Data integrity assessment for maritime anomaly detection

Clément Iphar, Cyril Ray, Aldo Napoli

Data integrity assessment for maritime anomaly detection

Clément Iphar

Centre for research on risks and crises (CRC). MINES ParisTech - PSL Research University. Sophia Antipolis, France

Cyril Ray

French Naval Academy Research Institute (IRENav). Brest, France

Aldo Napoli

Centre for research on risks and crises (CRC). MINES ParisTech - PSL Research University. Sophia Antipolis, France

Abstract

In the last years, systems broadcasting mobility data underwent a rise in cyberthreats, jeopardising their normal use and putting both users and their environment at risk. In this respect, anomaly detection methods are needed to ensure an assessment of such systems. In this article, we propose a rule-based method for data integrity assessment, with rules built from the system technical specifications and by domain experts, and formalised by a logic-based framework, resulting in the triggering of situation-specific alerts. A use case is proposed on the Automatic Identification System, a worldwide localisation system for vessels, based on its poor level of security which allows errors, falsifications and spoofing scenarios. The discovery of abnormal reporting cases aims to assist marine traffic surveillance, preserve the human life at sea and mitigate hazardous behaviours against ports, off-shore structures and the environment.

Keywords: Data falsification, integrity assessment, AIS.

*Corresponding author: Clément Iphar. Phone: +39 0187 527 391. 1 rue Claude Daunesse, 06560 Sophia Antipolis, France

Email addresses: clement.iphar@mines-paristech.fr (Clément Iphar), cyril.ray@ecole-navale.fr (Cyril Ray), aldo.napoli@mines-paristech.fr (Aldo Napoli)

1. Introduction

Volume, Velocity, Variety and Veracity are the four traditional challenges associated with Big Data. The Volume refers to the total amount of data to be processed, which is ever increasing, with each day companies collecting petabytes of data and the total amount of data created overcoming the exabyte (McAfee & Brynjolfsson, 2012). The Velocity concerns the ability to handle, gather and exploit data, and is more and more important as the volume of data increases. The Variety challenge covers the various formats that can be taken by data (images, text messages, signal, amongst others), and the rise of digital information, at the origin of the explosion of data volumes, generated data of various types that have to be handled in an efficient way. The Veracity challenge is linked to the inner value of data, representing the fact for a piece of information to be truthful, so to correctly depict the phenomena measured or represented in the way it is expected to do.

Nowadays, in the flows of data being created, the will to extract meaningful information though data science methods has risen. However, the quality of this information is tightly linked with the quality of the data this information is extracted from, and data quality assessment became key features in the conception of information systems. Trust and confidence in data are the cornerstone of the trust of the user in the outcomes of any analysis. Therefore, information systems must incorporate a layer of data integrity analysis, that will bring the user to a knowledgeable understanding of the pieces of information that are eventually presented to him or her.

The development of cybersystems creates the need for the development of means aiming at protecting those systems and being able to respond to an attack, as well as assessing the potential issues resulting from a variety of attacks. Those means, as a whole, constitute cybersecurity. A cyberattack usually consists in the access, the change, the diffusion or the destruction of potentially sensitive information (Toumsi & Rais, 2018). Money extortion, intelligence or

the interruption of usual business processes are usually the main reasons for cyberattacks, although in some cases the attribution of the attacks is not clear (Rid & Buchanan, 2015), despite its importance in implementing further security layers. Such attacks can be ideologically motivated (Holt et al., 2019), and it is difficult to measure the extent to which cyberattacks occur as firms tend to under-report such attacks, and only make it public when investors already suspect with a high likelihood its existence (Amir et al., 2018). The attacks can target critical infrastructures of countries (Maglaras et al., 2018) or the daily life of citizens with cyberthreats existing in various areas such as domotics (Arabo, 2015) or street furniture (Comert et al., 2018). Issues about cyberattacks on mobile objects have been thoroughly studied for cars (Petit & Shladover, 2015), airplanes (Waheed & Cheng, 2017) and vessels (Costé, 2018).

Grounded in the theory of Situation Awareness developed by Endsley (1995), which is based on a descriptive view of decision making, the detecting and the classification of abnormal behaviours is a key task of any situational awareness system, for several reasons such as the extraction of relevant contextual information and the proper monitoring of both self-reporting systems and non-cooperative systems. The eventual purpose of this data processing is to design a decision-making system that provides an operator, which is in charge of monitoring the system, with qualitative information in a quantitatively measured fashion. The qualitative factor represents the usefulness of each piece of information and the quantitative factor is the amount of information that will be presented to the operator. The operator must therefore get information with a quality which is good enough to make a decision but also to understand the underlying meaning of the data handled, through evaluation criteria; but the operator must at the same time only be presented an amount of information sufficient to take an informed decision but reasonable with respect to the cognitive capacities of a person.

Cooperative mobile data witnessed a recent rise in several fields such as pedestrians, goods transportation, cars, vessels, airplanes. These data are subject to anomalies, misuses and falsification, and anomaly detection methods can

in this respect be used in order to assess these data. Data streams from sensors have various qualities, and the assessment of this data quality with respect to the nature of the sensor is necessary in order to construct analysis frames that take into consideration all available information so that falsification issues and attack issues can be clearly discriminated. Since falsifications and attacks address different issues originating from different sources, it is particularly important to be able to differentiate them as soon as possible so that the relevant methods can be applied for data analysis.

In general, machine learning techniques are widely used for data analysis (Kotsiantis et al., 2006). Those techniques include regression, classification, clustering, deep learning, image processing or natural language processing. Since the topic of maritime cybersecurity issues has few available and usable data for the construction of a model for the training of an algorithm, this field of study is prone to the use of alternative methods that do not require such training dataset. In this work, a base of rules in description logics is used in order to assess data. The approach suits cases where the understanding of the situation must be contextualised in an inference-based system. Description logics, by their nature and their large use in ontology building (Baader et al., 2004), enables a formal and unambiguous description of expert rules. This base of rules enables a better interpretability and a better understanding of the results with respect to other techniques of machine learning, as it is possible to directly link one rule to an actual natural language situation. In this paper, a base of rules have been built with the help of several military experts.

With the multiplication of low-cost sensors, surveillance systems are on a rise, particularly collaborative systems that require little equipment. In the maritime domain, the Automatic Identification System (AIS) is a legally-enforced system put in place by the International Maritime Organization (IMO, 2003). As a large source of data on maritime navigation, this system is widely used for the understanding of the maritime situation. Its high rate of transmission and vast network of receiving antennas allow a large harvest of AIS messages that enable a precise tracking of vessels both on short and large geographic and

temporal scales. However, this system is very weakly secured and therefore is prone to issues and attacks such as erroneous information, data falsification and data spoofing. In spite of those issues, its data is largely used as a basis for maritime-based studies, without seeing its data quality questioned somewhat. In this respect, a data integrity analysis would allow to put into perspective the blind use of AIS information and highlight the main issues that the system face, so that action can be taken to mitigate the risks linked to an improper use of such maritime information and on a larger scale being in grade of assessing the type of issue faced by the system so that an user could take targeted action. Research have demonstrated that AIS is vulnerable, prone to spoofing (Bhatti & Humphreys, 2017), with missing (Lecornu et al., 2013), collided (Last et al., 2015), erroneous (Harati-Mokhtari et al., 2007) and falsified (Katsilieris et al., 2013) messages. Although few cases are reported (for example in gCaptain (2018) in East China sea, in Wired (2017) off the Russian coast and in Llyodslist (2019) in the strait of Hormuz), it has a concrete impact on maritime navigation.

The research question that arises and which is addressed by this paper is how it is possible to conceive an information system for decision support integrating anomaly detection and falsification-discovering mechanisms based on data quality in order to alert the user that the pieces of information displayed are possibly non-genuine. This research has been applied to an AIS dataset constituted of messages received by our antenna and parsed with an in-house parser.

In the following of this paper, Section 2 introduces the added-value of mechanism of anomaly detection in the decision-support system, addressing the issues of trust in information, data quality dimensions and methods for the detection of anomalous events. Section 3 presents the AIS (Automatic Identification System), which is the most important source of information on vessels at sea. As our study is based upon this system, its issues in relation with data quality are presented. Section 4 presents the proposed methodology for the design of an information system which assesses data, extracts relevant pieces of informa-

tion and presents them to an operator as a decision support tool, taking into consideration the specificities of the system studied, going from the assessment of data fields to the evaluation of some selected scenarios. Section 5 explains the implementation of the methodology and the way data is processed and results are presented from an architectural point of view. Section 6 illustrates, before concluding remarks, the results of some data analyses conducted with a 6-months AIS dataset, and a discussion on the conception, the results and the application of this system in a real-world case.

2. Anomaly detection for enhancing decision-support systems

In order to be efficient, an anomaly detection process must assess data within a predetermined frame which allows a classification of issues under a clear terminological framework. This section presents the basic definitions of trust, which leads to the selection of a subset of dimensions relevant for anomaly detection, presented and contextualised so that they can fit the frame of anomalous events.

2.1. Trust in information

The notion of trust in the source of information is important, as strong attention on the trustworthiness of the different sources is given by the users. Those sources can be a written document, humans or machines, easy or hard to access. The pieces of information can also be first-hand or second-hand. There is however no clear and straightforward definition of trust, and it tends to vary between people, or between domains (Blomqvist, 1997).

The simple access to the source is not sufficient to assess trustworthiness, and ideally the way in which sources are accessed by people must enable them to form an opinion about the source, and therefore to assess its trustworthiness (Hertzum et al., 2002). In a case where a user cannot collect information about the source, an absence of trust or even distrust can appear. With the development of computers and services, there is a tendency, in order to find information, to rely more and more on data and applications of the Internet. It

is possible to find an abundance of information, however in the wide spectrum of data sources, many information may present contradictory opinions about the same topic. So the users have to seek for hints in assessing the trustworthiness of online information.

Trust is fundamentally a social relation. As demonstrated by Denize & Young (2007), trust is thoroughly embedded in the processes of information exchange, communication and decision-making. Machines and sensors, which support these processes, shouldn't be trusted. However with the development of digital technologies, the users behave towards them in a way close to the one they would have behaved with another human being (albeit not similarly), and people rely more and more on information given by electronic devices (McKnight, 2005). So the use of technology is directly affected by the trust that the user has in it (Kelton et al., 2008). As technology is an addition of physical components and of programs encoded, both the digital (software) and the physical (hardware) parts can be assessed. As the digital part is composed of information, the trust in the technology corresponds to a trust in information (Kelton et al., 2008).

Trust in that context can be expressed through dimensions, representing data quality at large, which are presented in Section 2.2. For instance in (Costé et al., 2016), two dimensions have emerged as being important: the trustworthiness, which is the degree in which will the sensor be truthful, and the competence, which is the level of expertise of a sensor or system component in the proper subject.

2.2. Data quality and its dimensions

The quality of data can be divided in two parts, the external quality (the quality from the point of view of the user) and internal quality (qualities from the point of view of the supplier).

Internal quality generally lies on concision, clarity, generality, cohesion and simplicity (Devillers, 2004). For the transmission of data quality information, *metadata* are often used. Their use and understanding is however not easy, even for experts. A description of internal quality can be performed by answering

the question: "*how can I measure the quality of my data and how can I signify it?*". The internal quality is an absolute technical quality.

External quality covers, amongst others, ease of use, reliability, accuracy, conformity to the expectations, robustness and openness, so external quality can be considered as being the *fitness for use*, which worth answering the question: "*what are the needs of the user on data quality and information quality and how can I give it in order to prevent them from having an abusive use of them?*". Because of the multiple and various needs, external quality is more difficult to assess, as it implies the linking of data and its use, the expectations of the data users and the concerns of data producers (Vasseur et al., 2005). External quality is a relative use quality, measuring the ability to fulfill a particular need. Agumya & Hunter (1998) demonstrated that there is a strong link between the fitness for use, the acceptable risk and the risk response. Pierkot et al. (2011) defines external quality as "*the suitability of the specifications to the user's requirements. It is measured by the difference between the resource wished for by the user and the resource which has actually been produced*".

Data quality has been separated into twenty dimensions by Wang & Strong (1996), organised in four categories:

- the accessibility of data: accessibility degree, access security and cost-effectiveness
- the accuracy of data: accuracy degree, believability, completeness, objectivity, reputation, traceability, variety of data sources
- the relevancy of data: approximate amount of data, ease of operation, flexibility, relevancy degree, timeliness, value-added
- the representation of data: conciseness and consistency, ease of understanding, interpretability

For some activities, poor data quality can be a risk worsening factor, and endanger them. As the decision one takes is based on information that is available, poor quality data can then lead to poor decisions. In their use of information,

decision-makers can be influenced by several variables: their experience level, information overload and time constraints. Information overload happens when the amount of information is too important for the time available to respond. Consequently, the global quality decreases when there is not enough time for processing the incoming data. In this scope, it is particularly important to reduce the information load of the decision-maker, in order to draw the attention and focus on important features that need human operators.

