Detection of spoofed AIS: Simulated tracks vs. real maritime data

Alexandru POHONTU[1], Constantin VERTAN[1], Iancu CIOCIOI[2], Ciprian POPA[1,2]

[1] National University of Science and Technology Politehnica Bucharest, Romania

[2] "Mircea cel Batran" Naval Academy, Constanta, Romania

alexandru.pohontu@stud.etti.upb.ro, constantin.vertan@upb.ro,

iancu.ciocioi@anmb.ro, ciprian.popa@anmb.ro

Abstract: The Automatic Identification System (AIS) is a valuable tool for enhancing maritime safety and security, primarily through its vessel tracking and collision avoidance functions. However, AIS is vulnerable to various cybersecurity threats, with simulated spoofed AIS tracks emerging as a significant concern. This paper analyses the stochastic kinematics of multiple vessels recorded in the Black Sea. Additionally, several Machine Learning models are evaluated for their effectiveness in distinguishing between genuine and spoofed maritime tracks. Accuracies exceeding 98% were obtained. The main concept of this study arises from the recognition that predicting future trajectories of maritime ships is susceptible to measurement and process errors. Measurement errors are primarily induced by inaccuracies in Global Navigation Satellite Systems, while process errors stem from factors such as weather conditions, wind, currents, and inconsistencies in vessel steering. Mathematical models typically generate spoofed tracks that lack the error variations observed in genuine data when predicting future positions, velocities, and headings. By analyzing and understanding these sources of error, this study demonstrates the potential to distinguish genuine maritime trajectories from simulated ones, thereby enhancing the detection of spoofed AIS tracks.

Keywords: maritime anomaly detection, AIS spoofing, simulated trajectories, maritime surveillance.

Detectarea datelor AIS falsificate: trasee simulate vs. date maritime reale

Rezumat: Sistemul de Identificare Automată (AIS) reprezintă un instrument valoros în menținerea siguranței și securității maritime, în principal datorită funcțiilor sale de urmărire a navelor și de evitare a coliziunilor. Totuși, AIS este vulnerabil la diverse amenințări cibernetice, iar traseele navelor falsificate și simulate prin canale AIS reprezintă o preocupare semnificativă. Această lucrare analizează cinematica stochastică a mai multor nave maritime înregistrate în Marea Neagră. În plus, sunt evaluate mai multe modele de învățare automată care diferențiază traseele maritime autentice de cele falsificate. Au fost obținute performanțe de peste 98%. Conceptul principal al acestui studiu pornește de la recunoașterea faptului că predicția traiectoriilor viitoare ale navelor maritime este susceptibilă la erori de măsurare și de proces. Erorile de măsurare sunt generate în principal de inexactități ale sistemelor globale de navigație prin satelit, în timp ce erorile de proces provin din factori precum condițiile meteorologice, vântul, curenții și inconsistențele în direcția de guvernare a navelor. Modelele matematice generează, de obicei, trasee falsificate care nu prezintă variațiile de erori observate în datele autentice atunci când se prezic pozițiile, vitezele și direcțiile viitoare. Prin analiza și înțelegerea acestor surse de eroare, acest studiu demonstrează potențialul de a distinge traiectoriile maritime autentice de cele simulate, îmbunătățind astfel detectarea traseelor AIS falsificate.

Cuvinte cheie: detectarea anomaliilor maritime, falsificarea AIS, trasee simulate, supraveghere maritimă.

1. Introduction

The maritime transportation sector plays a vital role in global trade, representing the most cost-efficient way of transporting goods over large distances. Consequently, the increasing importance of this domain has made it a scenario for numerous illicit activities, ranging from illegal migration and piracy to illegal fishing, smuggling, and environmental pollution. In response, numerous maritime agencies have emerged to detect and deter these threats. Central to their operations is the collection of various maritime-related data from different sources (e.g., coastal

https://doi.org/10.33436/v35i1y202503

maritime radars, EO/IR cameras, satellite imagery), with one of the most prominent being the Automatic Identification System (AIS) (Newaliya & Singh, 2021).

According to international regulations, vessels engaged in international voyages weighing over 300 gross tons (GT), cargo vessels over 500 GT, and all passenger vessels must be equipped with AIS transponders. EU fishing vessels over 15 meters must also comply (Zheng et al., 2023). AIS was initially designed as a communication system to enhance navigation safety by enabling data exchange between ships and shore stations. Today, it serves broader maritime safety and security functions, including vessel tracking, fishing monitoring, maritime risk analysis, accident investigations, waterway planning and management, and research (Androjna et al., 2024).

AIS operates as a cooperative radio network where vessels voluntarily transmit their identification details, navigation, and positional data at regular intervals based on their speed. Subsequently, this information is captured by other vessels, as well as coastal or satellite receiving stations (Kontopoulos et al., 2018). The International Telecommunication Union (ITU) recommendation describes the system and employs the Time Division Multiple Access (TDMA) protocol. Also, it utilizes two dedicated wavelengths in the very-high frequency (VHF) band, primarily via terrestrial networks: 161.975MHz and 162.025MHz. Terrestrial AIS stations or onboard transceivers typically cover 15-20 nautical miles (1 Nm = 1.852 Km), limited by geodetic visibility and affected by location, transceiver altitude, type, and weather conditions. Satellite AIS (S-AIS) extends the communication range through low Earth orbit for enhanced data exchange over large distances (Androjna et al., 2021).

Although AIS has proven to be a valuable tool for maritime navigation, it was not primarily designed with security as a priority. Consequently, it is susceptible to malicious attacks that manipulate or disrupt AIS data and can potentially misinterpret vessel activities that might engage in illicit actions (Androjna & Perkovič, 2024). The vulnerability of AIS lies in its plaintext, unencrypted, and unauthenticated nature, making it prone to spoofing and manipulation by malicious actors. Thus, the absence of encryption and authentication makes AIS vulnerable to cyber-attacks, jeopardizing the integrity and reliability of the system (Coleman, Kandah & Huber, 2020). Multiple cyber-attacks threaten its operations. AIS jamming involves transmitting signals to disrupt the system's communication channel, while Global Positioning System (GPS) jamming targets the navigation systems of vessels. Attacks such as AIS spoofing and AIS impersonation involve the creation of false vessel tracks or altering a genuine vessel's identity. Additionally, there are situations where vessels turn off their transponders to hide their live position, thus allowing them to operate incognito ("*AIS dark activities"*) and potentially engage in illicit scenarios (Androjna & Perkovič, 2021).

