On the interest of data mining for an integrity assessment of AIS messages

Clément Iphar, Aldo Napoli, Cyril Ray

To cite this version:

Clément Iphar, Aldo Napoli, Cyril Ray. On the interest of data mining for an integrity assessment of AIS messages. 1st International ICDM Workshop on Maritime Domain Data Mining (MDDM 2016), Dec 2016, Barcelona, Spain. pp.Article number 7836690, Pages 368-373 - ISBN 978-150905472-5, 10.1109/ICDMW.2016.0059 . hal-01421878

HAL Id: hal-01421878
https://minesparis-psl.hal.science/hal-01421878

Submitted on 23 Dec 2016

On the interest of data mining for an integrity assessment of AIS messages

Clément IPHAR, Aldo NAPOLI
MINES ParisTech – PSL Research University
CRC – Centre for research on Risks and Crises
Sophia Antipolis, France
{clement.iphar, aldo.napoli}@mines-paristech.fr

Cyril RAY
Naval Academy Research Institute (IRENav)
Brest, France
cyril.ray@ecole-navale.fr

Abstract — **Put in place by the International Maritime Organization, the Automatic Identification System is a worldwide maritime electronic system that sends radio broadcasted messages at a high rate between the stations, either on board the vessels or on shores. However, some misuses of the system such as identity theft, localization spoofing or disappearances have been demonstrated. The high rate of transmission implies a considerable amount of data to process in order to point out those irregularities. This paper proposes a method based on data mining and clustering methods combined to an integrity assessment of AIS messages for anomaly detection, with a proposition of software architecture for a data processing done both on-the-fly and with archived data. The computation of confidence coefficients and the use of data mining techniques will lead to behaviour characterization with the purpose of enhance the maritime situational awareness.**

Keywords—Automatic Identification System ; data falsification ; integrity assessment

I. Introduction

The maritime traffic is globally dense, and as the traffic increases some regions are facing a particularly high density of traffic, particularly in Europe, South-Eastern Asia or specific locations such as canals or straits. For instance, Malacca strait and the Suez canal have respectively a yearly traffic of circa 50,000 and 20,000 vessels. Moreover, this traffic sometimes takes place in quite hazardous places such as Malacca strait [1] or the Suez canal [2].

For the enhancement of the security and safety of navigation some systems have been put in place. Amongst those systems (radar, Long Range Identification and Tracking for instance), the Automatic Identification System uses electronic devices to send and receive messages to or from other vessels or some coastal stations within its radio horizon range. However, this system is not fully reliable and some errors and data falsification have been demonstrated.

Spatial data analysis [3], [4] and data mining techniques [5] were used in some studies about maritime accidents, but without covering the fields of data falsification discovery. On the one hand, behavioural analysis of moving objects is treated by [6] and [7] and on the other hand, data mining techniques [8] for trajectories analysis are displayed in [9] and [10]. An overview of the clustering methods [11] presents the possibilities associated with the use of such methods in our case of study. The use of data mining for the enhancement of maritime situational awareness has been presented in [12], for the depiction of the current navigation state and forecasted positioning data.

In our study which is focused on falsification discovery, a methodology involving clustering and data mining methods for integrity assessment is proposed, that would lead to the determination of confidence coefficient for the message and the sender, and given the high amount of data available through this system, clustering methods for anomaly detection and behavioural characterization are proposed.

In this article, section II presents the Automatic Identification System, the volumetry of data and its weaknesses, then section III presents the structure and characteristics of data, section IV the proposed method for data processing, then section V presents the proposed architecture of the software.

II. AIS Falsifications

A. The principles of the system

The AIS was put in place in 2000 by the International Maritime Organization and its characteristics and deployment schedule are defined in the Safety of Live at Sea convention [13]. This convention, initiated two years after the sinking of the RMS Titanic in 1912, aims at defining the minimal requirements to which every vessel belonging to a signatory country should comply with. The convention deals with a large scope of subjects which range from the construction of vessels to the way radio-communications shall be done.

One of the themes of the convention is the safety and security of maritime navigation, and it is in this scope that the system was created, as it provides a real-time spatiotemporal positioning of a vessel to every vessels and shore station located in its radio range of action.

