

Carsten Hilgenfeld

JAKOTA Cruise Systems GmbH
FleetMon in Rostock, Germany

Nina Vojdani

Chair of Production organisation and Logistics
University of Rostock, Germany

Frank Heymann

German Aerospace Center (DLR)
Inst. Communications and Navigation in Neustrelitz, Germany

Evamarie Wiessner

German Aerospace Center (DLR)
Institute of Transportation Systems in Berlin, Germany

Bettina Kutschera

DNV-GL
GL Maritime Software GmbH in Rostock, Germany

Chris Bünger

JAKOTA Cruise Systems GmbH
FleetMon in Rostock, Germany

Generating a node in an AIS-based routing graph for improved Estimated Time of Arrival. (Big) Data challenge: using AIS for generating a routing graph

ABSTRACT

For the international exchange of goods, an exact estimated time of arrival (ETA), especially in case of delays, is of great importance. Using global data of the automatic identification system (AIS) a grid node is generated. The sum of such nodes and their connections form a routing graph. As an

example, with one node of in total more than 100,000 nodes it is described how this point gets the maximum vessel length and draft assigned.

Keywords: Routing Graph, maritime, generic, AIS, ETA

JEL Classification Codes: C53, C55, C61, C82, F17, L91, R41

1. Introduction

Most of the goods are transported across the world’s oceans. Since there are no fixed waterways by sea, with a few exceptions, every ship can freely choose its trajectory.

Operational disruptions or weather-related delays may cause deviations from the actual timetable. It is of considerable importance for stakeholders in a logistics chain to be informed about punctuality or deviations from the timetable [Mestl, 2016]. If there are delays, a continuously updated Estimated Time of Arrival (ETA) is particularly important. To determine this ETA, a routing graph that is as accurate as possible is a very important requirement [Toth, 2014].

An improved ETA leads to a significantly better predictability of supply chains and thus to a significantly improved cost structure. In [Loh, 2015, p. 331] the Cost Consequences of a Port-Related Supply Chain Disruption, the advantages of a more precise ETA are shown for Chinese ports. Comparability in relation of congestions was demonstrated in [Gidado, 2015, p. 161] by analyzing African ports. Both papers clearly identify the ETA as one of the most important factors to reduce costs.

2. MERMAID and its objectives

The MERMAID project, which was awarded the DLR IDEA AWARD 2018, aims to generate a fully generic routing graph. Based on experience in road traffic simulation and modelling [Filtsch, 2018], this network is generated from movement patterns. Prior to the start of the work, a cooperation agreement including extens specifications was agreed upon between the project partners.

Previous methods mostly used manually generated routing networks. These are very time-consuming to create, require a lot of maintenance and extensive navigational knowledge for each region. Alternatively, this navigational data can be stored at providers such as Admiralty Maritime Data Solutions [Admiralty, 2018], which are very expensive for a global approach. Therefore, the goal is to develop a method how such a graph can be generated automatically and processed with new data at any time.

2.1. Evaluation of AIS data

For the project well over 125×10^9 worldwide AIS messages from the year 2016 are evaluated. Even before the work began, it was clear that this data could only be evaluated highly automatically. Therefore, the first demand was that one of the project partners involved (*Institute Communications and Navigation*) will prepare this data and make it available to the other partners.

2.2. Generic creation of the network

Due to the development of new shipping routes or changing traffic management, one of the requirements was that the network could be generated anew at any time. It follows from this requirement that the network differs slightly during each new generation due to the technology used, although it may be based on the same data. However, this inaccuracy was accepted in favour of fully automatic generation.

2.3. Manual expandability and integration of ports

Currently, a less extensive routing graph with approx. 25,000 nodes is in productive operation. This is maintained and permanently extended with the help of a node editor. More than 1,000 ports are also connected to this network, which in most cases is the destination of the routing. As a result, one of the requirements was that the 5,000 ports existing in FleetMon's port database will be automatically integrated and, if, for example, the port area is expanded, or speed requirements change, this generated network can be adapted via the node editor. FleetMon is a brand of JAKOTA Cruise Systems GmbH.