Multiple researchers attempt to address the problem of AIS spoofing, often referring to different aspects of the issue. Some define AIS spoofing as the act of AIS impersonation, where real vessels change their identifiers (e.g., MMSI, ship class) to conceal their activities. Others consider AIS spoofing as the act of creating fictitious vessels to generate confusion at sea and mislead maritime authorities (Kontopoulos et al., 2018). It is important to note that a clear and universally accepted definition of AIS spoofing has not yet been adopted, leading to some confusion in the literature. This paper addresses the problem of AIS spoofing, where the challenge lies in distinguishing between simulated vessels and genuine ones. Such situations have been documented in real-world scenarios, such as Elba Island, the Black Sea, the Strait of Hormuz, and the Port of Shanghai (Androjna et al., 2021). It is also noted in electronic warfare contexts where specialized equipment, also known as W-AIS, is utilized.

The rest of this paper is structured as follows: Section 2 discusses the measures proposed for securing the AIS system, highlighting cryptographic and authentication-based solutions. Section 3 reviews relevant work in the field, analyzing existing indirect methods for detecting spoofed AIS data. Section 4 details the research methodology, including data collection, preprocessing steps, and the machine learning models that were implemented. The obtained results are presented and analyzed in Section 5, while Section 6 provides the conclusions and outlines future research directions.

2. Measures to secure AIS

The lack of authentication poses a significant vulnerability to the AIS system. Its inability to validate data sources opens the door to potential AIS spoofing by malicious actors, which can disrupt vessel navigation and compromise maritime security. In response to these challenges, numerous solutions have been proposed, many of which adopt asymmetric cryptography to digitally sign AIS messages. Cryptographic authentication, leveraging Public Key Cryptography (PKC), ensures message authenticity and verifies that it has been digitally signed by its sender. One backwards-compatible approach for securing AIS messages involves attaching a digital signature in a separate, follow-on message. This method allows messages to be broadcast in unencrypted format while enabling receivers to verify authenticity through a digital signature derived from the message's hash code (Wimpenny et al., 2022). Similarly, a proof-of-concept for Secure AIS has been reported, presenting a backwards-compatible solution that enables source authentication, encryption, and legitimate pseudo-anonymization functions (Goudosis & Katsikas, 2022).

Although several secure AIS variants have been proposed, their widespread adoption is limited by the need for international consensus on the distribution of public encryption keys (Coleman, Kandah & Huber, 2020). Even if a new secure AIS system were agreed worldwide, AIS spoofing would likely remain a threat for decades until all vessels - more than 100,000 globally - adopted the technology.

3. Related work

Until a secure version of AIS is adopted, various indirect solutions can be implemented to enable legitimate vessels to broadcast their identifiers and kinematic details. Advanced systems employing probabilistic models and Machine Learning (ML) techniques can inform and enhance expert analysis. Multiple solutions for AIS spoofing detection have been proposed, most involving the analysis of vessels' attributes and their associated kinematic data.

One relevant method addresses the challenge of determining the reliability of AIS data by leveraging radar measurements and tracking system information. This approach involves a novel application of the sequential log-likelihood ratio test in a generalized form. The method aims to identify scenarios in which real vessels initially transmit accurate AIS data but, at a certain point, begin transmitting falsified coordinates to enter restricted areas undetected. The probability of detecting such deception increases near 100% as the discrepancy between AIS-reported tracks and radar tracks grows, and as the number of radar systems used increases. (Katsilieris, Braca & Coraluppi, 2013). Although this method demonstrates high accuracy in identifying falsified AIS signals, it requires the use of expensive radar equipment.

Another methodology for analyzing AIS data emphasizes the integrity and signal strength of information as a crucial element. The authors propose different approaches for identifying AIS false tracks, such as controlling the integrity of each field in each AIS message individually or measuring the received power level. The latter approach allows for estimating the broadcast power based on the known distance (Ray, Iphar & Napoli, 2016). While accurate, the second solution requires specialized equipment for signal analysis and goniometry to detect whether the reported location of the vessel corresponds to the position where the signal was broadcast.

A recent approach to detecting AIS spoofing utilizes trajectory segmentation and the Isolation Forest algorithm to differentiate between legitimate and falsified AIS signals. This method segments ship trajectories based on time interval thresholds and extracts features such as speed and distance between consecutive points to distinguish between regular and spoofing ships. The Isolation Forest algorithm is then applied to identify anomalies in the trajectory data, leveraging the fact that spoofing ships exhibit distinct motion patterns compared to legitimate vessels. Experimental results indicate that this approach achieves identification accuracies of 88.4% to 93.3%, depending on the chosen segmentation time interval (Zheng et al., 2023).

Another approach for AIS spoofing detection integrates maritime vessel route extraction with clustering-based anomaly detection. The method preprocesses AIS data to remove noise and irrelevant information before applying a density-based clustering algorithm, DBSCAN, to detect spoofed transmissions. By analyzing vessel trajectories, waypoints, and deviations from expected maritime routes, the system identifies three main types of spoofing: dark period spoofing (AIS transmission shutdown), position-based spoofing (false coordinate reports), and MMSI-based spoofing (identity fraud). The technique effectively isolates anomalies through speed analysis, trajectory clustering, and geospatial filtering, achieving significant data reduction while preserving critical trajectory information (Prasad, Vatsal & Chowdhury, 2021).

4. Research methodology

This paper proposes a different approach for evaluating the accuracy of maritime trajectory predictions and identifying potential navigation deviations. Its objective is to distinguish between genuine and spoofed kinematic patterns of navigation. This is done by analyzing the stochastic kinematic errors between consecutive vessel positions and by implementing various ML models. This approach is based on the premise that spoofed tracks exhibit limited deviations due to their computer-generated precision. On the other hand, genuine vessels exhibit measurement errors, such as those introduced by GPS inaccuracies, and process errors, which may arise from environmental influences like wind, water currents, or minor variations in steering.

4.1. Data collection and pre-processing

The first phase of this study involved collecting seven days of live tracking data from multiple AIS transponders deployed near the Romanian shore. Additionally, live AIS feeds were obtained from other regions through various Maritime Situational Awareness (MSA) web platforms. For instance, MarineTraffic and VesselFinder provide on-demand raw AIS-NMEA data by configuring multiple Application Programming Interfaces (APIs), while AIS Hub serves as a free online collaborative AIS-NMEA data-sharing service. Furthermore, different databases offer raw or processed AIS data. For example, the AIS Exploratorium platform (available at http://ais.exploratorium.edu/) contains over 1.2 million records, whereas a processed AIS dataset is accessible on Kaggle (available at https://www.kaggle.com/datasets/aswinjose/ais-maritime-data).