Not all the vessels from the signatory countries are concerned with the AIS regulation, as the convention 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" [13].

The system uses transponders to send messages through Very High Frequency marine bandwidth on two worldwide dedicated frequencies: 161.975MHz and 162.025MHz. The transponder is linked with a Global Navigation Satellite System (GNSS) which computed its location as well as external sensors (electronic compass for instance) in order to fill in several data fields within the messages. The AIS can also transmit to satellite the messages (AIS-SAT), that makes possible to keep a track of a vessel even beyond the radio horizon. One kind of message (number 27) is shorter than all other messages and is especially dedicated to satellite reception.

The system, albeit being mainly initially designed for security purposes, the AIS has alternative uses such as 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.

Indeed, some examples of the use of AIS for alternative purposes can be found in studies in several subjects such as accident investigation [14], the detection of near miss collision between vessels [15], the behaviour understanding in a waterway through traffic monitoring [16] or the mapping of the fishing effort using AIS data [17].

B. Data Volumetry

All equipped vessels must transmit localization messages at a given rate, which ranges from 2 seconds to 12 seconds according to the speed of the vessel, the higher the speed, the higher the rate of transmission. Localization messages are sent every 3 minutes when the vessel is at anchor. As stated before, all messages are not localization messages, however they account for circa 90% of the total message number. 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, and as stated by [18], at the global level, "on a good day, approximately 400,000 ship position reports are received from more than 22,000 different ship identification numbers (Maritime Mobile Service Identity, or MMSI). In a summary made in Oct. 2011, the total number of position reports received exceeded 110 million messages from more than 82,000 different MMSI numbers". The total amount of data that can then be accumulated through the reception and storage of AIS data tends the treatment of those pieces of information towards big data issues.

C. A system with weaknesses

Three major cases of bad data quality can be distinguished: the errors (when false data in non-deliberately broadcasted), the falsifications (when false data is deliberately broadcasted) and the spoofing (when data is created or modified and broadcasted by an outsider) [19]. 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 pieces of information 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 pieces of information 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 [20], 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 [21], the disappearances [22], the broadcast of false GNSS coordinates or the statement of a wrong activity [23] are types of falsification. According to [20], 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 [24]. Those spoofing activities are done in order 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).

III. DATA

A. Data structure

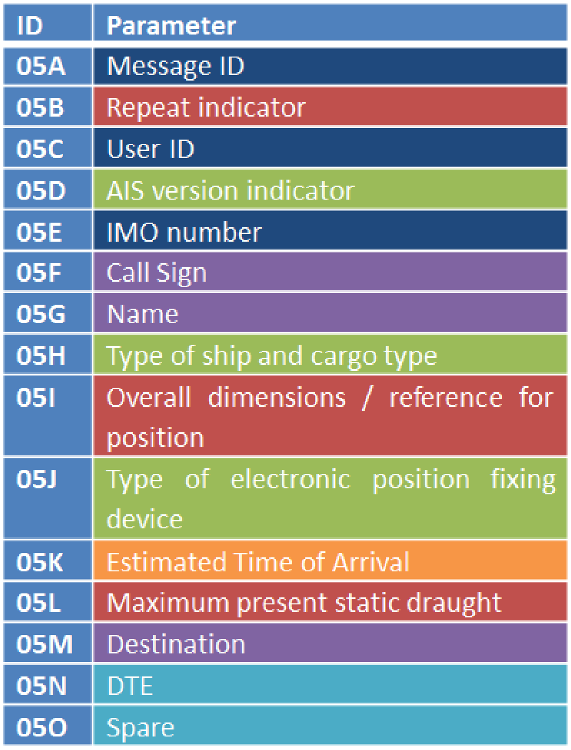
Due to the variety of communications, there are 27 different kinds of messages. Position report messages, information messages and check messages are the three main groups of messages [25]. All 27 kinds of messages are standardized by the International Telecommunications Union [26] and have its particular outline with data fields, each one being allocated a certain number of bits. The information within each field can take several forms: boolean, text, number representing a physical quantity, number representing a choice in a given list. The content of the messages do vary largely from one message to another: in a position report message the data fields are the speed, the position and the cape, amongst others, while an aid to navigation message will display the type of beacon, its name or the location of a hazard. Each message has then different data fields, accordingly with its type, each data field having a given location in the message and a given number of bits allocated. The meaning of the values of the data fields within the messages are also defined in an unambiguous way in the technical specifications [26].

