
# Τεχνητή Νοημοσύνη και κυβερνοασφάλεια: Έρευνα στην Ναυτιλία

## CSP007

ΠΑΡΟΥΣΙΑΣΗ ΑΠΟ: ΣΤΥΛΙΑΝΟΣ ΚΑΡΑΓΙΑΝΝΗΣ, (PDMFC, ΠΟΡΤΟΓΑΛΙΑ)

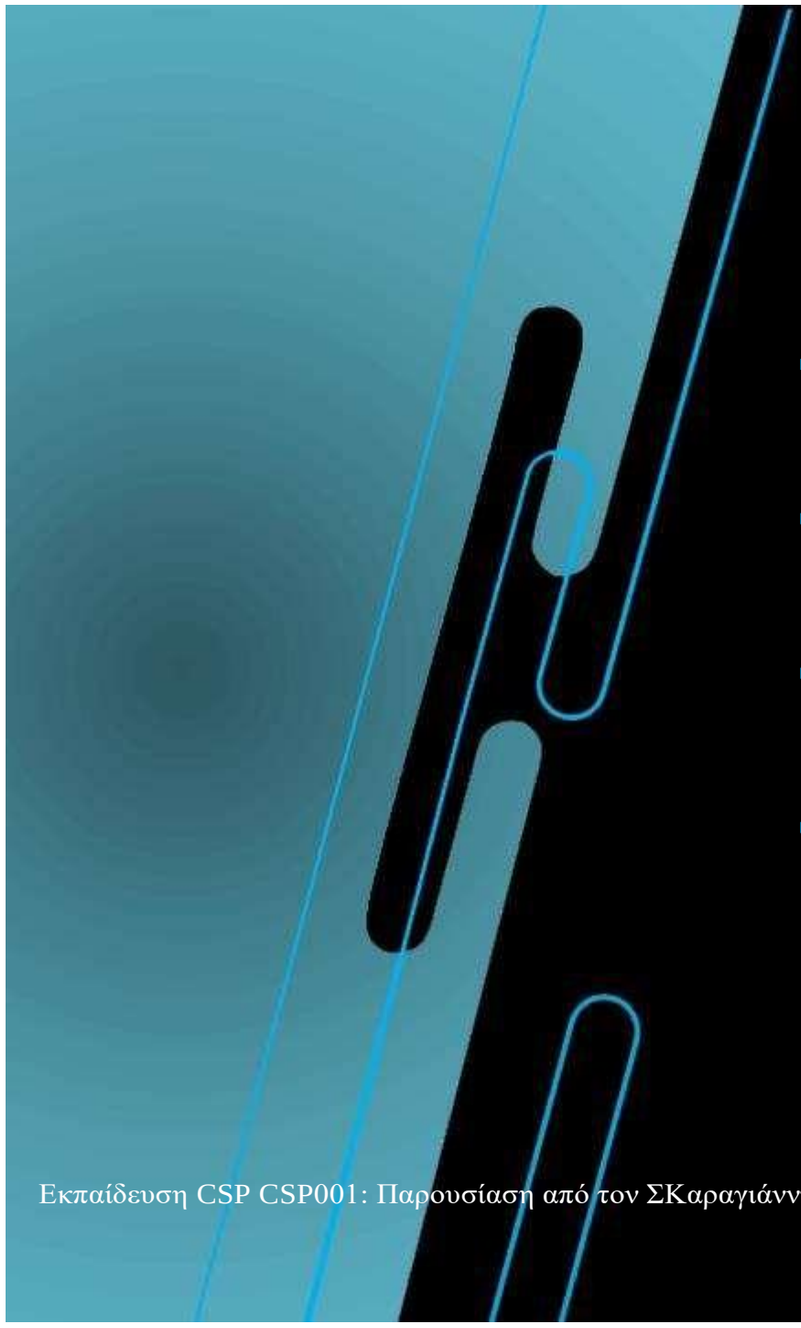
# Επιθέσεις Αντιστροφής Μοντέλου και Παρεμβολών

## Κατηγορία 1: Επίθεση Αντιστροφής Μοντέλου

- Ερώτηση του μοντέλου για εξαγωγή χαρακτηριστικών: Οι επιτιθέμενοι διερευνούν το μοντέλο για να αποκαλύψουν τα χαρακτηριστικά στα οποία βασίζεται.
- Αναγνώριση πληροφοριών για μελλοντικές επιθέσεις: Οι συλλεχθείσες πληροφορίες μπορούν να χρησιμοποιηθούν στον σχεδιασμό μελλοντικών επιθέσεων.

## Κατηγορία 2: Επίθεση Παρεμβολής

- Οι επιτιθέμενοι τροποποιούν τις εισόδους ώστε να ενεργοποιήσουν συγκεκριμένες αποκρίσεις από το μοντέλο. Κακόβουλες εισροές μπορεί να προκαλέσουν σοβαρά αποτελέσματα, όπως συγκρούσεις.
- Απαιτεί τροποποίηση των εισόδων του ML μοντέλου: Οι επιτιθέμενοι πρέπει να αλλάξουν τις εισόδους για να χειραγωγήσουν τη συμπεριφορά του μοντέλου.
- Η ακρίβεια επηρεάζεται από την ποιότητα των αρχικών δεδομένων: Η επιτυχία της επίθεσης εξαρτάται από την ποιότητα των δεδομένων στα οποία βασίζεται το μοντέλο.Η ακρίβεια επηρεάζεται από την ποιότητα των πρωτοτύπων δεδομένων: Η επιτυχία της επίθεσης προαπαιτεί την ποιότητα των πρωτοτύπων δεδομένων.