2.3. Data integrity

A distinction between the integrity assessment and the veracity assessment of information may arise, as in the case of AIS messages, the integrity assessment represents the value associated to the trust we have that information within an AIS message accurately depicts the behaviour of the vessel with respect to the other messages that we receive and process, whereas the veracity assessment represents the intrinsic trustworthiness that we associate with the fact that the message is genuine and its pieces of information are true. In this respect, integrity relates to the nature of a piece of information with respect to a reference whereas veracity is mainly linked to the relation of data to the world. Veracity represents the fact for a datum to be truthful, *i.e.* to correctly depict the World in a way it is expected to (Iphar et al., 2019). Consequently, the evaluation of integrity though data assessment techniques is a means for the understanding a the overall problem that is data veracity. Nevertheless, due to the semantic proximity of those terms, and given that this distinction is not the main locus of this contribution, both integrity and veracity assessments will be referred as integrity assessments in the remaining of this paper.

2.4. Anomalous events and anomaly detection

Anomaly detection is an important part of data-related studies and is often based on aforementioned data quality dimensions. Associated to any study, a normality must be established as the assessment of an anomalous thing is relative, and a distance must be chosen for distance computation. In addition,

threshold triggering criteria must be put in place, enabling an actual discrimination of anomalies.

Several anomalies are distinguishable: the point, contextual and collective anomalies. In a point anomaly, an individual instance is considered as being anomalous with respect to the rest of data, in a contextual anomaly, an instance is not anomalous in a general assessment but becomes anomalous when the context is cleared and in a collective anomaly the data considered separately are not anomalous by themselves, but their occurrence together makes an anomalous collection (Chandola et al., 2009).

In anomaly assessment, pattern discovery is crucial as a pattern is by definition constructed by recurring elements, the repetition of which is predictable (Martineau & Roy, 2011). The terms of anomaly, non-standard, outlier or unusual can be used for each piece of information out of the frame, so which does not belong or seem not to belong to one of the clusters formed by the pattern analysis. The patterns can be a statistical distribution, a succession of events as a sequence or a cluster. If the pattern evolves over time it follows a dynamic model, if it does not it is said static. Machine learning, statistical methods and neural networks are amongst the usable methods for pattern discovery.

3. Use and weaknesses of a maritime identification system

The application case of this paper relies on vessels and maritime data. More particularly, the data analysed is sent by a specific system, the Automatic Identification System, implemented by the International Maritime Organization and with enforced use worldwide. This section aims at presenting this system with its uses and its misuses. The section ends with a positioning of the system with respect to the anomaly detection features as developed in Section 2.

3.1. A system for maritime data broadcasting

The Automatic Identification System is an information system for vessels transmitting information about the position, the kinematics, the physical characteristics of the vessel, its identity and information related to the safety of

navigation. Today, besides its initial purpose of collision avoidance, it has a widespread use (Fournier et al., 2018). The AIS helps mariners to better know their environment, it is used by coastal authorities to be aware of the traffic off their coast, by countries to be able to know the location of the vessels having their pavilion, by companies in order to monitor their fleet and by analysts or by researchers as a useful tool for the understanding of maritime traffic and its various hazards.

The Automatic Identification System was put in place by the Safety Of Life At Sea (SOLAS) convention, and some ships from the signatory countries are concerned by the deployment of this system. The SOLAS 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*" (IMO, 2004). Following this definition, all seagoing vessels are not obliged to carry the AIS, therefore relying only on this system provides a partial view of the maritime traffic. However, it is possible for vessels to carry the system although it is not compulsory for them

The transmission of AIS data is done in the Very High Frequency (VHF) bandwidth, on two worldwide dedicated wavelengths: 161.975 MHz and 162.025 MHz. In order to transmit and receive AIS signals, some dedicated devices have been put in place since the introduction of the system. Four main kinds of devices can be distinguished: class A transceivers (on the vessels for which AIS is compulsory), class B transceivers (on the vessels for which AIS is not compulsory), multi-channel receivers and radio scanner receivers (Iphar, 2017).

At first, the system was only terrestrial, with transmission occurring from one vessel to another, or between a shore station and a vessel, in a range of distance which is limited by the curvature of the Earth (circa 40 nautical miles in optimal conditions (ESA, 2012) for class A vessels), or the transmission power (5 to 10 nautical miles (Serry & Lévêque, 2015) for class B vessels). Recently, the development of low orbit satellites enabled to receive messages even far from the coastline, as it uploads and stores the received messages then download

information as soon as a coast line and a shore station is reached.

The development of the Internet gave an even more important step forward in the knowledge of maritime situation as websites display AIS information from all over the world[1]. So where ships previously disappeared beyond the skyline from a terrestrial point of view, they can now be tracked in the whole world by every person who can access the Internet network.

The rate of transmission, or the reporting interval of AIS message largely varies according to the type of vessel, its speed and the type of message sent and ranges, for a class A vessel, from 2 seconds to 3 minutes for positioning report messages. In one day, the European Maritime Safety Agency (EMSA) receives about 9 million terrestrial AIS and 7 million satellite AIS messages, from over 96,000 vessels detected by more than one source (EMSA, 2019) and Natale et al. (2015) estimates that in a month, and 130,000 vessels of all categories are sending those messages.

AIS messages have been designed to carry messages of various types, each one carrying a given type of information. In this respect, 27 different messages have been designed, each one having its own layout of data fields nature according to the type of information it is supposed to carry. The study of Tunaley (2013) proposes a separation in six categories of messages, namely standard, aid to navigation, timing, safety, binary and others.

The data inside AIS messages can basically be divided into three main categories: static, dynamic and voyage-related (Lundkvist et al., 2008). Static data are data fields which are not intended to change, or at least to seldom change, such as call sign, name of the vessel, length and beam, or the type of ship. Dynamic data are the pieces of information contained in the data fields which are expected to change over time, displaying a physical motion, such as the latitude, longitude, course over ground or speed over ground. Voyage-related data are pieces of information that are expected to change often, at each new voyage, such as the draught, the destination, the estimated time of arrival or

[1] *e.g. marinetraffic.com*, *aishub.net*, amongst others

the hazardous nature of the cargo.

3.2. The weaknesses of AIS

The AIS is an open system conceived and motivated by international authorities so that it could be used by the greatest possible amount of users. However this openness led to the lack of control of the system, and there are several ways in which the AIS fails to transmit genuine data: (1) issues due to the intrinsic weaknesses of the system, (2) errors in the messages, (3) falsified data in the messages (Ray et al., 2015) and (4) AIS signal spoofing (Balduzzi et al., 2014b). Those four ways are presented in this subsection.

3.2.1. The AIS has intrinsic weaknesses

Those weaknesses are linked to the system itself, without implying human interaction. The two main families of those intrinsic issues are missing data and message collision (Iphar, 2017).

The system in itself can fail in transmitting information. Some transponders fail to reach all the requirements set by the International Telecommunications Union, and some ships display large blank areas. This missing data, as shown in Lecornu et al. (2013), weakens the exploitation of AIS data by decreasing the reliability, but does not prevent it. The AIS has some critical shortfalls in additions to problems such as limited bandwidth and range: limited retransmit capabilities for a few messages and no retransmit capabilities for the majority (McGillivary et al., 2009).

Message collision is another weakness of AIS. A message collision occurs when a message is overlapping another one, partially or completely. All AIS signals are not received by the receivers, as there is a loss percentage, particularly in the case of satellite transmission (Eriksen et al., 2006). When the installation is correct, with a good-level hardware and a good weather, most loss is due to VHF transmission. About 2% of messages are lost due to channel overload (Last et al., 2015). But the biggest reason for message loss is the shadowing due to

obstacles (Last et al., 2015), either be on board the vessel (masks), or other vessels hiding more distant ones.

3.2.2. The system broadcasts errors

A part of the information contained in AIS messages is entered manually by the crew, both at the initialisation of the system for permanent data and at every new journey for journey-related data (Iphar, 2017). According to the study of Harati-Mokhtari et al. (2007), both static and dynamic data are subject to errors, and as each human-filled field is subject to errors, as well in static data such as identification number of the ship, name of the vessel that in dynamic data such as the navigation status, the estimated time of arrival or the destination.

Thus, the Maritime Mobile Service Identity (MMSI) number (main ship identifier used by the AIS) is false in an estimated 2% of the cases (Harati-Mokhtari et al., 2007). Also, the type of the vessel is often unclear. As 6% do not define a type at all, 3% define their vessel simply as "vessel" (Windward, 2014).

The name of the vessel is another issue, as 0.5% does not have a registered name, and some others exceed the allocated space in the field, which is 20 characters. Globally, only 41% of the ships report their destinations (Windward, 2014).

3.2.3. The system presents falsification cases

Intentional falsification of the AIS signal can be done for instance 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 (Iphar, 2017).

At sea, only vessels, buoys and relevant aids to navigation features must broadcast AIS messages. However, cases of fishing vessels putting AIS transceivers on fishing nets have been demonstrated (gCaptain, 2018), in order to force other vessels to modify their course off those nets.

Identity theft also exists in the maritime domain (Windward, 2014). It corresponds to the fact to navigate with a MMSI number which is not the real

one, allocated and internationally recognised, but with the one of another vessel that actually exists somewhere else.

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 unauthorised area, or trade illegal goods (Katsilieris et al., 2013) with other ships or on coasts.

In this respect, five main issues are developed by Windward (2014): the identity fraud, the concealing of destination, the fact to voluntarily stop the broadcast, the GNSS manipulation and the spoofing of the system, as the ability of an attacker to control a vessel under autopilot by spoofing the GNSS signal has been analysed and demonstrated in Bhatti & Humphreys (2017).

3.2.4. The system undergoes spoofing

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., 2014a). 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).

In the scope of spoofing capabilities, several threats can be taken into consideration: ship spoofing, aid to navigation spoofing, collision spoofing, weather forecasting, AIS hijacking and availability disruption threats (Balduzzi et al., 2014a). Cases presented in this section have been implemented in a proposed software and self-built transmitter, with built AIS frames (Balduzzi et al., 2014a), the resulting trajectory of which is presented in Figure 1.

In Figure 1, the result was received and displayed on the website *marinetraffic.com*, as a station of this network received the signal. Other kinds of attacks or tests have occurred to appear on this platform, where fake data (*e.g.* ships)

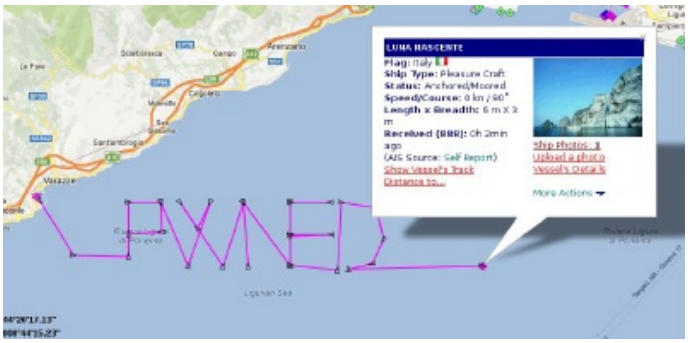

Figure 1: Example of a spoofed ship following a programmed path, from Balduzzi et al. (2014a) (in print, colour should be used for this Figure)

are persistent.

An attacker would be able to counterfeit information to blame someone else about an event, for instance a voluntary oil spill in the open sea, or the intrusion of a enemy vessel in the waters of another nation.

Availability disruption threats are three of a kind: slot starvation, frequency hopping and timing attacks. Slot starvation consists in impersonating the maritime authority to reserve all the slots, thus all stations within coverage have no slot available for reservation and emission. Frequency hopping is the fact to instruct the AIS transceivers to change their transmission frequency, as it is possible by protocol specification for given areas in the World. In timing attacks, the malicious user instructs transceivers to delay their transmission, by doing it repetitively, it prevents the system from functioning normally; and on the contrary, the attacker can command transceivers to send updates at a very high rate, thus overloading the channel.

3.3. Anomaly detection for AIS

As stated in Section 3.2, there is an issue about the Automatic Identification System in the way it transmits information in an unsecured way, with error, falsification and spoofing cases. As this system is widely used for navigation, security and evaluation of maritime domain activities such as fishing (Hu et al., 2016), vessel noise (Erbe et al., 2012), vessel emissions (Goldsworthy &

Goldsworthy, 2015), traffic modelling (Chen et al., 2015), emergency response (Schwehr & McGillivary, 2007) or animal collision (Wiley et al., 2011), one must ensure that the data used for such evaluations are genuine data, actually representing what it stands for. However, as AIS is open and multiple errors and misappropriate uses are possible, it is difficult to trust data transmitted by AIS.

