaluation of adversarial attacks.
- We assess the impact of attacks conforming to the proposed threat model.

- We show that even small changes can effectively fool state of the art ML approaches.

Assessment of Adversarial Examples in 5G Network Infrastructures

Huawei Innovation Workshop
on Artificial Intelligence for Cyber-Security
Friday, July 23rd, 2021

UNIVERSITÄT LIECHTENSTEIN

Dr. Giovanni Apruzzese
Prof. Pavel Laskov

Myopic Threat Model – Application Scenario

- Multiple heterogeneous devices that are receiving the services provided by the 5G network infrastructure
- The data generated by the environment is then received, collected and analyzed by the organization managing the 5G services
- Such data is analyzed by some ML method to assist some network function
- The environment receives the “feedback” of the 5G infrastructure
- Such feedback is the result of the combined effects of the entire 5G infrastructure to the operations performed by the entire environment
- The attacker is a *client* in the 5G network infrastructure, with full control on her UE but that is subject to realistic constraints; she has limited visibility into the 5G infrastructure.