After decoding the raw NMEA-0183 messages, the sequential locations and timestamps of the detected vessels were extracted. Additionally, specialized software was used to generate multiple simulated tracks ("bogus vessels"), each with predefined velocities, waypoints, and headings. A similar non-commercial tool for generating AIS virtual tracks is the NMEA Simulator (available at https://github.com/panaaj/nmeasimulator/releases).

The pre-processing step involved the calculation of various parameters for each *k* sequential position of recorded trajectories. This was done by analyzing pairs of sequential latitude and longitude coordinates (ϕ, λ) of the vessels: $k-1$ (previous recorded position), *k* (current position), and $k+1$ (upcoming recorded position) (see Figure 1). To understand the vessels' movement dynamics, the analysis was conducted multiple times at different minimal Δt periods between two consecutive recordings. These intervals included 1, 30, 60 seconds, 30 minutes, and 1-hour setups.

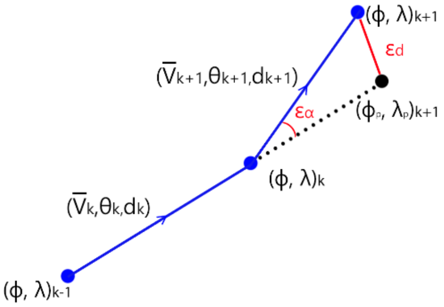

Figure 1. Representation of Sequential Vessel Coordinates

Different parameters were calculated in each *k*-sampling interval, including the magnitudes of velocities $\left|\vec{V}\right|$, travelled distances *d* (1), and bearing angles Θ (2) between consecutive locations.

d and *Θ* were determined as the distances and bearings between two consecutive points on the sphere's surface. Based on current positions, speeds and bearings, predicted upcoming coordinates $\left(\varphi_p, \lambda_p\right)_{k+1}$ were calculated to evaluate the errors for distance ε_d (3) and bearing ε_α (4) between the estimated and actual future coordinates.

$$d = 2r \cdot \arcsin\sqrt{\sin^2\left(\frac{\varphi_{k-1}-\varphi_k}{2}\right) + \cos(\varphi_{k-1})\cdot\cos(\varphi_k)\cdot\sin^2\left(\frac{\lambda_{k-1}-\lambda_k}{2}\right)} \tag{1}$$

$$\theta = \arctan(\sin(\lambda_{k-1}-\lambda_k)\cdot\cos(\varphi_k)\cdot\cos(\varphi_{k-1})\cdot\sin(\varphi_k) - \sin(\varphi_{k-1})\cdot\cos(\varphi_k)\cdot\cos(\lambda_{k-1}-\lambda_k)) \tag{2}$$

$$\varepsilon_d = d((\varphi,\lambda)_{k+1}, (\varphi_p,\lambda_p)_{k+1}) \tag{3}$$

$$\varepsilon_\alpha = \theta((\varphi,\lambda)_{k-1}, (\varphi,\lambda)_k) - \theta((\varphi,\lambda)_k, (\varphi,\lambda)_{k+1}) \tag{4}$$

where (φ,λ) represent latitude and longitude coordinates expressed in radians and *r* is the radius of the Earth.

Furthermore, a Kalman Filter was implemented on the available datasets. This filtering technique enabled the estimate of the trajectory of each vessel with greater precision, by accounting for various uncertainties (Zhang, Xining, Zhang, Xynian & Li, 2023). By extracting and analyzing all Kalman prediction errors ε_{dk} (5), the extent to which the vessels' trajectory measurements were affected by external factors was obtained. These factors included measurement (GPS inaccuracy) and process (e.g., wind, currents, steering) errors.

$$\varepsilon_{dk} = z_k - H x_k^{-} \tag{5}$$

where z_k is the actual measurement vector at time *k* and $H x_k^{-}$ is the predicted measurement vector from the Kalman filter, based on the predicted state x_k^{-} before updating with the new measurement z_k (see Figure 2).

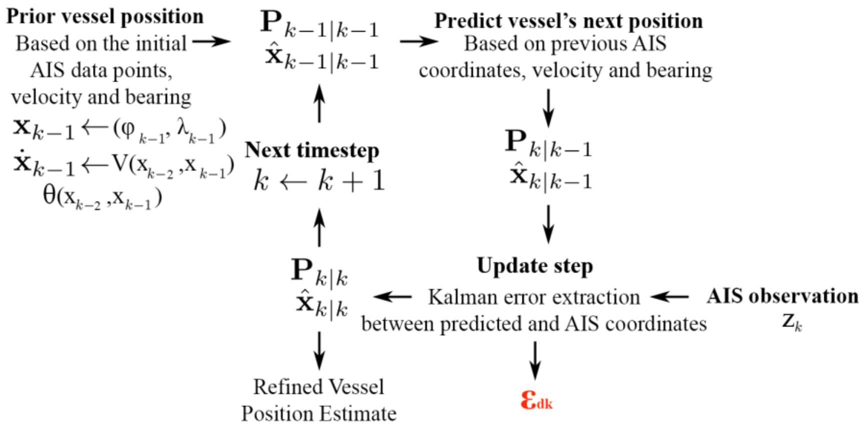
Prior vessel position
Based on the initial
AIS data points,
velocity and bearing
$\mathbf{x}_{k-1} \leftarrow (\varphi_{k-1}, \lambda_{k-1})$
$\dot{\mathbf{x}}_{k-1} \leftarrow V(x_{k-2}, x_{k-1})$
$\theta(x_{k-2}, x_{k-1})$
$\mathbf{P}_{k-1|k-1}$
$\hat{\mathbf{x}}_{k-1|k-1}$
Predict vessel's next position
Based on previous AIS
coordinates, velocity and bearing
Next timestep
$k \leftarrow k+1$
$\mathbf{P}_{k|k-1}$
$\hat{\mathbf{x}}_{k|k-1}$
$\mathbf{P}_{k|k}$
$\hat{\mathbf{x}}_{k|k}$
Update step
Kalman error extraction
between predicted and AIS coordinates
AIS observation
z_k
Refined Vessel
Position Estimate
ε_{dk}

Figure 2. Process for extracting sequential Kalman errors

4.2. Data analysis

The next step of the study involved calculating the mean values μ (6) and standard deviations σ (7) for the previously calculated parameters ($|\vec{V}|$, ε_d, ε_α and ε_{dk}) at different time interval setups t_i, in all vessel trajectories. Additionally, for the $t_i = 1\text{sec}$ setup, μ and σ values were also calculated for the $\Delta t_{i,i+1}$ periods between two sequential reporting times of vessels within the AIS. It is important to note that the reporting periods can differ significantly for vessels, depending on their velocity. For instance, vessels underway report every 2 up to 10 seconds based on their speed.