Several methods are used and have been implemented for anomaly detection of maritime traffic using maritime communicating sensors, such as clustering and classification in which different behaviours are discriminated in different classes (Zissis, 2016), Bayesian networks in which vessel behaviours are categorised following the statistical-based theory of Bayes (Hadzagic & Jousselme, 2016), data driven path-finding algorithms for vessel estimated time of arrival computation (Alessandrini et al., 2018), event calculus for pattern discovery (Pitsikalis et al., 2018), hidden Markov Models in which this probabilistic model is used in order to discriminate various vessel routes (Zouaoui-Elloumi, 2012) (Yaghoubi Shahir et al., 2014), unsupervised route extraction in which routes are extracted from raw data based on vessel trajectories (Pallotta et al., 2013), genetic algorithms (Chen et al., 2014) or low-likelihood behaviour which is based on the measure of the behaviour expectancy from a vessel (Alessandrini et al., 2016). The development of those methods was facilitated by the rise of open data available from sea-going vessels (Kazemi et al., 2013).

In addition, in the specific case of AIS messages, the maritime environment constitutes a complex environment of study, with a great amount of elements consisting of an important amount of agents, the capabilities of which are restricted. As an example, vessel tracking is an essential and relatively well developed task for the understanding of maritime environment. This tracking is in general based on the fusion of data from various sensors such as AIS signals, imaging devices or radar signals, but every single device has a coverage area that varies (because of masks or weather) and that is limited and thus limits the global knowledge of the situation. The perception of some elements that can be hazardous is however limited (cargo, identities of passengers, identities of mariners for instance) which implies a limit of the detection of anomalies,

because an hypothetically perfect analysis would require a perfect knowledge of the various components serving as information sources on a perfectly known interpretative framework.

In this perspective, the determination of quality dimensions as defined in Section 2.2 is necessary, so as to ensure a proper assessment of AIS data. The data quality dimensions of accuracy, currentness, completeness, precision, consistency (Huh et al., 1990), integrity (Fox et al., 1994) and reliability (Brodie, 1980) have been highlighted as particularly important in the analysis of AIS issues by Iphar et al. (2015), and represent the cornerstone of the methodology presented in Section 4.

4. A methodology for integrity assessment of maritime data

As shown in Section 3.2, AIS messages present vulnerabilities in their structure and data, such as falsification, and those vulnerabilities can increase or lead to the creation of maritime risks. In this section, a method for assessing integrity of AIS messages is presented. In this method, a thorough examination of AIS messages leads to the identification of 935 integrity items, which are elements in which AIS data may disagree. In the complex AIS structure, it would be an indicator of an integrity issue. A system of flags, based on the one hand on integrity items and on the other hand on non-AIS data (contextual data such as fleet registers), has been developed, the goal of which being to highlight humanly understandable anomalies about the AIS, in the frame of some specified scenarios. Those flags are raised when a combination of integrity assessment item results are gathered. In parallel, the conjunction of some given flags will trigger some specific scenarios. The final purpose is to deliver, in near-real-time, information with added-value to maritime authorities and rescue centres.

4.1. Integrity assessment of messages

4.1.1. A variety of message and data types

As mentioned in section 3.1, the AIS messages are various in their nature, they can therefore be discriminated in various families, each one gathering sim-

ilar kind of messages, which will undergo similar integrity assessments as they will present similar data fields. Figure 2 presents several ways to perform an AIS messages classification.

Figure 2: Variety of AIS messages (in print, colour should be used for this Figure)

The left-hand side column of the Figure 2 displays the different possible kind of senders of messages. Indeed, some of the messages are only sent by base stations (which are shore-based stations or other non-vessel stations), some others are only sent by mobile stations, while a large number of the messages can be sent by both base and mobile stations. Given this distribution, it is not expected that a single station (individuated by its MMSI number) sends messages which do not match its category. In addition, the same column shows the messages sent specifically by class A stations (*i.e.* violet and blue ovals, not circled) and those sent only by class B stations (circled ovals). As vessels are

not expected to change their class, any MMSI is not expected to send any pair of (class A, class B) messages.

The central column displays the variety of AIS messages, as several kinds of AIS messages exist, and all messages belonging to the same family will tend to undergo similar studies. Moreover, when it comes to assessments involving several messages, any pair of similar messages will tend to propose similar items, as the same data fields that are involved in the comparison of the messages will be found in both couples of messages.

The right-hand side column of the Figure 2 shows three of the main messages families: the messages in which static data is provided, the messages in which positioning is involved and the messages in which a communication between two vessels is involved. For a message, the fact to have a positioning data (*i.e.* latitude and longitude fields) enables all position-related assessments. Similarly, the fact to have static data enables identity-related assessments and the fact to have communication data (*i.e.* source and destination MMSI numbers) enables all kind of analyses linked to the identities and locations of those vessels. The messages in grey colour of Figure 2 do not belong to any of those three kinds of messages families.

Not only is there a diversity within AIS messages, but the data within can take several forms. Amongst the data fields, the diversity can be illustrated by the message number 5 (static and voyage related data message). The fields of the message with the parameter represented are presented in Table 1, alongside with the type of datum and the nomenclature value, the meaning of which will be explained in section 4.1.2.

Six data types are then discriminated, which are: numeric representing an identifier (such as identification numbers of the vessel), numeric representing a physical quantity (dimensions of the vessel, or speed in another message type), numeric representing a choice (in a list of choices, such as the navigational status, where amongst others "0" stands for *under way using engine*, or "1" stands for *at anchor*), textual, date and binary.

Those data types are described by the AIS specification, and can be found

Field	Data type	Nomenclature
Message ID	Numeric representing an identifier	05A
Repeat Indicator	Numeric representing a quantity	05B
User ID	Numeric representing an identifier	05C
AIS version indicator	Numeric representing a choice	05D
IMO Number	Numeric representing an identifier	05E
Call Sign	Textual	05F
Name	Textual	05G
Type of ship and cargo type	Numeric representing a choice	05H
Overall dimension / reference for position	Numeric representing a quantity	05I
Type of electronic position fixing device	Numeric representing a choice	05J
ETA	Date	05K
Maximum Present Static Draught	Numeric representing a quantity	05L
Destination	Textual	05M
DTE	Binary	05N
Spare	Binary	05O

Table 1: Different data types in AIS Message 5

during normal use conditions. However, two additional cases must be taken into consideration: empty fields and default values. Empty fields often occur when a field has no value allocated, constituting an issue of data completeness. Default values exist in AIS messages and are also described by the system specifications. Any field with no allocated value will display the default value. For instance, in the case of message number 1, "181" is the default value for the longitude field, or "511" for the true heading data field (Raymond, 2016).

4.1.2. Integrity assessment items

As displayed in Figure 3, four ways to discriminate the inner integrity of the data within the fields of the 27 AIS messages can be distinguished. The first level consists of the assessment of the integrity of each field of each message taken individually. The second level is found at the scale of one single message, and assesses, in this very message, the integrity of all the fields with respect to one another. Given that messages of the same type have the same fields, it is possible to assess their integrity by comparing them, which makes the third level. Eventually, the fourth level consists in the comparison and the integrity assessment of the fields of different messages. Although pieces of information can

come from different messages, it is indeed possible to assess their integrity, due to the fact that some fields are either the same or linked or comparable. Those four ways will, in the following, be respectively referred as first-order, second-order, third-order and fourth-order assessments. The first-order and second-order assessments rely on one single message, and are therefore invariant with the environment, whereas the third-order and fourth-order assessments need several messages in data history to be assessed (at least one other, up to an entire time series for one vessel), and the outcome of those assessments can vary according to the environment (which includes the sample size or the location of the message within the sample).

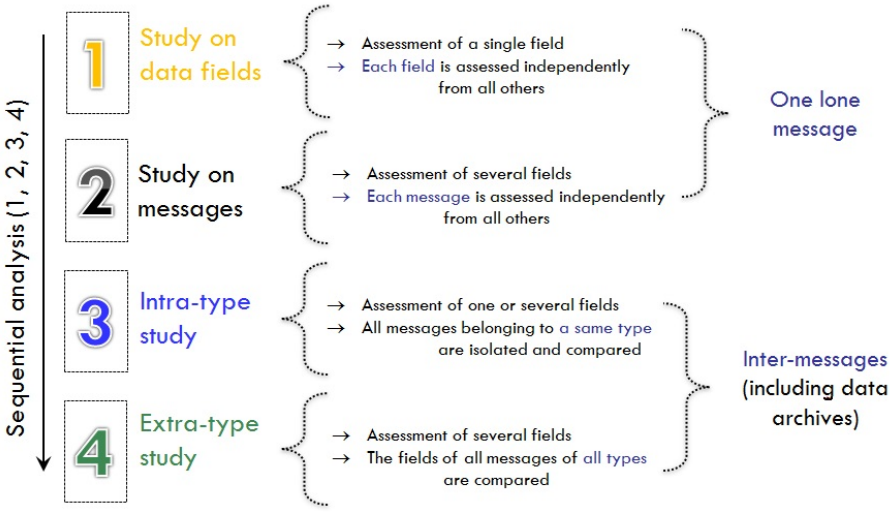

Figure 3: The four-order assessment (in print, colour should be used for this Figure)

The assessment of data integrity is performed through integrity items, which are statements, simple and unambiguous, involving one or several data fields. Each statement involves one field or several fields, either in the same message, in several messages in which data could be in discordance with specifications or in which several pieces of information within the fields could disagree, *i.e.* displaying two or more pieces of information that are not expected to be displayed in an expected functioning of the system.

In order to avoid any confusion in which data field is treated (as some fields are similar or identical in several message types) or which item is assessed (as some items, dealing with those similar or identical fields, will look alike), a nomenclature has been set to uniquely identify each data field from each message type, and each item from each order of assessment. Table 2 presents the message number 1 (scheduled class A position report) with all its data fields, their size represented by the number of bits allocated and their associated nomenclature (message number concatenated with a letter corresponding to the order the field in the message).

Nomenclature	N^O of bits	Field Name
01A	6	Message ID
01B	2	Repeat Indicator
01C	30	User ID
01D	4	Navigational Status
01E	8	Rate of turn
01F	10	Speed over ground
01G	1	Position Accuracy
01H	28	Longitude
01I	27	Latitude
01J	12	Course over ground
01K	9	True heading
01L	6	Time stamp
01M	2	Spatial manoeuvre indicator
01N	3	Spare
01O	1	RAIM-flag
01P	19	Communication state

Table 2: Nomenclature of data fields of message 1

4.1.3. Assessment classification

Two main families of assessments can be discriminated: those that assess conformity, *i.e.* the conformity to the AIS specifications of the AIS message, and those that assess coherence between different data fields in one or several messages. The integrity assessment of AIS messages uses both coherence and conformity items, as they are complementary items for the understanding of the maritime situation.

The conformity items encompass all the first order items and a marginal part of second order items (*e.g.* the message number 24, which is a message sent in two separate transmissions, so one can be received and not the other). In the first order items, the presence, in any field, of a default value does not constitute a conformity issue. However, what constitutes a conformity issues is the presence of an empty field where a value is expected.

The coherence items encompass all the remaining second order items and all the third and fourth order items. Within all coherence items, eleven families of items have been discriminated. Those families are presented in Table 3, with the orders to which those items can belong and a short description of their nature.

4.1.4. A logic-based formalism for integrity assessment

After the determination of the item list, each item must be rigorously assessed in order to check the conformity or the coherence of the fields within. A Boolean value is associated with the item to the message assessed, taking the value *True* or *False*, considering the assignment of this value as an answer to the question:

Is the statement expressed in the item demonstrating an AIS-data integrity violation?

Therefore, of the application of the item to a message demonstrates an integrity issue, then the value *True* is allocated to this item for this message, else it is *False*.

The essence of the item will not be assessed in some cases, for several reasons. Should it happen, as the integrity of the system has not been violated, the value associated to this item is *False*. For instance, third order algorithms require at least one former message of the same type from the same sender, if it is the first message received from this station, it does not constitute an integrity violation, in spite of the fact that the item cannot be assessed. The same reasoning applies for fourth order items with some rare messages: for instance, as the reception of a message number 13 is quite rare, the items involving message 13 data fields

Families #	O1	O2	O3	O4	Description
Conformity issues	X	X			Non compliance to the specifications
Inconsistent field values		X	X	X	Inconsistencies between two or more values are found, from the same message or from different messages
Data field evolution			X	X	The evolution of the value of a data field in several messages is not coherent
Motion evolution			X	X	Consecutive motion values between several data fields are not coherent
Unusual values			X	X	The value of one given field is not in accordance with the usual values this field takes when sent by this vessel in other messages
Overabundant reporting				X	The vessel sends a number of messages which to too important with respect to its kinematic values and the specification-defined transmission rate, in absence of any message 23
Overabundant communication				X	Two stations communicate too often between themselves
Remote communication				X	A communication between two stations which are supposed to be too far away from one another
Unexpected data field change		X	X	X	The value of one given field has unexpectedly changed with respect to the former message sent by this vessel
Position fixing device issue		X	X	X	The vessel displays whereabouts which are not compatible with the declared used position fixing device
Unexpected country location		X	X	X	The station is fixed and has whereabouts which are not in accordance with its country
Inconsistent response				X	Either the data field is part of a response message, however, the message that triggered this response is nowhere to be found, or the data field is an inquiry and the response is nowhere to be found

Table 3: Families of items and the assessment order(s) in which they are found

will be seldom assessed, and as a consequence each time that no message 13 shows up, the value *False* will be assigned.