Figure 1. Traffic density at Świnoujście entry: cargo and tanker only



Source: own studies [FleetMon].

2.4. Ship type-specific routing

The routing graph will only be used for commercial shipping. From practical operations, however, it can be seen that the ships use different trajectories depending on the type of ship (e.g. tanker or ferry). In particular, there are traffic separation areas, some of which are differentiated according to the vessel type or type of load. Therefore, the network or routing algorithm must be capable of guiding a tanker along its intended routes, while a ferry may take a different route [Pallotta, 2013, p. 2229].

3. Origin of the data

The terrestrial AIS data comes from FleetMon's own AIS network of thousands of stations, complemented by AIS data from three satellite networks. Special attention was paid to the highly complex DLA (Data License Agreements) of data transfer. All parties involved have committed themselves to protect this data, which has a significant value, with special care against unauthorized access. DLR provided a secure and highly available computing cluster for this purpose. It was agreed to exchange the data, which will exceed 10,000 gigabytes, mechanically via a data carrier.

4. Preparation and evaluation of the data

AIS data is designed for ships to exchange motion information automatically [IMEA, 2018] and are trustworthy in this content [Heymann, 2013]. In the AIS signal data such as speed, course or draft are transmitted [Raymond, 2016]. However, various studies have shown that these must be prepared for further processing [Vojdani, 2015]. Therefore, the project partners agreed on how the data should be prepared. Depending on the condition of maneuvers, this movement information is transmitted every 3 seconds by the ships which are required to carry such equipment.

Table 1. Frequency of AIS messages

Vessel	Message Interval [sec]
Class A, less than 3kn	3 min
Class A, more than 3kn (mooring modus)	10s
Class A, till 14 kn	10s
Class A, till 14 kn course change	3 $\frac{1}{3}$ s
Class A, 14 till 23 kn	6s
Class A, 14 till 23 kn course change	2s
Class A, more than 23 kn	2s

Vessel	Message Interval [sec]
Class B, till 2 kn	3 min
Class B, 2 till 14 kn	30 s
Class B, 14 till 23 kn	15 s (until 30 s)
Class B, more than 23 kn	5 s (until 15 s)

Source: [IMEA, 2018]

4.1. Data requirements

The requirements were defined as follows. The aim is that all data records are actually included in the generation of the network. Therefore, no plausibility logic should be implemented on the user side.

- All AIS data that is implausible due to its position should be removed. Reason: AIS data can sometimes send out very different position reports and a ship can thus be located on the shore, for example. It was agreed to remove a measuring point rather than to leave it in the data base, even if the plausibility was unclear.
- In some regions there are only very few positions reports due to the technology used (for example at the poles for satellite coverage). For example, if the speed is implausible, the data set would not be removed if the data density was very low, but corrected and marked separately. Thus, these corrected data records can be excluded from consideration in case it becomes apparent that the data basis is still sufficient.

4.2. Preparation challenges

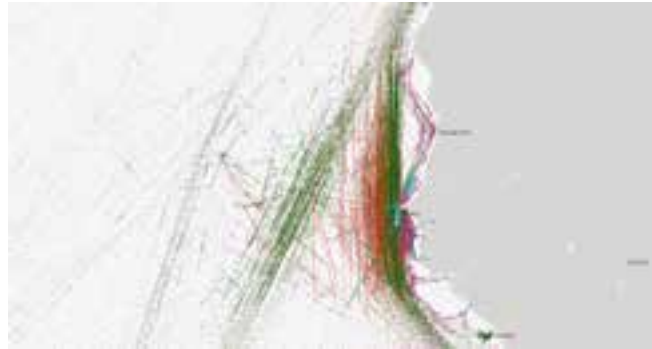
Prior to the start of the work, the quality of the data was examined in two random samples. This method made it possible to edit the developed methods on a regular desktop PC without being dependent on large computers. These samples included:

- 1) one day in the German Baltic Sea area off Rostock [Felski, 2015, p.702];
- 2) 15 days of data from all over the world with a limited number of data sources.