Vessels at anchor may report every 3 minutes. In real scenarios, these reporting intervals are not constant due to various factors such as signal loss, signal fading, and other environmental conditions, which can lead to even longer gaps between reports. Additionally, the radio channel's rate of occupation plays a crucial role; heavy AIS traffic in a particular area can lead to increased congestion and competition for available slots. This congestion may further delay transmissions, impacting the frequency and reliability of updates.

$$\mu_t = \frac{\sum x}{n} \tag{6}$$

$$\sigma_t = \sqrt{\frac{1}{n}\sum_x (\mathrm{x} - \mu_t)^2} \tag{7}$$

where *x* represents the values of each calculated parameter in vessel trajectories, *n* is the total number of recordings, and *t* is the time interval setup.

Subsequently, histograms and probability density functions (PDF) were obtained by applying a Kernel Density Estimation (KDE) application. The PDF *f(x)* (8) was obtained by placing a Gaussian kernel function ϕ (9) on each observation x_i of the training set of size *n*.

$$f(\mathrm{x}) = \frac{1}{nh}\sum_{i=1}^{n}\phi\left(\frac{x - x_i}{h}\right) \tag{8}$$

$$\phi\left(\frac{x - x_i}{h}\right) = \frac{1}{h\sqrt{2\pi}}e^{-\sqrt{\left(\frac{x-x_i}{h}\right)^2}} \tag{9}$$

Also, a correlation matrix was computed to assess the correlation ρ (10) between the mean $\left|\vec{V}\right|$, ε_d, ε_α and ε_{dk} parameters.

$$\rho_{X_i,\mathrm{X}_j} == corr(X_i,\mathrm{X}_j) = \frac{\mathrm{cov}(X_i,\mathrm{X}_j)}{\sigma_{X_i}\sigma_{Xj}} = \frac{E[(X_i - \mu_{X_i})(X_j - \mu_{Xj})]}{\sigma_{X_i}\sigma_{Xj}} \tag{10}$$

where $(\mathrm{X}_i, \mathrm{X}_j)$ represent the pair of variables on which to calculate the correlation.

4.3. Machine learning implementations

After analyzing the distributions of the $\left|\vec{V}\right|$, ε_d, ε_α and ε_{dk} parameters for all recorded trajectories, it was observed that genuine vessels exhibit distinct statistical patterns. These patterns enable the identification of spoofed vessels, which often show limited variations in velocities and predictable future positions due to nearly zero recorded errors between consecutive locations.

Building on this insight, more complex and performant machine learning (ML) models were explored to enhance detection accuracy. Mean values and standard deviations for parameters such as time intervals between AIS reports, velocities, distance and bearing prediction errors between consecutive locations, and Kalman filter distance errors were computed. These statistical features were then used as inputs for multiple ML models, classifying vessels as either genuine or spoofed based on these metrics.

After evaluating each implemented ML model, a permutation importance index I_{Xi} (11) was calculated for each input parameter to assess its individual contribution to model accuracy. This index was computed by shuffling the values of each parameter independently and measuring the resulting drop in model performance. For example, the permutation importance of feature X_i was determined by the change in accuracy ΔA when X_i was randomly permuted:

$$I_{X_i} = A_{baseline} - A_{shuffled\,(\mathrm{X}_i)} \tag{11}$$

5. Results

5.1. General aspects

On average, approximately 1200 distinct vessel trajectories were discovered daily from AIS recordings. These included both stationary ships and those actively transiting the region, as well as various Aids to Navigation (ATONs).

The identification of stationary spoofed vessels became evident after plotting the trajectories of all recorded platforms and calculating their mean speed, bearing, and the estimated errors between consecutive locations. The ATON tracks were used for this analysis. By definition, these tracks represent spoofed data, denoting different stationary platforms such as buoys, beacons, or lighthouses. Typically, these platforms are reported at locations apart from their actual physical positions by remote AIS stations, which transmit manually entered coordinates that remain unchanged over time. Figure 3 shows the recorded coordinates of two open-sea, non-propelled platforms (a drifting vessel and a vessel at anchor) that exhibit limited motion due to environmental factors like wind and currents.

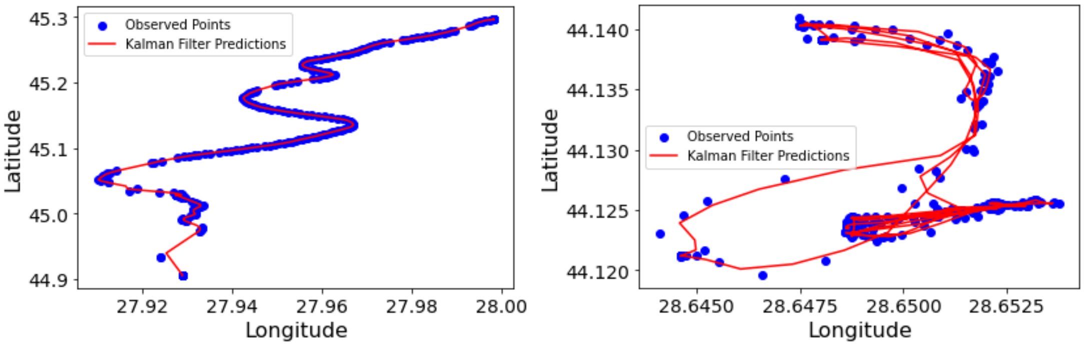

Figure 3. Drifting Vessel (Left) and Anchored Vessel (Right)

Figure 4 shows the recorded coordinates of two stationary platforms (a moored vessel in a harbor and an ATON). Unlike the ATONs, all stationary vessels exhibited position variations due to GPS inaccuracies and drifting.