Predicate logic present, under a formal form, the actions that lead to the integrity determination of an item in a rigorous and unambiguous way. Relying

on three main elements: the data fields values, the syntax and the expert knowledge values, a logic-based formalism based on predicate logic has been chosen for item assessment.

The data field values consist of the fields needed for the assessment of the item. According to Section 4.1.1, various data types can be involved, and their number depends on the assessed item, as it can require either few data or several fields.

The syntax is the whole of the logical elements that make the statements understandable and unambiguous. In this case, the selected elements are: $\exists$, the existential quantifier, !, the uniqueness indicator for existential quantifier, $\forall$, the universal quantifier, $\vdash$, the implication, $\neg$, the negation, $\leftarrow$, the attribution, $\in$, the affiliation, $\cup$, the union, $\cap$, the intersection, $\top$ the *True* statement and $\perp$, the *False* statement.

The expert knowledge consists of a set of values that have been set for each item in which it is necessary. Some items are straightforward, such as the ones assessing conformity, because with respect to the technical specifications, the data value is either in accordance or in disagreement. However, for the determination of items in which continuous data such as speed or location are used or for which distances are computed, a threshold value between the *True* and the *False* value must be determined. In this perspective, the knowledge of an expert of maritime navigation is used for the establishment of those thresholds.

From this point on, M_x stands for the set of all messages number x, m stands for a single message, R_m^z stands for the result of the assessment of item z on message m, D stands for the set of data field values (a list of fields, set in accordance with the need), T_R is a time interval representing the chosen assessment reference time (T_R standing for $T_{Reference}$), T_A is a time interval representing the current assessment time (T_A standing for $T_{Assessment}$)(*i.e.* in the analysis, all messages received during T_A are assessed, using all the messages received during T_R as our archived message database. An in-depth explanation of this mechanism will be presented in Section 5.2.3).

Two examples are provided here, one very simple and one more complex. In

the simple one, the purpose is to check if the field longitude (01I, as defined by nomenclature, *cf.* Section 4.1.2) is within $[-90, 90] \cup \{91\}$, which is its expected range of values (because the extent of longitude values is between -90 and 90 and the default value is 91). In the other one, the purpose is to check whether the whereabouts, represented by the longitude (01H) and the latitude (01I), are in accordance with the kinematic values of the messages which are the course over ground (01J), the speed over ground (01F) and the rate of turn (01E). This item uses additional functions, named f and g in this item, for trajectory planning (the description of which is not the purpose of this section).

Example 1: Item 01S05: *Value of the field 01I is less than -90 or greater than 90 and not equal to 91*

$$\forall m(D,t) \in M_1, D = \{id, lat\}, t \in T_A$$
$$((lat \in [-90, 90] \cup lat = 91) \vdash R_m^{01S05} \leftarrow \bot)$$
$$(\neg(lat \in [-90, 90] \cup lat = 91) \vdash R_m^{01S05} \leftarrow \top)$$

Example 2: Item 01I05: *01H and 01I positional field values evolution is not consistent with kinetic values in 01F, 01E, 01J and time*

$$\exists f : [-180, 180] \times [-90, 90] \times [0, 102.2] \times [0, 4.21] \times [0, 360] \times [-180, 180] \times [-90, 90] \rightarrow \mathbb{R}^+$$
$$\exists g : [0, 102.2] \times [0, 4.21] \times \mathbb{N}^* \times \mathbb{N}^* \rightarrow \mathbb{R}^+$$
$$\forall m(D,t) \in M_1, D = \{id, lon, lat, speed, rateturn, course\}, t \in T_A$$
$$((\exists! m'(D',t') \in M_1, t' \in T_R, t' < t, D' = (id', mmsi', lon', lat'), mmsi = mmsi', \min_{\forall t' \in T_a}(t' - t)) \vdash$$

$$(\Lambda = f(lon, lat, speed, rateturn, course, lon', lat'), \Omega = g(speed, rateturn, t', t) :$$
$$(\Lambda < \Omega \vdash R_m^{01I05} \leftarrow \bot),$$
$$(\neg(\Lambda < \Omega) \vdash R_m^{01I05} \leftarrow \top)))$$
$$(\neg(\exists! m'(D',t') \in M_1, t' \in T_R, t' < t, D' = (id', mmsi', lon', lat'), mmsi = mmsi', \min_{\forall t' \in T_R}(t' - t)) \vdash R_m^{01I05} \leftarrow \bot$$

Although only two examples are shown in this section, all 935 items from all 27 messages have been formalised under this logic-based formalism.

4.2. Falsification scenarios

4.2.1. Integration of contextual information

A sole focus on data coming from the AIS itself is interesting for the integrity evaluation of the AIS system, and would be sufficient if AIS were an isolated system. However as a vessel carrying this system evolves in an environment subject to changes, it is always useful to rely on additional data, allowing a more accurate study. A complete understanding of a situation sometimes needs several points of view, and one sensor might not be sufficient to discriminate a situation considered as normal from a situation considered as abnormal. Indeed, a situation considered as abnormal with respect to one given system data might be explained from another source, and *vice versa*, an expected situation from the point of view of the AIS can be highlighted as abnormal in light of external data.

Basically, every source having a common data field with the AIS can be used. As the system covers a wide range of information, those complementary sources can be varied, coming from different domains. In general, the establishment of an exhaustive list of such usable sources is not possible because of several reasons: the sources evolve, appear and go out of date in an unpredictable pattern, it is not possible to be aware of all available sources on a given subject, and the need for the use of a given source varies largely according to the type of study conducted.

Such contextual information can be split into vessel-oriented and navigation-oriented data (*cf.* Section 5.1). Vessel-oriented information typically contains reference information about ships (name, size, owner, ...) that enables a comparison with AIS data while navigation-oriented data focuses on geographic features helping to understand the ship's navigation. This includes, for instance, traffic separation schemes, aids to navigation such as the fairways or the navigational lines but also coastline or the location of ports.

4.2.2. Selected falsification scenarios

When an integrity assessment of the system is performed, falsifications scenarios can be considered. In general, systems can be falsified, therefore, pointing out the different cases in which such a falsification can happen is an important task. A falsification being the fact either to transmit erroneous data or to trick the system by making it behave in a way it is not supposed to, a falsification scenario can take several forms and will be either one particular falsification, one particular way to change data, the ingestion of false data or forcing the system to behave the wrong way.

A variety of scenarios are possible in the case of AIS falsification and spoofing. A selection of representative falsification scenarios is presented in this Section[2]. The scenarios are presented in the Table 4 with a short description of each of them.

Case #	Scenario name	Description
1.1	MMSI	Station has an irregular MMSI number
1.2	Identity issue	Vessel displays an identity incompatible with complementary data sources
1.3	Identity change	Vessel has changed one of its identity data fields
1.4	Ubiquity issue	Station displays various whereabouts at the same time
2.1	Wrong position	Vessel displays an impossible location
2.2	Kinematic inaccuracies	Vessel positional values are in disagreement with kinematic values
2.3	Disappearing/Reappearing vessel	Vessel has unexpectedly disappeared for an unusual time
2.4	Spontaneous unexpected appearing	Vessel has appeared in an unexpected area
3.1	Message 22 alert	Station broadcasts a message number 22
3.2	Message 23 alert	Station broadcasts a message number 23

Table 4: Considered falsification scenarios

[2]A few others have been implemented in the frame of the DéAIS project (Ray et al., 2015) in which this work is included, including scenarios linked to the number of messages received by one station by unit of time, the analysis of the signal or the number of messages received from one given MMSI number (saturation)

The first category of cases deals with static information and identity data of the vessels (*i.e.* scenarios 1.x of Table 4). In this category, we gathered the issues related to the MMSI number, the identity change (that might be normal but can be suspicious) and the ubiquity issues (which consists of the fact to receive positions that are too remote from one another, from one single MMSI, in a small timeframe). The second category gathers analyses upon all spatio-temporal information of AIS messages (*i.e.* scenarios 2.x of Table 4), and the scenarios selected deal with the wrong position of a vessel (*e.g.* inland reporting), the fact to disappear and reappear in unexpected location (*e.g.* in the case of a voluntary switch off of the system), kinematic inaccuracies (position values in consecutive messages not in accordance with speed, course and turn values) or the fact to spontaneously appear in an unexpected location. Third, a category considered in this paper concerns two AIS management messages which are amongst the most peculiar messages of the system: the message number 22 (channel management) and number 23 (group assign command). Those messages, only sent by base stations, send operational parameters to mobile stations which are of paramount importance: they assign and can change the frequency of transmission (more particularly the transmission channel) in the case of message 22, and force a transmission interval or a forced quiet time to mobile stations in the case of message 23. Those messages can be sent to specified vessels (assigned mode) or to all vessels in coverage (broadcast mode). In this latter case, several vessels can be affected by a single management message, that can heavily hinder the ability of the system to properly operate.

4.2.3. Definition of flags

Boolean flags based on expert-based inference rules are used to highlight anomalies. Each flag stands for a fundamental explicit case of integrity breach in the data assessed, and takes the value *True* if a problem is spotted according to the relevant associated items and *False* (default value) if no problem is spotted.

In the scope of the study of the AIS, four kinds of flags have been defined, two belonging to the family of the flags linked to integrity assessment items

and system data (the ones for which the number does not vary): the integrity assessment items flags and the vessel type flags; and another two belonging to the family of the flags directly linked to contextual data (so for which the number of flags varies with respect to available data): the scenario-specific flags and the maritime situational indicators flags. Two of those four classes of flags are presented in the following section.

4.2.4. Flag assessment

Flags linked to the integrity assessment items. In Section 4.1, a method for determining the integrity status of every single assessment item was defined. This method treated data fields separately, therefore it was not possible to easily extract any information from it. However, as it was showed in Section 4.1.3 that items can gather around categories, the extraction from each set of items (corresponding to each message type) of issues that are humanly easily understandable, which are the flags presented in section 4.2.3, is of interest.

Each flag stands for a specific issue in the analysis of AIS messages, and for each of the scenarios, a list of corresponding integrity assessment items have been established, the results of which must be evaluated in order to get the outcome of the flag computation. The list of integrity items for each flag is fixed, and the selection of flags which directly use integrity items results is fixed for each scenario. As a consequence, the list of integrity items needed for each scenario can be easily deduced by gathering all items of every single flag of the given scenario.

For example, in the case of the *remoteness* flag (excessive communication distance), there are 17 different items corresponding to this flag in the case of message type number 1. If only one of those items display a *True* value, then the *remoteness* flag will be set to *True*.

Scenario-specific flags. Those flags are totally dependent on the available external datasets, and each flag will be tied to the content of the database itself. Therefore it is impossible to set a fixed list of those scenario-specific flags, as

their number and nature vary according to the available databases. The fact to use such contextual information is particularly important in order to be aware of the environment of the system, and the assessments provided are as various as data coming from the system enable it.

Each flag is associated with one particular assessment type involving both AIS data and contextual information (*i.e.* it is necessary to query both system and non-system data before assessing the item), then the computation of the result is performed by a specially designed algorithm. As a consequence, it is not possible to write a general assessment program but it is needed to adjust the program to the data structure and type of the contextual dataset.

An example of such scenario-specific flags is presented in the following of this section, with the involvment of a fleet register.

Example: *f_fr_consistency*

This exemple assesses the conformity of AIS data with a given fleet register, in our case the European Union Fishing Vessel Fleet Register[3], which is publicly available and contains the list of EU fishing vessels. In this database, the fields in common with AIS are the call sign (which will serve as foreign key, usable for a join), the vessel name and the vessel dimensions (which will be the values to be compared).

