IA e sicurezza informatica: Ricerca in ambito marittimo

CSP007

PRESENTAZIONE DI: STYLIANOS KARAGIANNIS, (PDMFC, PORTOGALLO)

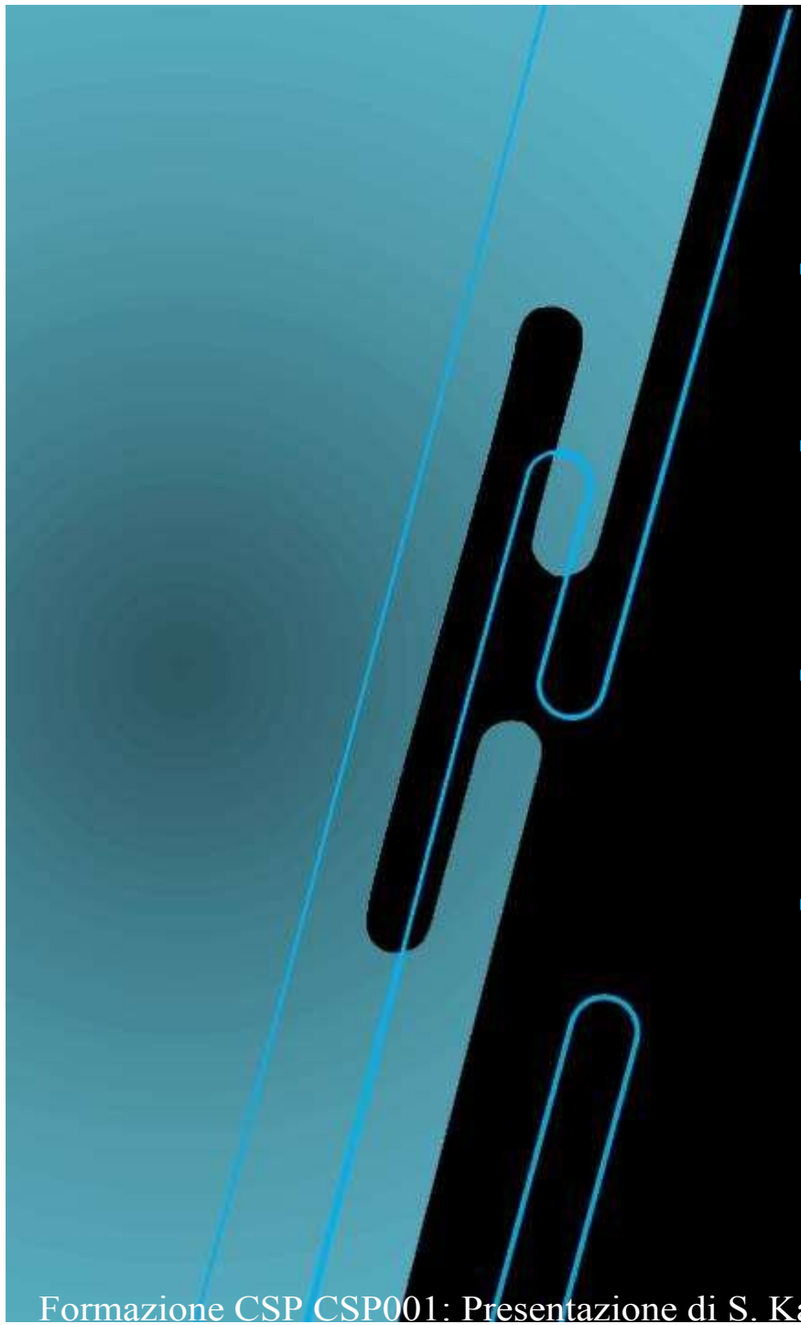
Inversione del modello e attacchi di perturbazione

Classe 1: Inversione del modello

- Interrogazione di modello per dedurre caratteristiche: Attaccant sondanoil modello per scoprire le caratteristiche cui si basa.
  - Ricognizione per attacchi futuri: Le informazioni raccolte possono essere utilizzate per pianificare attacchi futuri.

Classe 2: Attacco di perturbazione

- Manipolazione degli input per provocare risposte: Gli aggressori modificano gli input per provocare risposte specifiche da parte del modello. Gli input dannosi possono portare a risultati gravi come le collisioni.
- Richiede la modifica degli input del modello ML: Gli aggressori devono alterare gli input per manipolare il comportamento del modello.
- L'accuratezza è influenzata dalla qualità dei dati originali: Il successo dell'attacco dipende dalla qualità dei dati originali.



Set di dati Avvelenamento per AIS

- La manomissione dei dati AIS comporta l'alterazione del numero MMSI, l'identificativo univoco di un'imbarcazione, modificandone gli ultimi due valori e causando potenzialmente un'identificazione errata.
- L'attacco comprende anche la modifica della velocità dichiarata dell'imbarcazione utilizzando le deviazioni standard della velocità, che interrompe il processo decisionale e la consapevolezza della situazione dei sistemi di intelligenza artificiale.
- Gli aggressori possono inoltre sostituire le coordinate GPS dell'imbarcazione con quelle di altre imbarcazioni, creando confusione e ingannando i sistemi di intelligenza artificiale che si affidano a informazioni precise sul posizionamento.
- L'obiettivo dell'impiego di dati AIS avvelenati è quello di minare l'integrità dei modelli di intelligenza artificiale nelle operazioni marittime, compromettendo la loro efficacia durante l'addestramento e ingannando i sistemi di consapevolezza situazionale esistenti.