B. The unique identification of data fields

In order to get an all-encompassing assessment of the integrity of AIS messages, all the available information from all available messages shall be taken into consideration. However, as a lot of messages display localization information, there is a quite high redundancy of some data fields. This profusion of barely identical fields in different messages leads to the need of an ad-hoc field nomenclature so that one field in a given message cannot be confused with another one in another message. The nomenclature we propose matches each field of each message to a unique three-character string of type “XXY”, where the XX part stands for the number of the message from 01 to 27 and the Y part stands for a letter ranging from A to S (at most, as no message has more than 19 fields), the position of which in the alphabetical order indicating the position of the field in the given message. So in this nomenclature, the field “02F” corresponds to the sixth field (as F is the sixth letter in the alphabetical order) of the message number 2, i.e. the “Speed Over Ground” field.

C. The diverse data types within the messages

The data within the messages can take several forms, and the message number 5 is a good witness of the diversity of data types. The following table displays the layout of the message number 5, with the unique nomenclature identifier, the parameter in the field and the datum type.



ID	Parameter
05A	Message ID
05B	Repeat indicator
05C	User ID
05D	AIS version indicator
05E	IMO number
05F	Call Sign
05G	Name
05H	Type of ship and cargo type
05I	Overall dimensions / reference for position
05J	Type of electronic position fixing device
05K	Estimated Time of Arrival
05L	Maximum present static draught
05M	Destination
05N	DTE
05O	Spare

Figure 1: The data fields of message 5
Dark blue: num. repr. an identifier ; Red: num. repr. a physical quantity ; Green: num. repr. a choice in a given list ; violet: textual ; Orange: date ; Sky Blue: binary

Discrimination can be made between six types of data: numeric representing a physical quantity, numeric representing an identifier, numeric representing a choice in a given list, date, textual and binary, as shown in Figure 1.

For the numeric values, the physical quantity data type is a piece of information that has a sense in the physical world (such as latitude, longitude, speed or for message 5 the maximum present static draught). In general, as the physical quantities are continuous quantities, a given precision is associated with it, for instance here 0.1m for the 05L field. The identifier data type is a piece of information linked with a field number standing for a unique identification number, such as MMSI number of IMO number. The choice in a given list data type stands for the pieces of information of which the meaning has to be deduced from a given list of values are stated in the technical specifications [26]. For instance, the 05I field stands for the type of electronic position fixing device, that can range from 0 to 15. If the value is 1 then “GPS” is used, if it is 2 then “GLONASS” is used, and so on until the last possible value.

For the non-numeric values, a textual data type stands for the pieces of information in which the bits are converted by groups of 6 into ASCII characters, used in the case of “Name” or “Destination” fields. Date data type is a rare data type in AIS messages standing for a date representation and binary data type is related data types where a statement is declared as true or false. Alternatively, it can be used to state if the value of the physical quantity representing the field is superior or inferior to a given threshold, or a choice in a list of exactly two choices.

IV. Method for Data Processing

A. Selection of cases suitable for study

In the data comparison field, a set of messages must be selected in order to compare data within a scope where comparison makes sense. Four different relevant ways for data processing have been discriminated by our study: MMSI-based, station-based, location-based and route-based, with the hypothesis that the messages received from the same MMSI, by the same station, from the same route or location will tend to look alike.

The MMSI-based study takes only into consideration the messages from a single emitting station, that can be a vessel, an aircraft, a coastal station or even a buoy as all those devices have an assigned MMSI number. The station-based study concentrates on the messages received by a single station on-shore, and albeit there is a theoretical coverage of circa 40 nautical miles, it is possible that some messages from a way longer position are received because of the favourable transmission conditions or that some messages sent from a shorter location are not received because of masks. The location-based study takes into consideration the messages sent from a given spatial extent, for instance a circle around a point of interest, a fishing area or a particularly hazardous zone. The route-based study selects only the messages that are sent by the vessels which follow a given route, for instance sailing between two given ports or going through a given corridor (a Traffic Separation Scheme for instance).