The expected challenges have been confirmed. A total of six specific individual aspects could be identified, which must be subjected to treatment.

- Strongly different data density with satellite-based or terrestrial coverage:
Terrestrial antennas in a good technical configuration can receive over 90% of the signals emitted up to every 2 seconds per time period and transmit them correctly to the processing infrastructure. In contrast, even the best satellite configuration currently available can only receive less than 1% of the information sent by ships. As a result, areas with satellite coverage have a much lower data density in relation to the actual traffic.

Figure 2. TER coverage on the coast, SAT on the ocean (West Africa-Senegal)



Source: own studies [FleetMon].

- **MMSI doubling:**
The incorrect but occurring multiple use of the same MMSI (Maritime Mobile Service Identity) occurs when viewed globally. For example, FleetMon's data contains more than 1,000 different objects that have used MMSI 123456789 at least once.
- **Multiple MMSIs at the same time:**
There are regions of the world where spiritual symbolism is very important. Therefore, the AIS does not use the MMSI assigned to the ships, but a lucky number. For example, while "4" is consistently avoided in China as MMSI, they like to use 6, 8 or 9, which leads to a multiple use of MMSIs and thus to difficulties in assigning the position to a specific object. This also makes it more difficult to check the plausibility of individual objects.
- **Warp signals:**
This means that the signals that appear to come from the same object have jumped back and forth on the trajectory, while the adjustment of all other evaluable parameters appears correct.
- **Signals with time delay:**
A similar challenge to the previous problem but with a greater impact. Here it sometimes happens that a signal arrives several days to weeks too late in the system. The reason for this could apparently be people who have a conscious interest in feeding in erroneous data into the worldwide data traffic. But also several technical aspects could be the reason, e.g. when sending AIS data batch-wise at a fixed transmission time.
- **Positions on land or truncated position values:**
Due to various reasons, but also intent, it can happen that the position is strongly erroneous or truncated and thus e.g. on land. This mainly occurs in the case of incorrect latitude or longitude determination, where one of the two values is exactly zero. FleetMon filters out tens of thousands of signals per day, with one of which is 0.00 (for latitude or longitude).
- **Fishing vessels:**
These vehicles pose a special challenge, as they do not follow the flow of traffic and thus distort the routes. The special feature is that these vessels often do not communicate in the

AIS that they are fishing in order not to inform the competition about profitable fishing grounds. This is comparable to trying to interpret the traffic area using a traffic flow with construction site vehicles, but without knowing that they are construction site vehicles.

Figure 3. Moving fishing vessels (North-West-Australia)



Source: own studies [FleetMon].

- Signals in a grid:
Some cooperative AIS partners of FleetMon use non-company hardware. In this case, the position information is truncated at a certain position after the decimal point. This is due to the fact that the positions of the ships are no longer correctly reproduced and a misinterpretation of the traffic load at a point can occur. This results in a westward shift in the eastern hemisphere and an eastward shift in the western hemisphere (by cutting off the position). In Figure 4 more than 12,000 positions are combined on the arrow showing a ship entering the harbour.

Figure 4. TER coverage on the coast, SAT on the ocean (West Africa-Senegal)



Source: own studies [FleetMon].

