Research Note

AIS data profiling for error analysis in Indian Ocean Region (IOR)

Vatsal Maheshwari, Dr (Cdr) Arnab Das & Shridhar Prabhuraman

Motivation and Broad application:

The International Maritime Organization (IMO) objectives for implementation of Automatic Identification System are to enhance safety and efficiency of navigation, safety of life at sea, and maritime environmental protection [1]. The AIS unit receives data, which is processed by simple software and displayed on a chart display. Data received by the AIS unit are encoded in NMEA sentences [2].

Fast mandatory implementation of AIS equipment for SOLAS ships without adequate research on its use, may be having a negative impact on its success and hence endanger safety of marine navigation. The collision between Hyundai Dominion and Sky Hope is an example of AIS-assisted collision in which the OOW (Officer On Watch) aboard Hyundai Dominion used the AIS text facility to communicate with the Sky Hope [3].

Transmission of erroneous information by AIS is an important issue that can affect its usefulness. Thus, there is a need to target specific dominant errors present in AIS and forming regulations and suggestions on remedial actions to be taken for the prevention of errors in AIS. This might eventually help in preventing accidents.

There have been studies in the UK and Singapore strait to analyze the errors in AIS. There has not been a similar kind of research done specifically for Indian Ocean Region. This paves the path to take inspiration from the studies done outside and undertake error analysis for IOR.

Previous works:

Abbas Harati-Mokhtari et al. (2007) [3] conducted a VTS-based AIS study and the study was conducted over about one month, during September–October 2005 at Liverpool Vessel Traffic Service (VTS) station. The study was conducted on vessels leaving and approaching Liverpool Bay, and on vessels at anchor or alongside in port. The port information is updated from the database of Lloyds Register. Additionally, the AIS ships navigational status was checked against its radar plot.

Data-mining study has been conducted earlier for data recorded by AISLive Company of Lloyds Register-Fairplay Ltd. The data consisted of 400,059 AIS reports from 1st March to 17th March 2005, collected from a number of AIS receivers located in a worldwide geographical area.
The major errors were found in MMSI, Ship's type, name and call sign, ship's navigational status, ship's length and breadth, ship's draught, destination and ETA.

Bar graphs for each error and Percentage of ships observed with inconsistent AIS information in the VTS-based study were plotted to visualize the vulnerabilities of AIS.

Q Meng et al. (2014) [4] analyze the vessel traffic characteristics in the Singapore Strait by means of the big AIS data. The vessel traffic characteristics were extracted from the AIS data and statistical analysis techniques were applied.

First, it created an initiative for a comprehensive study of vessel traffic characteristics in the Singapore Strait. Maritime sectors could benefit from the findings (e.g., enhancement of navigational safety strategies in the Singapore Strait). Second, this study compared the AIS data with the VTS data, which should be useful to researchers in their selection of appropriate data sources for the analysis of navigational safety-related issues.

Rino Bošnjak et al. (2012) [1] analyzed AIS errors and proposed measures for improvement of its functioning. Several algorithms were proposed to rectify errors in AIS data and to provide adequate training to the seafarers. Abbas Harati-Mokhtari's UK based study was taken into consideration as well.

Specific domains critical to the work:

Automatic Identification System contains a lot of information on every aspect of the ship. It is required to get familiarized with each field to identify the types of errors that can occur in the data. A broad study of each field and the possible anomalies can help in error analysis and their correction in future endeavors.

To analyze AIS data, one needs to be familiar with various Python modules like NumPy, Pandas and Matplotlib which are specialized for visualizing data and making interactive plots and chartings. One can also use Seaborn for better visualization.

As mentioned specifically in the Previous works section, UK and Singapore were involved in VTS-based AIS study. It is important to search for a shipping registry which is containing most of the data of the ships in the region. A web-crawler can be designed to scrap data present on the databases and then further validation can be performed. Thus, the knowledge of open source tools like Scrapy and Selenium can be handy for web-scraping and data analysis.

Challenges and opportunities:

This project involves real-time analysis of the data which can be a challenge as AIS data updates after every 6 minutes for Class A and every 2-30 seconds for Class B.

