Occurrence of Unknown Sensor Data within AIS Dynamic Messages

Pojava nepoznatih podataka senzora unutar AIS dinamičkih poruka

Paweł Banyś
German Aerospace Centre (DLR)
Neustrelitz, Germany
E-mail: PAWEL.BANYS@DLR.DE

Frank Heymann
German Aerospace Centre (DLR)
Neustrelitz, Germany
E-mail: FRANK.HEYMANN@DLR.DE

Maciej Gucma
Maritime University of Szczecin
Szczecin, Poland
E-mail: M.GUCMA@AM.SZCZECIN.PL

Summary

For more than a decade, the Automatic Identification System (AIS) has contributed to increasing the safety of navigation at sea. Despite the benefits of the system, AIS messages shared between vessels and the AIS dynamic data transferred to the Portable Pilot Units may contain unknown values of sensor data if the sensor data on board becomes either unavailable or undeliverable for any reason. In this paper, an experiment is conducted to analyse the performance of an AIS transponder during a virtual sea voyage. By altering the sensor data rate it is possible to cause the AIS transponder to output AIS messages with unknown sensor data. After performing the experiment, a generic approach is used in order to establish a correlation between the sensor data rate and the relative occurrence frequency of AIS unknown values. This leads to the formulation of a simple equation which describes the interdependence between the interval of sensor data provision, the age of the sensor data allowed by the AIS transponder and the percentage of unknown data within the AIS data output.

Sažetak

Više od desetljeća sustav automatske identifikacije (AIS) pridonosi povećanju sigurnosti plovidbe na moru. Unatoč prednostima sustava, AIS poruke dijeljene između plovila, kao i AIS dinamički podaci preneseni na prijenosne peljarske jedinice, mogu sadržavati nepoznate vrijednosti podataka senzora ako ti podaci na ploči iz bilo kojeg razloga postanu nedostupni ili neisporučivi. U radu je proveden eksperiment za analizu performansi AIS transpondera tijekom virtualnog putovanja na moru. Promjenom brzine podataka senzora moguće je podesiti izlaz AIS poruka s nepoznatim podacima senzora s AIS transpondera. Nakon izvođenja eksperimenta koristio se generički pristup kako bi se uspostavila povezanost između brzine podataka senzora i relativne učestalosti pojavljivanja nepoznatih vrijednosti AIS-a. Time se dolazi do jednostavne jednadžbe koja opisuje međuovisnost intervala pružanja senzorskih podataka, starosti senzorskih podataka koje dopušta AIS transponder i postotka nepoznatih podataka u izlazu AIS podataka.

1. INTRODUCTION / *Uvod*

One of the most important technical innovations introduced into the maritime world during the first decade of the twenty first century is the Automatic Identification System (AIS). The main advantage of using AIS is the ability to continuously exchange navigational data between vessels carrying the AIS equipment via two reserved very high frequency (VHF) marine channels. Thanks to AIS the officers of the watch at sea as well as the authorities ashore can obtain the basic characteristics of the vessels within their current radio range. As of 2016, according to Equasis, there were 89,804 merchant vessels worldwide [4, p.6]. Taking into consideration the compulsory carriage requirements enforced by the International Convention for the Safety of Life at Sea (SOLAS) [7], about 70% of the above mentioned world fleet should be expected to have AIS equipment on board because of their vessel type and size.

A typical Class A AIS transponder designed for SOLAS vessels gathers data from the on board sensors it is connected to, and transforms the readings into a binary AIS message. The two main types of AIS messages relevant for traffic assessment are dynamic and static messages. The former contain automatically derived sensor-based variables that describe the ship position and her movement, such as geographic latitude (LAT), geographic longitude (LON), true heading (THDG), rate of turn (ROT), speed over ground (SOG), and course over ground (COG). The latter contain human configured parameters like vessel name, call sign, vessel size, and destination harbour, to name a few [8, p.108-113].

Since the introduction of AIS, maritime pilots also benefit from the system. During pilotage duty on board they are able to use a dedicated serial data interface called Pilot Plug

(PP), which provides an exact copy of the dynamic message transmitted by the AIS transponder. To the PP, the pilots connect a Portable Pilot Unit (PPU), a small tablet based computer which provides navigational information needed by the pilot during manoeuvres [10]. The PPU software is able to display the own vessel position and vector on a nautical chart using the provided PP data at the rate of one dynamic message per second.

Furthermore, the AIS hardware provides an important data input for the Voyage Data Recorder (VDR)which SOLAS vessels use to store all relevant information generated by their navigational instruments aboard [11]. During investigations following an accident, it becomes essential to know the dynamic data of other vessels involved in order to assess manoeuvres leading to calamity and to identify responsible parties.

The design of AIS allows some data to be unavailable prior to VHF transmission of the AIS message. For not wasting the time slot reserved for the AIS broadcast, the missing readings of the parameters are replaced by special, so called default values, that explicitly indicate which sensor data are currently unknown. The same applies to the PP output. The default values of selected dynamic AIS parameters are shown in Table 1.

Table 1 The selected default values of the parameters stored within an AIS dynamic position report
Tablica 1. Odabrane zadane vrijednosti parametara pohranjenih u AIS izvješću o dinamičkoj poziciji

AIS variable	LAT	LON	THDG	ROT	SOG	COG
Raw default value	54600000	108600000	511	−128	1023	3600

Source: [8, p.109]

The more unknown values within the dynamic message are transmitted via the PP, the less effective the navigational support for the pilots on the PPU may become. The same principle applies to the usability of the VDR dataset: the more incomplete the AIS data from external vessels is, the more difficult it is to obtain a reliable picture of a hazardous situation that is being investigated post factum.

2. OCCURRENCE OF AIS DEFAULT DATA UNDER REAL CONDITIONS / *Pojava zadanih podataka AIS-a u stvarnim uvjetima*

The topic of default values within AIS messages has already been covered in the literature. For instance, the results obtained in Taiwan indicate that AIS data recorded in 2004 off the coast of Taiwan contained default values of the dynamic parameters, with unknown value of THDG reaching the highest score of 6% of all position reports [3, p.2250-2251].

Another study devoted to this subject was published in 2012 [1]. AIS data, covering the Baltic Sea along the German coast, recorded during September 2011 was spatially divided into two exemplary categories: harbour area and open sea area, as outlined in Figure 1.

All AIS dynamic position reports containing any default values of the AIS variables listed in Table 1 were counted. The statistics of their occurrence within the Class A transponders is shown in Table 2.

Table 2 Relative occurrence frequency of default values of specific dynamic AIS variables in September 2011
Tablica 2. Relativna učestalost pojavljivanja zadanih vrijednosti specifičnih dinamičkih AIS varijabli u rujnu 2011

AIS variables set to default value	Relative occurrence frequency in Class A messages [%]		
	Harbour area	Open sea area	Harbour & open sea area
ROT	30.2	8.8	19.2
THDG	30.1	7.2	18.3
COG,SOG	2.2	<0.1	1.1
LON,LAT	<0.1	<0.1	<0.1

Source: [1]

The results show that the highest occurrence of the default values can be observed in case of the ROT and the THDG AIS variables.

An alternative study, which focused on the AIS completeness, was conducted in 2015. The researchers used AIS dynamic position reports of the Gulf of Gdańsk over many weeks [5]. It was shown that the number of incomplete AIS messages during a 24-hour recording, which is presented in Figure 2.

The results indicate that ROT and THDG (originally abbreviated in Figure 2 as "HDG") are the two AIS variables, which are most affected by the occurrence of AIS default values. The other AIS variables also show signs of unknown sensor data, but they are obviously exceeded by the default values for both ROT and THDG.

The analysis of 2012 has been recreated to check whether the results presented in Table 2 have changed after more than 4 years. For this purpose, AIS data was recorded in March 2016. An identical spatial division into harbour area and open sea area was made. The occurrence statistics of AIS default values within Class A transponders are presented in Table 3.

