PICTURE: Physical and Intrinsic Security of Embedded Neural Networks

Pierre-Alain Moellic

PICTURE

Physical and Intrinsic Security of Embedded Neural Networks

anr agence nationale de la recherche

Appel : AAPG (CE39)

Année : 2020

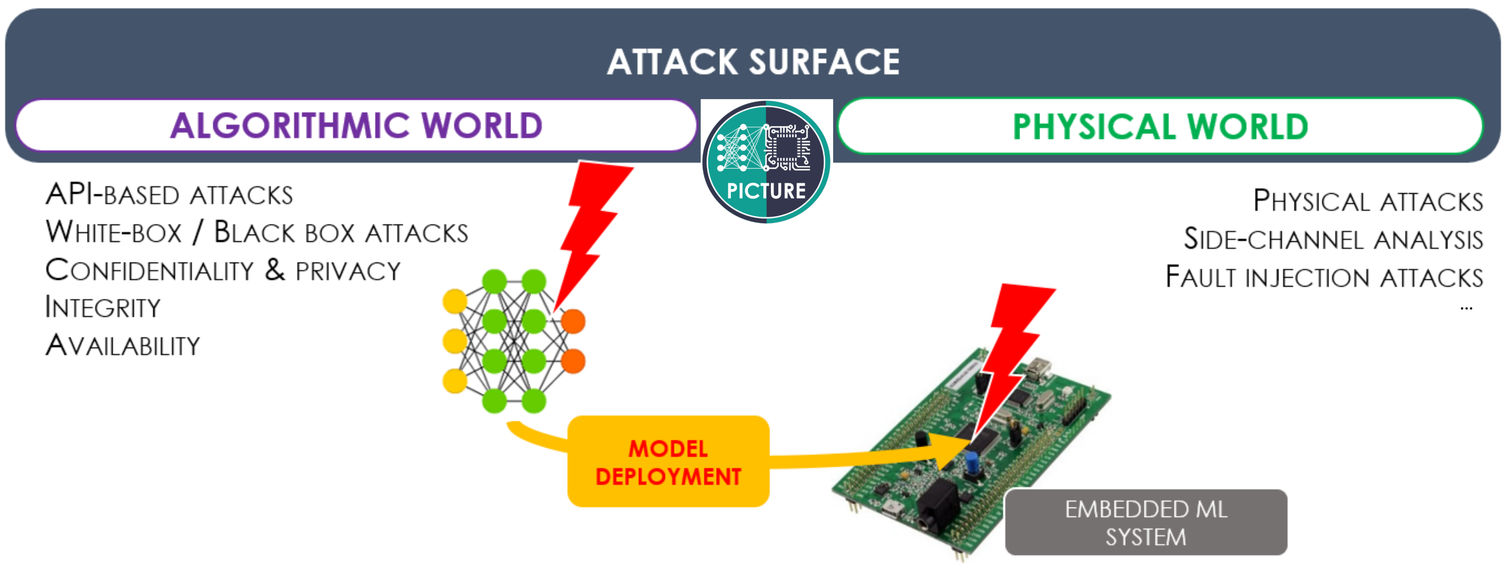
Instrument : PRCE

Contact : Pierre-Alain MOELLIC

COORDINATEUR : Pierre-Alain MOELLIC (CEA LETI)

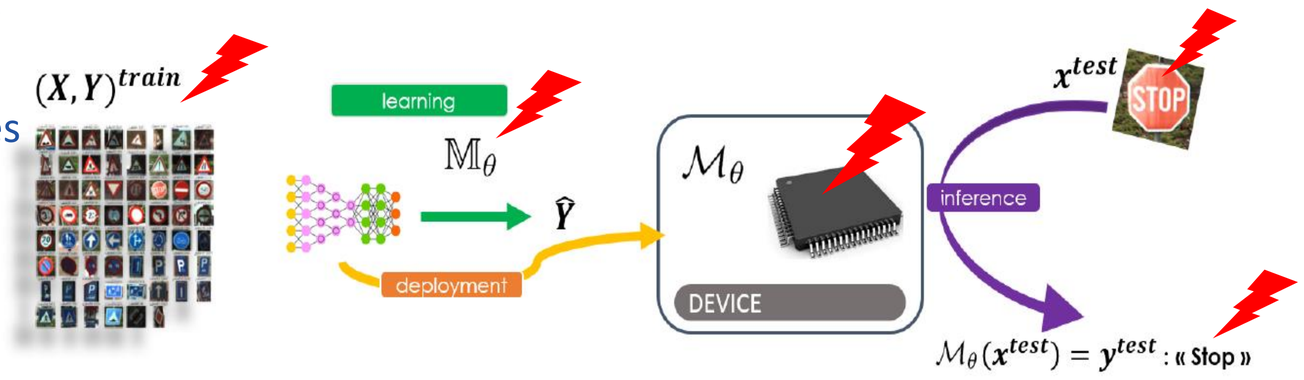
PARTENAIRES :
CEA-LETI, Mines Saint-Etienne, IDEMIA, STMICROELECTRONICS

