Pattern Recognition and Applications Lab

University of Cagliari, Italy

Università di Genova

Cybersecurity and AI: The PRAlab Research Experience

Maura Pintor, *Giulia Orrú, Davide Maiorca, Ambra Demontis, Luca Demetrio, Gian Luca Marcialis, Battista Biggio and Fabio Roli*

Ital-IA, Italy, May 29th, 2023

PRALab People / Research / Projects

Faculty members:
Battista Biggio
Ambra Demontis
Luca Didaci
Giorgio Fumera
Giorgio Giacinto
Davide Maiorca
Gian Luca Marcialis
Giulia Orrù
Maura Pintor
Lorenzo Putzu
Fabio Roli (Lab Director)

PhD students:
Daniele Angioni
Sara Concas
Simone Maurizio La Cava
Srishti Gupta
Emanuele Ledda
Gianpaolo Perelli
Giorgio Piras
Alessandro Sanna

Post-doc:
Rita Delussu
Angelo Sotgiu
Marco Micheletto
Roberto Casula

Lab fellows:
Carlo Cuccu
Doriano Edosini
Andrea Panzino

Research fields
AI Security and Safety
Cybersecurity
Biometrics
Multimedia Analysis, Video Surveillance and Ambient Intelligence
Brain and Medical Signal Processing

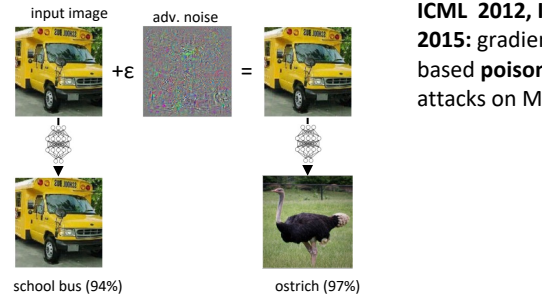
- 25+ research projects funded in 2012-2022
- 8 EU projects (2 coordinated by PRA Lab)
- 1.5 M€ EU funding
- More than 3M€ overall funding
- 400k€ yearly turnover

Recent projects on AI Security
- HE Sec4AI4Sec 2023-2025
- HE ELSA 2022-2024
- PRIN 2017 RexLearn
- FFG Comet Module S3AI

Pioneers of Machine Learning Security

- Our team is internationally recognized among the pioneers of AI/ML security
  - we have been the first to discover the impact of gradient-based attacks on ML models
  - we have been the first to discover and systematize adversarial attacks on AI/ML, prior to their application to deep learning

ECML-PKDD '13: gradient-based **evasion** attacks on ML (**one year before adversarial examples**)



ICML 2012, ICML 2015: gradient-based **poisoning** attacks on ML

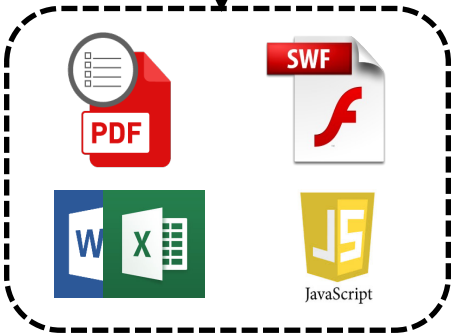
Attacker's Capability	Attacker's Goal: Misclassifications that do not compromise normal system operation	Misclassifications that compromise normal system operation	Querying strategies that reveal confidential information on the learning model or its users
	Integrity	**Availability**	**Privacy / Confidentiality**
Test data	Evasion (a.k.a. adversarial examples)	Sponge attacks	Model extraction / stealing Model inversion (hill climbing) Membership inference
Training data	Backdoor poisoning (to allow subsequent intrusions) – e.g., backdoors or neural trojans	DoS poisoning (to maximize classification error)	-

B. Biggio and F. Roli, *Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning*, Pattern Recognition, 2018 - ***2021 Best Paper Award and Pattern Recognition Medal***

Machine Learning for Cybersecurity

Problem I: is it effective to **detect Malware with Machine Learning**?