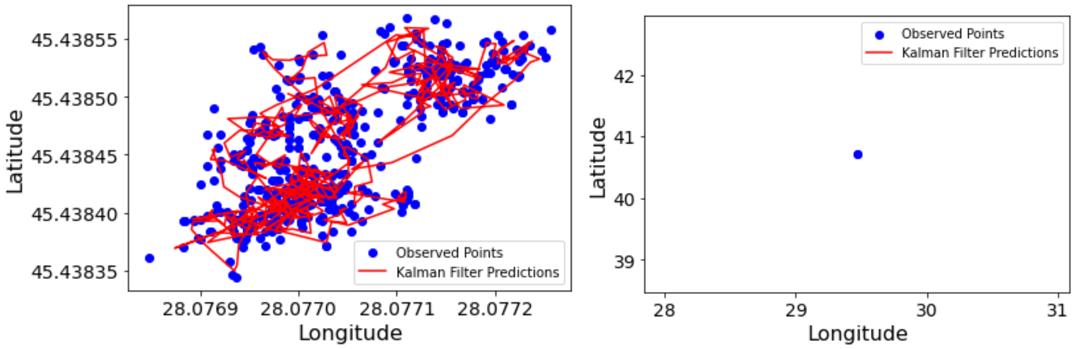

Figure 4. Moored Vessel (Left) and Stationary ATON (Right)

Upon further investigation of the recorded data, it became evident that distinguishing between genuine and spoofed moving vessels is feasible. One method of differentiation involves analyzing the vessels' speed over ground (SOG) and course over ground (COG) variations, particularly with smaller time interval setups.

Figure 5 illustrates the evolution of speed and COG over 15 minutes for a genuine vessel endeavoring to maintain a speed of 11 knots (1 kt = 1 Nm/hour) and a COG of 243° relative to geographic North. The vessel recorder showed continuous variations in velocities (up to 2 kt) and courses (up to 2.5°) during this period.

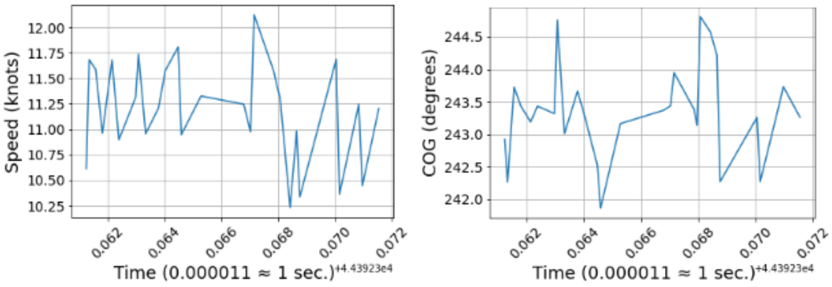

Figure 5. Speed and COG variations for a genuine vessel

Figure 6 presents the evolution over 15 minutes for a non-maneuvering spoofed vessel simulated to maintain a constant velocity of 8.05 kt and a constant COG of 140.9°. However, very small, insignificant errors may have been induced by the chosen projection for geographic coordinates, limited decimal precision in calculations, and the limited resolution of the AIS positional messages.

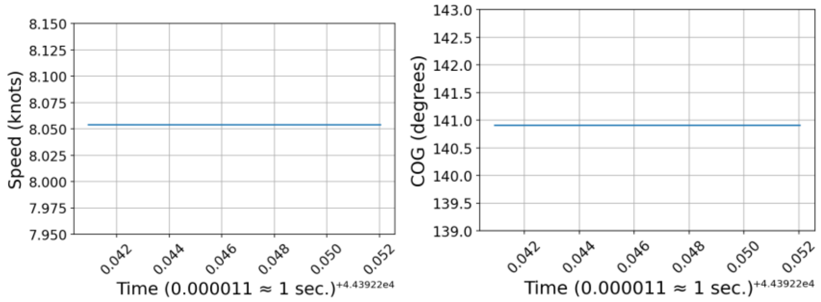

Figure 6. Speed and COG variations for a non-maneuvering spoofed vessel

Figure 7 presents the evolution of a maneuvering spoofed vessel over a 12-minute period. This vessel was configured to change its speed and course over ground (COG) in a manner which is similar to real spoofed vessels recorded in the Black Sea in 2021 (Thomas & Chiego, 2022). Initially, the vessel maintained a perfectly constant speed of 20 kt and a steady COG of 0°. A notable deviation occurred as the vessel underwent sudden, multiple changes in speed, reaching an unrealistic maximum value of 800 kt (≈ 1500 km/h). Additionally, the changes in COG were executed abruptly, without intermediate values, suggesting significant shifts in COG at unrealistic angular speeds.

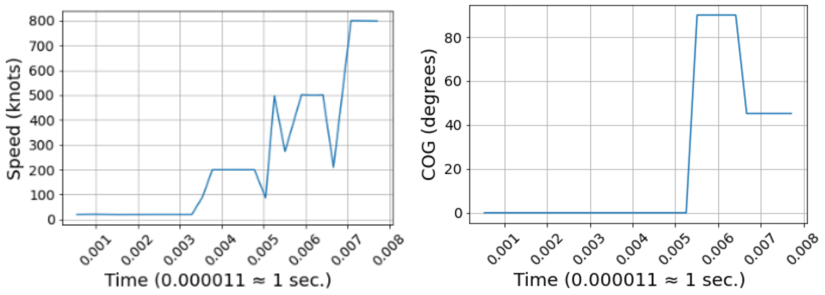

Figure 7. Speed and COG variations for a maneuvering spoofed vessel

5.2. Statistical results

Significant differences were observed between genuine and spoofed vessels when analyzing the mean discrepancies between predicted and travelled distances for consecutive locations. At a time difference of 60 seconds, genuine vessels recorded a mean distance estimation error $\mu_{\varepsilon d} = -0.19$ Nm, while spoofed vessels recorded a mean value almost equal to 0 ($\mu_{\varepsilon d} \approx 10^{-6}$). It's worth noting that this value was not 0 for spoofed vessels due to estimation errors induced by the resolution of the AIS messages for positional data (*res = 0.0001'*).

Table 1 presents the mean values of the averages and standard deviations that were calculated for $\left|\vec{V}\right|$, ε_d and ε_α for all genuine and spoofed vessels with mean speeds above 1 kt, at different time setups. It was observed for genuine vessels that the variation in distance and bearing errors ($\mu_{\varepsilon d}$ and σ_α) tends to decrease with increasing time intervals between two consecutive locations (1 second, 30 seconds, 60 seconds, 30 minutes). However, in the case of a 1-hour setup, these variations tend to increase again. One probable explanation for this phenomenon is that vessels can significantly change their speed and bearing during these longer periods.