PICTURE s'intéresse à la sécurité des réseaux de neurones embarqués (logiciels) en considérant une surface d'attaque regroupant les menaces algorithmiques (*adversarial examples, model extraction…*) et physiques (*side-channel*, *fault injection*) contre l'intégrité, la confidentialité et la disponibilité des modèles.



CONTEXTE ET OBJECTIFS

- Déploiement massif des modèles de Machine Learning
- Données / Tâches / Plateformes HW critiques
- Le *ML Pipeline* traditionnel est menacé à tous les étages
- Etat de l'art conséquent des attaques *algorithmiques: Adversarial Examples, Poisoning Attacks, Model Extraction…*



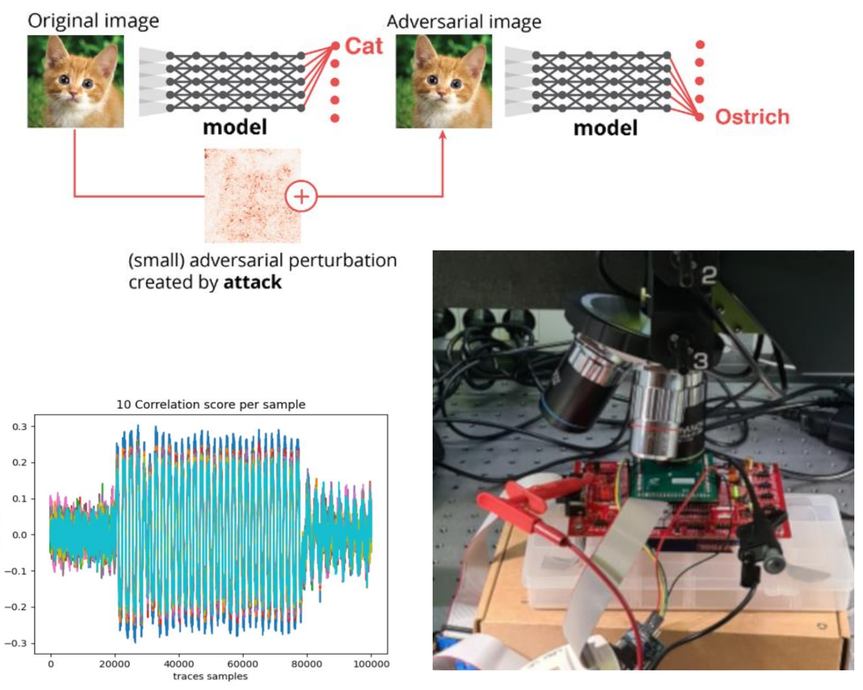
Nos objectifs

1. **Démontrer la criticité d'attaques combinant failles théoriques et physiques**
2. **Evaluer des contre-mesures physiques combinées à l'état de l'art des défenses algorithmiques et proposer des nouvelles défenses.**
3. **Disséminer des « bonnes pratiques » et suivre les actions de régulation/standardisation/certification.**

MÉTHODOLOGIE ET RÉSULTATS

Méthodologie

- Considérer une surface d'attaque globale et analyser conjointement failles physiques et algorithmiques:
  - ✓ *Fault injection analysis + adversarial perturbation*
  - ✓ *Side-channel analysis + model extraction*
- Analyser des modèles et plateformes réelles sur des cas d'usage critiques:
  - ✓ *Face Recognition systems*
  - ✓ *IoT*
- Méthodologie incrémentale WP Attaque ⟷ WP Défense + Evaluation



Dissémination

- PUBLIQUE "*State-of-the-art: Attacks & Threat Models*"
- 3 Publications 2021 (IEEE IJCNN, IEEE World Forum IoT)

https://picture-anr.cea.fr
@AnrPicture