B. Integrity assessment

As it was stated in [27], 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 [26]. 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.

In our method, the first and second levels are within a message and can be done on-the-fly whereas the third and fourth levels are between 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. Besides, a nomenclature for unique identification of assessment items has been performed. Figure 2 presents a sample of those items.

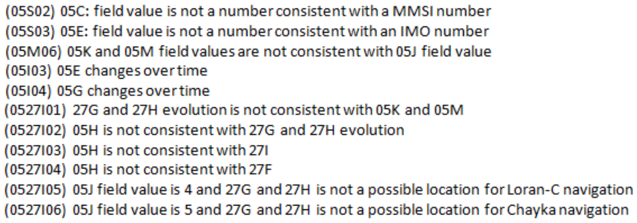

(05S02) 05C: field value is not a number consistent with a MMSI number
(05S03) 05E: field value is not a number consistent with an IMO number
(05M06) 05K and 05M field values are not consistent with 05J field value
(05I03) 05E changes over time
(05I04) 05G changes over time
(0527I01) 27G and 27H evolution is not consistent with 05K and 05M
(0527I02) 05H is not consistent with 27G and 27H evolution
(0527I03) 05H is not consistent with 27I
(0527I04) 05H is not consistent with 27F
(0527I05) 05J field value is 4 and 27G and 27H is not a possible location for Loran-C navigation
(0527I06) 05J field value is 5 and 27G and 27H is not a possible location for Chayka navigation

Figure 2: Integrity assessment items

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 pieces of information assessed in this particular item).

The experimental validation of this method is being done in ongoing work and uses data we have been collecting using antennas located in Brest harbour. The use of some additional data will be used in future works.

C. *Clustering methods for behaviour charecterization*

As the subset of items that can be used for an assessment depends on the type of assessment, it is then possible to compute several kinds of coefficients according to the purpose of the assessment. Some items are more useful to highlight disappearances whereas some other items are more likely to highlight identity fraud.

Once all different kind of assessment selected, and each kind being allocated a subset of items, a vector of coefficients can then be computed, and each vector compared then to archived data for pattern matching.

In addition, for the fields in the messages, a measure of the rareness of the fields can lead to the use of statistical methods. For a given set of values, typically the subset of messages extracted by one of the methods proposed in section IV.A, a statistical study on how do pieces of information tend to appear more than other. Of course, such an assessment is tightly linked to the type of data, as data can be of very diverse types, as shown in section III.C.

A strict and simple statistical distribution can be used in the case of binary and numeric representing a choice in a given list cases, whilst a buffer is used for numeric data representing a physical value and date (which means if we concentrate on value x and take a buffer y, all values between x-y and x+y will be taken into consideration). A textual distance such as Levensthein distance or a translation table can help in the case of textual fields, particularly for the destination field case, as the names are often written in various languages and sometimes with spelling mistakes. This process leads to a clustering of data under groups that represent the same port, and a statistical method can then be used in order to assess the frequency of each group.

In addition, similarity and dissimilarity measures are able to compare vectors of data extracted from AIS messages. Clustering methods with data of quite high dimensionality (in general about 15 data fields, where each data field worth one dimension) can then be applied. Moreover, conditional probabilities can be extracted from the set of messages, by using subsets of messages where a certain number of conditions are gathered (for instance, where a combination of field values occurs). This will be the purpose of future works.

V. Software Architecture

A. *Integration of external data*

If AIS was an isolated system, the sole study of its data would be enough for an integrity assessment. However, the AIS is set on vessels and of course vessels have diverse physical characteristics, and themselves evolve in diverse environments with particular features. And as the environment evolves, it is possible to enhance the understanding of the behaviour of vessels with external sources of information. Indeed, on the one hand, data can look anomalous on the basis of the sole AIS data and turned out as normal with the study of external data, and on the other hand data can look normal with the sole AIS data and turned out as anomalous with the assistance of external data.