Let B be the EU fishing vessel database, b be an element of B, ϵ be a Boolean standing for the fact for B to be exhaustive ($\top$ = exhaustive), $Dist^{\alpha}$ be a semantic distance (here an Edit distance), $Dist^{\beta}$ be a Minkowski distance (here a Manhattan distance), Ξ and Υ be the respective expert-defined thresholds for semantic and Minkowski distance for data compliance.

$$\forall m(D,t) \in M_5, D = \{id, callsign, name, dimensions\}, t \in T_A$$
$$((\exists! b(D_b) \in B, D_b =$$
$$\{callsign_b, name_b, dimensions_b\}, Dist^{\alpha}(callsign, callsign_b) = 0) \vdash$$
$$((Dist^{\alpha}(name, name_b) < \Xi \cup Dist^{\beta}(dimensions, dimensions_b) < \Upsilon) \vdash$$

[3] http://ec.europa.eu/fisheries/fleet/index.cfm

$$(f\_fr\_consistency \leftarrow \bot),$$
$$(\neg(Dist^{\alpha}(name, name_b) < \Xi \cup Dist^{\beta}(dimensions, dimensions_b) < \Upsilon)) \vdash$$
$$(f\_fr\_consistency \leftarrow \top)),$$
$$(\neg(\exists! b(D_b) \in B, D_b =$$
$$\{callsign_b, name_b, dimensions_b\}, Dist^{\alpha}(callsign, callsign_b) = 0) \cup \epsilon = \top) \vdash$$
$$f\_fr\_consistency \leftarrow \top$$
$$(\neg(\exists! b(D_b) \in B, D_b =$$
$$\{callsign_b, name_b, dimensions_b\}, Dist^{\alpha}(callsign, callsign_b) = 0) \cup \epsilon = \bot) \vdash$$
$$f\_fr\_consistency \leftarrow \bot$$

Other flags. In our analysis, all vessels must not be considered the same, as a fishing vessel is very different from a cargo vessel. Therefore, vessel type flags allow to discriminate vessels, so several vessel types have been set, and each one of those types has a flag which is *False* if the vessel is not of the type in question and *True* if the vessel is of the type in question. As the data type is part of AIS static message information, it is possible to assess it easily. In addition, flags have been set to describe maritime situations occurring at the time of the message. Those flags, backed on Maritime Situational Indicators (MSIs, defined in Jousselme et al. (2016) as descriptive patterns of maritime activity such as "*vessel in under way*" or "*vessel loitering*") allow to take into consideration the environment of the vessel, its location and the surrounding environment in order to get a more comprehensive analysis of the situation assessed.

5. Implementation

This section presents the implementation of the methodology introduced in Section 4. The first part of this section describes the reference dataset designed for the experiments. Then the architecture of the system is described. Finally, the data processing method and the workflow of data within the developed information system are presented.

5.1. Data

The dataset contains three categories of data: AIS data, vessel-oriented data and geographic data (mainly linked to navigation) and provides ship messages issued from the Celtic sea, the north Atlantic ocean, the English Channel and Bay of Biscay (France).

AIS data. The core of the data used for experiments is based on the 27 AIS messages types received by a terrestrial station located in Brest roadstead (France). The receiving station (VHF antenna, AIS receiver, Linux computer) collects AIS messages from a great part of the roadstead, from the entering and exiting traffic and on the passing-by traffic in the Ushant Traffic Separation Scheme (TSS). Figure 4 shows on the left, the location of the receiver (yellow star) and its theoretical range (blue polygon). The right part of the figure shows the real spatial extent of localised AIS messages during a time span of six months. The data (all messages) received by this antenna from October 1^{st}, 2015 to March 31^{st}, 2016 is used for our study.

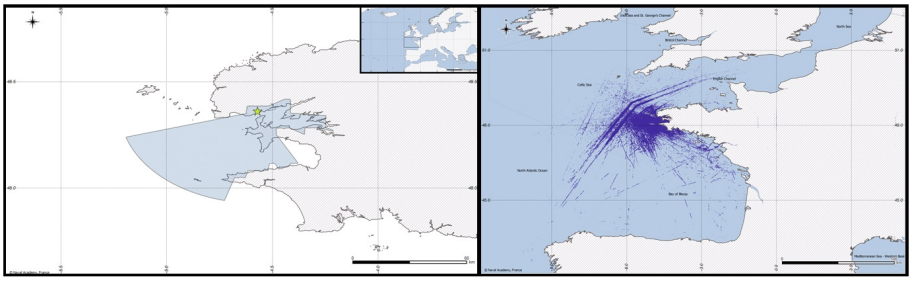

Figure 4: A view of the location of the geolocalised points in our AIS dataset messages (in print, colour should be used for this Figure)

The dataset consists of circa 24 million messages, 94% of them being geolocalised messages and 5% being static information messages, as shown in Table 5, which also display the number and percentage of messages per type of emitting station and families of messages. Message number 1 is by far the most used, representing 62% of all messages, before messages number 3 (13%), number 4 (12%), number 18 (4%) and number 5 (4%), which are the only messages to

have a frequency greater than 3%. In our dataset, 71% of the messages number 1 are within 10 km of the reception antenna.

Message Family #	Number	%
Total	24,033,893	100
Per Content		
Geospatial	22,493,074	93.6
Management	2,798	0.01
Static	1,084,275	4.5
Per Emitter		
Mobile station only	20,369,720	84.8
Base station only	2,803,972	11.7
Mobile and base stations	860,201	3.6
Per Message Type		
Standard	20,570,972	85.6
AToN	505,764	2.1
Timing	2,807,055	11.7
Safety	46	ϵ
Binary	150,044	0.6
Other	12	ϵ

Table 5: Number of messages received by a terrestrial receiver

Falsified AIS data. Genuine AIS messages of the dataset natively contain errors and misconfigurations. They also contain several falsifications (*cf.* Section 6). However, some behaviours involve rare or never received messages, other require a condition on data which is rare (for instance a weird-looking trajectory involving AIS location on shore such as the one presented in Figure 1). In order to test, evaluate and validate algorithms and specific scenario cases under reference data, controlled degradation of data has been also performed (Iphar et al., 2019). Our approach relies on two degradations: first, original AIS data has been manually or automatically modified. Second, some AIS frames or sequences AIS frames were created intentionally and injected in the dataset as described in Figure 6. Figure 5 shows a typical fake trajectory specially designed to activate many items and flags at once (wrong speed, heading, ubiquity, ...).

The building and use of those fictive frames allows to generate any falsification scenario. Thanks to an emitter platform based on a Software Defined

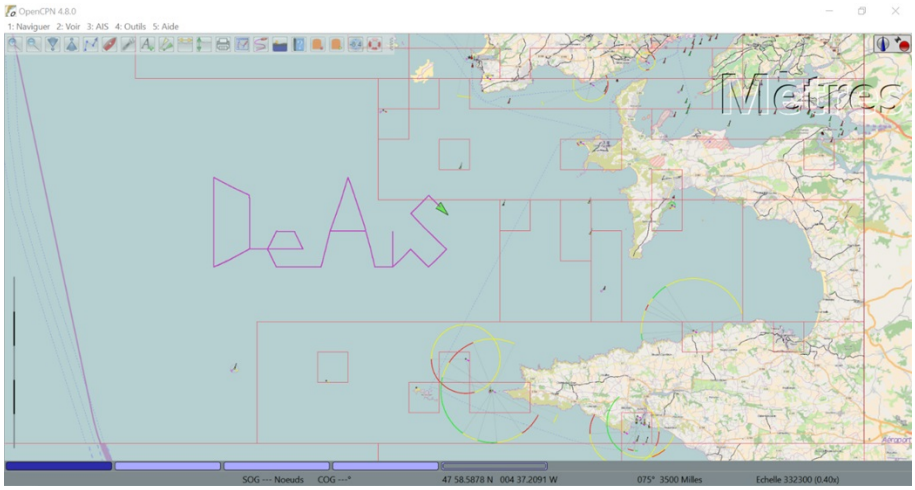

Figure 5: A fake trajectory (*visualisation based on OpenCPN (opencpn.org) and OpenStreetMap (openstreetmap.org)*) (in print, colour should be used for this Figure)

Radio (SDR) we designed similarly to Balduzzi et al. (2014a), false messages can also be broadcast live with real AIS flow (Alincourt et al., 2016). During experiments, because of their potential threat to navigation, all falsified messages have been either broadcast within a laboratory platform with very low power or piped directly within our reference database (in the middle of real historical messages).

Vessel-oriented and geographic data. As stated in section 4.2.1, two kinds of complementary data are discriminated: vessel-oriented and navigation-oriented. In our study, because we use data from a Brittany-based station, some sets of data have a limited spatial extent around our point of interest, whereas other are at larger scale, even worldwide. All contextual data prepared for the dataset has been temporally (when applicable) and spatially aligned with the extent of the AIS data assessed. First, receptor-specific information has been added (*cf.* Figure 4), such as the coverage areas of the receptor (the theoretical one, with respect to the local topography models and the Earth curvature, or the real one, with respect to data reception, which may vary with the meteorological conditions, the season or the time in the day), as well as the location of AIS data

receptor itself. Additional data prepared for our study includes, for instance, Natura 2000 protected areas, anchorage and restricted areas, polygons of Brest port and roadstead, two fleet registers, the location of ports of Brittany, the coastline and the Ushant traffic separation scheme. An extended release of this dataset also including local weather conditions and sea state is available (Ray et al., 2019).

5.2. Architectural principles

Based on the methodology presented in section 4, an information system has been developed for the detection of AIS falsifications. The system is designed to handle both real-time asynchronous and offline analysis of messages and works with both streamed and historical data. While in the experiments, only one receiver has been used, the information system has been designed to cope with multiple AIS receivers and one central database. Figure 6 shows the different components of the system.

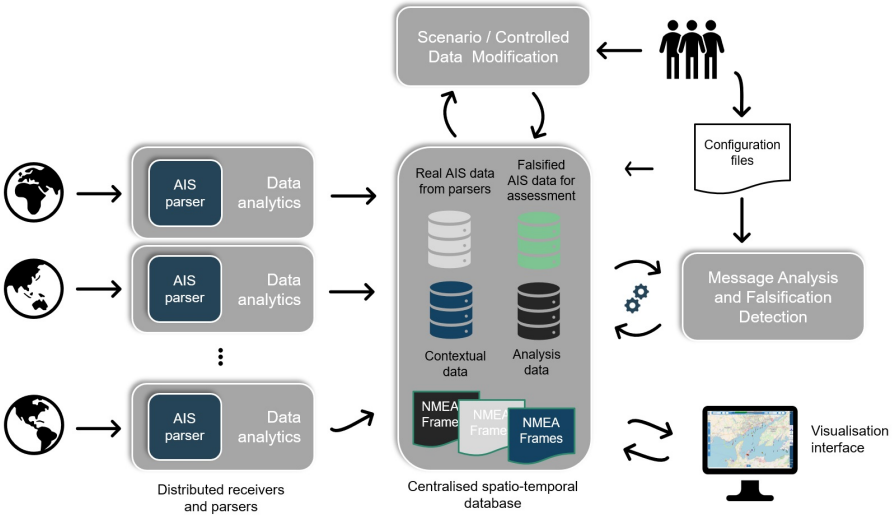

Figure 6: Information system (in print, colour should be used for this Figure)

5.2.1. Data integration

Most AIS messages come without any timing information because the AIS has been initially designed as an anti-collision system to be used in real-time. First, a receiving station timestamps the messages in UTC format immediately upon reception. In a second step, the parser reads AIS frames, extracts parameters of the message and stores parsed information in the central database. Additionally, the parser exports the timestamped messages, in the raw and unparsed format received to the central server which also stores all the messages (*i.e.*, all the 27 different message types described in the ITU-R.M 1371-4 or NMEA 4.0 specification) in text files (one file per day). This parser, written in Java, extends and adapts the CC-BY-NC-SA 3.0 parser aismessages[4]. The additional functionalities developed include: the connection to the database, ingestion, and outputting of UDP (User Datagram Protocol) streams and files (used by the AIS), automatic folder import, data logger for unparsed messages, data export translated to the standardised TAG Block format, data analytics about the receiving flow of messages.

5.2.2. Data management

For data storage and manipulation, a database management system was used, because of the ability of such systems to find, write, sort, modify or transform data in complex databases, while ensuring the user a level of robustness of the analysis by avoiding partial assessment or information loss. The choice of the widespread and open source relational database management system PostgreSQL was made, using the SQL querying language, with the adjunction of the PostGIS extension, for the treatment of spatial features. The main contents of the database are presented in Figure 7.

The database gathers several main elements organised per database schemas: The AIS messages schema contains 27 tables corresponding to each message type and a table for all unparsed messages (error table).

[4]https://github.com/tbsalling/aismessages

Contextual data schema consists of all vessel-oriented and geographic data useful for the analyses of scenario cases, as described in section 4.2.1. Amongst data in this schema, the receptor data table store all information about the receiver used, such as the type of material used, the receiver location and its theoretical coverage. While theoretical coverage of a receiver is fixed, its practical coverage change every day especially because of weather conditions. A specific practical coverage called black hole has been proposed to highlight daily uncovered areas from where no AIS messages are expected (Salmon et al., 2016).