Non-Binary Malware

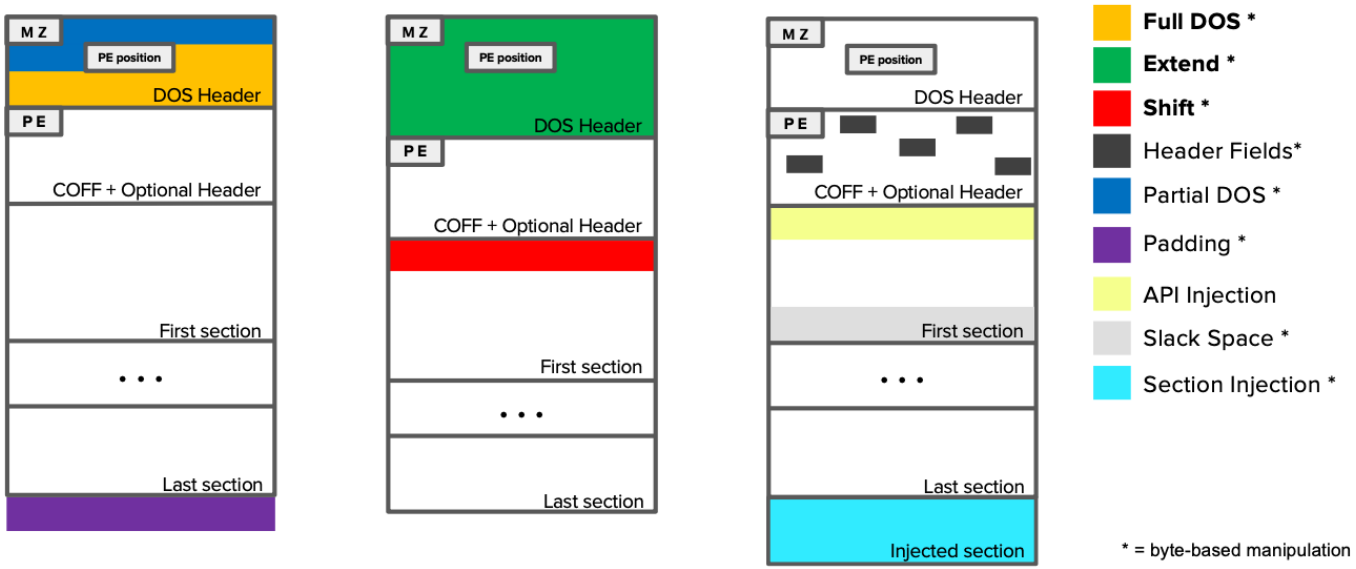


Binary Malware

Problem II
Security Issues of ML Malware Detectors
*Can attackers **evade** or **mislead** malware detectors based on machine learning?*

Problem III
Advanced Program Analysis
How can attackers conceal hidden (and malicious) functionalities in programs?

Adversarial EXEmples: Practical Attacks on Machine Learning for Windows Malware Detection



Demetrio, Biggio, et al., Adversarial EXEmples, ACM TOPS 2021
Demetrio, Biggio, et al., Functionality-preserving ..., IEEE TIFS 2021

Machine Learning Security Publication Highlights

Attacks on Machine Learning

ECML '13 / ICML '12, '15: Pioneering work on gradient-based evasion and poisoning attacks

USENIX Sec. '19: Transferability of evasion and poisoning attacks

IEEE TDSC '19, IEEE TIFS/ACM TOPS '21: Adversarial perturbations on Android and Windows malware

ECML '20: Poisoning attacks on algorithmic fairness

NeurIPS '21: Fast minimum-norm attacks

NeurIPS '22: Indicators of attack failure

WACV '23: Phantom Sponges

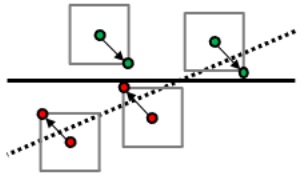
Robust Learning and Detection Mechanisms