Figure 1 The spatial categories of AIS data: harbour (left) and *open sea (right)*
Slika 1. Prostorne kategorije podataka AIS-a: luka (lijevo) i otvoreno more (desno)
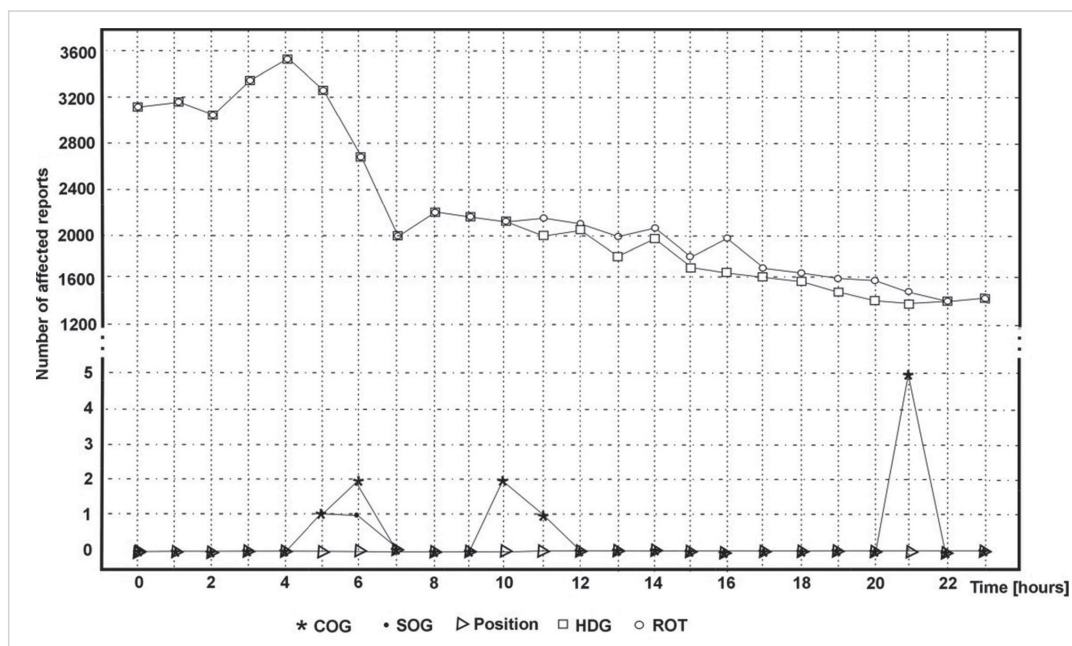Source: [1]

Figure 2 Hourly outcomes of incomplete AIS messages for selected AIS data components
Slika 2. Rezultati nepotpunih AIS poruka svaki sat za odabrane komponente AIS podataka

Source: [5]

Table 3 Relative occurrence frequency of default values for specific dynamic AIS variables in March 2016 compared with the results (in parentheses) of September 2011
Tablica 3. Relativna učestalost pojavljivanja zadanih vrijednosti za određene dinamičke AIS varijable u ožujku 2016. u usporedbi s rezultatima (u zagradama) iz rujna 2011.

AIS variable set to default value	Relative occurrence frequency in Class A messages [%]		
	Harbour area	Open sea area	Harbour area & open sea area
ROT	41.8 (30.2)	2.6 (8.8)	22.5 (19.2)
THDG	41.7 (30.1)	1.7 (7.2)	22 (18.3)
COG	0.3 (2.2)	<0.1 (<0.1)	0.1 (1.1)
SOG	<0.1 (2.2)	<0.1 (<0.1)	<0.1 (1.1)
LON	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)
LAT	<0.1 (<0.1)	<0.1 (<0.1)	<0.1 (<0.1)

Source: Author's research

The comparison in Table 3 shows that the AIS variables ROT and THDG remain the top two sources of AIS default values within the areas of investigation. It is also worth mentioning that about 3.4% of shipborne AIS transmitters, whose AIS data was acquired in March 2016 (3% in September 2011), broadcast a mixture of known and unknown (default) values of their external sensor data. It is expected that if an on-board sensor breaks, its data will be marked as unknown until the sensor is repaired some time later. It is less likely that such a sensor will repair itself and fail repeatedly within seconds or minutes. Therefore such a mixture of known and unknown values of the sensor data could be better explained by a faulty data connection between the sensor and the AIS transponder. In this respect, it does not matter which standard of data connection could be involved, be it NMEA 0183, NMEA 2000 or any other proprietary link. Any of them could be disrupted and lead to such an AIS broadcast.

To illustrate that this phenomenon occurs in harbours as well as on the open sea, the positions of vessels that transmitted a mixture of known and unknown (default) values within their dynamic AIS messages in March 2016, are shown in Figure 3.

This indicates that the existence of AIS default values of a single on-board sensor chosen for analysis cannot be perceived only in terms of two categories of vessels: those that do not transmit AIS default values at all, and those that broadcast only AIS default values. Results obtained from other researchers, like [6] or [12], to mention a few, confirm this assumption.

The results above show that there may be circumstances where some components of a ship's bridge system are occasionally unable to deliver sensor data to the AIS transponder with no conclusion that any element of the bridge system sustained permanent damage.

3. INDUCING AIS DEFAULT VALUES DURING A LAB EXPERIMENT ON AIS TRANSPONDER / *izazivanje zadanih vrijednosti AIS-a tijekom laboratorijskog eksperimenta na AIS transponderu*

The findings of the above mentioned studies indicate that the AIS parameters most affected by the presence of AIS default values are ROT and THDG. The most significant feature of those two AIS variables is that their values are generated by external sensors (namely a gyrocompass and a ROT indicator also abbreviated as ROTI) which are connected to the AIS transponder. In contrast, LAT, LON, COG and SOG are usually derived from an internal positioning device, typically a Global Navigation Satellite System (GNSS) receiver, integrated into the AIS transponder. As long as no obstruction of sensor data flow is obstructed, the AIS transponder is able to generate a complete dynamic position report without using the default values. It should be noted that establishing a stable connection between external sensors like the gyrocompass and the AIS transponder is a certain technical challenge related to the bridge design and is clearly more susceptible to disruptions than the internal data links that function within the AIS transponder firmware, such as the delivery of GNSS data.
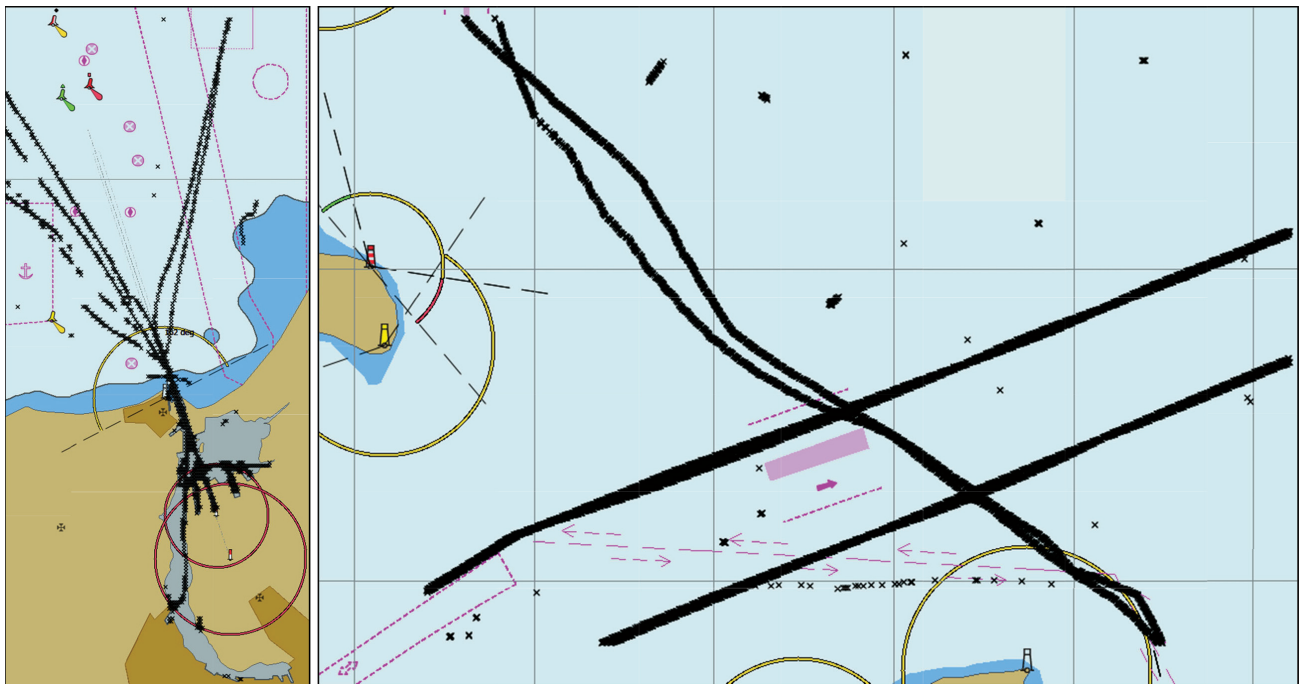
Figure 3 Locations of vessels (black dots) broadcasting AIS data mixture of known and unknown values: in the Harbour of Rostock (left) and at open sea north of Arkona (right)
Slika 3. Lokacije plovila (crne točke) koja emitiraju miješane AIS podatke poznatih i nepoznatih vrijednosti: u luci Rostock (lijevo) i na otvorenom moru sjeverno od Arkone (desno)
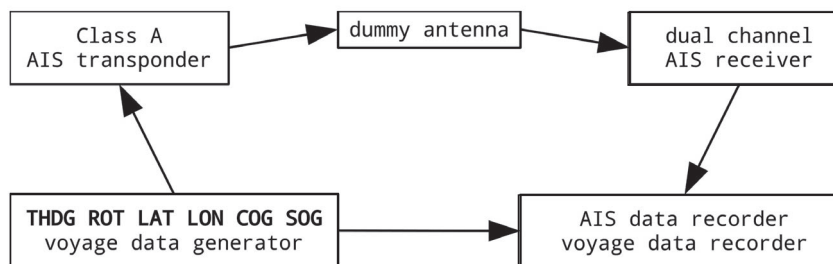Source: Author's research