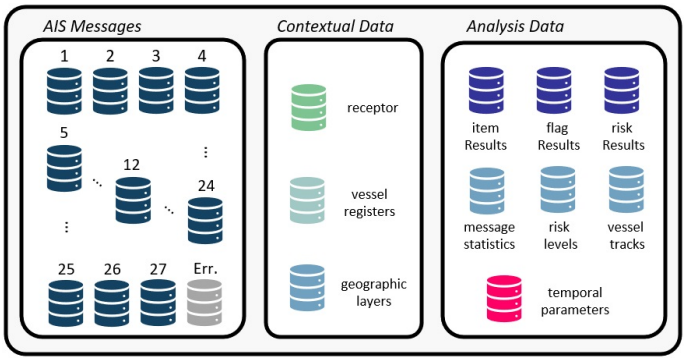

Figure 7: Database content (AIS message tables combine both real AIS data and falsified AIS data for assessment) (in print, colour should be used for this Figure)

A third schema (analysis data) contains several "working tables" that gather all information related to data treatment (especially temporal parameters describe below), statistics about message flow and the results of analysis. Vessel tracks maintain an exhaustive list of seen vessels through a quadruplet of nominative information (*i.e.* MMSI, name, IMO and callsign) especially for the flag *f_quadruplet* (*cf.* Section 6.1.1). It was chosen to store every intermediate result in the database right away after each assessment. This enables the systematic database querying by the software particularly for comparisons between assessments. In total, the database stores: the item computation results (Boolean values) and the flags after scenario computation (Boolean values). Tables related to risk levels and risk results (integer values), unrelated to the content of this article, are also present, in the prospect of a further analysis of the risk on

maritime navigation.

5.2.3. Asynchronous batch data processing

Message analysis and falsification detection constitutes the core component of the information system. The Python programming language was chosen for its development. Beyond the availability of several libraries (*e.g.* maths libraries for statistical computations) that favour our data analysis, Python is easy to handle and it enables database querying with embedded SQL. The program works using an incremental and a sliding temporal window in which items and scenarios are assessed. The processing follow the four-order assessment model described in Figure 3 and is organised in two steps further detailed in this section. In a first step, the program queries the database to fetch all the needed information and messages in a given time window. In a second step, it executes a batch processing of data to assess integrity issues and detect falsification cases by computing flags. Those two steps constitute a processing loop, as the process is triggered again after a waiting time, and a new computational first step occurs. The corresponding data analysis workflow is described in Section 5.2.4.

AIS messages are collected continuously (average velocity of our receiver is about 77 messages per minute), parsed and stored in a database together with analytic values of the flow. While it would be possible to stream and analyse every single message on-the-fly, it has been chosen to perform an asynchronous batch processing within temporal windows working on the database.

This choice is adapted to a centralised database architecture where duplicates can occur in the cases where the same message is received by different stations. It also enables the creation of temporal series of messages from the emitter, easing series-based analyses compared to historical data. The need for this approach is also directly drawn from the data treatment process, where a group of messages with timestamps between given bounds (for each of the 27 message types) are consecutively assessed for the items, the scenarios and for the flags.

In such batch processing, newly arriving messages are collected into a group

of AIS messages. The whole group is then processed at a future time. The time when each group is processed can be determined in different ways. For instance, it can be based on a fixed time interval (*e.g.* every minute) or on some triggered conditions (*e.g.* process the group according to the frequency of received messages, *i.e.* once a given amount of messages has arrived).

Each processing loop is characterised by three timestamps, defining two temporal windows: the incremental and the sliding temporal windows. The lower bound of the incremental temporal window consists in the lower temporal bound for queries on historical data (which are required in some items). The lower bound of the sliding window is the lower temporal bound of the timespan considered for data processing during the current loop. The upper bound of both incremental and sliding temporal windows is the upper temporal bound of the timespan considered for data processing during the current loop. At each new loop, the bounds of the sliding window are renewed, in such a fashion that consecutive loops have consecutive temporal windows, in order to ensure a thorough assessment of all received messages.

The system depends on a series of variables stored in configuration files, that are to be set beforehand (cf. Figure 6). Those parameters are mainly the item list, the scenario list and the temporal parameters (*e.g.* waiting time).

As the AIS system itself it not fixed, the program is evolutive and was conceived in a way enabling an easy enhancement and evolution. Indeed, the list of items is not fixed over time, as the AIS system still evolves, new items and analysis might come up and be included in the program.

5.2.4. Data Analysis Workflow

Message analysis and falsification detection component loops on several steps: first it updates temporal windows, then it computes item assessment and flag assessment which are detailed in this section.

Item assessment. Amongst formalised integrity items (935 in total), a total number of 666 have been successfully implemented into our system. These

items are spread amongst the different levels of the four-order assessment model. Several elements of the database are involved in this assessment process, namely the AIS messages, the configuration tables (*e.g.* temporal window parameters) and results tables.

Each assessment starts reading a configuration file in which a set of items to be computed are listed (up to 666 valid ones). The reading of each item triggers one of the four following cases depending on the item level:

1. The read item is either of order 1 and 2, so the querying to the database will involve only one message;
2. The read item is either of order 3 and 4 so several messages, possibly of different types will be queried;
3. The read item has a bad format (non-existent item). The algorithm, in this case, returns the information that this item has not been treated;
4. The read item has no value. This ends the loop (the program stops or another loop is going to start with the next temporal window).

In both *(1)* and *(2)* cases, a table of results for the item is created in the database. Then the relevant AIS message database is queried for the data fields of interest for this item and for the temporal span defined by the working window. Once the values are returned, a loop occurs on it, treating all the messages within the working temporal window one-by-one. From this point on, the processing varies with the order of the algorithm.

In the *(1)* case, the values are directly assessed by the corresponding algorithm, and the result is filled in the item result table once all messages have been treated.

In the *(2)* case, another query to AIS messages tables is necessary in order to get all the necessary pieces of information from other messages, which can come either from the same message type (in the case of order 3 item) or from another message type (in the case of order 4 item). Once the result of the query stored, all necessary data are present and the assessment can occur, followed by the filling in of the table gathering all results in the database, once all the

messages have been treated.

Flag assessment. A flag assessment is based on the analysis of AIS messages (item assessment as described above) together with contextual data sources (geographic, navigation-related data...). This assessment is driven by temporal windows and by the configuration file describing the list of algorithms to run for each scenario. It also specifies the data sources to involve in computations (AIS messages, external information, various results of the item algorithms). The flag assessment results are stored in dedicated database tables.

First, the temporal extent of the working window is obtained so that the messages to be assessed are known. A assessment is then performed on all falsification scenarios selected by the user.

For each of the scenarios, a flag table (in the database) corresponding to the scenario in question is created, and the relevant item results are queried and stored in the program, enabling further flag studies. Then, for each of the messages in the working window, and for each of the various assessments that lead to the determination of a flag, the same process is repeated.

In each of those processes, the scenario function calls the algorithms that correspond to the given assessment, and this algorithm successively calls relevant AIS and contextual data sources in order to perform a computation that leads to the assessment of the flag, and its eventual raising if the conditions are gathered. The flag results are then returned to the scenario function, and once all assessments for each message have been performed, all the computed flags are stored in the database, in a purposely created database table.

5.2.5. Visualisation interface

Geographic views and visual analytic capabilities have been demonstrated as offering a solution to the display of relevant information to the user, amongst the ever-growing amount of data that the systems have to process (*e.g.* Varga et al. (2017)). As cybersecurity issues are growing, visualisation emerges as a solution and it is particularly important to use it in a right way so that the

relevant data is presented to the user in an unambiguous way minimising false positive information.

The visualisation interface introduced in Figure 6 is designed to highlight integrity issues detected by the system. The web-based interface (Figure 8) enables a dual display showing a map of the maritime traffic but also the list of AIS messages with detected anomalies and associated risks. The interface includes a cartographic layer, a few data analytics and a text-based listing of detected features.

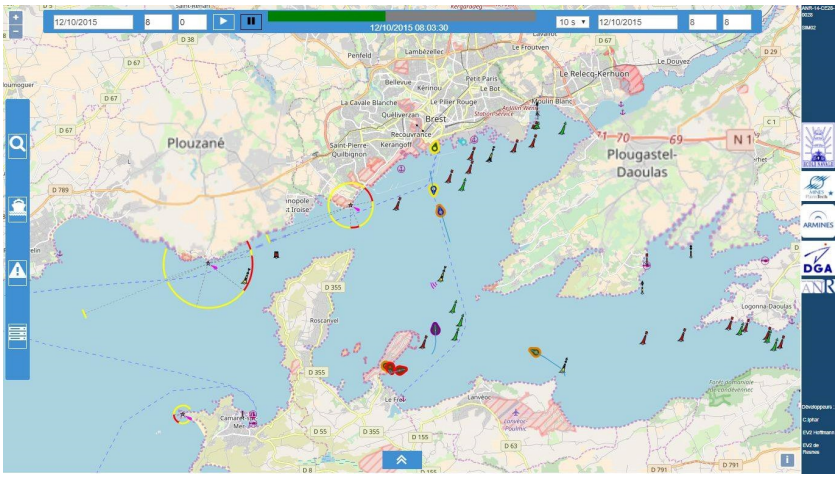

Figure 8: Web-based interface. Ships are coloured depending on the analysis of their AIS messages (in print, colour should be used for this Figure)

The map relies on two layers: the cartographic layer and the data layer. The cartographic layer constitutes the background of the interface, consisting of Open Street Map (OSM) tiles, enhanced by Open Sea Map (OSeaM) features. The data layer is made of points which are the vessels that have been selected as deserving particular attention by the program. This interface takes as input data the results of item and flag assessments described in Section 5.2.4.

6. Results

This section presents some results of computations done on a subset of the real dataset presented in Section 5.1. First a selection of all the flags that have been implemented is presented. The result of the analysis on those subsets is presented, with the number of corresponding flags raised, and the visual feature that has been implemented in order to visualise alerts raised by the detection system.

6.1. On Various Scenarios

6.1.1. Selected flags

As it is not possible to display the results of all the flags, a subset of flags has been selected, in order to cover the diversity of scenarios put in place. In total, 23 flags have been implemented and 8 have been chosen for testing and validation of results. Table 6 summarises the characteristics of those chosen flags. In addition to those flags, the flags linked to the vessel type have also been computed. In Table 6, the scenario column displays the number of the scenario described in Table 4.

Flag name	Sce.	Description of anomaly
f_country	1.1	MMSI has an invalid country code
f_fr_consistency	1.2	AIS inconsistent with available fleet register data
f_quadruplet	1.3	One element of identity quadruplet (MMSI number, IMO number, Callsign, Name) has changed
f_ubiquity	1.4	Vessel displays two distinct locations at the same time
f_outOfScope	2.1	Invalid vessel location coordinates
f_nextposition	2.2	Position not compatible with positional and kinematic (speed, course, turn) of former AIS message
f_disapreap	2.3	Unexpected disappearance and reappearance after an unexpectedly long time
f_suddenapp	2.4	First apparition of a vessel in a location where it is not expected for a vessel to appear for the first time

Table 6: Description of the selected list of flags

6.1.2. Selected data

Time slices of 6 hours have been chosen for data analysis. Data from seven consecutive days, covering a full week, has been randomly chosen, between Wednesday 14^{th} October to Tuesday 20^{th} October, 2015, for each day between 06:00 and 12:00 hours. For each of those seven periods, the corresponding historical data preceding the beginning of the analysed time bracket are added to the dataset. In order to avoid a time-demanding computation during the assessment phase, historical data have been limited to two full days (48 hours). Required AIS and contextual data are also extracted from the dataset presented in section 5.1.

6.1.3. Data computation and discussion

The results of the computation is presented in table 7, the lines representing the scenarios, the columns the days, and the values the number of flags raised. Table 8 shows the number of vessels of each type for each time section studied.

