Data Quality Assessment for Maritime Situation Awareness
Clément Iphar, Aldo Napoli, Cyril Ray

To cite this version:

HAL Id: hal-01269684
https://hal-mines-paristech.archives-ouvertes.fr/hal-01269684
Submitted on 5 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
DATA QUALITY ASSESSMENT FOR MARITIME SITUATION AWARENESS

C. Iphar a,*, A. Napoli b, C. Ray b

a Centre for Research on Risks and Crises, MINES ParisTech, PSL – Research University - Sophia Antipolis, France - (clement.iphar, aldo.napoli)@mines-paristech.fr
b Naval Academy Research Institute (IRENav), Brest, France – cyril.ray@ecole-navale.fr

KEY WORDS: maritime positioning data, data quality, data falsification

ABSTRACT:

The Automatic Identification System (AIS) initially designed to ensure maritime security through continuous position reports has been progressively used for many extended objectives. In particular it supports a global monitoring of the maritime domain for various purposes like safety and security but also traffic management, logistics or protection of strategic areas, etc. In this monitoring, data errors, misuse, irregular behaviours at sea, malfeasance mechanisms and bad navigation practices have inevitably emerged either by inattentiveness or voluntary actions in order to circumvent, alter or exploit such a system in the interests of offenders. This paper introduces the AIS system and presents vulnerabilities and data quality assessment for decision making in maritime situational awareness cases. The principles of a novel methodological approach for modelling, analysing and detecting these data errors and falsification are introduced.

1. INTRODUCTION

Although being a place full of wildlife richness, the sea is also a major place for human activities. As a matter of fact, many human activities, such as fishing, transportation of goods, cruising and sailing take place at sea, all having an impact, more or less important, on the worldwide economy. All around the world, fishing activities enable millions of fishermen to live from this activity, goods transportation by sea represent 90% of total goods transportation and cruising generates more than 315 thousand jobs in Europe as of 2013, for a global number of more than 20 million passengers (MEDDE, 2014). The sea is an important place for energy transportation, with for instance gas tankers, the importance of which being ever increasing (Napoli, 2014). The ever increasing maritime traffic (MEDDE, 2014) leads to more and more crowded areas. The risks in coastal areas near the harbors are numerous, as it is a place where an important amount of vessels are gathered, each one of them trying to optimize its trajectory, leading to conflicting behaviors.

Rules were put in place in order to prevent collisions at sea, which endanger the crews, can be hazardous for the environment (e.g. in case of oil spill), for the local economy and lead to important earnings and money losses for the involved people or companies.

Surveillance systems were also put in place by the states willing to improve safety and security at sea off their coasts, and identify hazardous areas and suspicious ships, which could have illegal interests. One of them is the AIS system, which uses VHF radio frequency to transmit messages to other fitted out vessels and shore-based stations. It is then possible to gather all those real-time positions, which is done by the authorities (harbors, Vessel Traffic Services) or Internet dedicated website (such as marinetraffic.com). Although the different surveillance systems are meant to be complementary, the AIS system has a wider range of detection than radar, and enables to get rid of the potential masks due to the land in coastal areas. However, it is a system based on signals sent by a device on board the vessel, the use of it being at the crew’s discretion. The messages sent can be falsified, the signal can be spoofed and errors can be committed during the filling in of some information in the system. A mathematical processing of data could bring information on the reliability and integrity of data by providing a messages-based coefficient by methods inspired by the Information Theory. Such a coefficient would rate information. People in Vessel Traffic Services (VTSs) and on-board vessels can be led to take decision on the basis on available information they have. The knowledge of the quality of information they handle is of paramount importance in the situational awareness in which the decisions are taken. This article presents a contextualization of AIS falsifications and the methodological approach in the scope of the assessment of the quality of data, however, quantitative assessment of data, data quality standards and results of any implementation are not the purpose of this paper.

2. THE AIS SYSTEM

2.1 The genesis of the system

The Automatic Identification System (AIS) was put in place by the Safety Of Life At Sea (SOLAS) convention, in its 2002 version (IMO, 2004). This convention, initiated in 1914 by the sinking of the RMS Titanic two years beforehand, has the purpose of defining the minimal requirements to which every vessel from signatory countries should comply with. The SOLAS convention deals with a lot of subjects, ranging from the construction of vessels to the way radio-communications shall be done. One of those subjects is the security and safety of
maritime navigation, and the AIS system was created in this scope, in order to provide real-time spatiotemporal positioning to the vessels.