Figure 4 Exchange of data between the components that were used during the AIS transponder experiment
Slika 4. Razmjena podataka između korištenih komponenti tijekom eksperimenta AIS transpondera
Source: Author's research

In order to investigate how often the AIS default data is transmitted by AIS transponder in relation to the external sensor data rate, a lab experiment was conducted with a Class A AIS transponder and a collection of simulated sensors. The arrangement of the components used in the experiment is presented in Figure 4.

The core element of the above setup was a custom written Python-based voyage data generator that was able to deliver an NMEA 0183 data stream from three artificial sensors: GNSS receiver, gyrocompass and ROTI. Throughout the whole experiment, the voyage data generator delivered its data to the AIS transponder and the voyage data recorder. The AIS transponder transmitted the AIS messages via a dummy VHF antenna link to the AIS receiver, which transferred the AIS data to the AIS data recorder.

The simulated sensor data completely depicted a fictitious vessel that was on a voyage of about 2600 nautical miles for 22 days. During the voyage the vessel was performing a continuous series of identically commanded manoeuvres. Each manoeuvre took 2 hours and comprised 1 hour of rhumb line navigation

with a constant COG, 30 minutes of constant turn to port and 30 minutes of constant turn to starboard. The maximum absolute ROT was arbitrarily set to 15 degrees per minute. Throughout a single manoeuvre the SOG stayed constant and non-zero.

The current geographic position of the ship during her simulated voyage was calculated with the PyProj Python module (version 1.9.5.1). For simplicity, the influence of wind, waves, tides and currents was neglected, so that THDG of the ship was always equal to her COG. ROT as a function of time was defined by:

$$R(T) = \begin{cases} 0 & : \quad T < t_1 \cup T \geq t_4 \\ r \cdot \sin^2\left(\dfrac{T - t_1}{t_2 - t_1} \cdot \dfrac{\pi}{2}\right) & : \quad t_1 \leq T < t_2 \\ r & : \quad t_2 \leq T < t_3 \\ r \cdot \sin^2\left(\dfrac{T - t_3}{t_4 - t_3} \cdot \dfrac{\pi}{2} + \dfrac{\pi}{2}\right) & : \quad t_3 \leq T < t_4 \end{cases} \tag{1}$$

$$T \in \mathbb{R} \wedge t_n \in \mathbb{R} \wedge n \in \{1,2,3,4\} \wedge r \in \mathbb{R}$$

where r is the maximum ROT, the time from t_1 to t_2 is the turn acceleration stage, the time from t_2 to t_3 is the constant turn stage, and the time from t_3 to t_4 is the turn deceleration stage. ROT defined by Equation (1) is continuous and differentiable everywhere.

The current value of THDG at any given moment of the turn manoeuvre was calculated using the following equation:

$$H(T) = h + \int_{t_1}^{T} R(T) \, dT \tag{2}$$

$$T \in \mathbb{R} \wedge h \in \mathbb{R} \wedge t_1 \in \mathbb{R} \wedge t_1 \leq T$$

where h is the initial value of THDG prior to the turn. The time t_1 is the beginning of the turn manoeuvre. The integral defined by Equation (2) could be calculated by the following equation:

$$\int R(T)\,dT = \begin{cases} \dfrac{r \cdot \left(\pi \cdot (T - t_1) + (t_1 - t_2) \cdot \sin\left(\dfrac{\pi \cdot (T - t_1)}{t_2 - t_1} \right) \right)}{2 \cdot \pi} + h & : \quad t_1 \leq T < t_2 \\[4mm] r \cdot T + h & : \quad t_2 \leq T < t_3 \\[4mm] \dfrac{r \cdot \left((t_4 - t_3) \cdot \sin\left(\dfrac{\pi \cdot (T - t_4)}{t_3 - t_4} \right) + \pi \cdot (T - t_4) \right)}{2 \cdot \pi} + h & : \quad t_3 \leq T < t_4 \end{cases} \tag{3}$$

$$T \in \mathbb{R} \wedge r \in \mathbb{R} \wedge h \in \mathbb{R} \wedge t_n \in \mathbb{R} \wedge n \in \{1,2,3,4\} \wedge r \in \mathbb{R}$$

An example plot of the ROT function describing a 60 second turn to starboard at 15 degrees per minute, bounded by 10 second long steady THDG, together with the corresponding values of THDG changing from the initial 225° to 237.5° during the turn, is shown in Figure 5.

The stepwise characteristic of the single two-hour-long manoeuvre is shown in Table 4.

Table 4 The parameters of the single two-hour-long manoeuvre
Tablica 4. Parametri jednog dvosatnog manevara

Step	Duration [s]	Maximum ROT [°/minute]	Timing of the turn [s]		
			$t_2 - t_1$	$t_3 - t_2$	$t_4 - t_3$
1	1200	0	not applicable		
2	1800	−15	10	1780	10
3	1200	0	not applicable		
4	1800	15	10	1780	10
5	1200	0	not applicable		

Source: Author's research

The duration of the manoeuvre steps listed in Table 4 was chosen so that the simulated vessel was either keeping a constant course over ground for one hour or turning for one hour. According to the ITU specification, a Class A AIS transponder can nominally transmit up to 1800 dynamic AIS messages per hour which constitutes a suitable collection of data for further evaluation [8, p.8].

Although the hydrodynamics of vessel motion as investigated by [2] or [9] were completely ignored, the data feed from the voyage data generator to the AIS transponder created a sufficient illusion of AIS equipment as an operational part of the real shipborne bridge system. During the lab experiment, the AIS transponder operated in so-called autonomous mode. This means that the device determined its own schedule for VHF transmission and its time slots did not conflict with other AIS transmitters. The main advantage of this approach is that the AIS transponder was not forced to change its schedule in order to respond to the arbitrary data interrogation messages broadcast by other AIS stations, as specified in [8, p.30]. The autonomous mode guaranteed that the VHF transmission intervals chosen by the AIS transponder, as specified in [8, p.8], depended only on SOG and ROT obtained from the simulated sensors.

The set of values of SOG, expressed in knots and used in the course of the experiment, was defined by:

$$V = [0.5, 17, 26] \tag{4}$$

The set of the external sensor data rates, at which both the gyrocompass and the ROTI delivered their simulated THDG and ROT to the AIS transponder during the experiment, was defined according to:

$$G = [20, 10, 5] \cup \left\{ \frac{1}{x} \,\middle|\, 1 \leq x \leq 60 \wedge x \in \mathbb{N} \right\} \tag{5}$$

It offered the possibility to simulate an external sensor with a data rate ranging from a high performance output of 20 measurements per second down to a low performance output of one measurement per minute. Each manoeuvre was repeated in a two-level nested loop over the values of SOG defined by Equation (4) and the sensor data rates defined by Equation (5).