4.3. Single steps in the evaluation

Once the challenges were identified, an easy way to clean up this data was sought. This was achieved essentially through three components:

- 1) Holding and checking the vessel state. For example, if the Course over Ground or the position (which could be transferred into speed) deviates from the previous value within a short time, this value is subjected to a plausibility check. This would in particular remove individual positions on land and signals with sporadically incorrect time stamps.
- 2) Tracking the ship based on natural movements. The aim of the methodology was to track a ship's movement and to avoid warping or time outliers, for example. Anomalies in the movements were viewed locally and could thus be removed. A local view also made it possible to track several objects with the same MMSI.
- 3) Filtering fishermen by checking whether the vessels have moved out of the box within a certain period of time. It can be assumed that a vehicle carrying out business trips for the transport of goods tries to drive as fuel-efficiently as possible. Even with the most diverse marginal parameters such as evasion of a storm area or avoidance of pirate areas, a commercial ship will hardly circumnavigate several times. The system, therefore, checks whether each object has left a certain area at a certain point in time. If this was not the case, the data was removed. This means that a fisherman on his way to the fishing grounds is valid, while he was removed during the fishing itself (see Figure 3).

5. Assigning attributes to the routing graph

It was agreed that individual routing points would be assigned traffic types, speed and draft parameters. The example of a point in front of Świnoujście in Poland shows that separation must be made after the incoming and outgoing traffic. As a result, the vehicles will accelerate as they leave the port, while at the point of entry the speed limit will be observed.

Figure 5. Area for a routing point of the routing graph: Świnoujście entry



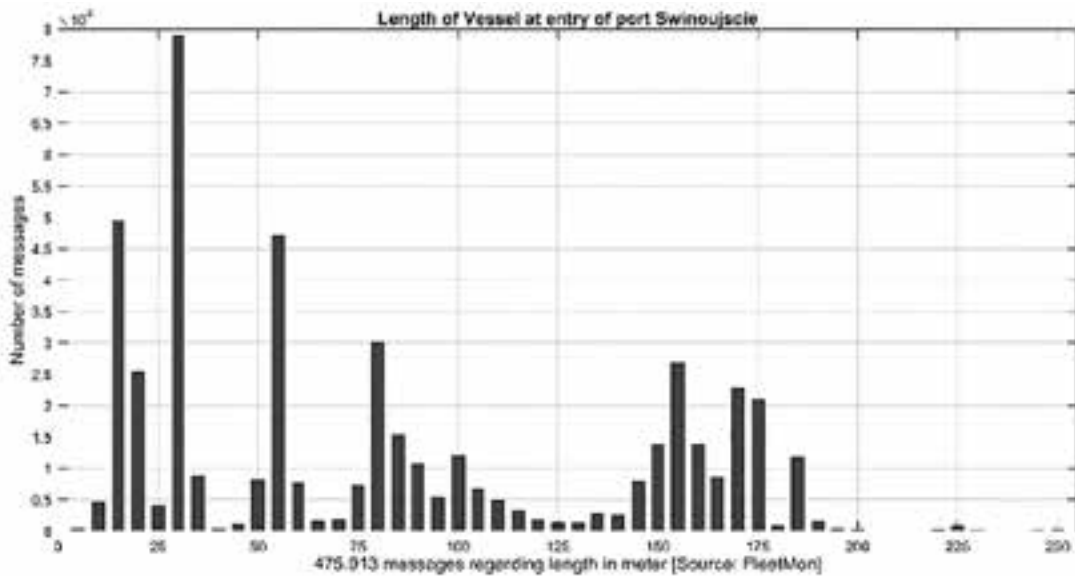
Source: own studies [FleetMon].