2.2 The concerned vessels

All the vessels from the signatory states are not concerned with this system. Indeed, the SOLAS convention, in its fifth chapter, nineteenth rule, paragraph 2.4, states 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). The deadline for setting up the system on-board ships depends on the ship itself: from the 1st of July 2002 for new ships and ships engaged in international voyages at the soonest to the 1st of July 2008 for some ships not engaged in international voyages at the latest.

2.3 The transmission mode of AIS

The AIS messages use the Very High Frequency (VHF) bandwidth, especially the dedicated Marine VHF bandwidth (4 distinct bandwidths for a total range of 2.225MHz). Two worldwide wavelengths are dedicated to the transmission of AIS messages: 161.975MHz and 162.025MHz. In order to transmit and receive signals, dedicated devices have been built since the introduction of this system. Four kinds of devices can be distinguished:

- Class A transponders: they equip the vessels for which the use of AIS is compulsory. They can emit and receive simultaneously on both channels.
- Class B transponders: in general, they equip the vessels for which the use of AIS is not mandatory but which want to be equipped.
- Multi-channel receivers: those devices can only receive messages, on both channels simultaneously.
- Radio scanner receivers: those devices can only receive messages, switching from one channel to the other one when wanted.

Two kinds of transmission are possible: terrestrial, and by satellite. At first, the system was intended to be only terrestrial, for transmission from one vessel to another or between a vessel and a coastal beacon, in a range of distance limited by the curvature of the Earth (up to circa 40 nautical miles in the best conditions (NASA, 2012)). Then, the development of ad hoc satellites enabled the harnessing of the signal even far from the coast line. The satellite uploads and stores up the received messages then downloads all the stored messages when it meets a coastal beacon. In the last years, with the development of the Internet, such data can be displayed in almost real-time on dedicated websites, and led to the knowledge of behaviors at sea. Where formerly ships literally disappeared beyond the skyline, they can now be tracked all around the world by every person having an Internet access. In a standard position report of class A transponder, the geographical information in the messages is sent every 2 to 12 seconds when the vessel moves (depending on the speed of the vessel, the higher the speed, the closer the messages). When the ship is anchored, the messages are sent only every three minutes (Redoutey et al., 2008). In a mean day, 400 000 messages are sent from more than 22 000 vessels (NASA, 2012). The main data given in the message are (ITU, 2014):

- The number of the message
- A unique identification number of the user
- The navigational status
- The turning rate
- The speed
- The longitude
- The latitude
- The course
- The true heading
- The time
- The communication of the date of the next message

2.5 How the data flow is handled

As formerly stated, an AIS message includes the time of the next communication. This means that the system is self-governing for the attribution of the time slots. The system selects an empty slot in which it wants to communicate, it chooses it on the basis of the previously received messages in order to prevent to use an already booked slot and prevents all others ships from using the slot it books for its next communication (ITU, 2014). This protocol is call SOTDMA, which stands for Self Organized Time Division Multiple Access, and it is especially designed for communication networks at sea. In case of conflict (when a meeting with a new station which has booked the same slot), the stations can modify their own allocations. Other protocols exist, for class B transponders and some messages having priority.
2.6 The uses of AIS

The system has several uses, albeit initially designed for safety and security purposes, some people use the system another way.

The AIS can be used for the prevention of boarding (alarm triggering when a small closest point of approach is computed), investigation in case of accident, control of fishing fleets, cargo fleets, global traffic, traffic in specific hazardous areas, maritime safety (for a state), aid to navigation or search and rescue operations.

3. A VULNERABLE SYSTEM

3.1 A system for decision making

As stated before, the AIS system is used for decision making in maritime situations, on-board or ashore. Those decisions can be important as they can bring into consideration up to the lives of crew members and the wholesomeness of the environment. However, the AIS system is characterized by its neither authenticated nor encrypted nature. A part of the information (the static information, the one which does not evolve with time, such as the name of the ship or its length) is entered manually at the initialization of the device, and this information is not controlled by any authority. Moreover, some dynamic information is provided continuously by the GNSS antenna paired up with the AIS device. It is the combination of all this information that would help the user in making a decision.