Table 1. Probabilistic Results of Genuine Kinematic Data

Genuine vessels						Simulated vessels				
	1 sec	30 sec	60 sec	30 min	1 hour	1 sec	30 sec	60 sec	30 min	1 hour
μ_V	4.89	3.99	3.80	3.05	2.85	12.15	11.45	11.30	8.96	5.57
σ_V	27.61	4.47	2.97	1.28	1.15	2.34	2.22	2.17	1.72	1.72
$\mu_{\varepsilon d}$	-0.20	-0.19	-0.19	-0.17	-0.28	$5\cdot10^{-6}$	10^{-4}	10^{-4}	0.15	0.78
$\sigma_{\varepsilon d}$	0.65	0.66	0.71	1.69	2.28	$4\cdot10^{-4}$	$7\cdot10^{-3}$	0.01	1.13	1.69
μ_α	-1.35	-1.08	-0.98	0.27	-1.13	0.26	1.02	1.89	16.23	16.43
σ_α	75.32	77.26	79.59	91.10	93	3.74	6.40	8.45	21.9	20.55

Table 2 presents the correlation results between the normalized $\left|\vec{V}\right|$, ε_d , ε_α and ε_{dk} values at the 60-second time interval setup. It can be observed that the mean velocities μ_V of the vessels are inversely correlated with their mean error $\mu_{\varepsilon d}$ for predicted travelled distance, indicating that these values decreased while the speed increased. Additionally, the mean Kalman error $\mu_{\varepsilon dk}$ was positively correlated with μ_V . In the case of the mean bearing error for consecutive locations μ_α , the mean speed of the vessels did not have a significant impact on this parameter.

Table 2. Correlation Matric for the Kinematic Parameters

	μ_V	$\mu_{\varepsilon d}$	$\mu_{\varepsilon dk}$	μ_α
μ_V	1	-0.33	0.33	-0.04
$\mu_{\varepsilon d}$	-0.33	1	-0.19	0.09
$\mu_{\varepsilon dk}$	0.33	-0.19	1	-0.06
μ_α	-0.04	0.09	-0.06	1

Additionally, it was observed that differentiating between genuine and spoofed vessels is possible by analyzing their Kalman filter errors distributions. Spoofed vessels tend to have perfectly linear trajectories with near-zero Kalman errors, while genuine vessels exhibit irregularities. Figure 8 presents two genuine vessels: one that attempts to maintain a constant COG and another showing multiple changes in COG, while Figure 9 displays a spoofed vessel behavior.

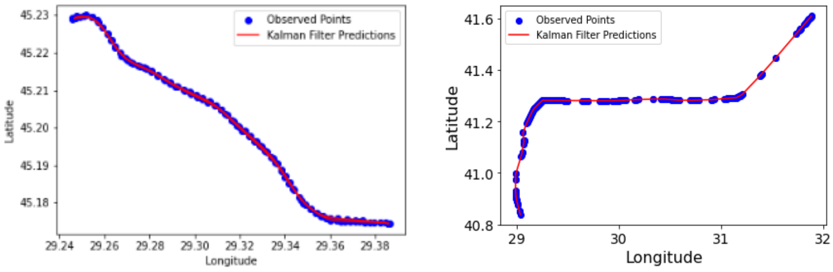

Figure 8. Genuine Trajectories

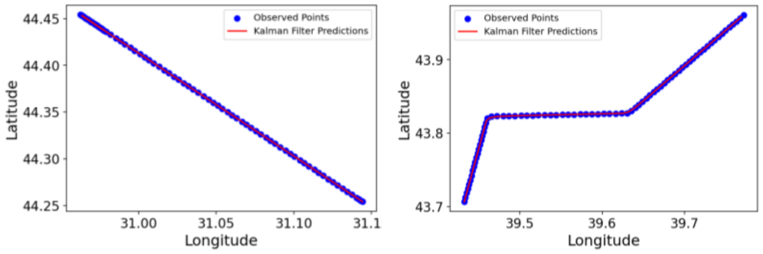

Figure 9. Spoofed Trajectories

Figure 10 presents the KDE plots for two vessels that attempted to maintain a linear course: one genuine and one simulated. For the genuine vessel, the course errors between consecutive actual and predicted positions exhibited significant variance, with most values ranging from -20 to 20 degrees. Additionally, limited course errors, ranging from 40 to 80 degrees, were also observed, likely due to abrupt changes or external factors affecting the trajectory. In contrast, the simulated vessel's course was perfectly rectilinear, resulting in all measured course errors between recorded and predicted trajectories being consistently 0, highlighting their predictable nature.

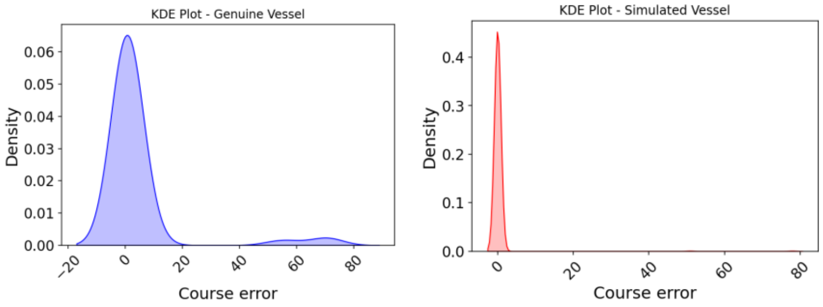

Figure 10. KDE plots for “course_error” parameters

5.3. Analyzing Machine Learning models' effectiveness

For this study, multiple machine learning models were trained using datasets collected at different time intervals: 1 second, 30 seconds, 60 seconds, 30 minutes, and 1 hour. The input features for these models included the calculated averages and standard deviations of various parameters, such as vessel velocities, time intervals between AIS reports, and the distance and bearing errors between consecutive locations. The performance of each ML model was evaluated based on these input features, with the results summarized in Table 3.

Overall, Random Forest, the Ensemble Model, Decision Trees and KNN achieved the highest accuracy across all setups, with values near 99%, even at longer time intervals. Deep Neural Networks also performed well, maintaining high accuracy above 0.96 across all setups, showing their capacity for handling complex patterns in the data. Conversely, all other models exhibited slightly lower accuracies, particularly at longer intervals.