Table 2. Border of an example node: Boundingbox

Lower Left	Upper Right
LAT 53.922 – LONG 14.2775	LAT 53.9265 – LONG 14.2855

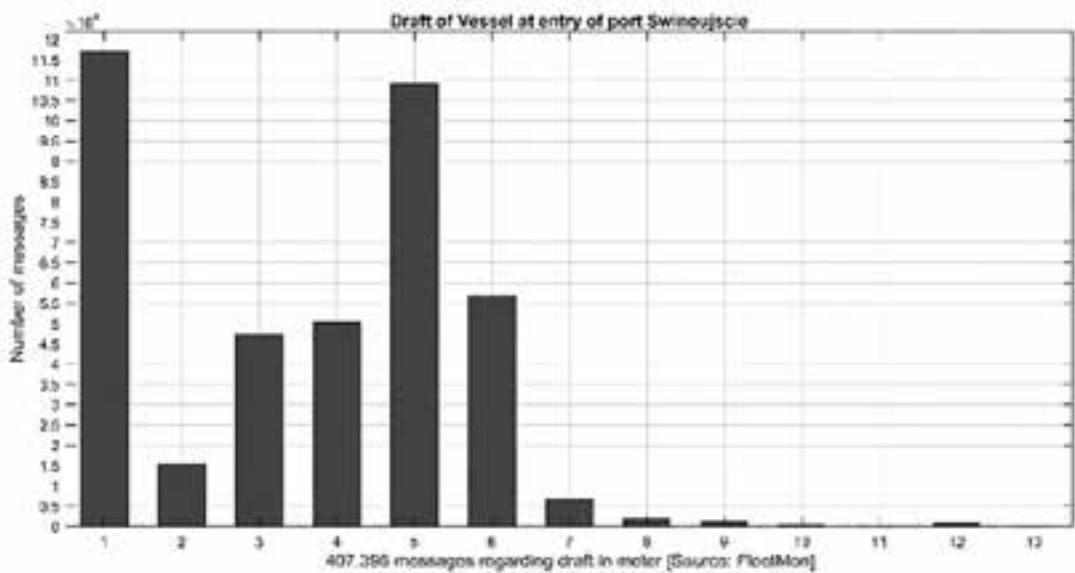
Source: own studies.

Figure 6. Length of vessels at Świnoujście entry/departure port



Source: own studies [FleetMon].

Figure 7. Draft of vessels at Świnoujście entry/departure port



Source: own studies [FleetMon].

For 2016, 502,820 individual position reports from 1824 different MMSIs were evaluated within the limits of this bounding box. Of this data, 200 reports had no heading and 191 no speed (0.004%). Of these, 35 messages were removed, as they did not pass the plausibility check with ship lengths of more than 450 meters.

The harbor access of Szczecin and Świnoujście has a draught of 13.5 meters [SSSPA, 2018] and thus coincides with the data from the AIS, which can be seen in Figure 7.

By applying this method to all locations in the world, it was possible to assign the required drafts, speeds and ship types to each individual node of the routing graph.

6. Discussion and interpretation of the results

The project is now almost completed. The results already show that the equally distributed and above all worldwide network has increased the accuracy of prediction in many cases by more than 25%. These results were confirmed by many customers through practical evaluations. The reason for this is that ships sailing off the main shipping routes find a more detail routing network and the algorithm, therefore, leads them closer to maritime reality. Thus, these can be predicted more finely granulated and the ETA prediction accuracy can benefit.

However, it is also shown that not everything can be mapped automatically. As described in the previous chapters, there is sometimes a very high divergence in the coverage. Therefore, despite scaling of the AIS data, certain maritime traffic areas are strongly underrepresented so that no graph could be generated for these areas. The requirement formulated in the task that the graph can be maintained by the existing modules eliminates this deficiency by manual work.

The extremely large amounts of data posed a challenge. It must be made clear once again that the usual calculation on a DLR server would have taken more than 1 year. For this reason, the main computer center in Gottingen and Braunschweig [DLR, 2018] was consulted. The over 13,000 computing cores of the Super Computer were used to verify the plausibility of the AIS signals. In the course of the rest of the project, the code is now optimized so that it can be calculated by standard mainframe computers of the client company.

7. Related work

Fleet management software *ShipManager*[™] provides solutions to support management of fleets. If a time-critical order for spare parts is triggered by a ship within a fleet management software, the position and expected arrival at any port can be communicated via FleetMon's software using the technological basis described here without the help of third-party applications for the employees in the shipping company office.

