AI and Cybersecurity: Research in Maritime

CSP007

PRESENTATION BY: STYLIANOS KARAGIANNIS, (PDMFC, PORTUGAL)

Model Inversion and Perturbation Attacks

Class 1: Model Inversion

- **Querying model to deduce features:** Attackers probe the model to uncover the features it relies on.
- **Reconnaissance for future attacks:** Information gathered can be used to plan future attacks.

Class 2: Perturbation Attack

- **Manipulating inputs to provoke responses:** Attackers modify inputs to trigger specific responses from the model. Malicious inputs can lead to severe outcomes like collisions.
- **Requires modifying ML model inputs:** Attackers need to alter inputs to manipulate model behavior.
- **Accuracy influenced by original data quality:** The success of the attack depends on the quality of the original data.

Dataset Poisoning for AIS

- Tampering with AIS data involves altering the MMSI number, a vessel's unique identifier, by modifying its last two values, potentially causing misidentification.
- The attack also includes modifying the vessel's reported velocity using velocity standard deviations, which disrupts AI systems' decision-making and situational awareness.
- Attackers may further replace the vessel's GPS coordinates with those of other vessels, introducing confusion and misleading AI systems relying on accurate positioning information.
- The objective of deploying poisoned AIS data is to undermine the integrity of AI models in maritime operations, compromising their effectiveness during training and deceiving existing situational awareness systems.

RadarPWN

What is RadarPWN?

- Comprehensive, easy-to-use, and documented dataset
- Includes cyberattacks against radar communication

Data Collection Environment

- Utilizes BridgeCommand ship simulator
- Incorporates Radar Attack Tool
- Integrates OpenCPN chart plotter

Data Contents

- Navico BR24 and NMEA 0183 network traffic
- Benign and attack scenarios recorded

Available Forms of Dataset

- Contains network captures (pcaps)
- Includes configuration files for simulation runs and logs
- Adds screen recording videos of simulation runs
- Visualizes attacks in network captures

How RadarPWN can be used?

- **Training AI Models:** Provides labeled data for training AI, enabling detection and classification of cyberattacks on radar systems.
- **Testing and Validation:** Diverse attack scenarios aid in evaluating AI algorithms for cybersecurity, ensuring effectiveness and robustness.
- **Defense Mechanism Development:** Helps develop AI-driven defense systems by analyzing attack patterns and vulnerabilities.
- **Simulation and Scenario Generation:** Offers a simulation environment for creating and testing different attack strategies, enhancing AI adaptability.
- **Real-time Threat Detection:** Integration with onboard radar enables AI to analyze data, detect anomalies, and alert against cyber threats, bolstering maritime network security.

Thank you

Organization: PDMFC

Please send all questions to:
stylianos.karagiannis@pdmfc.com