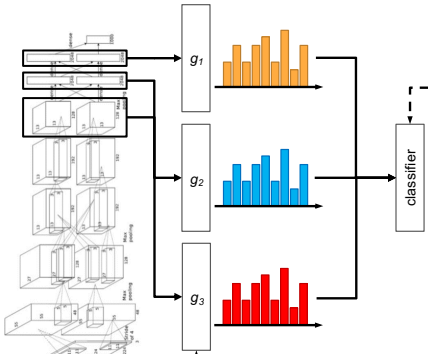
IEEE Symp. S&P '18: Robust learning against training data poisoning

IEEE TDSC '19: Optimal/robust linear SVM against adversarial attacks (use case on Android malware)

NEUCOM '21: Fast adversarial example rejection

IEEE TPAMI '21: Learning with domain knowledge to improve robustness against adversarial examples





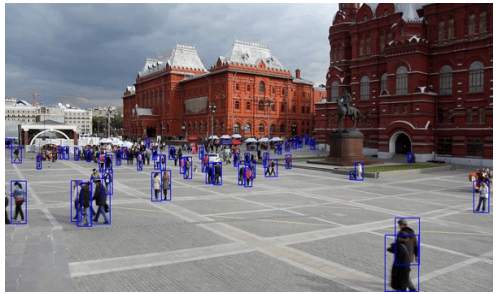
The PRALab Biometric Unit

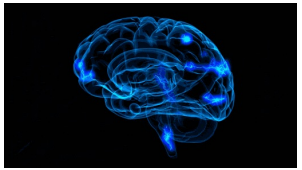
Basic issues

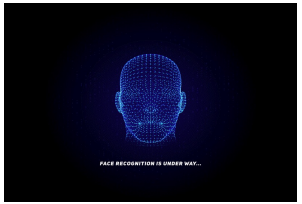
Feature extraction
Supervised learning
Adaptive learning
Deep learning
Decision fusion
Adversarial classification
Explainable AI

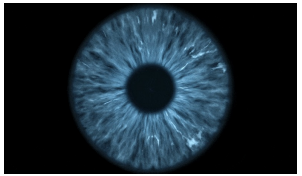
Practical Issues

Face recognition
Fingerprint recognition
Multimodal recognition
Adaptive biometric systems
Spoofing attacks
Deepfake detection
Crowd analysis









Biometrics Publication Highlights

Fingerprints

IEEE TIFS '21: Fingerprint recognition with embedded presentation attacks detection

PR '22: Towards realistic fingerprint presentation attacks

Handbook of Biometric Anti-Spoofing '23: Review of the Fingerprint Liveness Detection (LivDet) competition series

Deepfakes

ICIP '22: Tensor-Based Deepfake Detection In Scaled And Compressed Images.

ICIAP '22: Experimental Results on Multi-modal Deepfake Detection

Applied Sciences '22: Analysis of Score-Level Fusion Rules for Deepfake Detection

Other Biometrics

ICPR '21: Detecting anomalies from video-sequences

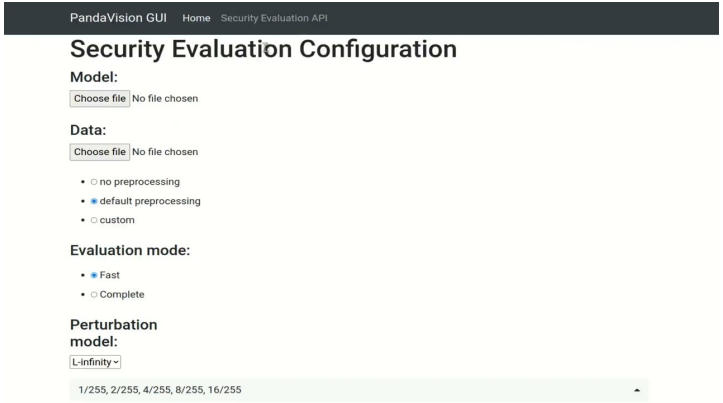
ICPR '22: 3D Face Reconstruction for Forensic Recognition

