A method for integrity assessment of information in a worldwide maritime localization system

Clément Iphar, Aldo Napoli, Cyril Ray

A method for integrity assessment of information in a worldwide maritime localization system

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 enabling vessels to send localization messages. Those messages are used for several uses such as fleet monitoring, traffic control or boarding prevention. The messages sent contain errors, falsifications and undergo spoofing due to the unsecured channel of transmission, and that weakens the whole system and the safety of navigation. This paper introduces the methods for the integrity assessment of messages and the discovery of anomalous data, particularly based on spatial information, which is the cornerstone of AIS messages. This will lead to the determination of non-genuine messages and the highlighting of falsifiers, with the objectives to discover the falsifications, point out the falsifiers, remove the falsified messages from the following studies and thus improve the effectiveness of the system.
Keywords: Automatic Identification System; data falsification; integrity assessment; anomaly detection.

1 Introduction

As a major place of human activities such as transportation of goods, fishing, cruising and sailing, the ocean has an important impact on the worldwide economy, and a central place in some particular domains such as energy transportation and goods transportation where 90% of worldwide traffic is done by sea. This ever increasing traffic leads to navigation difficulties and risks in coastal and crowded areas where numerous ships exhibit different movement objectives which can be conflicting.

The international traffic is globally dense, but some regions (such as Northern Europe or South-Eastern Asia) and some locations (such as straits or canals) have a particularly dense traffic. For instance Malacca strait has 50 000 vessels per year, the Panama canal 14 000, the Bab-el-Mandeb strait 10 000 and the Suez canal 20 000.

The comprehension of vessels intentions and of maritime situation comes through the analysis of ships localizations, with a long or a short time span of data. On the one hand, short-term studies concentrate on instantaneous behaviours while on the other hand long-term studies concentrate on the trajectories and the identification of specific behaviours.

Currently, one of the most widely deployed systems is the Automatic Identification System (AIS) that helps in completing the radar picture and providing information beyond the radar horizon. Section II presents the systems and the vulnerability problems it undergoes and examples of AIS falsification and section III presents the proposed methodology for the analysis of AIS messages and the detection of falsifications.

2 The Automatic Identification System

2.1 A system for maritime safety

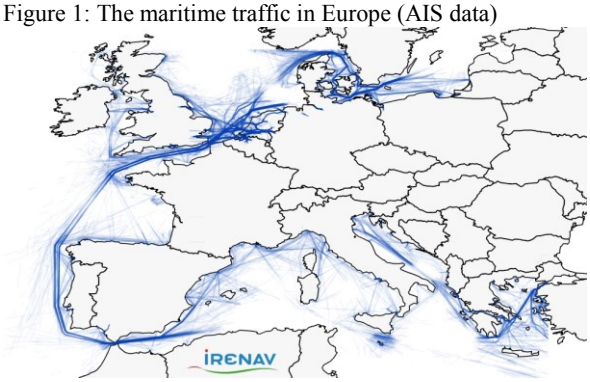
The Automatic Identification System (AIS) was put in place by the International Maritime Organization, in the Safety Of Life At Sea (SOLAS) convention, in its 2002 version [4]. All the vessels from the signatory states are not concerned with this system. Indeed, the SOLAS convention, in its fifth chapter, 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” [4].

At first, messages were transmitted (radio-transmission in the Very High Frequency bandwidth) from vessel to vessel and from vessel to coastal station within the radio horizon range (circa 40 nautical miles), then with the development of satellite technology, messages are also received by satellites and their position and trajectory can be displayed in dedicated websites. The understanding of the worldwide traffic then switched from local to global.

There are 27 different kinds of messages depending on the type of transmission, but the most used ones are the numbers 1, 3 and 5 [12], used by transponders of vessels for which the use of AIS system is compulsory, for spatiotemporal position reports, respectively scheduled, special and static.

In a standard position report of those transponders, 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 [12]. The number of messages is important: in a mean day, 400 000 messages are sent from more than 22 000 vessels [8], and in the sole waters of the European Union, circa 200 million messages are sent every year. The main data given in the message are the number of the message, the 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 and the destination.

Figure 1: The maritime traffic in Europe (AIS data)



2.2 The spatial components within the messages

As AIS messages are primarily localization messages, the spatial components of it 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.

The total number of bits allocated in the message number 1 for latitude and longitude is important (27 and 28, respectively), thus the elementary unit of the measure is the one ten thousandth of minute of arc (1/10 000 arcmin). The longitudes are given in a scale going from -180 to 180 deg, positive values going towards east and the latitudes are given in a scale going from -90 to 90 deg, positive values going towards north. For all longitudes and for the latitude close to the equator, this basic unit is worth circa 20cm on the surface of the Earth, this value progressively deceasing as latitudes grow. As the GNSS computation is mainly done 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 the message number 5, the destination of the vessel has a dedicated textual field of 20 6-bits ASCII characters. This voyage-based spatial information has nevertheless the drawback to be filled widely inappropriately.