Table 3. Accuracy of implemented ML models

	1 sec.	**30 sec.**	**60 sec.**	**30 min.**	**1 hr.**
SVM	0.97	0.96	0.95	0.90	0.89
Logistic Regression	0.96	0.95	0.93	0.87	0.89
Decision Trees	0.99	0.99	0.99	0.99	0.97
Random Forest	0.99	0.99	0.99	0.99	0.98
KNN	0.99	0.99	0.99	0.96	0.96
XGBoost	0.96	0.95	0.95	0.90	0.89
Naïve Bayes	0.99	0.98	0.98	0.90	0.81
Ensemble model	0.99	0.99	0.99	0.99	0.98
Deep Neural Network	0.99	0.99	0.99	0.98	0.96

After analyzing the permutation importance of all input parameters across the implemented models, it was observed that the following features had the greatest influence on the models' efficiency: the mean and standard deviations of the AIS reporting intervals, the standard deviation of the velocities, the mean course error, and the mean distance error between measured and predicted positions. Figure 11 presents the 3D plot of 4000 genuine and bogus vessels (in equal ratios) based on three selected calculated features. As observed, the simulated vessels exhibited no variations in AIS reporting intervals and distance errors. However, limited variations were observed in the course errors, as many of the simulated vessels were programmed to display multiple sudden and abrupt changes in their course.

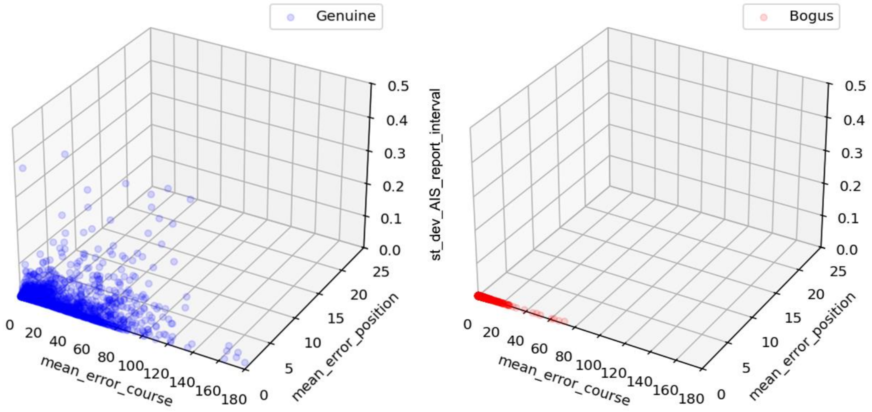

Figure 11. 3D plots of genuine and bogus parameters

6. Conclusions

AIS is essential for maritime safety, enabling vessel tracking and collision avoidance. However, the system's vulnerabilities, including susceptibility to spoofing and manipulation, underscore the need for robust detection measures. Spoofed AIS tracks pose significant risks, potentially leading to misinterpretations of vessel activities and compromising navigation safety.

Until a more secure version of AIS is globally adopted, various statistical and machine learning (ML) models can be employed to detect instances in which fake vessel tracks are transmitted through the system. This study demonstrates the effectiveness of using stochastic methods and ML models to differentiate between authentic and computer-generated tracks. The central premise is that predicting trajectories for genuine maritime vessels is subject to measurement and process errors, while spoofed tracks exhibit limited deviations due to their computer-generated precision.

By analyzing recorded kinematic data, distinct patterns can be extracted to differentiate genuine vessel trajectories from simulated ones. Furthermore, most ML models tested in this study achieved accuracies above 99%, thus highlighting the relative ease of detecting spoofed vessels and the effectiveness of these techniques in safeguarding maritime navigation.

REFERENCES

Androjna, A. & Perkovič, M. (2021) Impact of Spoofing of Navigation Systems on Maritime Situational Awareness. *Transactions on Maritime Science*. 10(2), 361–373. doi: 10.7225/toms.v10.n02.w08.

Androjna, A. & Perkovič, M. (2024) AIS Data Falsification - How Long Will It Be Before We Can No Longer Trust AIS. In *2024 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea), 14-16 October 2024, Portorose, Slovenia*. IEEE. pp. 301-306. doi:10.1109/MetroSea62823.2024.10765645.

Androjna, A., Pavić, I., Gucma, L., Vidmar, P. & Perkovič, M. (2024) AIS Data Manipulation in the Illicit Global Oil Trade. *Journal of Marine Science and Engineering*. 12(1), 6. doi:10.3390/jmse12010006.

Androjna, A., Perkovič, M., Pavic, I. & Mišković, J. (2021) AIS Data Vulnerability Indicated by a Spoofing Case-Study. *Applied Sciences*. 11(11), 5015. doi:10.3390/app11115015.

Coleman, J., Kandah, F. & Huber, B. (2020) Behavioral Model Anomaly Detection in Automatic Identification Systems (AIS). In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC), 6-8 January 2020, Las Vegas, NV, USA*. IEEE. pp. 0481-0487. doi:10.1109/CCWC47524.2020.9031248.

Goudosis, A. & Katsikas, S. (2022) Secure Automatic Identification System (SecAIS): Proof-of-Concept Implementation. *Journal of Marine Science and Engineering*. 10(6), 805. doi:10.3390/jmse10060805.

Katsilieris, F., Braca, P. & Coraluppi, S. (2013) Detection of Malicious AIS Position Spoofing by Exploiting Radar Information. In *Proceedings of the 16th International Conference on Information Fusion, 9-12 July 2013, Istanbul, Turkey*. IEEE. pp. 1196-1203.

Kontopoulos, I., Spiliopoulos, G., Zissis, D., Chatzikokolakis, K. & Artikis, A. (2018) Countering Real-Time Stream Poisoning: An Architecture for Detecting Vessel Spoofing in Streams of AIS Data. In *2018 IEEE 16th International Conference on Dependable, Autonomic and Secure Computing, 16th International Conference on Pervasive Intelligence and Computing, 4th International Conference on Big Data Intelligence and Computing, and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), 12-15 August 2018, Athens, Greece*. IEEE. pp. 981-986. doi:10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00139.

Newaliya, N. & Singh, Y. (2021) A Review of Maritime Spatio-Temporal Data Analytics. In *2021 International Conference on Computational Performance Evaluation (ComPE), 1-3 December 2021, Shillong, India*. IEEE. pp. 219-226. doi:10.1109/ComPE53109.2021.9751726.

Prasad, P., Vatsal, V. & Roy Chowdhury, R. (2021) Maritime Vessel Route Extraction and Automatic Identification System (AIS) Spoofing Detection. In *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT),19-20 February, Bhilai, India*. IEEE. pp. 1–11. doi:10.1109/ICAECT49130.2021.9392536.