Moreover, some cases of possible falsifications such as disappearances must be investigated as the fact to receive or not to receive a message is more complex than the presence or absence of the given vessel within the line of sight. Anomaly detection in such case [28] requires extra processing. Indeed, on the one a vessel can sail beyond the line of sight and yet the messages sent from it are received, on the other hand a vessel can sail within the line of sight distance and yet the messages sent by it are not received for several reasons including masks, message collision or system failure.

An approach on multiple frames is enabled by the diversity of possible databases, and we selected three main families of databases: environmental, vessel-based and navigation-based. However, there is no exhaustive list of possible integrable databases as databases appear, evolve and go out of date through time. Their purpose is to bring expert data to enlighten specific situations.

Environmental data would embrace all kind of databases with a relation with the environment in which the vessel evolves. Meteorological databases with information such as pressure, temperature, wind force, rainfalls and wave height can be used for behaviour understanding. Tidal information and bathymetric maps can also be useful for the understanding of coastal navigation.

Vessel-oriented databases would gather insurance databases or the different fleet registers available, from the freely online available European Union Fishing Fleet Register to charged Lloyd's Register.

Navigation-oriented databases would embrace all kind of information linked to navigation and vessels' routes. Some other sensors (radar for instance) can be included, as well as regular origin and destination ports for each vessel.

B. The proposed architecture

A synoptic diagram of the proposed architecture can be found in the Figure 3 below. Various sources can provide the signal, the parser gives message parameters, the data processing of the signal provides some signal parameters and two different steps of data processing. All this architecture is built around the database in order to fill it and use it for knowledge discovery.

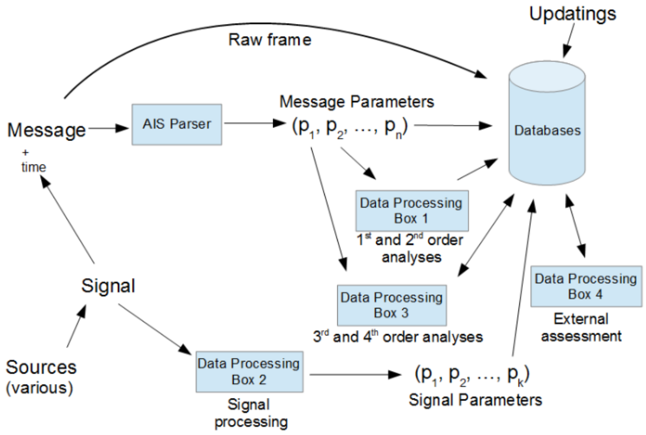


Figure 3: Proposed Architecture

It is very important in our study that the parser we chose and modify at our convenience does not discard any message, as we want to process every single frame going through the reception system.

The data processing box number two corresponds to a signal processing for the computation of four characteristics figures of a transponder [29], which are the rise time (power from 10% to 90%), the time before modulation (from 50% of power to first bit of the training sequence of the message), the fall time (power from 90% to 10%) and the time after modulation (from the last bit of the message to 50% of power). Some patterns are recognizable and it is then possible to distinguish the diverse emitter transponders with those characteristics. Although it is not possible to determine the identity of the vessel of origin only with those characteristics, it is however possible to determine whether or not two or more messages are sent by the same vessel.

The data processing box number one is in charge of on-the-fly analysis of the first and second levels of data assessment, in order to have as output computed coefficients to store in the database. Similarly, the data processing box number three is in charge of the analysis of the third and fourth levels of data assessment, in order to have as output computed coefficients to be stored in the database. This second part of the study, unlike the first, needs some requests to the database to be done.

In the database itself, each new entry will lead to the creation of a new item with a unique identifier, the time of reception, the raw frame, all the message field values and the various computed coefficient obtained through assessments (the four orders and the signal parameters) as attributes.

The data processing box number four is in charge of integrity assessments between AIS data and external data, i.e. with data coming from the external providers. Of course, the types of processing will vary accordingly with the type of external information available, and no strict process is proposed. A list of assessment items, similar to the ones describes in section IV.B for inner information integrity assessment can be created for each new database when its specifications are known (such as its fields, their source, their precision and reliability).