As there are seven or more broad AIS fields needed for this research, it becomes complex to write Python script for the same. For the validation part, one needs a shipping registry containing the data of most of the ships in that region. As described in the Previous works, UK has its own shipping registry named as Lloyds Register [3]. But Indian Ocean Region does not have any Shipping register for its own. This provides a challenge to search for a shipping registry which can be reliable enough to contain most of the data of Indian Ocean Region. Multiple shipping registry can be found for validation, and if some data is not found, one can switch to that, but this will bring even more complexity into the Python code. In addition to this, the data that has not been found might deviate us from the correct analysis of errors in AIS. Thus, it is important to use databases which cover each ship's data.

AIS data that we take as input is usually an uncleansed one, there could be several "Unnamed" columns present in the file. This requires proper data cleansing before using it for Visualization part.

Analysis of seven fields and their obvious errors, with data validation through web scraping will increase the time and space complexity of the model. This provides an opportunity to optimize the present algorithms and eventually, develop fast algorithms for the same. Otherwise, it could take minutes or even few hours to get the result.

Research Directions:

It is important to gain profound knowledge of the two types of AIS system: Class A and Class B. Our focus must be on Class A type AIS system, since ships are required to get installed this one and furthermore, Class B systems are short-ranged as well as less-featured [5, 13].

There are various types of AIS data such as Static data, Dynamic data, Voyage related data and VTS data. Research should be done to understand all these types of AIS data and the fields that come inside them are MMSI, Ship's name and Call sign, Ship's type [14, 15], Vessel Navigational Status, Length and beam, Ship's position, Draught and Destination & ETA. These various AIS fields and their errors have been discussed in a brief way.

MMSI (Maritime Mobile Security Identification) number:

The Maritime Mobile Service Identity (MMSI) is a number that uniquely identifies a vessel (or shore station). It can be regarded more or less as a phone number for maritime purposes. The number can be used to make a call to a specific vessel, or a group call to a group of vessels. A MMSI number contains 9 digits. Part of the MMSI is a country code. These digits are called the Maritime Identification Digits (MID). The MID is typically a three-digit code between 200 and 799 [6].

Errors in MMSI can be obvious such as number of digits not equal to 9 or multiple ships having the same MMSI number. On the other side, we can validate the data to get the incorrect values from VTS-databases.

Ship's name and call sign:

Generally, Vessel Names are required to contain at most 20 characters. If that exceed the AIS's 20 character limit, it should be shortened (not truncated) to 15 character-spaces, followed by an underscore {_}, thence the last 4 characters-spaces of the vessel name, e.g. GRAND JOLLY ROGER OF THE SEA to GRAND JOLLY OF _ SEA, THE GRAND JOLLY ROGER to THE GRAND JOLLY_OGER. Call-sign must reflect the call-sign assigned to the vessel by the FCC [7].
Some obvious ones can be blank values or numbers instead of characters.

Vessel Navigational status:

Navigation Status is a one- or two-digit numbers used for denoting the status of the ships. It must always be up-to-date, i.e. at anchor, underway, engaged in fishing, etc. It should be remembered to change update your status when at anchor or moored, which reduces AIS reporting rates to every 3 minutes; thus, mitigates network congestion and improves overall AIS efficiency and range [7, 8].

Apart from obvious errors such as blank values or digit-based errors, data can be validated using VTS-databases to get the incorrect values (if any).

Length and beam of the ship:

A ship's **Length Overall [LOA]** is measured in feet and inches from the extreme forward end of the bow to the extreme aft end of the stern.

A ship's extreme breadth, commonly called beam, is measured in feet and inches from the most outboard point on one side to the most outboard point on the other at the widest point on the ship. [9]

Apart from obvious errors like length shown to be greater than the beam etc., data can be validated to get incorrect values as well.

Ship's position:

By taking compass bearings of suitable objects on the shore and transferring these bearings on to the chart, the point of intersection of the bearings, called a fix, gives the ship's position. [10]

Ship's position is denoted by latitude and longitude.
Obvious errors in this include null values, latitude to shown greater than 90 degrees etc., further, data can be verified to get incorrect values as well.

Draught:

Draught is defined as the distance from the keel to the waterline (WL), as measured at the forward and aft ends of the ship. [11]
Obvious errors include draught shown as 0 meters, draught greater than the length of ship etc. while incorrect values can be found using validation.

Destination and ETA (Estimated Time of Arrival):

Destination and your origination must be encoded using 5- character UN location codes (UN/LOCODE) for (per IMO SN/Circ.244) or 4-character U.S. Geographic Unique ID (US/GUID) codes.