Ray, C., Iphar, C. & Napoli, A. (2016) Methodology for Real-Time Detection of AIS Falsification. In: Vespe, M. & Mazzarella, F. (eds.) *Maritime Knowledge Discovery and Anomaly Detection Workshop, JRC 2016, 5-6 July 2016, Ispra, Italy*. pp. 74-77. doi:10.2788/025881.

Thomas, A. & Chiego, C. (2022) Maritime Cybersecurity: AIS Manipulation Motivations in the Maritime Domain. *Department of Global Studies & Maritime Affairs, California State University Maritime Academy*. https://wpsanet.org/papers/docs/ Thomas_Aurora_AIS_spoofing.pdf [Accessed 09 February 2025].

Wimpenny, G., Šafář, J., Grant, A. & Bransby, M. (2022) Securing the Automatic Identification System (AIS): Using public key cryptography to prevent spoofing whilst retaining backwards compatibility. *Journal of Navigation*. 75(2), 333-345. doi:10.1017/S0373463321000837.

Zhang, Xining, Zhang, Xynian & Li, P. (2023) Prediction Model of Ship Arrival Time using Neural Network and Kalman Filter. In *2023 3rd International Conference on Neural Networks, Information and Communication Engineering (NNICE), 24-26 February, 2023, Guangzhou, China*. IEEE. pp. 84–88. doi:10.1109/NNICE58320.2023.10105708.

Zheng, H., Hu, Q., Yang, C., Mei, Q., Wang, P. & Li, K. (2023) Identification of Spoofing Ships from Automatic Identification System Data via Trajectory Segmentation and Isolation Forest. *Journal of Marine Science and Engineering* 11(8), 1516. doi:10.3390/jmse11081516.

Alexandru POHONTU holds a bachelor's degree in Marine Engineering from the ″Mircea cel Batran″ Naval Academy, Constanta and a master's degree in robotics from the Romanian Military Technical Academy ″Ferdinand I″ of Bucharest. He is currently a Ph.D. student at the National University of Science and Technology Politehnica Bucharest. His thesis focuses on implementing AI algorithms in the field of Maritime Domain Awareness, with expertise in maritime surveillance technologies, maritime C2 systems, and maritime intelligence gathering. For his contributions to developing solutions for Maritime Domain Awareness systems, he was awarded the Romanian Honor Medal of Communications and Informatics.

Alexandru POHONTU deține o diplomă de licență în Inginerie Maritimă de la Academia Navală „Mircea cel Bătrân" din Constanța și o diplomă de master în robotică de la Academia Tehnică Militară din România „Ferdinand I" din București. În prezent, este doctorand la Universitatea Națională de Știință și Tehnologie Politehnica București. Teza sa se concentrează pe implementarea algoritmilor de inteligență artificială în domeniul Conștientizării Domeniului Maritim (Maritime Domain Awareness), având expertiză în tehnologii de supraveghere maritimă, sisteme de comandă și control maritim (C2) și colectarea de informații maritime. Pentru contribuțiile sale la dezvoltarea soluțiilor pentru sistemele de Conștientizare a Domeniului Maritim, a fost distins cu Medalia de Onoare a Comunicațiilor și Informaticii din România.

Constantin VERTAN is a professor at the National University of Science and Technology Politehnica Bucharest, with extensive experience in image processing, computer vision, and artificial intelligence. He holds a Ph.D. in Electronic Engineering, and he is author/ co-author of more than 150 journal, conference papers and patents in the field of image processing and its applications. His expertise encompasses both theoretical and applied research, focusing on developing innovative AI-driven solutions for challenges in engineering, medical imaging, and industrial automation. His work has been recognized through various awards and memberships in professional organizations.

Constantin VERTAN este profesor la Universitatea Națională de Știință și Tehnologie Politehnica București, având o vastă experiență în procesarea imaginilor, viziunea computerizată și inteligența artificială. Deține un doctorat în Inginerie Electronică și este autor/coautor a peste 150 de articole științifice, lucrări de conferință și brevete în domeniul procesării imaginilor și al aplicațiilor acesteia. Expertiza sa acoperă atât cercetarea teoretică, cât și cea aplicată, concentrându-se pe dezvoltarea de soluții inovatoare bazate pe inteligență artificială pentru provocări din inginerie, imagistică medicală și automatizare industrială. Activitatea sa a fost recunoscută prin diverse premii și apartenențe la organizații profesionale.

Iancu CIOCIOI is a Senior Lecturer in the Department of Electrical and Electronic Engineering at the ″Mircea cel Batran″ Naval Academy, Constanta. Before his retirement from the Romanian Navy, he served in various executive and management positions as an officer, particularly in the field of education. He holds expertise in analog and digital electronics, power electronics, radar, radiocommunication, and hydroacoustic equipment. A seasoned educator, Iancu Ciocioi combines his extensive operational experience with a passion for teaching, fostering a dynamic and innovative learning environment at the academy.

Iancu CIOCIOI este lector în Departamentul de Inginerie Electrică și Electronică al Academiei Navale „Mircea cel Bătrân” din Constanța. Înainte de retragerea sa din Marina Română, a ocupat diverse funcții executive și de conducere ca ofițer, în special în domeniul educației. Deține expertiză în domenii precum electronica analogică și digitală, electronica de putere, tehnologii radar, radiocomunicații și echipamente hidroacustice. Educator experimentat, Iancu Ciocioi îmbină vasta sa experiență operațională cu pasiunea pentru predare, contribuind la crearea unui mediu de învățare dinamic și inovator în cadrul academiei.

Ciprian POPA holds a bachelor's degree in Marine Engineering and a master's degree in Electrical Engineering. He is currently a Ph.D. student at the National University of Science and Technology Politehnica Bucharest, in the field of Electrical Engineering. He teaches at the ″Mircea cel Batran″ Naval Academy, Constanta and has extensive experience in maritime operations and navigation. For his contributions, he was awarded the Romanian Navy Medal of Honor.

Ciprian POPA - deține o diplomă de licență în Inginerie Maritimă și o diplomă de master în Inginerie Electrică. În prezent, este doctorand la Universitatea Națională de Știință și Tehnologie Politehnica București, în domeniul Ingineriei Electrice. Predă la Academia Navală „Mircea cel Bătrân” din Constanța și are o vastă experiență în operațiuni maritime și navigație. Pentru contribuțiile sale, a fost distins cu Medalia de Onoare a Marinei Române.