Some updates to the external databases will, in some cases, be necessary, as to ensure to have data that are not outdated to lead to a reliable data quality assessment. This implies the fact that databases must have a time stamp, some information on the provider and all kind of metadata that could be useful for the processing of data.

C. Use case

In the scope of the discovery of anomalous data, it is important that the selected use cases are in relationship with the weaknesses of the system as discussed in section II.C. Taking into consideration this analysis, we selected in a first draft five scenarios for which peculiar messages types can and will be used. Those scenarios are identity theft (messages 5, 24 and spatiotemporal messages), use of a fictitious identity (message 5, 24 and spatiotemporal messages), frequency hopping (message 22), disappearances (spatiotemporal messages) and GNSS spoofing (spatiotemporal messages and messages for which an answer is expected, for instance the pair 10/11). The messages that are considered as spatiotemporal are the ones in which the latitude and longitude are given, i.e. messages number 1, 2, 3, 9, 18, 19 and 27.

As for the use case of disappearances, theoretical coverage of terrestrial receivers have been computed, the practical coverage have been extracted from received data and an official vessel fleet register have been downloaded for the first external information implementations.

D. Implementation details

This article focuses on methodology and does not provide implementation details, which are ongoing research. The AIS parser, already in operation, is in Java language and is adapted from AISMESSAGE parser available online. Also in operation is the signal data processing. In addition, the database architecture has been built. It is based on a relational database using a PostgreSQL / PostGIS architecture, with one table for each kind of message, tables for the storage of analysis results and additional tables receiving data such as, for instance, weather, fleet registers or signal information. The analysis process will make use of this database in continuous mode using Python language programs for the extraction of information linked to the processes presented in this paper.

VI. Conclusion

This article proposes a methodology for anomaly assessment and falsification discovery in AIS messages, using data mining techniques and an integrity-based assessment, enhanced by external databases. The purpose is to assign to each message and to each single user of the system a confidence coefficient, leading to behavioural characterization of trajectories that are spatial, temporal and semantic. An architecture for the implementation of this method is proposed, that will be able to lead to an integrity-based assessment, itself leading to an alert triggering system in decision-support cases, for the eventual enhancement of the maritime safety.

Acknowledgment

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

References

[1] N. Afrida, “Malacca Strait rampant with pirates”, The Jakarta Post, 2nd January 2015. [Online]. Available: http://www.thejakartapost.com/news/2015/01/02/malacca-strait-rampant-with-pirates.html

[2] Armed Conflict Location and Events Dataset, “Real-time Analysis of African Political Violence”, Monthly Report, Conflict Trends (NO. 45), January 2016.

[3] L.A. Le Blanc and C.T. Rucks, “A multiple discriminant analysis of vessels accidents”, Accidents Analysis & Prevention, vol. 28, no. 4, pp. 501-510, July 1996.

[4] C. Marven, R. Canessa and P. Keller, “Exploratory Spatial Data Analysis to Support Maritime Search and Rescue Planning”, in J. Li, S. Zlatanova and A. G. Fabbri, eds. “Geomatics Solutions for Disaster Management”, New York: Springer Berlin Heidelberg, pp. 271–288, 2007.

[5] A. Torun and S. Düzgün, “Using spatial data mining techniques to reveal vulnerability of people and places due to oil transportation and accidents: a case study of istanbul strait”, in “ISPRS Technical Commission II Symposium”, Vienna, pp. 12 – 14, 2006.

[6] J. Roy, “Anomaly detection in the maritime domain”, in C. Halvorson, D. Lehrfeld and T. Saito (eds.), “Optics and Photonics in Global Homeland Security IV”, 2008.

[7] J. Roy and M. Davenport, “Exploitation of Maritime Domain Ontologies for Anomaly Detection and Threat Analysis” in proceedings of Waterside Security Conference, Carrara, Italy, November 3-5, 2010.