After completing the simulated voyage, the recorded AIS data was analysed for the presence of AIS default values for THDG and ROT. The relative occurrence frequency of unknown THDG data in respect to different SOG classes is shown in Figure 6.
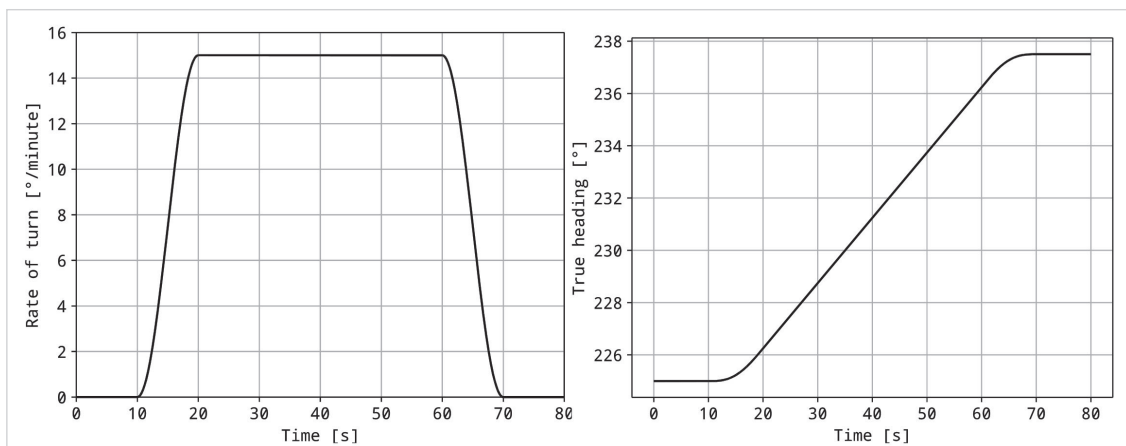

Figure 5 Example of ROT function describing a turn to starboard starting at T = 10 and ending at T = 70 (left) and the corresponding value of THDG during that turn (right).
Slika 5. Primjer funkcije rotacije koja opisuje skretanje na desno i počinje na T = 10 a završava na T = 70 (lijevo) i odgovarajuću vrijednost THDG-a tijekom skretanja (desno).
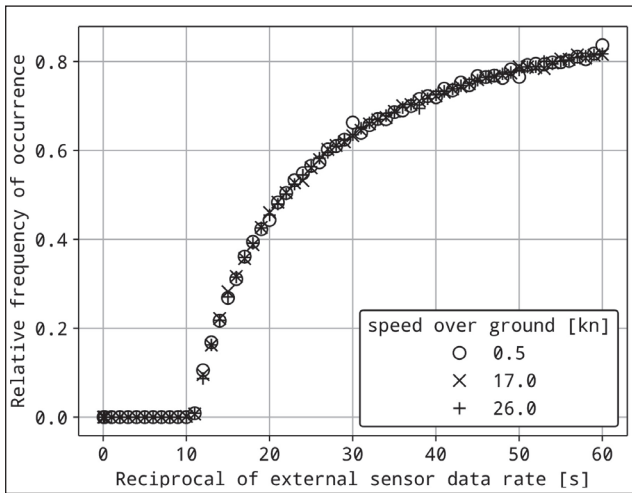Source: Author's research

Figure 6 Relative occurrence frequency of AIS default values for THDG observed during the lab experiment at three levels of SOG

Slika 6. Relativna učestalost pojavljivanja AIS zadanih vrijednosti za THDG uočena tijekom laboratorijskog eksperimenta na tri razine SOG-a

Source: Author's research

It can be seen that the AIS transponder did not transmit any default data for THDG as long as the gyrocompass delivered THDG values more frequently than one data report per ten seconds. After the output data rate from the gyrocompass slowed down beyond this limit, the AIS transponder started to generate default values of THDG. The abscissa of 10 seconds can be considered as the oldest age of the sensor data allowed by the AIS transponder. The age of the sensor data is the delay between the moment when the sensor data value is made available to the AIS transponder, and the instant when the sensor data value is broadcast via VHF. The smaller the sensor data age, the more up-to-date the sensor data transmitted by AIS or delivered to the PPU becomes. It should be noted that the sensor data age allowed by the internal configuration of the AIS transponder is not defined by AIS standards. Therefore the choice of the value remains at the discretion of the hardware manufacturer. According to the author's personal inquiry at several leading makers of AIS transponders, it is concluded that their choice of the sensor data age is either 5 seconds or 10 seconds.

The simulation results depicted in Figure 6 show that the AIS transponder, prior to the upcoming VHF transmission, considered the measurement data from the external sensor as up to date if the sensor data was not older than 10 seconds. All sensor data older than 10 seconds was evaluated as unknown and the appropriate AIS default value was then set in the AIS dynamic position report. The plot in Figure 6 shows a high similarity in the relative occurrence frequency of the AIS default data for the three different classes of SOG. The relative occurrence frequency of unknown ROT data in relation to the three SOG classes was identical to the relative occurrence frequency of THDG, therefore the presentation of ROT was omitted in Figure 6.

4. RELATIONSHIP BETWEEN THE SENSOR DATA RATE AND THE OCCURRENCE OF AIS DEFAULT VALUES / *Odnos između brzine podataka senzora i pojave zadanih vrijednosti AIS-a*

The lab experiment on the AIS transponder produced an extensive collection of dynamic messages containing AIS default data. However, despite its duration of 22 days, the experiment was too short to create a more comprehensive collection of the timeline interactions between the times when external sensor data was sent to the AIS transponder, and the times when the AIS data was transmitted over VHF radio link and delivered to the PPU.

To simulate the timing of the data exchange between the external sensors and the AIS transponder, which broadcasts its data in a constant interval, a custom written program was used to generate the operational timelines of an external sensor input and the transmission schedule of the AIS transponder. The program used a three level nested loop to iterate over the AIS broadcast intervals, the external sensor data rates, and the time shifts of the timeline starting points. The basic idea behind the timeline simulation is to generate a series of times when the AIS transponder is about to broadcast the AIS message, together with a series of times when an external sensor delivers its data to the AIS transponder. By calculating the time differences between these moments, it is possible to figure out how old the external sensor data is upon its delivery to the AIS transponder. The age of the sensor data determines whether or not the AIS transponder should use the AIS default values in the AIS message.

In order to illustrate the concept of the timelines, an example of ROTI providing ROT measurement every 12 seconds to an AIS transponder transmitting AIS data every 10 seconds with a sensor data age set to 5 seconds and the total AIS runtime of 40 seconds is shown in Figure 7.
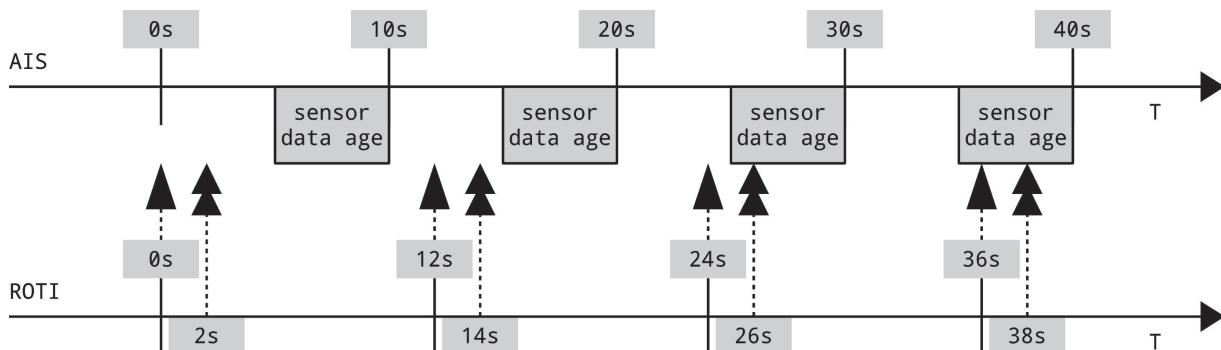

Figure 7 Example timelines of the AIS broadcast schedule and the ROT sensor data delivery: without time shift (single arrows) and shifted by 2 seconds relatively to the AIS timeline (double arrows)

Slika 7. Primjeri vremenskog okvira rasporeda emisije AIS-a i isporuka podataka senzora rotacije: bez vremenskog pomaka (pojedinačne strelice) i s pomakom od 2 sekunde u odnosu prema vremenskom okviru AIS-a (dvostruke strelice)

Source: Author's research

In case of the sensor data delivery without time shift, only one AIS dynamic position report, transmitted at $T = 40s$, would contain the actual value of ROT. The other three AIS broadcasts would mark the ROT variable as unknown and use its corresponding default value because the external sensor data would be considered obsolete.

The starting points in time of the external sensor operation and the AIS broadcast schedule are not always the same and their perfect alignment, as shown in Figure 7 (single arrows), might be a rarity. Therefore it was also necessary to simulate the timelines with shifted starting points. In order to exemplify the effects of a time shift on the change of the occurrence of the AIS default data, the timeline starting point of the external sensor in Figure 7 is moved forward by 2 seconds. The result of this adjustment is marked by double arrows. Due to this time shift the AIS dynamic position reports, sent at $T = 30s$ and at $T = 40s$, would contain the actual value of ROT. The other two AIS transmissions would have to apply the default value to the ROT variable. This means that the relative occurrence frequency of AIS default data dropped from $\frac{3}{4}$ to $\frac{2}{4}$ only because of altering the relative time shift of the sensor timeline.