The INTERREG project “*RTF – Using ferry real time information to optimize intermodal transport chains in the Baltic Sea Region*” [RFT, 2018] will implement innovative information

technologies to support stakeholders in intermodal transport chains in the case of unforeseen events such as disruptions, breakdowns or delays through the use of real-time ferry information.

References

1. Admiralty (2018). Retrieved from: www.admiralty.co.uk/charts/standard-nautical-charts [accessed: 14.12.2018].
2. DLR (2018). Retrieved from: www.dlr.de/dlr/desktopdefault.aspx/tabid-10081/151_read-8655/#/gallery/12860 [accessed: 14.12.2018].
3. Felski, A., Jaskolski, K., Banys, P. (2015). Comprehensive Assessment of Automatic Identification System (AIS) Data Application to Anti-collision Manoeuvring. *The Journal of Navigation*, Vol. 68, Iss. 4, July, pp. 697–717.
4. Flitsch, C, Kastner, K., Bósa, K., Neubauer, M. (2018). *Calibrating Traffic Simulation Models in SUMO Based upon Diverse Historical Real-Time Traffic Data – Lessons Learned in ITS Upper Austria*. SUMO 2018 – Simulating Autonomous and Intermodal Transport Systems, 14–16 May, Berlin.
5. Gidado, U. (2015). *Developing Country Studies, Consequences of Port Congestion on Logistics and Supply Chain in African Ports*. Vol. 5, No. 6, Lagos, Nigeria. Retrieved from: www.iiste.org [accessed: 14.12.2018].
6. Heymann, F., Noack, T., Banys, P. (2013). *Plausibility analysis of navigatin related AIS parameter based on time series*. ENC2013 The European Navigation Conference, April 23–25, Vienna, Austria.
7. International Marine Electronics Association, IMEA (2018). NMEA 0183 V.4.1.1 Standard for Interfacing Marine Electronic Devices.
8. Loh, H.S., Thai V.W. (2015). Cost Consequences of a Port-Related Supply Chain Disruption. *The Asian Journal of Shipping and Logistics*, Vol. 31, Iss. 3, September, pp. 319–340.
9. Mestl, T., Dausendschön, K., (2016). *Port ETA Prediction based on AIS Data*. 15th International Conference on Computer and IT Applications in the Maritime Industries, Lecce, 9–11 May.
10. Pallotta, G., Vespe, M., Bryan, K. (2013). *Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction*. NATO Science and Technology Organization (STO), Centre for Maritime Research and Experimentation (CMRE), 15(6), pp. 2218–2245.
11. Raymond, E.S. (2016). *AIVDM/AIVDO Protocol Decoding: AIS-NMEA-Msg Typs, Coding and Interpretation*. Retrieved from: <http://catb.org/gpsd/AIVDM.html> [accessed: 01.07.2018].
12. RFT (2018). Retrieved from: <https://projects.interreg-baltic.eu/projects/rft-150.html> [accessed: 17.07.2018].
13. SSSPA (2018). Szczecin and Swinoujscie Seaports Authority S.A. Retrieved from: www.port.szczecin.pl/en [accessed: 01.07.2018].
14. Toth, P., Vigo, D. (2014). *Vehicle Routing – Problems, Methodes and Applications*., ISBN: 978-1-611973-58-7, SIAM – series on optimization. New York, United States: Society for Industrial and Applied Mathematics.

15. Vojdani, N., Ahn, M., Hilgenfeld, C. (2015). *Aufbereitung von AIS-Daten der WSV-N für die Simulation von Seeverkehrssystemen*. [EN: Preparation of AIS Data of the WSV-N for Simulation of traffic in marine systems]. Workshop der ASIM/GI-Fachgruppen. 18–19 June, Stralsund, Germany. ISBN: 978–3-9813334-9-7.