As it is not possible to assess if the data is true, a big issue of data quality is arisen. As the AIS is a self-supplying system, some actors could want to put into the system deliberately erroneous data in order to hide their activities. Other actors could make unintentional errors when they fill in their device, and others could want to use the lack of protection of the signal to spoof it and mislead mariners (Balduzzi et al., 2014a). The purpose of such an operation is to lead to mistaken decision from decision-makers. In this scope, the assessment of the fact to know whether or not the message is genuine is closely linked to the assessment of the quality of the received data, and to the eventual decision taken.

3.2 Errors in AIS data

Some data that AIS messages contain are entered manually at the system initialization (first use). A study of errors in those data is done by (Harati-Mokhtari et al., 2007). These errors are not intentional, they can be done by underestimating the importance of a proper fulfillment of the fields or by ignorance of the way the system works.

For the unique identification number of the ship (also called MMSI number), 2% appear to be erroneous, as they follow impossible patterns (e.g. the first three numbers must represent the state of the ship) or characteristic numbers such as “0”, “1”, “999999999” or “1193046” which is supposed to be a default value of the field for a manufacturer.

The type of the vessel is sometimes unclear. As 6% have no specified type, 3% are only described as “vessels”. The type can also be unclear for the user, as three identical boats used as service boats in a harbor were set with three different types (Harati-Mokhtari et al., 2007).

The name of the vessel is not properly filled in in 0.5% of the cases, as the name is either missing or too long: only 20 letters are allocated. Problems linked to geographical string names are developed in (Mazzola et al., 2013).

For the physical characteristics of the vessel, between the available databases and the AIS messages, 47% of the vessels have discrepancies in length and 18% in beam. As for the draft, 17% are “0” values and 14% are greater than the length of the vessel.

In the dynamic data, errors are also widespread, as stated by (Harati-Mokhtari et al., 2007); as for the position, 1% of vessels have a latitude greater than 90° or a longitude greater than 180°; at least 30% have an erroneous navigation status, e.g. say they are sailing while having a discordant speed; the destination is unclear in about half the cases. For instance, people using a vague name, a country, an abbreviation, a blank space….

On the website marinetraffic.com where a part of the international traffic is displayed, some cases of erroneous destination fields are shown. Six examples of such problematic data are: “ATLANTIC OCEAN”, where the destination is too vague, “HOME”, where the destination is perhaps true but not precise, “FOS SUR MER”, where the vessel clearly seems to come from this French city, and not go to it, “CH 16 FOR DESTINATION”, where the pilot asks for a communication in the maritime channel number 16, “TBA”, where the destination seems not to be known yet and “ANYWHERE BUT HERE”, where the pilot seems to joke without having bad intentions.

Moreover, the system in itself can fail in transmitting information. Some transponders fail to reach all the requirements set by the ITU, and some ships display large blank areas. This missing data, as shown in (Lecornu et al., 2013), weaken the exploitation of AIS data by decreasing the reliability, but does not prevent it (as a meaningful statistical study is needed in order to judge the quality of data).

3.3 Intentional falsification of AIS data

Intentional falsification of the AIS signal is done by the crews on board the ships in order to modify or stop the message they send, in the very particular purpose of misleading the outside world.

Identity theft also exists in the maritime domain (Windward, 2014). It corresponds to the fact to navigate with a Maritime Mobile Service Identity (MMSI) number which is not the real one, allocated and internationally recognized, but with the one of another vessel that actually exists somewhere else. Hundreds of ships are disguised this way. As the MMSI number changes, there is no way to assess a priori whether the vessel one is looking at is the right one. As stated in (gCaptain, 2012) and (The Maritime Executive, 2012), Iran used to falsify the MMSI number of some ships in order to trade with Syria, then under embargo. The Iranian ship Millionaire took the identity of the Syrian ship Lady Rasha. At some time, there were two declared Lady Rasha, one in the Mediterranean Sea and one in the Indian Ocean.

Destination masking is also sometimes a falsification (Windward, 2014). As sometimes it can be considered as an error, some other cases are about a voluntary deficiency of
information, done in order to sidestep the overview of the global ships flows.

Disappearances are also a kind of falsification, as ships turn off their AIS transponder in order to hide some of their activities, such as fishing in an unauthorized area, or trade illegal goods (Katsilieris et al., 2013) with other ships or on coasts.

