Integrity Assessment of a Worldwide Maritime Tracking System for a Geospatial Analysis at Sea

Clément Iphar, Aldo Napoli, Cyril Ray

Integrity Assessment of a Worldwide Maritime Tracking System for a Geospatial Risk Analysis at Sea

Clément Iphar
MINES ParisTech
PSL Research University, CRC
Rue Claude Daunesse
Sophia Antipolis, France
clement.iphar@mines-paristech.fr

Aldo Napoli
MINES ParisTech
PSL Research University, CRC
Rue Claude Daunesse
Sophia Antipolis, France
aldo.napoli@mines-paristech.fr

Cyril Ray
IRENav
Ecole Navale
Brest, France
cyril.ray@ecole-navale.fr

Abstract

The Automatic Identification System (AIS) is an electronic system set on board vessels, and transmits its location, amongst many other data. As messages are sent and received by vessel and coastal stations within the radio horizon, it enables a better understanding of the surroundings for vessel and coastal states. However, this system has weaknesses, as errors in data and falsification such as identity theft or disappearances have been demonstrated. This paper presents a methodology based on the notion of data integrity, in which data coming from AIS is processed in different ways in order to produce confidence coefficient on the genuineness and truthfulness of the data within. Such a study is done using an architecture built around a relational database using geospatial predicates. As AIS falsification brings maritime risks, a geospatial analysis of risks is proposed, based on both the AIS integrity assessment and geographic features around the studied vessel, leading to a better response from competent authorities in case of need, an overall improvement of maritime situational awareness and an enhancement of the safety and security of maritime navigation.

Keywords: Automatic Identification System; data falsification; integrity assessment; geospatial risk analysis

1 Introduction

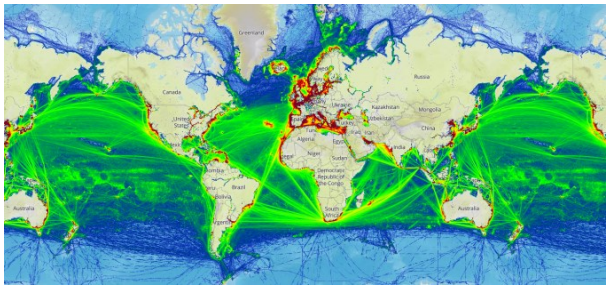
The ocean, crossroads of international issues, is facing a growing pressure of human activities. In some domains such as goods transportation or energy transportation, up to 90% of the world traffic is done by sea. Fishing, sailing and cruising are amongst the domains of maritime activities, generating an important and ever-increasing traffic. As the number of vessels using these routes, entering or leaving ports and crossing heavily occupied areas increases, navigation difficulties arise. In addition, the high number of sailing vessels, each one exhibiting its own movement with its own objective, may lead to conflicting situations, which may evolve into an hazardous situation.

In order to enhance the security and safety of navigation, several systems have been put in place by coastal states or international organizations. Those systems are either active (such as Vessel Management System) or passive (such as radar). Their purpose is to localize vessels and give coastal states and vessels at sea a genuine overview or the surroundings. One of those systems is the Automatic Identification System (AIS), in which vessels broadcast localization messages to other vessels and coastal stations in their neighbourhoods. However this system undergoes some problems such as errors in data and falsification cases.

As AIS data is used by competent authorities as a decision support tool, the study of the genuineness of information sent in the messages is of paramount importance for the assessment of geospatial risks of vessels, thus increasing the level of maritime situational awareness, for which the scope of research and interest is large (Claramunt *et al.*, 2017).

Figure 1 presents the density of the worldwide vessel traffic, using AIS positions acquired by satellite.

Figure 1: Worldwide AIS Traffic Data in 2015 (marine traffic)



In this article, we present a method for the assessment of the quality of AIS messages and the evaluation of geospatial risks associated. In section 2, the AIS is presented along with its geospatial nature and its weaknesses. Then section 3 presents the method for genuineness assessment of the message. Next, section 4 focuses on the spatiality of risks associated with AIS falsification, followed by a conclusion.

2 The Automatic Identification System

2.1 A tool for mariners

The introduction of the Automatic Identification System was decided in 2000 by the International Maritime Organization, and implemented in the 2002 version of the Safety Of Life At Sea convention. All the vessels are not concerned with this regulation, as it is stated in the convention that “All ships of

300 gross tonnage and upwards engaged on international voyages and cargo ships of 500 gross tonnage and upwards not engaged on international voyages and passenger ships irrespective of size shall be fitted with an automatic identification system" (IMO, 2004).