For the purpose of generating the AIS broadcast timelines, the set of all possible Class A AIS transmission intervals specified by [8, p.8] was defined in seconds by:

$$A = \left\{ 2, \frac{10}{3}, 6, 10, 180 \right\} \tag{6}$$

In order to create the set of the external sensor data rates, the following data set was used:

$$B = \left\{ 20, 10, 5, 2 \right\} \cup \left\{ \frac{1}{x} \middle| 1 \leq x \leq 200 \wedge x \in \mathbb{N} \right\} \tag{7}$$

Set B covers sensor data rates ranging from a high performance feed of 20 measurements per second down to a near failure data provision of 1 measurement per 200 seconds.

The set of the relative time shifts of the timeline starting points in seconds was provided by:

$$C = \left\{ \frac{x}{10} \middle| 0 \leq x \leq \lfloor 10 \cdot a - 1 \rfloor \wedge x \in \mathbb{N}_0 \right\} \tag{8}$$

where a is the AIS broadcast interval selected by the loop iterating over the set defined in Equation (6).

The set of the sensor data age values applied to the configuration of the simulated AIS transponder was specified in seconds by the following:

$$D = \left\{ 1, 3, 5, 10, 15, 20 \right\} \tag{9}$$

The usage of various sensor data ages D can be compared to the performance of the lab experiment on six different AIS transponders under real conditions. It is expected that diverse AIS transponder configurations may produce different results.

The sets A, B, C and D were looped through by the timeline generator. Each pair of timelines created in the innermost loop, which described the simulated AIS broadcast schedule and the simulated sensor data delivery, had the total AIS runtime of 24 hours. In the course of the simulation the alignment of timelines was analysed to compute the values of the sensor data age and to count the occurrence of AIS default values of the sensor data, as shown exemplarily in Figure 7. It is worth mentioning that if, instead of the simulative approach, the above timeline generation was conducted in real time as part of another AIS transponder experiment, it would take about 6,720 years to complete.

Figure 8 shows the relative occurrence frequency of AIS default sensor data with respect to different values of sensor data age, which was obtained during the timeline analysis.
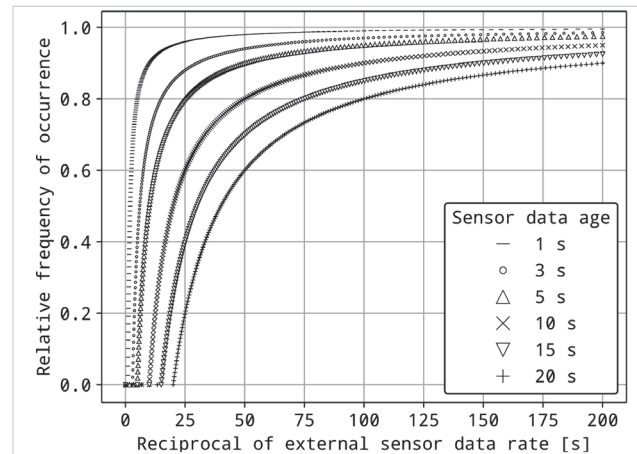

Figure 8 Relative occurrence frequency of the AIS default values of the sensor dataat different levels of the sensor data age

Slika 8. Relativna učestalost pojavljivanja AIS zadanih vrijednosti podataka senzora na različitim razinama starosti senzorskih podataka
Source: Author's research

It can be observed that the simulated AIS transmissions are free from default values of the external sensor data, as long as the interval, in which the external sensor data is delivered, remains less than or equal to the sensor data age allowed by the AIS transponder. The occurrence of AIS default data becomes visible after the sensor data age is exceeded by the external sensor data interval. The value of the sensor data age has a major influence on what percentage of AIS default values may be expected during the provision of AIS data, both VHF and PP. In the example above, a sensor operating at an interval of one measurement per 50 seconds would contribute to a relative occurrence frequency of default data at about 60%, if the AIS transponder had a relaxed sensor data age setting of 20 seconds. A rigorous setting of sensor data age at 3 seconds would produce 90% of AIS transmissions with a default value of the same sensor data feed. It can also be noticed that the behaviour of the real AIS transponder shown in Figure 6 bears a strong similarity to the one in Figure 8 when the sensor data age equals 10 seconds, which corresponds to the data age configuration of the AIS transponder subjected to the experiment.

5. CURVE FITTING OF THE RELATIVE OCCURRENCE FREQUENCY OF AIS DEFAULT DATA / *Prilagodba krivulje relativne učestalosti pojavljivanja zadanih podataka AIS-a*

In order to find a mathematical function that can describe the relation between the external sensor data interval and the relative occurrence frequency of the AIS default data at any given constant sensor data age setting (cp. Figure 8), the following general form of a piecewise defined function was used:

$$F(k) = \begin{cases} 0 & : & 0 < k \leq g \\ G(k) & : & g < k \end{cases} \tag{10}$$
$$k \in \mathbb{R}_{>0} \wedge g \in \mathbb{R}_{>0}$$

where F(k) is the overall relative occurrence frequency of AIS default data, k is the interval of the external sensor data, g is the sensor data age allowed by the AIS transponder, and G(k) is a function describing the actual positive occurrence frequency

of AIS default data, when the sensor data interval exceeds the sensor data age.

The following important features of G(k) had to be taken into consideration, before a suitable mathematical function could be thought of for the curve fitting attempt:

$$\lim_{k \to \infty} G(k) = 1 \tag{11}$$

$$G(g) = 0 \tag{12}$$

$$\forall\, k > g : \frac{\partial}{\partial k} G(k) > 0 \tag{13}$$

$$\forall\, k > g : 0 < G(k) < 1 \tag{14}$$

When the external sensor data feed is close to breakdown and rarely provides the sensor measurements, the relative occurrence frequency of AIS default data will reach almost unity, which is consistent with Equation (11). Equation (12) guarantees that function F(k) is continuous at $k = g$, and Equation (13) indicates that function G(k) should be strictly increasing. Finally, Equation (14) enforces that function G(k) must not intersect its horizontal asymptote, since the relative occurrence frequency can be neither greater than unity nor negative.

One of the least complicated mathematical functions that could match the above features belongs to the family of rational functions and is defined by the following equation:

$$G(k) = 1 - \frac{e^p}{k + e^p - g} \tag{15}$$

$$k \in \mathbb{R}_{>0} \wedge g \in \mathbb{R}_{>0} \wedge g < k \wedge p \in \mathbb{R}$$

where p is the real number parameter influencing the curvature of the function. The usage of e^p ensures that the numerator in Equation (15) is always greater than zero.

The process of curve fitting was accomplished with the Python module curve_fit, which originates from the scipy.optimize open source collection of numerical algorithms. The curve_fit module uses nonlinear least squares method to fit a given function to a dataset. It was applied to find the value of p for each sensor data age defined by Equation (9). The plot shown in Figure 8 superimposed with the best fit dashed line curves is presented in Figure 9.
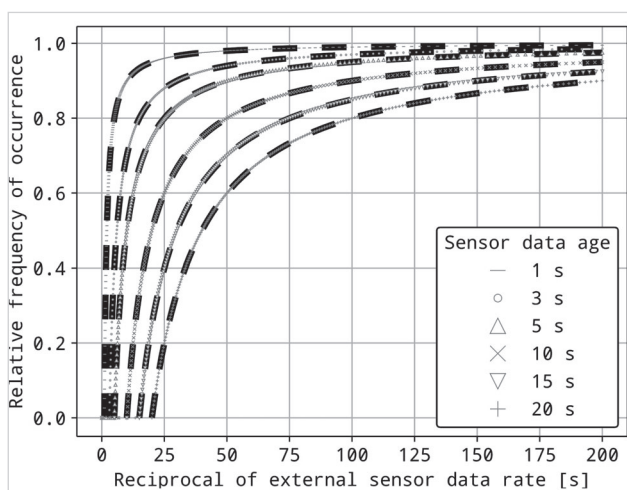

Figure 9 Relative occurrence frequency of AIS default values of the sensor data at different values of the sensor data age superimposed with the best fit curves (dashed lines)
Slika 9. Relativna učestalost pojavljivanja AIS zadanih vrijednosti podataka senzora pri različitim vrijednostima starosti podataka senzora sa superponiranim najbolje prilagođenim krivuljama (isprekidane linije)
Source: Author's research

The computed parameters of the best fit curves used in Equation (15) are listed in Table 5.