[8] J. Chen, C. Lai, X. Meng, J. Xu and H. Hu, “Clustering moving objects in spatial networks”, in proceedings of the 12th international conference on Database systems for advanced applications, Bangkok, Thailand, April 9-12, 2007.

[9] F. Giannotto, M. Nanni, F. Pinelli and D. Pedreschi, “Trajectory pattern mining”, in proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD’ 07). San Jose, CA, USA, August 12-15, 2007.

[10] B. Idiri, “Méthodologie d'extraction de connaissances spatio-temporelles par fouille de données pour l'analyse de comportements à risques : application à la surveillance maritime”, PhD Thesis, Ecole Nationale Supérieure des Mines de Paris, december 2013.

[11] A.K. Jain, M.N. Murty and P.J. Flynn, “Data Clustering: A Review”, ACM Computing Surveys, vol. 31, no. 3, pp. 264-323, sept. 1999.

[12] V. Fernandez Arguedas, F. Mazzarella and M. Vespe, “Spatio-temporal Data Mining for Maritime Situational Awareness”, in proceedings of the OCEANS’15 GENOVA conference, Genova, Italy, May 18-21, 2015.

[13] International Maritime Organization, “International convention for the safety of life at sea”, 2004.

[14] Y; Wang? J. Zhang, X. Chen, X. Chu and X. Yan, “A spatial-temporal forensic analysis for inland-water ship collision using AIS data”, Safety Science, vol. 57, pp. 187-202, 2013.

[15] W. Zhang, F. Goerlandt, J. Montewka and P. Kujala, “A method for detecting possible near miss ship collisions from AIS data”, Ocean Engineering, vol. 107, pp. 60-69, 2015.

[16] F. Xiao, H. Ligteringen, C. van Gulijk and B. Ale, “Comparison study on AIS data of ship traffic behavior”, Ocean Engineering, vol. 95, pp. 84-93, 2015.

[17] F. Natale, M. Gibin, A. Alessandrini, M. Vespe and A. Paulrud, “Mapping Fishing Effort through AIS Data”, PloS ONE, vol. 10, no. 6, 2015.

[18] National Aeronautics and Space Administration, “Space station keeps watch on world’s sea traffic”, NASA, 2012.

[19] C. Ray, C. Iphar, A. Napoli, R. Gallen and A. Bouju, “DeAIS project: Detection of AIS Spoofing and Resulting Risks”, in proceedings of the OCEANS’15 GENOVA conference, Genova, Italy, May 18-21, 2015.

[20] A. Harati-Mokhtari, A. Wall, P. Brooks and J. Wang, “Automatic Identification System (AIS): Data Reliability and Human Error Implications”, Journal of Navigation, vol. 60, no. 3, pp. 373-389, september 2007.

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

[22] Windward, “AIS data on the high seas: an analysis of the magnitude and implications of growing data manipulation at sea”, October 2014.

[23] 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, pp. 1196-1203, Istambul, Turkey, July 9-12, 2013.

[24] 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, LA, USA, December 7-12, 2014.

[25] J.K.E. Tunaley, “Utility of Various AIS Messages for Maritime Awareness”, presented at the 9th ASAR Workshop, Longueuil, Canada. October 15-18, 2013.

[26] International Telecommunication Union, “Recommendation ITU-R M.1371-5 (02/2014) – Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band”. ITU, 2014.

[27] C. Iphar, A. Napoli and C. Ray, “Detection of false AIS messages for the improvement of maritime situational awareness” in proceedings of the OCEANS’15 WASHINGTON conference, Washington DC, USA, October 19-22, 2015.

[28] F. Mazzarella, M. Vespe, D. Tarchi, G. Aulicino and A. Vollero, “AIS Reception Characterisation for AIS on/off Anomaly Detection” in proceedings of the 19th International Conference on Information Fusion, Heidelberg, Germany, July 5-8, 2016.

[29] E. Alincourt, C. Ray, P.M. Ricordel, D. Dare-Emzivat and A. Boudraa, “Methodology for AIS Signature Identification through Magnitude and Temporal Characterization”, in proceedings of the OCEANS’16 SHANGHAI conference, Shanghai, China, April 10-13, 2016.