Messages are broadcasted from and to vessels and coastal stations, at first within the radio horizon range, which is circa 40 nautical miles. But in the 2000's, the development of both satellite technology and the worldwide spread of the Internet permitted messages to be received by satellites, centralized in servers and displayed on dedicated websites.

Depending on the type of transmission, 27 different kinds of messages (Tunaley, 2013) are defined, of which 11 are geolocalized, accounting for circa 90% of the total number of messages. The most sent message is the scheduled position report message, accounting for circa 65% of all the messages. Standard scheduled position reports are done every 2 to 12 seconds when the vessel moves, the frequency depending on its speed, and every 3 minutes at anchor. This high rate of transmission implies a high number of messages, for instance, in the waters of the European Union, there are circa 10,000 unique vessels per day and about 100,000,000 messages per year.

2.2 A geospatial system

As AIS messages are primarily localization messages, therefore their spatial component are of foremost importance. In a classical position report message, the purely spatial fields are the latitude and the longitude, while the spatial-linked fields are the rate of turn, the speed over ground, the course over ground and the true heading.

In message number 1, the total number of bits allocated for latitude and longitude is quite important (27 and 28, respectively), thus the elementary unit of the measure is the one ten thousandth of minute of arc. For all of the longitudes and for all the latitudes close to the equator, this basic unit is worth circa 20cm on the surface of the Earth, and progressively decreases as latitudes grow. As the GNSS computation is mainly performed when the vessel is moving, the accuracy of this computation shall be at least of the order of magnitude of the meter, thus the size of the basic unit shall not be the limiting element of the study.

In message number 5, the destination of the vessel has a dedicated textual field of 20 6-bits ASCII characters. The drawback to this voyage-based spatial information is the fact that it may be filled inappropriately across a wide proportion of messages.

Other data fields including the user ID and the time stamp are not spatial data but will be useful in our study. The user ID will be used to define trajectories by tracking one single user over time while the time stamp will be used to reconstruct the trajectory and transfer spatial data into spatiotemporal data.

2.3 The weaknesses of AIS

Three major cases of bad data quality can be distinguished: the errors (when false data is non-deliberately broadcast), the falsifications (when false data is deliberately broadcast) and the spoofing (when data is created or modified and broadcast by an outsider) (Ray *et al.*, 2015). Data contained in AIS messages can be erroneous, falsified or spoofed for several reasons: there is no strong verification of the transmission, the transmission is done using a non-secured channel, some data might not be well known by the crew or the crew may want to hide some data from other people's knowledge. Those operations modify and handicap the understanding of the maritime traffic.

The errors, by nature unintentional, can be caused by transponder deficiency, a wrong input of manual data, an input of manual data of poor quality, erroneous pieces of information that come from external sensors, and can have an impact on the name of the vessel, its physical characteristics, the position or the destination for instance. Those data can then be false, incomplete, impossible according to the norm or impossible according to the physics (for instance, a latitude field value shall be inferior to 90°). According to (Harati-Mokhtari *et al.*, 2007), circa 50% of the messages contain erroneous data.

A falsification is the fact to voluntarily degrade a message by the modification of a genuine value by a false value, or by stopping the broadcast of messages, made in order to mislead the outer world. Identity theft, the disappearances (Windward, 2014), the broadcast of false GNSS coordinates or the statement of a wrong activity (Katsilieris *et al.*, 2013) are types of falsification. According to (Harati-Mokhtari *et al.*, 2007), circa 1% of the vessels broadcast falsified data.

The spoofing of messages is done by an external actor by the creation ex nihilo of false messages and their broadcast on the AIS frequencies (Balduzzi *et al.*, 2014). Those spoofing activities are done to mislead both the outer world and the crews at sea, by the creation of ghost vessels, of false closest point of approach trigger, a false emergency message or even a false cape (in the case of a spoofed vessel).

3 Assessment of AIS messages

3.1 Data structure

Each one of the 27 different kinds of messages has its own standardized outline defined by the International Telecommunications Union (ITU, 2014) under the form of successive data fields, in which each field is allocated a given number of bits. Throughout all the fields of the 27 kinds of messages, the information that lies within these fields can take six different forms: 1. Boolean, 2. text, number representing: 3. a physical quantity; 4. a choice in a given list; 5. an id number, or 6. date. From one message to another, the content varies significantly: from data fields that represent position, speed and cape in position reports to data fields that represent name, dimensions and destination of the vessel in the static information message. The kind of field data varies in accordance to message type. The meaning of the data field values within the messages are also defined in an unambiguous way in the technical specifications (ITU, 2014).