Flag	Oct 14^{th}	Oct 15^{th}	Oct 16^{th}	Oct 17^{th}	Oct 18^{th}	Oct 19^{th}	Oct 20^{th}
Message number	25,340	23,294	24,316	14,749	17,063	22,564	20,537
f_country	3	10	849	6	5	33	484
f_fr_consistency	44	62	38	13	57	47	51
f_quadruplet	0	0	0	0	0	0	0
f_ubiquity	0	0	0	0	0	0	0
f_outOfScope	0	0	0	0	0	0	0
f_nextposition	55	30	31	20	37	72	83
f_disapreap	3	4	4	5	3	2	2
f_suddenapp	2	0	0	0	0	0	0

Table 7: Number of flags raised by session

From the results, albeit only a fraction of all flags are presented, some considerations can be drawn. Some of the flags (in our case *f_quadruplet*, *f_ubiquity* and *f_outOfScope*) have no occurrence during those seven days, which means that: no vessel changed identity, no vessel was present in another location (the use of this flag would be more useful with a worldwide network of stations, but two stations displaying the same identity at the same time in the Brest roadstead or off the Brittany coasts would have been detected) and no message from

Flag	Oct 14^{th}	Oct 15^{th}	Oct 16^{th}	Oct 17^{th}	Oct 18^{th}	Oct 19^{th}	Oct 20^{th}
Vessel number	2,410	2,968	2,445	1,801	1,905	1,644	1,602
Cargo	522	583	552	409	406	294	313
Hazardous cargo	127	221	133	90	97	45	54
Passenger	412	529	384	278	221	253	361
Pl/f/s	1029	1049	1016	756	928	778	651
Other	194	309	179	120	98	196	105
Incorrect	126	277	181	148	155	78	118

Table 8: Number of vessel type flags raised by session, pl/f/s = pleasure, fishing or service

vessels with out of bounds coordinate values. The flag *f_country* presents large discrepancies from one day to another, with the Oct 16^{th} and Oct 20^{th} values presenting outstandingly high values with respect to the other days. This is due to the presence of military vessels in the Brest bay, cruising under the MMSI number "777777777". The other flags remain in the same order of magnitude throughout the seven days of the study. As the flag counter is based on the messages received, sometimes the number of flags can be quite high, but only involving few vessels, possibly only one. From the Table 8 we can see that the proportions of vessels of each type remain consistent throughout the week, and that, for each day, the statistical mode is the pleasure, fishing and service class, probably because this class encompasses a large number of vessels.

6.1.4. Results visualisation

The result of the flag analysis, stored in a dedicated table in the database, is the input for the visualisation interface presented in section 5.2.5. This feature displays all the vessels for which at least one flag have been raised on a map (Figure 9). The vessels are shown in different colours according to their vessel type and the user has several options, being able to display all the vessels in the neighbourhood of the selected vessel or the elements relative to the vessel itself. The user is also able to discard the vessel if he/she judges that the raising of the flag does not demonstrate a situation to look after. The corresponding entry in the database is not erased from the table, but is tagged as discarded and is not shown on the screen any longer.

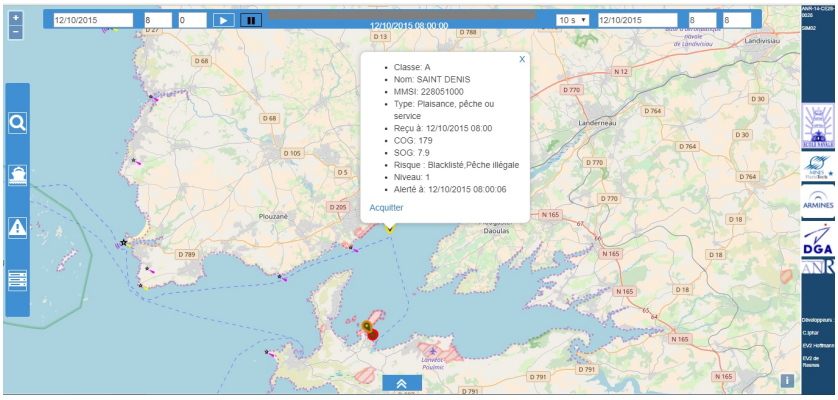

Figure 9: Detection and visualisation of an alert (in print, colour should be used for this Figure)

This interface aims at offering the people in charge of maritime monitoring a comprehensive overview of the maritime situation in their area of watch.

6.2. Discussion

This information system has been made as a decision-support tool for the user, which could be a private ship-owning company, but more probably a state-established civilian or military facility in charge of the monitoring of the maritime traffic off the coasts of the country and in inland waters. The purpose is to bring to people in charge of maritime traffic supervision to a concrete understanding of the situation, and a good means to achieve this goal is to present the results of the computation under a visual form, as shown in section 5.2.5.

The software proposed does not take decisions, its purpose is to be a tool in the hands of traffic monitoring personnel, made in order to warn the personnel on a specific amount of data they have to handle and process. The maritime traffic becoming even more important, the people in charge of monitoring face an ever-increasing number of data to assess, this tool, allowing the graphical display of suspicious vessels and their environment, helps the personnel to take a decision based only on useful information, and thus decreases their cognitive load. As the program is built on thresholds for item and flag raising, expert

knowledge is necessary to set them, and those thresholds can be adapted to local situations in case of an *in situ* software use. As the flags stand for maritime possible issue, the program helps in presenting those issues in an optimised way for the people in charge, however it remains the duty of the personnel to determine the normality or the abnormality of a situation, to consider a case as particularly hazardous or to discard it.

The use of description logics enables an inference system to assess data quality. The computation of items is straightforward and follows rules set by domain experts, thus computational speed is reduced, allowing a real-time use of the system. Maritime experts were also involved in the description of the scenarios and the modelling of the interface, bringing their expertise in the generation of several use cases. Experiments based on a dataset collected by our means showed the pertinence and the efficiency of the approach for solving simulated falsification cases, generated from reported real cases.

A limitation of this logic-based approach is its deterministic nature. Indeed, the outcome of item assessment are logical *True* or *False* values, allowing a fast data processing and triggering of alerts but not taking into consideration the uncertain nature of some pieces of information. Beyond fuzzy logic, the use of probabilistic models would enable a more precise modelling of the understanding of the maritime picture. However, the implementation of such a probabilistic approach is not straightforward and would require further research work in this field, involving a quantity of domain experts. The main obstacle towards such a system is the very nature of cyberthreats, for which few cases are reported, and therefore the construction of a knowledge base using machine learning methods is hardly realisable. In this respect, although the handling of uncertainty would undoubtedly enhance the understanding of maritime situations, a rule-based deterministic approach remains the most reliable solution as far as it is designed jointly with experts.

7. Conclusions

The work presented in this paper is part of the research in the fields of data integrity assessment, knowledge discovery and data science, with a domain exemplification in maritime situational awareness and maritime safety. The operational issue is a consequence of research questions raised after the demonstration that cyber systems were prone to attacks, and a global understanding of data that those systems provide must be provided. In our use case, a global maritime location system which is intended to provide additional safety to navigation as well as useful information to the surroundings vessels and coastal stations was easily falsified. The objective was then to propose a methodology in order to point out cases of non-genuine data and provide a risk assessment of those cases.

In order to do so, an approach based on the data quality dimensions was studied. Indeed, as information systems are data-based, they natively have data quality dimensions available to assess them. More precisely, in the diversity of data quality dimensions, integrity was discriminated as particularly important for a reliable assessment of data-based systems, and the assessment methodology is based on the development of integrity-based features assessing data veracity.

As such an integrity-based assessment requires a profound understanding of the mechanisms that rule the system in question, a thorough analysis of the system have been done, taking into consideration the primary purpose of the system and the uses that have later appeared in order to understand the wills of the people that wrote the specifications. The technical part of the system was studied as it provides precious information about the inner construction thereof, and the data part of the system was scrutinised in order to find any kind of combination of pieces of information that could result in an integrity breach.

From those integrity study results, and with the addition of non-system data such as fleet register data or navigation zones, flags were created, with the purpose pointing out data with issues with explicit statements, enabling the

displaying of those vessels in a interactive map, allowing the user to concentrate on those vessels and use visual analytics tools to find a proper solution to the problem displayed.

In the frame of this work, expert knowledge from the fields of civil activities such as merchant navy and military activities has been involved, with the collaboration of officers of the French navy and cadets of the French naval academy, and with the collaboration of Cerema, a French cluster of public experts. This heterogeneous group of experts elaborated falsification cases which have been implemented and presented in this paper.

Although the approach has been designed in an iterative way with professional domain expert, a limitation of this work is that no tests with operational personnel were performed, which would be necessary for an operational validation. However, this paper validated the approach in terms of performance or response quality, in which all inventoried falsification cases are linked to their corresponding detectors, enabling the assessment of the scenarios presented in Section 6.1.

The next step of this study will consist in the enhancement of this analysis with the notion of risk, and both the database and the program have been designed in foresight of this extension. Indeed, the various cases of problems pointed out by the system will end in different levels of risk and thus different levels of alerts to be set and presented to the operators, that will tend to change with respect to the type of vessel, the type of cargo, the location and the kinematics of the vessel and of the surrounding vessels, amongst other. This would be another step forward in the support of operators for decision-making at sea.

Acknowledgments

This research has been supported by The French National Research Agency (ANR) and co-funded by DGA (Directorate General of Armaments) 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

Agumya, A., & Hunter, G. J. (1998). Fitness for use: reducing the impact of geographic information uncertainty. In *Proceedings of the URISA 98 Conference* (pp. 245–254).

Alessandrini, A., Alvarez, M., Greidanus, H., Gammieri, V., Fernandez Arguedas, V., Mazzarella, F., Santamaria, C., Stasolla, M., Tarchi, D., & Vespe, M. (2016). Mining vessel tracking data for maritime domain applications. In *Proceedings of the 1st International ICDM Workshop on Maritime Domain Data Mining (MDDM 2016)* (pp. 361–367). Institute of Electrical and Electronics Engineers - IEEE. doi:10.1109/ICDMW.2016.20.

Alessandrini, A., Mazzarella, F., & Vespe, M. (2018). Estimated time of arrival using historical vessel tracking data. *IEEE transactions on intelligent transportation systems*, . doi:10.1109/TITS.2017.2789279.

Alincourt, E., Ray, C., Ricordel, P.-M., Dare-Emzivat, D., & Boudraa, A. (2016). Methodology for AIS signature identification through magnitude and temporal characterization. In *Proceedings of the OCEANS 2016 SHANGHAI Conference*. Institute of Electrical and Electronics Engineers (IEEE). doi:10.1109/oceansap.2016.7485420.

Amir, E., Levi, S., & Livne, T. (2018). Do firms underreport information on cyber-attacks? evidence from capital markets. *Review of Accounting Studies*, *23*, 1177–1206. doi:10.1007/s11142-018-9452-4.

Arabo, A. (2015). Cyber security challenges within the connected home ecosystem futures. *Procedia Computer Science*, *61*, 227–232. doi:10.1016/j.procs.2015.09.201.

Baader, F., Horrocks, I., & Sattler, U. (2004). Description logics. In S. Staab, & R. Studer (Eds.), *Handbook on Ontologies* (pp. 3–28). Springer-Verlag Berlin.

Balduzzi, M., Pasta, A., & Wilhoit, K. (2014a). A security evaluation of AIS Automated Identification System. In *Proceedings of the 30th Annual Computer Security Applications Conference* ACSAC'14 (pp. 436–445). New York, NY, USA: ACM. doi:10.1145/2664243.2664257.

Balduzzi, M., Wilhoit, K., & Pasta, A. (2014b). *A Security Evaluation of AIS*. Technical Report Trend Micro.

Bhatti, J., & Humphreys, T. E. (2017). Hostile control of ships via false gps signals: Demonstration and detection. *NAVIGATION, Journal of The Institute of Navigation*, *64*, 51–66.

Blomqvist, K. (1997). The many faces of trust. *Scandinavian Journal of Management*, *13*, 271 – 286. doi:10.1016/S0956-5221(97)84644-1.

Brodie, M. L. (1980). Data quality in information systems. *Information & Management*, *3*, 245 – 258. doi:10.1016/0378-7206(80)90035-X.

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, *41*. doi:10.1145/1541880.1541882.

Chen, C.-H., Khoo, L. P., Chong, Y. T., & Yin, X. F. (2014). Knowledge discovery using genetic algorithm for maritime situational awareness. *Expert Systems with Applications*, *41*, 2742–2753. doi:10.1016/j.eswa.2013.09.042.

Chen, J., Lu, F., & Peng, G. (2015). A quantitative approach for delineating principal fairways of ship passages through a strait. *Ocean Engineering*, *103*, 188–197. doi:10.1016/j.oceaneng.2015.04.077.

Comert, G., Pollard, J., Nicol, D. M., Palani, K., & Vignesh, B. (2018). Modeling cyber attacks at intelligent traffic signals. *Transportation Research Record*, *2672*, 76–89. doi:10.1177/0361198118784378.

Costé, B. (2018). *Détection contextuelle de cyberattaques par gestion de confiance à bord d'un navire*. Ph.D. thesis IMT Atlantique Bretagne-Pays de la Loire.

Costé, B., Ray, C., & Coatrieux, G. (2016). Modèle et mesures de confiance pour la sécurité des systèmes d'informations. *Ingénierie des systèmes d'information*, *2*, 1–24. doi:10.3166/ISI.22.2.1-24.

Denize, S., & Young, L. (2007). Concerning trust and information. *Industrial Marketing Management*, *36*, 968 – 982. doi:10.1016/j.indmarman.2007.06.004.