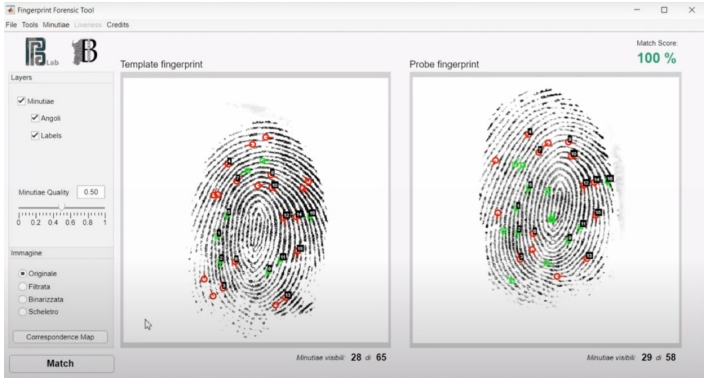
IET Biometrics '22: EEG personal recognition based on 'qualified majority'over signal patches

Practical applications and tools

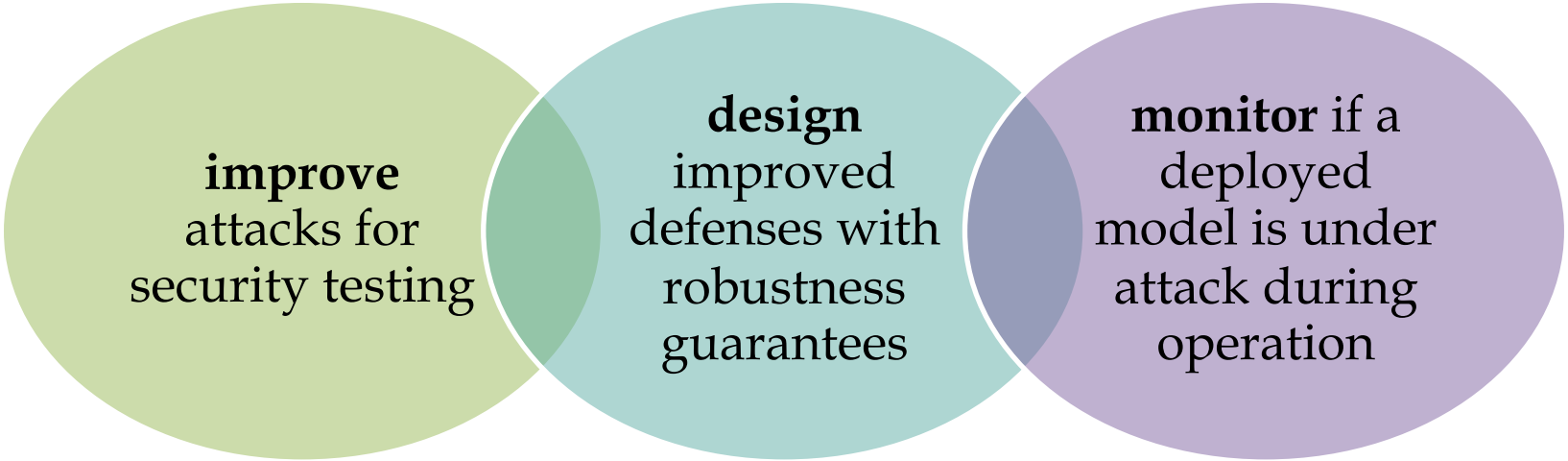
- MLSec
  - SecML: assess security evaluation of AI/MML technologies
  - SecML Malware: ad-hoc extension for security evaluation of malware classifiers