Table 5 Parameters determined for the best fit curves of the relative occurrence frequency of AIS default data at different values of the sensor data age
Tablica 5. Parametri određeni za najbolje prilagođene krivulje relativne učestalosti pojavljivanja AIS zadanih podataka pri različitim vrijednostima starosti podataka senzora

Sensor data age g [s]	Computed curvature p	Numerator e^p [s]	Pearson correlation coefficient
1	−0.001	0.999	$1\text{-}10^{-6}$
3	1.098	2.999	$1\text{-}10^{-6}$
5	1.609	4.999	$1\text{-}10^{-6}$
10	2.302	10.000	$1\text{-}10^{-6}$
15	2.708	15.000	$1\text{-}10^{-6}$
20	2.995	20.002	$1\text{-}10^{-6}$

Source: Author's research

It can be seen that the value of e^p stays very close to the value of the sensor data age. Therefore the equation component e^p was replaced by g and the final simplified definition of the function proposed in Equation (10) could be defined by the following equation providing the overall relative occurrence frequency of the AIS default data:

$$F(k) = \begin{cases} 0 & : \quad 0 < k \le g \\ 1 - \dfrac{g}{k} & : \quad g < k \end{cases} \tag{16}$$

$$k \in \mathbb{R}_{>0} \wedge g \in \mathbb{R}_{>0}$$

where k is the interval of the external sensor data provision, and g is the sensor data age allowed by the AIS transponder. The Pearson correlation coefficient was calculated between the values of the occurrence frequency of AIS default data acquired during the simulation and the corresponding values obtained from Equation (16). Its value, which is close to unity indicates that the function F(k) fits the occurrence of the AIS default data very closely.

Equation (16) may be helpful in explaining a possible mechanism behind the results presented in Table 2 or Table 3. For example, if the AIS transponder used during the lab experiment were to broadcast on average 22% of its AIS dynamic position reports with the default value of THDG, then the gyrocompass would have to deliver its data to the AIS transponder at the rate of one measurement per 12.8 seconds. Another possibility of reaching the same relative occurrence frequency of the default value of THDG would require an AIS transponder to be attached to a gyrocompass output of 1 measurement per second and to have the sensor data age set to 0.78 seconds. The relative occurrence frequency of AIS default data defined in Equation (16) applies to both, the AIS broadcasts over VHF and the PP output used by the pilots on board. It has to be emphasised that Equation (16) does not explain the occurrence of AIS default data thoroughly. There may be other factors like software bugs inside the transponder firmware, which may affect the proper operation of AIS equipment, or out of schedule external data interrogations, which may coincide with outdated external sensor data, or radio interference. However, such situations are beyond the scope of this work.

6. DETERMINING THE EXPECTED VALUE OF SENSOR DATA AGE / *Određivanje očekivane vrijednosti starosti podataka senzora*

Another aspect of using Equation (16) is related to the expected value of the sensor data age which defines how current the sensor data can be upon gathering by the AIS transponder. In order to explain the idea of the expected value of the sensor data age, the timeline configuration presented in Figure 7 served as a foundation to create a timing schema for two cases defined in Equation (16). The first configuration, shown in Figure 10, is the example of the expected value of the sensor data age of an external sensor delivering its data every 3 seconds to the AIS transponder which accepts the sensor data age of 5 seconds at most.

The sensor data interval k determines the maximum sensor data age. Since the external sensor data is provided every 3 seconds, there is no way the sensor data could be older than k. The sensor data age g belongs to the internal configuration of the AIS transponder. The data age values have a rectangular distribution within the interval $[0, k]$. Therefore the expected value of the sensor data age, which corresponds to the mean value in case of the rectangular distribution, is given by the following equation:
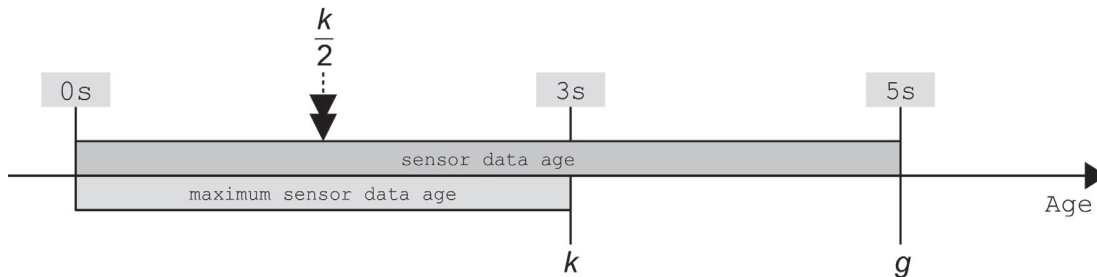
$$E = \frac{k}{2} \qquad (17)$$

The second configuration, shown in Figure 11, involves an external sensor delivering its data every 12 seconds to the AIS transponder which accepts the sensor data age of 5 seconds max.

The data age values follow a rectangular distribution within the intervals $[0, g]$ and $[g, k]$. That is why $\frac{g}{2}$ is the mean value of the external sensor data age when the sensor data is considered to be up to date, whereas $\frac{g+k}{2}$ is the mean value of the external sensor data age when the sensor data is outdated. Due to the existence of AIS default values in the above example, the calculation of the expected value of the sensor data age had to take into account the relative occurrence frequency of AIS default values, and was carried out using the following weighted equation:

$$E = \frac{g}{2} \cdot \frac{g}{k} + \frac{g+k}{2} \cdot \left(1 - \frac{g}{k}\right) = \frac{k}{2} \qquad (18)$$

where $\frac{g}{k}$ is the relative occurrence frequency of AIS non default data, and $1 - \frac{g}{k}$ is the relative occurrence frequency of AIS default data. It can be noticed that the expected value of the sensor data age is independent of the value of the sensor data age g controlled internally by the AIS transponder.

The same relation between the sensor data interval and the expected value of the sensor data age can be observed in the results obtained during the timeline simulation. Figure 12 provides the illustration.

Figure 10 Exemplary timing of the expected value of the sensor data age for *0 <k≤g* producing no AIS default values
Slika 10. Primjeri vremenskog rasporeda očekivane vrijednosti starosti podataka senzora za 0 <k≤g koji ne proizvode zadane vrijednosti AIS-a
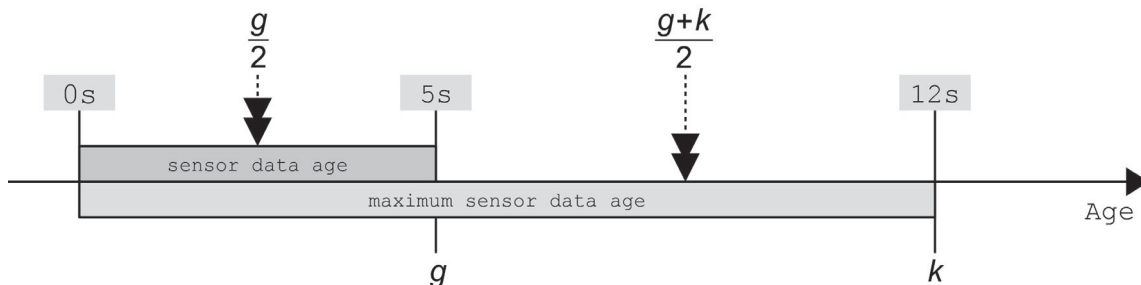Source: Author's research

Figure 11 Exemplary timing of the expected value of the sensor data age for *g < k* producing AIS default values
Slika 11. Primjeri vremenskog rasporeda očekivane vrijednosti starosti podataka senzora za g < k koji proizvode zadane vrijednosti AIS
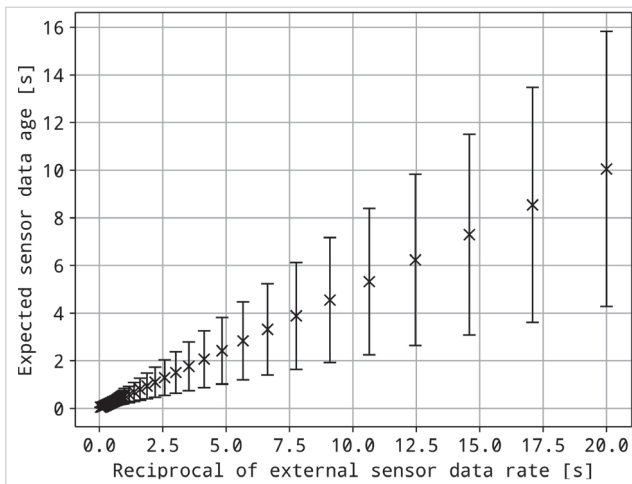Source: Author's research