# Δηλητηρίαση συνόλου δεδομένων για συστήματα τεχνητής νοημοσύνης (AIS)

- Η αλλοίωση των δεδομένων AIS περιλαμβάνει την αλλοίωση του αριθμού MMSI, του μοναδικού αναγνωριστικού του σκάφους, τροποποιώντας τις δύο τελευταίες τιμές του, δυνητικά προκαλώντας εσφαλμένη ταυτοποίηση.
- Η επίθεση περιλαμβάνει επίσης την τροποποίηση της αναφερόμενης ταχύτητας του σκάφους με τη χρήση τυπικών αποκλίσεων ταχύτητας, η οποία ανατρέπει τη λήψη αποφάσεων και την επίγνωση της κατάστασης των συστημάτων Τεχνητής Νοημοσύνης.
- Οι επιτιθέμενοι μπορούν επίσης να αντικαταστήσουν τις συντεταγμένες GPS του σκάφους με εκείνες άλλων σκαφών, προκαλώντας σύγχυση και παραπλανώντας τα συστήματα Τεχνητής Νοημοσύνης που βασίζονται σε ακριβείς πληροφορίες εντοπισμού θέσης.
- Η διαδικασία εγκατάστασης και εκτέλεσης εφαρμογών σε περιβάλλον παραγωγής ολοκληρώνει την αποσύνδεση της ακεραιότητας των μοντέλων Τεχνητής Νοημοσύνης στις ναυτικές επιχειρήσεις, θέτοντας σε κίνδυνο την αποτελεσματικότητά τους κατά την εκπαίδευση και εξαπατώντας τα υπάρχοντα συστήματα επίγνωσης της κατάστασης.

# RadarPWN

## Τι είναι το RadarPWN;

- Πλήρες, εύχρηστο και τεκμηριωμένο σύνολο δεδομένων
- Περιλαμβάνει κυβερνοεπιθέσεις κατά της επικοινωνίας μέσω ραντάρ

## Περιβάλλον συλλογής δεδομένων

- Χρησιμοποιεί προσομοιωτή πλοίου BridgeCommand
- Ενσωματώνει εργαλείο επίθεσης ραντάρ
- Ενσωματώνει τον χαρτογράφο OpenCPN

## Περιεχόμενα δεδομένων

- Navico BR24 και κίνηση δικτύου NMEA 0183
- Καταγράφονται σενάρια κανονικής λειτουργίας και επιθέσεων

## Διαθέσιμες φόρμες συνόλου δεδομένων

- Περιέχει αρχεία καταγραφής δικτύου (PCAP)
- Περιλαμβάνει αρχεία διαρθρώσεων για εκτελέσεις προσομοίωσης και αρχεία καταγραφής
- Προσθέτει βίντεο καταγραφής οθόνης των εκτελέσεων προσομοίωσης
- Οπτικοποιεί τις επιθέσεις σε καταγραφές δικτύου

# Πώς μπορεί να χρησιμοποιηθεί το RadarPWN;

- Εκπαίδευση μοντέλων Τεχνητής Νοημοσύνη: Παρέχει επισημασμένα δεδομένα για την εκπαίδευση της Τεχνητής Νοημοσύνης, ενεργοποιώντας την ανίχνευση και ταξινόμηση κυβερνοεπιθέσεων σε συστήματα ραντάρ.
- Δοκιμές και επικύρωση: Ποικίλα σενάρια επιθέσεων βοηθούν στην αξιολόγηση αλγορίθμων Τεχνητής Νοημοσύνης για την ασφάλεια στον κυβερνοχώρο, εξασφαλίζοντας αποτελεσματικότητα και ευρωστία.
- Ανάπτυξη λογισμικού αμυντικού μηχανισμού: Βοηθά στον προγραμματισμό λογισμικού με βάση την Τεχνητή Νοημοσύνη, αναλύοντας μοτίβα επιθέσεων και Ευπάθειες.
- Προσομοίωση και δημιουργία σεναρίων: Προσφέρει ένα περιβάλλον προσομοίωσης για τη δημιουργία και τη δοκιμή διαφορετικών στρατηγικών επίθεσης, ενισχύοντας την προσαρμοστικότητα της Τεχνητής Νοημοσύνης.
- Ανίχνευση απειλών σε πραγματικό χρόνο: Η ολοκλήρωση με το ραντάρ επί του σκάφους ενεργοποιεί την Τεχνητή Νοημοσύνη να αναλύει δεδομένα, να ανιχνεύει ανωμαλίες και να προειδοποιεί για απειλές στον κυβερνοχώρο, ενισχύοντας την ασφάλεια των θαλάσσιων δικτύων.

# Σας ευχαριστώ

Οργανισμός: PDMFC

Παρακαλούμε στείλτε όλες τις ερωτήσεις στη διεύθυνση:
stylianos.karagiannis@pdmfc.com