RadarPWN

Che cos'è il RadarPWN?

- Set di dati completo, facile da usare e documentato
- Include attacchi informatici contro le comunicazioni radar.

Ambiente di raccolta dati

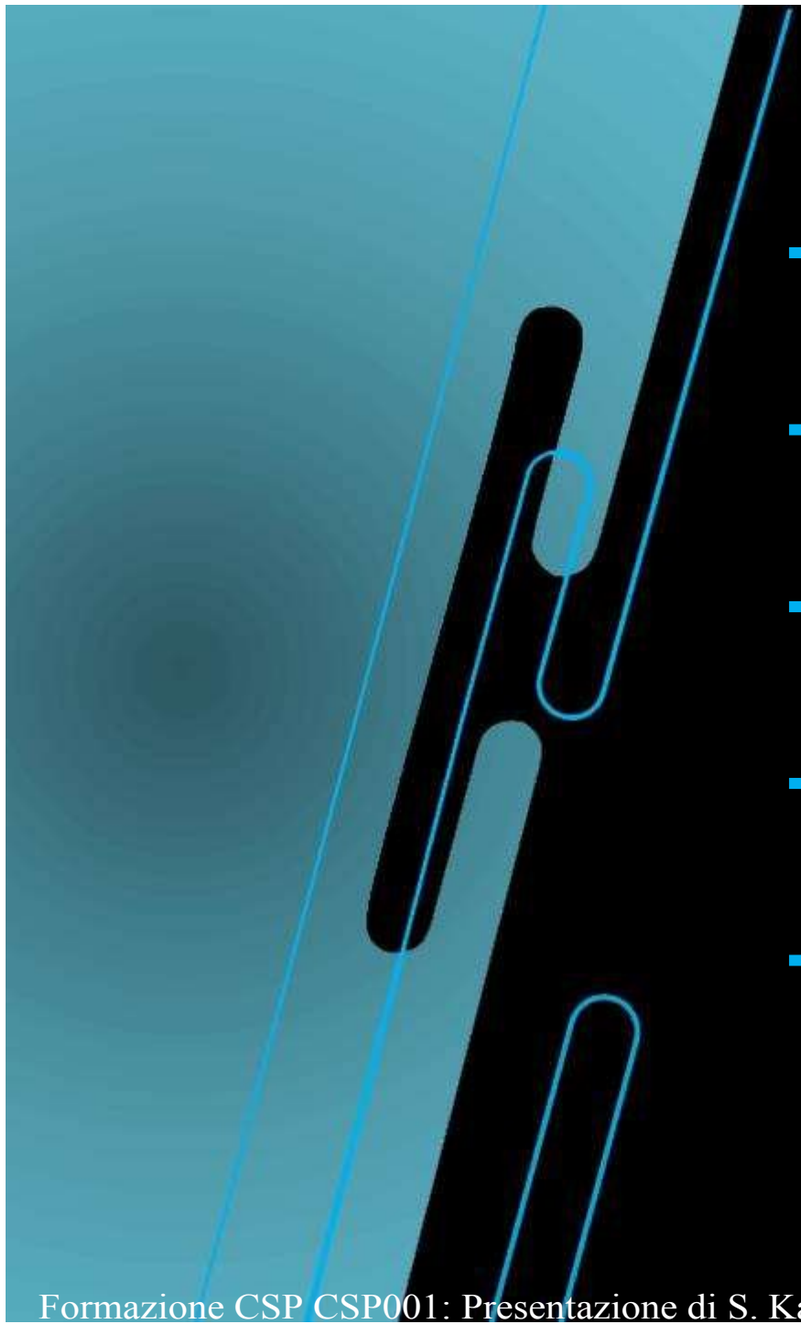
- Utilizza il simulatore navale BridgeCommand
- Incorpora lo strumento di attacco radar
- Integra il plotter cartografico OpenCPN

Contenuto dei dati

- Navico BR24 e traffico di rete NMEA 0183
- Scenari benigni e di attacco registrati

Forme disponibili del set di dati

- Contiene catture di rete (pcaps)
- Include i file di configurazione per le simulazioni e i log.
- Aggiunge video di registrazione delle schermate delle simulazioni
- Visualizza gli attacchi nelle acquisizioni di rete



Come può essere utilizzato RadarPWN?

- Formazione di modelli di intelligenza artificiale: Fornisce dati etichettati per l'addestramento dell'intelligenza artificiale, consentendo il rilevamento e la classificazione degli attacchi informatici ai sistemi radar.
- Test e convalida: Diversi scenari di attacco aiutano a valutare gli algoritmi di IA per la sicurezza informatica, garantendo efficacia e robustezza.
- Sviluppo di meccanismi di difesa: Aiuta a sviluppare sistemi di difesa basati sull'intelligenza artificiale analizzando i modelli di attacco e le vulnerabilità.
- Simulazione e generazione di scenari: Offre un ambiente di simulazione per creare e testare diverse strategie di attacco, migliorando l'adattabilità dell'IA.
- Rilevamento delle minacce in tempo reale: L'integrazione con il radar di bordo consente all'intelligenza artificiale di analizzare i dati, rilevare le anomalie e lanciare allarmi contro le minacce informatiche, rafforzando la sicurezza delle reti marittime.

Grazie

Organizzazione: PDMFC

Si prega di inviare tutte le domande a:
stylianos.karagiannis@pdmfc.com