Figure 12 Relation derived from the timeline simulation between the sensor data rate and the expected sensor data age together with its standard deviation
Slika 12. Odnos izveden simulacijom vremenskog okvira između brzine podataka senzora i očekivane starosti podataka senzora s njegovom standardnom devijacijom
Source: Author's research

The timeline simulation confirmed the analytical approach of calculating the expected value of the sensor data age E. Additionally, the plot in Figure 12 shows error bars of E based on its standard deviation. The lengths of the error bars noticeably follow a linear pattern. Since, as stated above, the data age values follow a rectangular distribution, it was possible to define the coefficient of variation of the expected value of the sensor data age by the following equation:

$$u = \frac{\sigma}{E} = \frac{\frac{k}{\sqrt{12}}}{\frac{k}{2}} = \frac{1}{\sqrt{3}} \approx 0.577 \qquad (19)$$

where σ is the standard deviation of the expected value of the sensor data age. Furthermore, the constant indicates that about 57.7% of all possible values of the sensor data age lie in the range between $\frac{k}{2} - \frac{k}{\sqrt{12}}$ and $\frac{k}{2} + \frac{k}{\sqrt{12}}$.

In view of how outdated the sensor data transmitted by the AIS might become, the ability of the system to broadcast default values of the sensor data gains in importance. For the navigator it is better to realise that certain parameters received from the vessels in proximity remain currently unknown rather than to believe that the AIS data obtained from other vessels reflects the reality. For the pilot using the PP the same principle applies to own vessel under pilotage. The faster a ship, either in terms of her SOG or her ROT, the more up-to-date data describing her navigational state vector is necessary. If, for instance, the AIS transponder on board a high speed craft underway at the speed of 32 knots were able to provide her position data aged 10 seconds on average without resorting to the AIS default

values, she would in fact make 165 meters without up-to-date AIS dynamic position report, but to the pilot observing the PPU and to other AIS participants she would remain stationary for 10 seconds. Although the bridge personnel can still fulfil their duties regardless of the current performance of AIS, it becomes crucial for the computer aided sensor fusion systems, especially those performing the radar target tracking aided by AIS data that the expected value of the sensor data age is kept small.

7. VERIFICATION OF THE RELATION UNDER REAL CONDITIONS / *Provjera odnosa u stvarnim uvjetima*

The relation between the external sensor data rate and the relative occurrence frequency of AIS default data which was formulated in Equation (16) was confronted with the reality by means of data obtained on board a multipurpose underwater operations vessel BALTIC TAUCHER II (IMO 9096387). She was taking part in a GNSS jamming experiment carried out at sea by the German Aerospace Centre (DLR) back in October 2015 [13]. During a twodaylong research voyage, data from all onboard sensors was being gathered by data recording equipment installed on the bridge. One of the data acquisition elements crucial for this verification is presented in Figure 13.

During a period of about 40 hours the sensor data recorder was storing THDG data concurrently at two separate nodes: the data output port of the gyrocompass and the data input port of the AIS transponder dedicated exclusively for datagrams containing the heading information. By timestamp comparison of those two data streams it was possible to assess the percentage of THDG data generated by the gyrocompass was reaching the AIS transponder, and to figure out the average sensor data rate of the THDG provision after traversing through the bridge sensor data network.

Moreover, the analysis of AIS data covering the vessel traffic of October 2015 throughout the whole German territorial waters indicated that 8.1% of dynamic AIS messages transmitted by BALTIC TAUCHER II contained unknown value of THDG parameter. According to the on-board technical documentation the vessel was equipped with FA-150 Class A AIS transponder manufactured by Furuno Electric. The device had the sensor data age set to 10 seconds. After transforming Equation (16) it may be possible to obtain the theoretical interval of THDG sensor data provision on board the vessel:

$$\begin{aligned} [g \quad y] &= [10 \text{ s} \quad 0.081] \\ k &= \frac{g}{1-y} \\ k &\approx 10.881 \text{ s} \end{aligned} \qquad (20)$$

where y is the observed occurrence of transmissions of unknown THDG from the ship, g is the sensor data age allowed by her AIS transponder and k is the computed interval of the THDG sensor
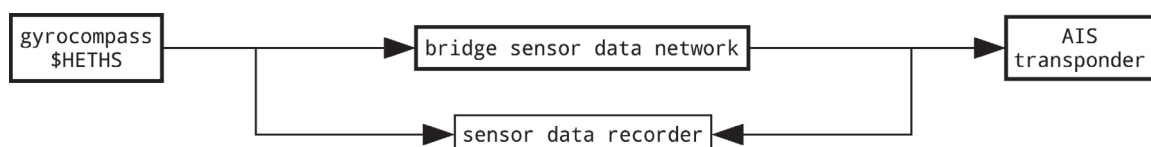

Figure 13 Acquisition of THDG data distributed through the bridge sensor data network on board BALTIC TAUCHER II
Slika 13. Prikupljanje THDG podataka distribuiranih putem mreže podataka senzora mosta na brodu BALTIC TAUCHER II
Source: Author's research

data provision on board.

The sensor data recorder acquired a total of 1298203 records of THS datagrams containing THDG information being output by the gyrocompass at a nominal frequency of 10 Hz. At the AIS transponder data entry the sensor data recorder gathered merely 12527 records of THS datagrams. It means that 99% of THDG data had been lost along the gyrocompass—transponder link. Interestingly, no indication of such a data loss was observed at the radar device which was receiving data from the gyrocompass at the interval of about 0.1 s. The general characteristics of occurrence of THDG output from the gyrocompass is shown in Figure 14.

It is noticeable that the internal software of the gyrocompass on board BALTIC TAUCHER II complied accurately with the specification of maintaining an output data frequency of 10 Hz. The particulars of occurrence of THDG input at the AIS transponder is presented in Figure 15.

Following the analysis of THDG data input at the AIS transponder it can be noticed that the mode of data delivery interval equalled 10.898 seconds. Generally speaking, the device was obtaining up-to-date heading information almost every 11 seconds. The difference between the above value and the length of interval derived from Equation (20) is only 17 milliseconds. Its magnitude is small enough to draw the conclusion that the verification of the relation between the external sensor data rate and the relative occurrence frequency of AIS default data was positive, based on example of the analysis carried out on board BALTIC TAUCHER II. Therefore the findings formulated in Equation (16) may be expected to reflect the real conditions in other cases of vessels contributing to the occurrence of unknown sensor data with AIS transmissions.
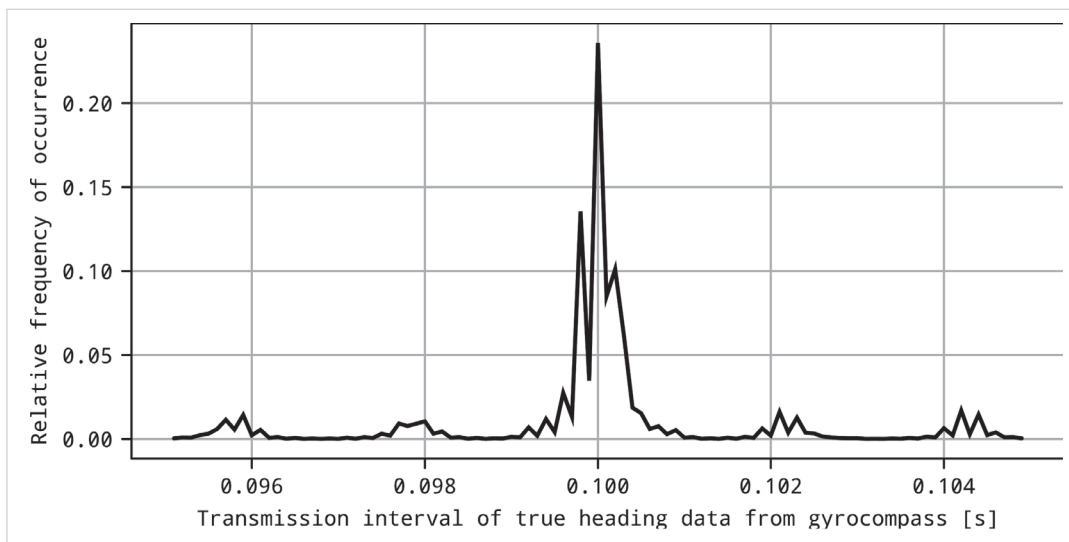

