Secure Systems Group, Aalto University & University of Waterloo

Sebastian Szyller

ML Security at SSG: Past, Present and Future

"Security is a process, not a product." -- Bruce Schneier

ML-based systems are wildly successful!

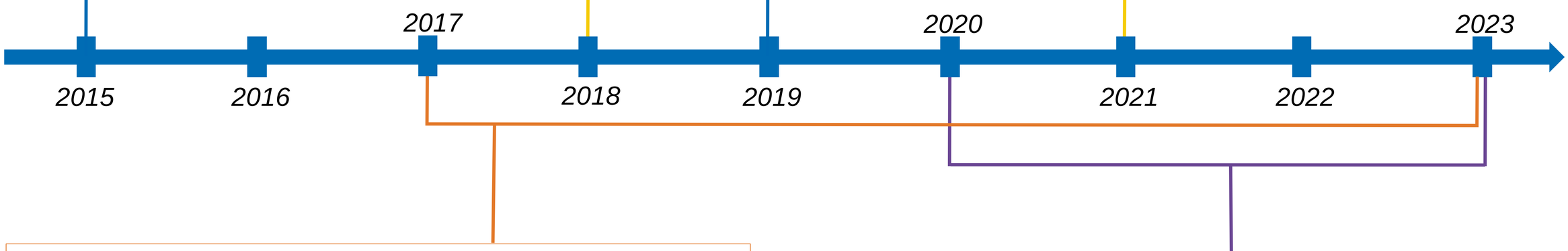
How to protect them against novel attacks?

ML for security:

- *Off-the-hook*: phishing website detection
- *DÏoT*: IoT anomaly detection
- *RecAgglo*: financial fraud detection

ML systems are brittle:

- *PRADA*: novel model extraction & detection
- *PRISM*: efficient evasion attacks
- *AD*^{3}: real-time RL evasion & detection
- NLP: evading hate-speech detection; ML-generated fake reviews



Real-world confidentiality:

- *MiniONN*: private inference
- *Boogeyman*: extraction with realistic adversary models
- *DAWN*: watermarking neural networks
- GANS: extracting image-translation models
- *Conflicts*: protecting against multiple threats

Effective ownership resolution:

- Dataset ownership schemes are not robust:
  - Dataset watermarking
  - Dataset Inference
- *False Claims*: inducing false positives against model ownership

Current work: defenses vs. other attacks, ML property attestation, …

Industry collaborations

Intel: *ICRI-SC*, *ICRI-CARS* & *PrivateAI* institutes

Zalando: fraud detection with real-world data

Output

- **5 doctoral dissertations**
- **16 Master's theses**
- **ACSAC, ACM CCS, ACM MM, AAAI, ESORICS, EURO S&P, ICDCS, SRDS, ...**

Training Experts

Industry: WithSecure, Intel, Nokia; **Academia**: Uni. Of Helsinki, Zhejiang Uni.; **Public sector**: KELA, CCC

Aalto University

https://ssg.aalto.fi/research/projects/mlsec/

UNIVERSITY OF WATERLOO