3.4 Spoofing of AIS data by surroundings actors

Spoofing of AIS data consists of an action made by an external actor in order to mislead the crew of the ship and the outside world on the behavior of the proper ship.

One of the trickery is a false closest point of approach alert (Balduzzi et al., 2014b). An alert is triggered and the vessel is forced to change its heading and perhaps be guided to hazardous places in order to avoid a hypothetical boarding by a ghost vessel.

Moreover, (Balduzzi et al., 2014b) implemented a spoofing program imitating a fake ship which is following a spatiotemporal path which made it spelling a word in the Mediterranean Sea. It was possible to see it displayed on the website marinetraffic.com.

Another similar trickery is to simulate a search and rescue alert, forcing all the surrounding vessels to change their course. The creation of a false vessel, or of a false aid to navigation can also be done, in order to provide the vessel with erroneous data and influence its behavior. Availability disruptions (Balduzzi et al., 2014b) are another kind of spoofing, preventing the vessel to transmit its data by several means.

An example of spoofing is given in (Bhatti and Humphreys, 2015), with a pirate who emits a GNSS false signal with a higher power than the genuine one, covering it. The system is then misled on its whereabouts, transmits a false AIS and forces the pilot or the autopilot to maneuver in order to return towards the cape he or it believes to be the right one, but actually the vessel moves away its real destination.

Serious consequences of these spoofing events can be foreseen in case malicious actors, such as the terrorists, want to cause a catastrophic incident. Back in 2011, the port of Antwerp underwent an attack from a drug cartel (Le Marin, 2014a). Generally speaking, the merchant navy and the ports do not follow quickly enough the new cybercrime breakthroughs (Le Marin, 2014b).

4. DATA QUALITY ASSESSMENT FOR DECISION MAKING

4.1 Using quality of data for AIS integrity measurement

As the purpose is to detect vessels that are undergoing attacks or that are emitting false messages on purpose, a means for the assessment of such vessels must be developed. As information comes from messages, it is possible to treat and analyze every message independently from the others, a group of messages as a whole, or a single message with respect to a group of messages.

Data contained in AIS message are various, and the messages themselves are numerous. Such an analysis would bring a numerical value on a message, a group of messages or a message with respect to a group of messages.

Inspired by the Information Theory developed by Shannon in (Shannon, 1948), a mathematical processing of data can be done in order to assess the usefulness of a message, its reliability and its integrity.

The integrity of data is of major importance, and can be assessed for a single message (integrity of the information within), for a group of messages (integrity between the data between the messages) and for a message with respect to a group of messages.

A coefficient for data integrity could then be computed, and with a specified threshold, detect the dubious messages on the basis of data quality, with many parameters taken into account, at several levels: internal to the system, a single message, a group of messages, or external to the system. Such a coefficient would be an indicator for data quality assessment.

4.2 The quality of data in the case of AIS messages

A minority of users actually falsifies their data, but they take a considerable importance in the monitoring of maritime awareness situation as they are the ones that profit from their falsification. Most of the data are sent in goodwill and in good intentions, but a certain amount is false, and sometimes, in case of spoofing, a vessel emits or receives messages that are not true, independently from the will of the crew. These three cases put the stress on the importance of the assessment of the quality of the data of such an unauthentic and unsecured system.

The lack of reliability of the system could guide people into a lack of belief in the truthiness of data, and thus of the whole system. The blacklists can be weakened as undesirable ships hide some activities, and the global view of the world traffic is distorted, leading to wrong decisions.

The purpose of such an analysis would also be to differentiate intentional falsifications from unintentional errors, by investigating the fields likely to receive erroneous or falsified data, and distinguish (in case of dubious data) the kind of pieces of information likely to be an error from the pieces of information likely to be a falsification. This can be addressed by a study on the field itself and its likeliness to be subject to a will of falsification.

4.3 The people using AIS messages in their activities

Fishermen use AIS in order to locate themselves, locate other ships in the area and know, thanks to their ECDIS (Electronic Chart Display and Information System), if they are in a legal fishing area. However, only ships longer than 20 meters are obliged to be equipped by AIS.

Mariners, especially cargo transportation ship crews, use AIS for several purposes. Safety and security of the navigation are done, but also surveillance of surroundings vessels, at the time of arrival at a harbor can be important for the availability of an unloading spot. AIS can also be used to ask a question to another ship or declare oneself in danger.