- Biometrics
  - Fingerprint Forensic tool
  - Deepfake detection tool

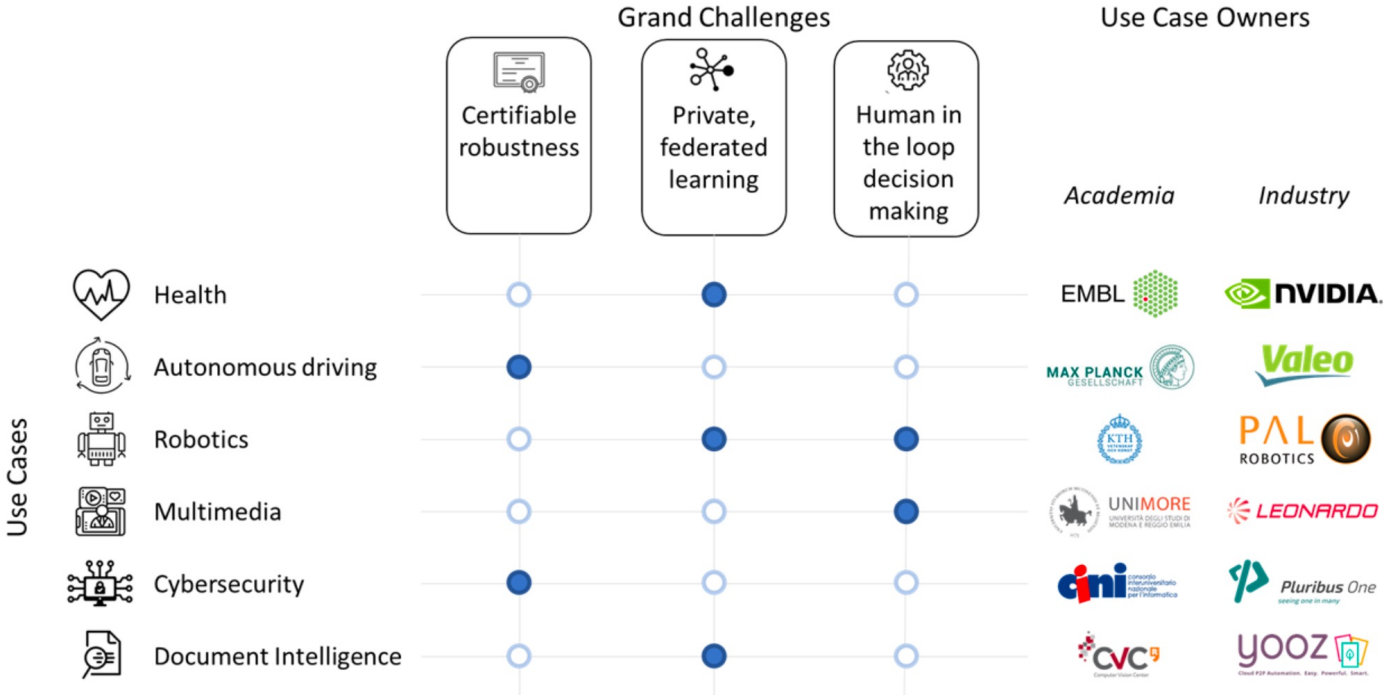




Challenges and Perspectives: towards ML*Sec*Ops



European Lighthouse on Secure and Safe AI (ELSA)



elsa
European Lighthouse on Secure and Safe AI

Useful links

Open Course on MLSec

https://github.com/unica-mlsec/mlsec

Software Tools

https://github.com/pralab

Machine Learning Security Seminars

https://www.youtube.com/c/MLSec

Pattern Recognition and Applications Lab

University of Cagliari, Italy

Università di Genova

Thanks!

Maura Pintor
maura.pintor@unica.it

Special thanks to Battista Biggio for sharing with me some of the material used in these slides.