Figure 14 Characteristics of THDG data output interval from the on-board gyrocompass
Slika 14. Značajke intervala izlaza podataka THDG-a iz brodskog žiroskopa
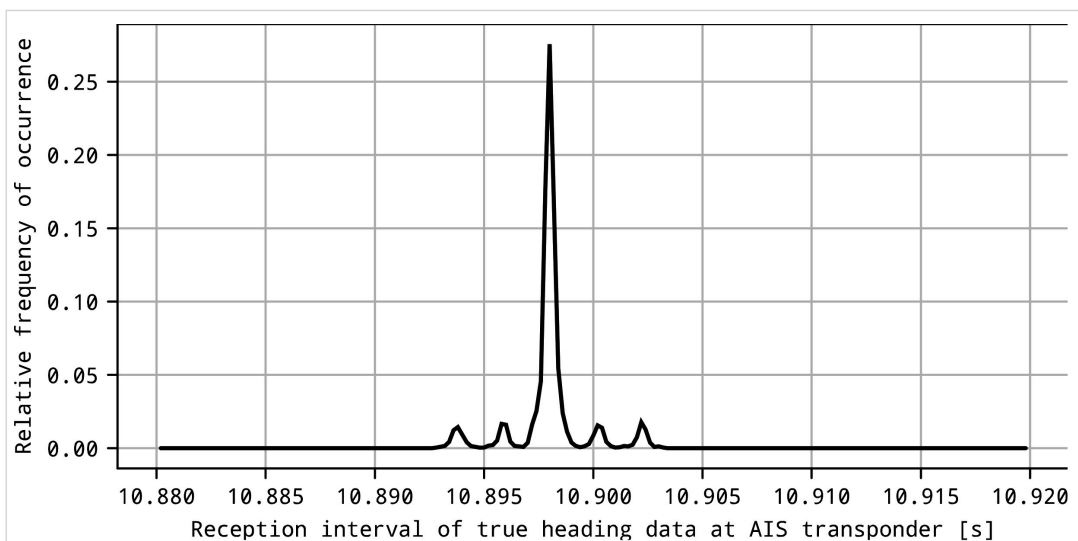
Source: Author's research

Figure 15 Characteristics of THDG data input interval at the on-board AIS transponder
Slika 15. Značajke THDG intervala unosa podataka na brodskom AIS transponderu

Source: Author's research

8. SUMMARY AND OUTLOOK / *Sažetak i perspektiva*

The benefit of the Automatic Identification System lies in its ability to provide a basis for data exchange between various system participants, both on board and ashore. The AIS specification allows the AIS transponders to transmit special so-called default values if certain parameters are unknown prior to data broadcast or data delivery via the PP. Various studies have been conducted on AIS default data. They have shown that the majority of AIS default data within the Class A dynamic position reports originate from external sensor data like THDG of a gyrocompass or ROT of a ROTI. The lab experiment on the Class A AIS transponder has confirmed the solid relation between external sensor data rate and relative occurrence frequency of AIS default data. Additionally, a concept of sensor data age was introduced to explain the choices made upon external sensor data by the AIS transponder. Further investigations, including a comprehensive simulation of the timelines of both an AIS transponder and an external sensor, have helped in finding a simple rational function that is able to describe the interdependence between the interval of external sensor data provision, sensor data age allowed by the AIS transponder and the percentage of AIS default data within the simulated AIS transmissions. In future research, more complex cases of external influences on AIS transponder concerning data interrogations or nonautonomous mode, may be investigated. Additionally, further connectivity and timeline tests between the sensors and the AIS transponder on board other SOLAS vessels might be an interesting subject for future investigation to verify the factual occurrence of unknown AIS data compared to the generic approach in order to evaluate the phenomenon.

9. IMPLICATIONS / *implikacije*

AIS technology is expected to remain one of the most important support for navigators at sea. As the number of transponders has increased worldwide, users of the system must face various problems, which are not only limited to the occurrence of unknown sensor data assessed in this paper. They concern the reception properties of the data, too. AIS signal congestion has already become a reality, especially in busy traffic areas, and administrative measures are taken to reserve additional marine VHF channels to operate exclusively for AIS data sharing. Another challenge related to AIS data reception is the satellite based monitoring of maritime traffic, which requires the application of sophisticated algorithms to decode a large number of overlapping AIS signals. Unfortunately, it is beyond the scope of this paper to cover all these important aspects of AIS infrastructure and operation. Nevertheless, it is important to keep in mind the complexity of AIS, which requires appropriate attention and approach. In the future, AIS might be approved for use together with radar for collision avoidance at sea. Future navigators may notice that only latitude and longitude are needed to plot targets, making other dynamic AIS parameters less important. The results of this paper may be seen as a warning that in case of problems with sensor connectivity on-board vessels. Any AIS parameter could become unknown, which could make the whole dynamic AIS message useless and unreliable from the navigational point of view.

Moreover, it might be worth mentioning, that there are generally two major groups of AIS data users. Firstly, the officers of the watch or VTS operators, who consider only the latest dynamic AIS data message important because it describes the current traffic situation. Any unknown AIS parameters become irrelevant and are disregarded as soon as new AIS message with a complete content arrives. They are not significantly affected by the occurrence of unknown AIS data, but problems related to the phenomenon cannot be completely excluded with respect to the prevailing conditions. Secondly, AIS data users are also all institutions and organisations, mostly shore-based, which are engaged in AIS data acquisition and data mining. All unknown AIS parameters stored in their databases will stay there indefinitely and might affect future data analyses and projects. For this group of AIS users, the less unknown AIS parameters are stored in their systems, the better their work can be done. Therefore all steps taken to limit the occurrence of unknown AIS data are welcome.

REFERENCES / *Literatura*

[1] Banyś, P. Noack, T. and Gewies, S. (2012). Assessment of AIS vessel position report under the aspect of data reliability. Annual of Navigation, 19, 5 16. https://doi.org/10.2478/v10367-012-0001-0

[2] Bertram, V. (2000). Practical Ship Hydrodynamics. Butterworth Heinemann

[3] Chang, S. J. (2004). Development and analysis of AIS applications as an efficient tool for vessel traffic service. Proc. MTTS/IEEE TECHNO OCEAN, 2249 2253. https://doi.org/10.1109/oceans.2004.1406499

[4] Equasis. (2016). The world merchant fleet in 2016. http://www.equasis.org/EquasisWeb/public/PublicStatistic. Accessed on 1 March 2018.

[5] Felski, A. Jaskólski, K. and Banyś, P. (2015). Comprehensive Assessment of Automatic Identification System (AIS) Data Application to Anti-collision Manoeuvring. Journal of Navigation, 68(4), 697 717. https://doi.org/10.1017/s0373463314000897

[6] Hennepe, F. te. Rinaldo, R. Ginesi, A. et al. (2010). Space based Detection of AIS Signals. Advanced Satellite Multimedia Systems Conference, 2045. https://doi.org/10.1109/asms-spsc.2010.5586862

[7] IMO. (2009). SOLAS (Consolidated Edition 2009). London: International Maritime Organization.

[8] ITU R M.1371 5. (2014). Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band. International Telecommunication Union.

[9] Journée, J.M.J. and Massie, W.W. (2001). Offshore Hydromechanics. https://ocw.tudelft.nl/wp-content/uploads/OffshoreHydromechanics_Journee_Massie.pdf. Accessed on 1 March 2018.

[10] Marucco, G. Defina, A. Mikkelsen, T.G. Hodgson, B. (2012). The SafePort project: Concepts and results of trials. IEEE First AESS European Conference on Satellite Telecommunications, 349 353. https://doi.org/10.1109/estel.2012.6400192

[11] Ren, Z. Huang, J. (2010). The Information Reconstruction System of VDR & AIS Data Fusion. International Conference on Anti Counterfeiting, Security and Identification, 181 183. https://doi.org/10.1109/icasid.2010.5551504

[12] Tu, E. Zhang, G. Rachmawati, L. et al. (2018). Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey From Data to Methodology. IEEE Transactions on Intelligent Transportation Systems, 1561. https://doi.org/10.1109/tits.2017.2724551

[13] Ziebold R., Romanovas M., Gewies S. (2016). Experimental Evaluation of the Impact of Jamming on Maritime Navigation. Proceedings of the 29th International Technical Meeting of the Satellite Division of The Institute of Navigation. https://doi.org/10.33012/2016.14810