Some other fields such as the user ID or the time stamp are not spatial data but will be useful in our study. The user ID will be used to be able to define trajectories by tracking one single user over time and the time stamp will be used to reconstruct the trajectory and shift spatial data into spatiotemporal data.

2.3 A vulnerable system

As stated before, the AIS system transmits through VHF radio band, however the transmission is free and the channels of transmission are not secured, which brings problems of both quality of information (as the messages are freely sent) and genuineness of reception (as anyone can cover genuine messages by falsified ones). Three major cases of bad data quality can be highlighted: the errors (unintentional broadcast of false information), the falsifications (intentional broadcast of false information) and the spoofing (intentional coverage of a genuine message) [9]. Moreover, the system transmits a high quantity of messages, which brings spatial big data issues in their study.

Indeed, some data contained in AIS messages are fulfilled manually during the system initialization (first use), they can be done by underestimating the importance of a proper fulfilment of the fields or by ignorance of the way the system works, and each human-filled field is subject to errors [3], such as the destination field.

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. The fact to falsify one's whereabouts is a kind of falsification allowing the user, displaying false GNSS coordinates, to be somewhere else and to hide its activity. 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 behaviour of the proper ship.

2.4 Exemplification of falsification

In the maritime domain, identity theft [13] 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. 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 [11], Iran used to falsify the MMSI number of some ships in order to trade with Syria, then under embargo. The Iranian ship *Millionnaire* 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 [13]. 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 their activities, such as fishing in an unauthorized area, or trade illegal goods [6] with other ships or on coasts.

One of the trickery is a false closest point of approach alert [1]. An alert is triggered and the vessel is forced to changed its heading and perhaps be guided to hazardous places in order to avoid a hypothetical boarding by a ghost vessel. Moreover, [1] 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 then possible to see it displayed on the website marinetraffic.com.

An example of spoofing is given in [1], with a pirate who emits a false signal with a higher power than the genuine one, covering it. In other vessels the system is then misled as it receives a message different than the transmitted one. This can force the pilot or the autopilot to manoeuver in order to bring the vessel away from a non-existing hazard, and possibly leading it to others actually existing hazards.

3 Detection and analysis of errors and falsifications

The uses of the AIS are numerous, albeit initially designed for safety and security purposes such as the prevention of boarding, the AIS can be used for investigation in case of accident, analysis of global traffic and traffic in specific hazardous areas, control of fishing fleets, cargo fleets or aid to navigation. The diversity of those uses makes necessary the fact for the users to use real data. Moreover, the confidence of the users in the system highly depends on the truthfulness of the messages, and the more unswayed is the user, the less confident will be the system.

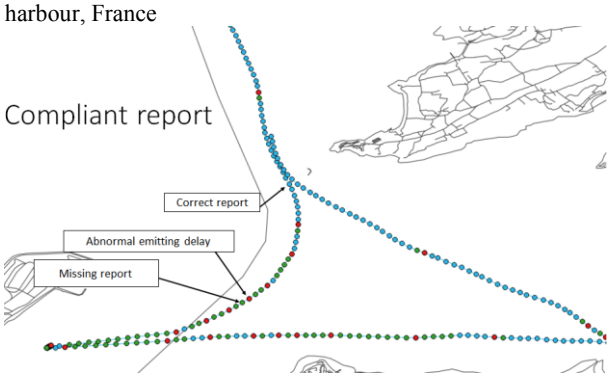
Besides the falsification detection, which is the purpose of the next part, one of the purposes of the analysis of AIS messages is to model and detect behaviours by the establishment of behavioural patterns and the comparison between actual data and those patterns. This moving object assessment with the techniques of data mining shall be done with reliable data, cleaned up of all erroneous, falsified or spoofed messages.

3.1 Integrity assessment for AIS messages

Amongst the data quality dimensions, integrity has a leading position for the determination of the reliability of sets of data. The assessment of data quality can also be done under the dimensions of accuracy, precision, reliability, currentness, completeness and consistency [5], but for the overall understanding and assessment of the coherence of data, integrity covers them all, the coherence we use being defined by [2], as: "when we gather information from less than fully reliable sources, then the more coherent the story that materializes is, the more confident we may be, ceteris paribus".

Such an assessment of integrity can be done within one lone message, between messages sent by the same vessel (according to their MMSI number), between messages sent by several vessels. An exemplification of the integrity assessment is, for a lone message: "Is the speed consistent with the declared type of vessel?" or "Are the declared GNSS coordinates compatible with the existence of a navigable area?"; for several messages: "Are the declared GNSS coordinates compatible with the time stamps, the speed and the heading?" or "Is the trajectory consistent with the declared destination?"; and for several messages sent by several vessels doing the same trip: "Is the trajectory of one vessel coherent with the mean trajectory of all vessels?" or "Is the speed of one vessel consistent with the mean speed of all vessels?".