3.2 Integrity assessment

As stated in (Iphar *et al.*, 2015), integrity is the most important of all data quality dimensions when it comes to the veracity assessment of AIS messages. The method we propose is based on the integrity of AIS data at several levels. The first level consists in the assessment of each single data field, taken apart from the others, which consists of checking whether the

field value is consistent with the possible field values given by the technical specifications. The second level consists of assessing the integrity of data within a single message, thus apart from all other messages, and check if there is any discording data between the fields. The third level is an assessment between messages of the same type (for field value evolution for instance) and the fourth level is an assessment between the fields values of different kind of messages.

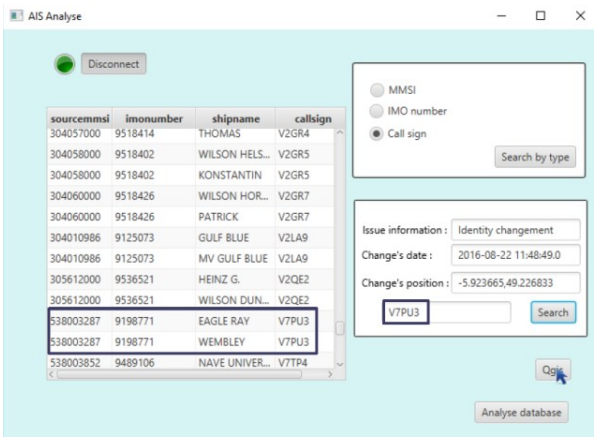
The following is an example of a simple assessment for the first level: "Is latitude between -90 and 90 degrees?", for the second level: "Are the speed and the rate of turn compliant?", for the third level: "Is the evolution of the position relevant?" and for the fourth level: "Is the fact that message 11 has been sent from a given coastal station to a given vessel consistent with message 1 position values of this vessel?".

In our method, the first and second levels are performed within the same message and can be accomplished on-the-fly, whereas the third and fourth levels are performed between multiple messages, and require database queries. We established a list of more than six hundred items, with each item corresponding to a single data integrity check. In parallel, a nomenclature for unique identification of assessment items has been performed.

According to the message we have or to the situation we want to assess, a subset of those items is selected and the data integrity assessment takes place with the selected items and the selected messages. A confidence coefficient is then computed taking into consideration a weight factor (how important is the item in the assessment) and an assessment factor (how integer are the data assessed in this particular item).

The experimental validation of this method is being performed in ongoing work and uses data we have been collecting by antennas located in the Brest harbour. Additional data will be used in future works. Figure 2 is an example of AIS analyse we performed, and where a vessel changed its identity during travel.

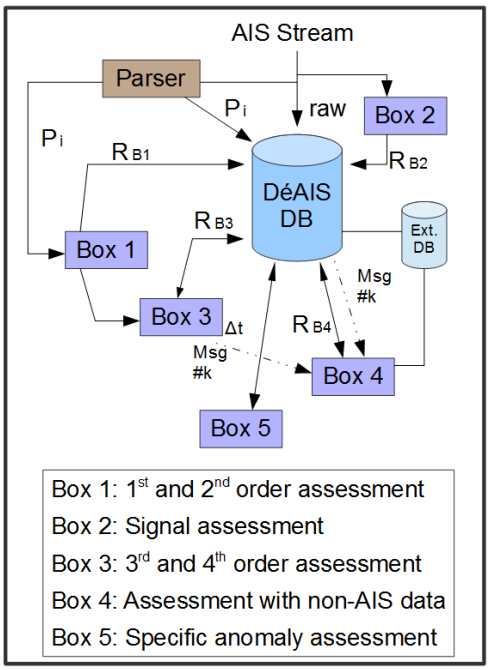
Figure 2: A Detected Identity Change



3.3 Software architecture

The architecture we propose is built around a Postgresql/Postgis database in which we store the messages, the computed coefficients, and all other information available that might be relevant in our study. As it is shown in Figure 3, when a message is received, the signal is processed (box 2) and the signal parameters extracted, they will be useful for the identification of the sender, as shown in (Alincourt et al., 2016).

Figure 3: Proposed architecture



The signal is then parsed, and the parameters of those messages, that is the data within the data fields, are processed using the method for integrity assessment presented in section 3.2. This method takes place in box 1 (for first and second levels) and in box 3 (for third and fourth levels).