Devillers, R. (2004). *Conception d'un système multidimensionnel d'information sur la qualité des données géographiques*. Ph.D. thesis Université Laval, Canada / Université de Marne-la-Vallée, France.

EMSA (2019). Emsa facts and figures 2018. Report, European Maritime Safety Agency, 44p.

Endsley, M. R. (1995). Toward a theory of situation awareness in synamic systems. *Human Factors*, *37*, 32–64.

Erbe, C., MacGillivray, A., & Williams, R. (2012). Mapping cumulative noise from shipping to inform marine spatial planning. *The Journal of the Acoustical Society of America*, *132*, 423–428. doi:10.1121/1.4758779.

Eriksen, T., Høye, G., Narheim, B., & Meland, B. J. (2006). Maritime traffic monitoring using a space-based AIS receiver. *Acta Astronautica*, *58*, 537–549. doi:10.1016/j.actaastro.2005.12.016.

ESA (2012). Space Station Keeps Watch on World's Sea Traffic. URL: http://www.esa.int/Our_Activities/Space_Engineering_Technology/Space_Station_keeps_watch_on_world_s_sea_traffic.

Fournier, M., Casey Hilliard, R., Rezaee, S., & Pelot, R. (2018). Past, present, and future of the satellite-based automatic identification system: areas of applications (2004–2016). *WMU Journal of Maritime Affairs*, *17*, 311–345. doi:10.1007/s13437-018-0151-6.

Fox, C., Levitin, A., & Redman, T. (1994). The notion of data and its quality dimensions. *Information Processing and Management*, *30*, 9–19. doi:10.1016/0306-4573(94)90020-5.

gCaptain (2018). Ais problems revealed in east china sea. published the 27 December 2018, by Laura Kovary. URL: https://gcaptain.com/ais-problems-revealed-in-east-china-sea/.

Goldsworthy, L., & Goldsworthy, B. (2015). Modelling of ship engine exhaust emissions in ports and extensive coastal waters based on terrestrial AIS data – An Australian case study. *Environmental Modelling & Software*, *63*, 45–60. doi:10.1016/j.envsoft.2014.09.009.

Hadzagic, M., & Jousselme, A.-L. (2016). Contextual anomalous destination detection for maritime surveillance. In M. Vespe, & F. Mazzarella (Eds.), *Proceedings of the Maritime Knowledge Discovery and Anomaly Detection Workshop* JRC Conference and Workshop Reports (pp. 62–65).

Harati-Mokhtari, A., Wall, A., Brooks, P., & Wang, J. (2007). Automatic Identification System (AIS): A Human Factors Approach. *Journal of Navigation*, *60*, 373–389.

Hertzum, M., Andersen, H. H., Andersen, V., & Hansen, C. B. (2002). Trust in information sources: seeking information from people, documents, and virtual agents. *Interacting with Computers*, *14*, 575 – 599. doi:10.1016/S0953-5438(02)00023-1.

Holt, T. J., Stonhouse, M., Freilich, J., & Chermak, S. M. (2019). Examining ideologically motivated cyberattacks performed by far-left groups. *Terrorism and Political Violence*, . doi:10.1080/09546553.2018.1551213.

Hu, B., Jiang, X., de Souza, E., Pelot, R., & Matwin, S. (2016). Identifying fishing activities from ais data with conditional random fields. In *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE.

Huh, Y. U., Keller, F. R., Redman, T. C., & Watkins, A. R. (1990). Data quality. *Information and Software Technology*, *32*, 559–565. doi:10.1016/0950-5849(90)90146-I.

IMO (2003). *Guidelines for the installation of a shipborne automatic identification system (AIS)*. Circular IMO.

IMO (2004). *International Convention for the Safety of Life at Sea*. Technical Report IMO.

Iphar, C. (2017). *Formalisation of a data analysis environment based on anomaly detection for risk assessment – Application to Maritime Domain Awareness*. Ph.D. thesis PSL Research University - MINES ParisTech.

Iphar, C., Jousselme, A.-L., & Ray, C. (2019). Pseudo-synthetic datasets in support to maritime surveillance algorithms assessment. In *proceedings of the VERITA Workshop, 19ieme Journées Francophones Extraction et Gestion des Connaissances (EGC) 2019*.

Iphar, C., Napoli, A., & Ray, C. (2015). Detection of false AIS messages for the improvement of maritime situational awareness. In *Proceedings of the Oceans'2015 Washington Conference*. Marine Technology Society and the IEEE Oceanic Engineering Society IEEE.

Jousselme, A.-L., Ray, C., Camossi, E., Hadzagic, M., Claramunt, C., Bryan, K., Reardon, E., & Ilteris, M. (2016). Maritime use case description, H2020 datAcron deliverable D5.1.

Katsilieris, F., Braca, P., & Coraluppi, S. (2013). Detection of malicious AIS position spoofing by exploiting radar information. In *Proceedings of the 16th International Conference on Information Fusion*.

Kazemi, S., Abghari, S., Lavesson, N., Johnson, H., & Ryman, P. (2013). Open data for anomaly detection in maritime surveillance. *Expert Systems with Applications*, *40*, 5719–5729. doi:10.1016/j.eswa.2013.04.029.

Kelton, K., Fleischmann, K. R., & Wallace, W. A. (2008). Trust in digital information. *Journal of the American Society for Information Science and Technology*, *59*, 363–374. doi:10.1002/asi.v59:3.

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, *26*, 159–190. doi:10.1007/s10462-007-9052-3.

Last, P., Hering-Bertram, M., & Linsen, L. (2015). How automatic identification system (AIS) antenna setup affects AIS signal quality. *Ocean Engineering*, *100*, 83–89. doi:10.1016/j.oceaneng.2015.03.017.

Lecornu, L., Montagner, J., & Puentes, J. (2013). Reliability evaluation of incomplete AIS trajectories. In *Proceedings of the COST MOVE Workshop on Moving Objects at Sea*.

Llyodslist (2019). Seized UK tanker likely 'spoofed' by iran. published the 16 August 2019, by Michelle Wiese Bockmann. URL: https://lloydslist.maritimeintelligence.informa.com/LL1128820/Seized-UK-tanker-likely-spoofed-by-Iran.

Lundkvist, M., Jakobsson, L., & Modigh, R. (2008). Automatic identification system (ais) and risk-based planning of hydrographic surveys in swedish waters. In *Proceedings of the FIG Working Week 2008*.

Maglaras, L., Ferrag, M. A., Derhab, A., Mukherjee, M., Janicke, H., & Rallis, S. (2018). Threats, countermeasures and attribution of cyber attacks on critical infrastructures. *EAI Endorsed Transactions on Security and Safety*, *5*. doi:10.4108/15-10-2018.155856.

Martineau, E., & Roy, J. (2011). *Maritime Anomaly Detection: Domain Introduction and Review of Selected Literature*. Technical Report Defence Research and Development Canada. Technical Memorandum - DRDC Valcartier TM 2010-460 - October 2011.

McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, *90*, 60–66.

McGillivary, P. A., Schwehr, K. D., & Fall, K. (2009). Enhancing ais to improve whale-ship collision avoidance and maritime security. In *Proceedings of the OCEANS 2009 Biloxi Conference*. IEEE.

McKnight, H. (2005). Trust in Information Technology. *The Blackwell Encyclopedia of Management*, *7*, 329–331.

Natale, F., Gibin, M., Alessandrini, A., Vespe, M., & Paulrud, A. (2015). Mapping Fishing Effort through AIS Data. *PLOS ONE*, *10*. doi:10.1371/journal.pone.0130746.

Pallotta, G., Vespe, M., & Bryan, K. (2013). Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction. *Entropy*, *15*, 2218–2245. doi:10.3390/e15062218.

Petit, J., & Shladover, S. E. (2015). Potential cyberattacks on automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *16*, 546–556. doi:10.1109/TITS.2014.2342271.

Pierkot, C., Zimányi, E., Lin, Y., & Libourel, T. (2011). Advocacy for external quality in gis. In *Proceedings of the 4th International Conference on GeoSpatial Semantics* GeoS'11 (pp. 151–165). Berlin, Heidelberg: Springer-Verlag. URL: http://dl.acm.org/citation.cfm?id=2008664.2008678.

Pitsikalis, M., Kontopoulos, I., Artikis, A., Alevizos, E., Delaunay, P., Pouessel, J.-E., Dréo, R., Ray, C., Camossi, E., Jousselme, A.-L., & Hadzagic, M. (2018). Composite event patterns for maritime monitoring. In *Proceedings of the 10th Hellenic Conference on Artificial Intelligence*.

Ray, C., Dréo, R., Camossi, E., Jousselme, A.-L., & Iphar, C. (2019). Heterogeneous integrated dataset for maritime intelligence, surveillance, and reconnaissance. *Data in Brief*, *25*. doi:10.1016/j.dib.2019.104141.

Ray, C., Iphar, C., Napoli, A., Gallen, R., & Bouju, A. (2015). DeAIS project: Detection of AIS Spoofing and Resulting Risks. In *Proceedings of the OCEANS 2015 Genova Conferece*. IEEE. doi:10.1109/OCEANS-Genova.2015.7271729.

Raymond, E. S. (2016). Aivdm/aivdo protocol decoding. URL: http://catb.org/gpsd/AIVDM.html.

Rid, T., & Buchanan, B. (2015). Attributing cyber attacks. *Journal of Strategic Studies*, *38*, 4–37. doi:10.1080/01402390.2014.977382.

Salmon, L., Ray, C., & Claramunt, C. (2016). Continuous detection of black holes for moving objects at sea. In *Proceedings of the 7th ACM SIGSPATIAL International Workshop on GeoStreaming* IWGS '16 (pp. 2:1–2:10). New York, NY, USA: ACM. URL: http://doi.acm.org/10.1145/3003421.3003423. doi:10.1145/3003421.3003423.

Schwehr, K. D., & McGillivary, P. A. (2007). Marine Ship Automatic Identification System (AIS) for Enhanced Coastal Security Capabilities: An Oil Spill Tracking Application. In *Proceedings of the OCEANS Vancouver 2007 Conference*. IEEE. doi:10.1109/OCEANS.2007.4449285.

Serry, A., & Lévêque, L. (2015). Le système d'identification automatique (AIS). *Netcom*, *29*, 177–202. doi:10.4000/netcom.1943.

Toumsi, W., & Rais, H. (2018). A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & Security*, *72*, 212–233. doi:10.1016/j.cose.2017.09.001.

Tunaley, J. K. (2013). Utility of Various AIS Messages for Maritime Awareness. In *8th ASAR Workshop*. Longueuil, Canada.

Varga, M., Winkelholz, C., & Träber-Burdin, S. (2017). The application of visual analytics to cyber security. In *Proceedings of the NATO STO IST-143 Lecture Series on Cyber Security Science & Engineering*.

Vasseur, B., Jeansoulin, R., Devillers, R., & Frank, A. U. (2005). Evaluation de la qualité externe de l'information géographique : une approche ontologique. In R. Devillers, & R. Jeansoulin (Eds.), *Qualité de l'information géographique : traité IGAT* (pp. 285–301). Hermès Science.

Waheed, M., & Cheng, M. (2017). A system for real-time monitoring of cyber-security events on aircraft. In *Proceedings of the IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*, 5–33. doi:10.1080/07421222.1996.11518099.

Wiley, D. N., Thompson, M., Pace, R. M., & Levenson, J. (2011). Modeling speed restrictions to mitigate lethal collisions between ships and whales in the Stellwagen Bank National Marine Sanctuary, USA. *Biological Conservation*, *144*, 2377–2381. doi:10.1016/j.biocon.2011.05.007.

Windward (2014). *AIS Data on the High Seas: An Analysis of the Magnitude and Implications of Growing Data Manipulation at Sea*. Technical Report Windward.

Wired (2017). When a tanker vanishes, all the evidence points to russia. published the 21 September 2017, by Matt Burgess. URL: https://www.wired.co.uk/article/black-sea-ship-hacking-russia.

Yaghoubi Shahir, H., Glasser, U., Nalbandyan, N., & Wehn, H. (2014). Maritime Situation Analysis: A Multi-vessel Interaction and Anomaly Detection Framework. In *Proceedings of the 2014 IEEE Joint Intelligence and Security Informatics Conference* (pp. 192–199). IEEE. doi:10.1109/JISIC.2014.36.

Zissis, D. (2016). Detecting anomalies in streams of ais vessel data. In M. Vespe, & F. Mazzarella (Eds.), *Proceedings of the Maritime Knowledge Discovery and Anomaly Detection Workshop* JRC Conference and Workshop Reports (pp. 36–38).

Zouaoui-Elloumi, S. (2012). *Reconnaissance de comportements de navires dans une zone portuaire sensible par approches probabiliste et événementielle : Application au Grand Port Maritime de Marseille*. Ph.D. thesis Mines ParisTech.