Estimated Time of Arrival (ETA) must be encoded in Universal Time Coordinated (UTC), and not local time [7].
Obvious errors are such as country name instead of port name or not following the IMO standards or writing a vague name etc.

Causes for these errors:

The causes for these errors can be attributed to a bad or a lack of declaration regarding the current status of the ship, a wrong installation (position, power, connection to sensors) or a wrong configuration of the transponder (frequency, type of messages sent) may have an impact on the mutual detection of ships resulting in absence of detection or low detection rate, bad handling of the risk of collision, false alarms and also have an impact on the detection of abnormal behavior by maritime surveillance systems.
Proper compulsory training and mandatory regulations can be imposed on the installers of shore systems. The responsibility of the correct installation, connection and configuration of these devices rely solely on the technical units in charge of the operations. With proper training and awareness, one can reduce the errors in AIS [12].

References:

[1] Bošnjak, R., Šimunović, L., & Kavran, Z. (2012). Automatic Identification System in Maritime Traffic and Error Analysis. *Transactions on Maritime Science*, *1*(2), 77–84. https://doi.org/10.7225/toms.v01.n02.002

[2] *FAQs - allaboutais.com*. http://www.allaboutais.com/index.php/en/faqs. (Accessed 29 May 2020)

[3] Harati-Mokhtari, A., Wall, A., Brooks, P., & Wang, J. (2007). Automatic Identification System (AIS): Data Reliability and Human Error Implications. *Journal of Navigation*, *60*(3), 373–389. https://doi.org/10.1017/s0373463307004298

[4] Meng, Q., Weng, J., & Li, S. (2014). Analysis with Automatic Identification System Data of Vessel Traffic Characteristics in the Singapore Strait. *Transportation Research Record: Journal of the Transportation Research Board*, *2426*(1), 33–43. https://doi.org/10.3141/2426-05

[5] Tucker, S. Class B. http://www.allaboutais.com/index.php/en/aisbasics1/ais-product-types/103-ais/basic-ais-product-types/135-class-b-transceiver. (Accessed 31 May 2020)

[6] *MMSI*. Maritime Mobile Service Identity (MMSI) - Telecom ABC. http://www.telecomabc.com/m/mmsi.html. (Accessed 25 May 2020)

[7] Navcen.uscg.gov. 2020. [online] Available at: https://www.navcen.uscg.gov/pdf/AIS/AISGuide.pdf (Accessed 29 May 2020).

[8] Sofrona, K. (2017, January 10). *What is the significance of the AIS Navigational Status Values?* MarineTraffic Help. https://help.marinetraffic.com/hc/en-us/articles/203990998-What-is-the-significance-of-the-AIS-Navigational-Status-Values-. (Accessed 29 May 2020)

[9] Pike, J. *Military*. Shipboard Measurements. https://www.globalsecurity.org/military/systems/ship/measurement.htm. (Accessed 30 May 2020)

[10] n.d. OBTAINING A SHIP's POSITION. [eBook] Available at: https://www.pfri.uniri.hr/bopri/documents/28-ME-tal.pdf (Accessed 3 June 2020).

[11] *What is the Draught of the ship?* MarineGyaan. (2017, August 17). http://marinegyaan.com/what-is-the-draught-of-the-ship/. (Accessed 5 June 2020)

[12] Ray, C., Gallen, R., Iphar, C., Napoli, A., & Bouju, A. (2015). DeAIS project: Detection of AIS spoofing and resulting risks. *OCEANS 2015 - Genova*. https://doi.org/10.1109/oceans-genova.2015.7271729

[13] Bhattacharjee, S., Bhattacharjee, S. B. S., & Tuicakau, I. (2019, November 8). *Automatic Identification System (AIS): Integrating and Identifying Marine Communication Channels*. Marine Insight. https://www.marineinsight.com/marine-navigation/automatic-identification-system-ais-integrating-and-identifying-marine-communication-channels/. (Accessed 24 May 2020)

[14] *Default Safety regulations for different types of ships*. Safety regulations for different types of ships. http://www.imo.org/en/OurWork/Safety/Regulations/Pages/Default.aspx. (Accessed 25 May 2020)

[15] *AIS Ship Types*. AIS Ship Types - AIS Data - VT Explorer. https://api.vtexplorer.com/docs/ref-aistypes.html. (Accessed 27 May 2020)