In addition, a fourth box is proposed, in which a similar confidence coefficient is computed with information coming from other sources than AIS: for instance a list of items can be put in place with a given fleet register, and we can then check whether or not the data in AIS messages is compliant with data in this fleet register. The Postgis extension allows us to assess geospatial predicates on the AIS messages, for instance to know if some vessel has crossed a maritime exclusion area, or when a vessel pretends to be located on land. The additional box 5 we propose is intended to gather the specific anomaly assessment which are the scenarios chosen for study, such as the appearance of a vessel on land, or a voluntary switched-off of the AIS.

4 The risks associated with AIS falsification

4.1 The spatiality of risks at sea

In the assessment of maritime risks, the degree of gravity associated with every type of risk will vary with respect to the location of this risk. The types of problems encountered by vessel range from sick people on board to explosion, with collision, loss of goods or illegal trade. Those issues can be

encountered in a port, near the coasts or on a busy maritime route. For instance an oil spill will be more serious near the coasts than in the high sea; a fallen container will be more hazardous on a busy maritime route; a fire in a vessel will be more hazardous in a poorly dense area than on a busy route, as there will be more vessels in the neighbourhood to rescue the crew.

4.2 Towards a comprehension of geospatial risk evolution

Once the integrity assessment of AIS messages is done, a use case analysis is performed, in order to consider the likelihood of one vessel undergoing a given situation. Use cases such as identity theft, disappearances or GNSS spoofing have been discriminated. A geospatial assessment of the risks, enhanced by an evaluation on the risk of evolution of this threat can then be performed, leading to a reliable assessment of spatiotemporal risks and thus allowing a better comprehension of geospatial risk and geospatial risk evolution. Once this assessment is done, all relevant information can then be delivered to competent authorities (which will vary according to the case), so that they will be able to take proportionate actions to mitigate the risk.

5 Conclusion

This article presents the AIS, a worldwide maritime localization system, its limits, and proposes a methodology for falsification discovery in the messages sent by this system. This is based on the notion of data integrity, and exposes the implementation of this methodology through a postgres/postgis relational database. By assigning a confidence coefficient to each message and to each user, within the message itself or with respect to a given use case, relevant information can be handed over to competent authorities to mitigate the risks associated with maritime navigation in general, and to the activities linked to the falsification of the AIS, with the overall purpose of enhancing the safety and security of maritime navigation.

Acknowledgments

The research presented in this paper is supported by The French National Research Agency (ANR) and co-funded by DGA under reference ANR-14-CE28-0028, in the frame of the DéAIS project, labelled by French clusters Pôle Mer Bretagne Atlantique and Pôle Mer Méditerranée.

References

Alincourt E., Ray C., Ricordel P.M., Dare-Emzivat D. and Boudraa, A. (2016) Methodology for AIS Signature Identification through Magnitude and Temporal Characterization. In: proceedings of OCEANS 2016; 10-13 April 2016

Balduzzi M., Pasta A. and Wilhoit K. (2014) A security evaluation of AIS automated identification system. In: proceedings of ACSAC 2014; 7-12 December 2014

Claramunt C., Ray C., Salmon L., Camossi E., Hadzagic M., Jousselme A-L., Andrienko G., Andrienko N., Theodoridis Y. and Vouros G. (2017) Maritime Data Integration and Analysis: Recent Progress and Research Challenges. In: proceedings of the EDBT conference, 192-197, 21-24 March 2017

Harati-Mokhari A., Wall A., Brooks P. and Wang J. (2007) Automatic Identification System (AIS): Data Reliability and Human Error Implications. Journal of Navigation; 60(3),373-389

IMO (2004) International Maritime Organization, International convention for the safety of life at sea, 2004

Iphar C., Napoli A. and Ray C. (2015) Data quality assessment for maritime situation awareness. In: Proceedings of the 9th ISPRS International Symposium on Spatial Data Quality (ISSDQ 2015), Volume II-3/W5, pages 291-296, 29-30 September 2015

ITU (2014) International Telecommunication Union Recommendation ITU-R M.1371-5

Kastilieris F., Braca P. and Coraluppi S. (2013) Detection of malicious AIS position spoofing by exploiting radar information. In: proceedings of the 16th international conference on information fusion; pp. 1196-1203, 9-12 July 2013

Ray C., Iphar C., Napoli A., Gallen R. and Bouju A. (2015) DeAIS project: Detection of AIS Spoofing and Resulting Risks. In: proceedings of OCEANS 2015; 18-21 May 2015

Tunaley J.K.E. (2013) Utility of Various AIS Messages for Maritime Awareness. In: Proceedings of the 9th Advanced SAR Workshop. 15-18 October 2015

Windward (2014) AIS data on the high seas: an analysis of the magnitude and implications of growing data manipulation at sea. Windward Company; October 20th 2014