Figure 2: Example of a case of compliant report in the Brest harbour, France



The last two questions can be applied to each field, and the study can go further and ask: "If not, why?". In this case, the use of external information such as external databases (weather for instance) is necessary to the understanding of behaviours that are anomalous-looking, in order to determine whether they are anomalous or not.

3.2 Anomaly detection in AIS-derived information

Anomaly detection is used in the field of data analysis, with the three elementary steps that are: (1) the identification of the "normality" characteristics by computation and determination of data classical signatures (for instance trajectory clusters in the AIS messages case), (2) the determination of metrics for the computation of the distance of the studied behaviour from the standard behaviour and (3) the determination of threshold criteria for the distance to the standard behaviour, allowing the normality, the abnormality, the magnitude of normality and the magnitude of abnormality for a datum.

The determination of a normal behavioural pattern can be done using diverse methods such as statistical methods, neural networks or machine leaning [7], that will depend on the nature of the data to study: the created patterns can be a sequence of events, a cluster or a statistical distribution. Statistical methods are fitted with the data in which extreme values are anomalous (the speed for instance), and do not work in the cases where the anomalies are evenly distributed. Neural networks are fitted for the discovery of hidden patterns with complex boundaries but their black box behaviour and their need of a learning phase. The principle of machine learning is to automatically learn complex structures and take data-based decisions. Amongst those methods are the genetic algorithms, Bayesian networks or the clustering.

The distance determination will depend on the type of data and distances such as Euclidian distance in dimension 1 (in the case of speed), Euclidian distance in dimension 2 (in case of localization coordinates with small distance), orthodromic distance (in case of long distance localization coordinates), mean, Hausdorff or Fréchet distances (in the cases of computation of distance between trajectories), or edition distance (in the case of distance with textual data, for the field "destination" for instance). The choice of the right distance for each field is of major importance in the anomaly detection process.

The third step in this anomaly detection is the thresholding for outlier determination in different cases, among which those aforesaid. This will be the purpose of future work.

4 Conclusion

This paper introduces issues on AIS, the system itself and its vulnerabilities that lead to multiples errors in the messages, falsification and spoofing of them. The spatial characteristics of the messages such as the GNSS coordinates are an important part of falsification, thus the development of an ad hoc methodology is necessary, based on messages integrity assessments and anomaly detection, to discover the falsified messages with the purpose of analysing the data between the raw information stream and the use for diverse purposes by the users.

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

[1] M. Balduzzi, A. Pasta and K. Wilhoit. A security evaluation of AIS automated identification system". In proceedings of the 30th annual computer security applications conference. New Orleans, 2014.

[2] S. Hartmann and L. Bovens. A probabilistic theory of the coherence of an information set. In proceedings of the 4th international congress of the society for analytical philosophy, pages195-206. Bielefeld, 2001.

[3] A. Harati-Mokhtari, A. Wall, P. Brooks and J. Wang. Automatic Identification System (AIS): a human factors approach. The Journal of Navigation, 60(3), 2007.

[4] International Maritime Organization. International convention for the safety of life at sea. 2004.

[5] C. Iphar, A. Napoli and C. Ray. Data quality assessment for maritime situation awareness. In proceedings of the 9th International Symposium on Spatial Data Quality. La Grande-Motte, 2015.

[6] F. Katsilieris, P. Braca and S. Coraluppi. Detection of malicious AIS position spoofing by exploiting radar information. In proceedings of the 16th international conference on information fusion. Istambul, 2013.

[7] E. Martineau and J. Roy. Maritime Anomaly Detection: Domain Introduction and Review of Selected Literature. Defence R&D Canada, Valcartier, 2011.

[8] National Aeronautics and Space Administration. Space station keeps watch on world's sea traffic. 2012

[9] C. Ray, C. Iphar, A. Napoli, R. Gallen and A. Bouju. DeAIS project: Detection of AIS Spoofing and Resulting Risks. In proceedings of OCEANS'15. Genova, 2015

[10] M. Redoutey, E. Scotti, C. Jensen, C. Ray and C. Claramunt. Efficient Vessel Tracking with Accuracy Guarantees. In proceedings of the 8th International Symposium on web and wireless geographical information systems. Shanghai, 2008.

[11] The Maritime Executive. Iran, Tanzania and falsifying AIS signals to trade with Syria. 7 December 2012.

[12] J.K.E. Tunaley. Utility of Various AIS Messages for Maritime Awareness. In proceedings of the 8th ASAR Workshop. Longueuil, 2013.

[13] Windward. AIS data on the high seas: an analysis of the magnitude and implications of growing data manipulation at sea. 2014.