Law enforcement authorities use the AIS in several ways, taking it away from its initial purpose. The surveillance of fishing areas and of coastal ship behaviors is made, as well as the
surveillance of traffic separation schemes (TSS) set by the IMO. They look for typical behaviors that can help us to find out AIS manipulators. Methods to compute the theoretical position of a ship, in order to improve maritime awareness, when all AIS data are not available are developed (Vanneschi et al., 2015).

Ship-owners, who want to follow the courses of their ships, thanks to the worldwide coverage this system has, can have a continuous global overview of their fleet.

4.4 Quality elements taken into account in a quality assessment of AIS messages

For the knowledge of the maritime situation and a better global maritime awareness, the quality of the data sent by the AIS must be improved. As this system is not reliable and used by many people for various purposes, the awareness of the users on the proper nature of the system shall be clear.

The integrity of data must be investigated, as some information is clearly impossible (out framed GNSS coordinates for instance) or physically incompatible (two vessels with the same number at the same time, a speed that does not match with the type of the vessel). Such a determination of the reliability of information can be done using a scale of belief, disbelief and uncertainty based on a statistical study (Ceolin et al., 2013).

The trust in a user will depend on the past behavior of this user, as the opinion about the source is a key part of trust (Hertzum et al., 2002). This part is mainly made by VTSs and harbors, with the blacklisting of suspicious vessels.

There is a dependence of users on the system, and the credibility and trust of the users will depend on the ability of the system to display information that is true and verifiable.

In this scope, the completeness of data is important, as well as its accuracy, in order to assess the confidence on the received and transmitted messages. This confidence should be acquired by the ascertainment of suspicious vessels, in order to consider their messages as subject to doubt; by the reduction of unintentional errors in the messages and by the detection of spoofing events.

As this research is at its beginning, the quality components to use in particular cases and the way they will be used is yet to be defined, as well as the modeling and analysis of data quality parameters in AIS messages are yet to be elaborated.

5. PRINCIPLES OF THE APPROACH

5.1 Towards a real-time analysis of data quality

The detection of errors, bad practices, suspicious or abnormal behaviors is done by the monitoring of specified areas, which depend on the purpose of the surveillance. In order to do that, the identification and a classification of the behaviors and of position reporting falsification techniques, as well as knowledge extraction methods, must be done. The detection is based on a rule-based engine (Riveiro and Falkman, 2009), (Ray et al., 2013), based on recorded trajectories and a panel of typical suspicious behaviors, for instance. As some information in AIS data is dynamic, and the flow of data is continuous, it is important to process the messages on-fly, with two parallel processes: one which looks for specific patterns and defined spatiotemporal behaviors and another one which stores the recently acquired data in order to aggregate it to the database (Ray et al., 2015). A real-time analysis of the data could offer an immediate ascertainment of suspicious or attacked vessels and lead to a quicker response.

In a global approach, a second step is needed, in order to have a system which could help for decision making. In order to move forward, a model of risks is needed.

5.2 The modelling of risks

The risks are to be modelled with ontologies, especially three of them: a trajectory ontology for a geometric definition of trajectories, a geographic ontology for the concepts which are specific to territory description and a domain ontology for the particular domain of AIS analysis. The analysis will be made in four steps, the first one being the technical and functional analysis of the system, in order to put a model on the studied system, the second one being the quantitative analysis, for a study of threat processes, the third one being the quantitative analysis, i.e. the analysis of the seriousness of the consequences of the various feared events, and the fourth one being the synthesis, in order to point out the most critical components of the system (Ray et al., 2015).

6. CONCLUSION

This paper introduces issues on AIS data quality, regarding the various problems of errors, falsification and spoofing of the messages. As the displayed information influences the point of view and analysis of the situation for decision-makers, the situation awareness of actors in a place as hazardous as sea is hampered. An assessment of the data quality would make the detection of AIS problems easier, enhance the situation awareness of decision makers and improve safety and security at sea.

ACKNOWLEDGEMENTS

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

REFERENCES


Bhatti, J., Humphreys, T.E., 2015. Covert Control of Surface Vessels via Counterfeit Civil GPS Signals. University of Texas, in Preparation


NASA, 2012. Space Station Keeps Watch on World’s Sea Traffic http://www.nasa.gov/mission_pages/station/research/benefits/se a_traffic.html#.VNTZoWNQqA0


