Data Reduction Techniques Applied on Automatic Identification System Data

Claudia Ifrim[1], Iulian Iuga[2], Florin Pop[1], Manolis Wallace[3(✉)],
and Vassilis Poulopoulos[3]

[1] Faculty of Automatic Control and Computers Computer Science Department,
University Politehnica of Bucharest, Bucharest 060042, Romania
claudia.ifrim@hpc.pub.ro, florin.pop@cs.pub.ro
[2] Bucharest, Romania
[3] 🐾 Knowledge and Uncertainty Research Laboratory
Department of Informatics and Telecommunications,
University of the Peloponnese,
22 131 Tripolis, Greece
{wallace,vacilos}@uop.gr
http://acs.pub.ro/
http://gav.uop.gr

Abstract. In recent years, the constant increase of waterway traffic generates a high volume of Automatic Identification System data that require a big effort to be processed and analyzed in near real-time. In this paper, we analyze an Automatic Identification System data set and we propose a data reduction technique that can be applied on Automatic Identification System data without losing any important information in order to reduce it to a manageable size data set that can be further used for analysis or can be easily used for Automatic Identification System data visualization applications.

Keywords: AIS data · Data reduction techniques · Data analysis
AIS data visualization · Intelligent transport systems

1 Introduction

The Automatic Identification System (AIS) is an automated tracking system used on ships and by vessel traffic services (VTS) for identifying and locating vessels by digitally broadcasting information such as unique identification of the ship, position, course, speed or navigation status in an interval of seconds to other nearby ships, AIS base stations and satellites [1].

AIS is intended to assist a vessel's watch-standing officers and allows maritime authorities to track and monitor vessel movements. On board, it is integrated with other navigation aids, such as: Global Positioning System (GPS),

---

I. Iuga—Independent Researcher.

Radio Detection And Ranging (RADAR), Electronic Chart Display Information System (ECDIS), Voyage Data Recorder (VDR), Automatic Radar Plotting Aid (ARPA) and other electronic navigation sensors.

Despite the constant increase of waterway traffic, we still don't have a global integrated monitoring and surveillance policy (or a unique standard) for sea traffic. This policy could offer us a solution for an efficient management of the increasing traffic and a better planning of resources. Tracking and monitoring commercial and recreation vessels (that requires to have an AIS equipment on board) on waterways represents an important issue for national security, economic and environmental sectors.

AIS data retrieved by AIS base stations or satellites is stored in data stations for current and future use. The massive amount of AIS data can easily overload a database and the storage system in a short period of time, leading to an increase of the processing and querying time.

In our paper we propose a reduction technique that can be applied on AIS data sets in order to reduce its size without compromising important information about the monitored vessels on waterways. Using a manageable sized AIS data set, applications that use historic and real-time AIS data can improve their performances by decreasing their response time to queries by processing fewer records that provide the same quality for their responses.

The rest of the paper is organized as follows: Sect. 2 describes what types of errors can be detected on AIS data and how can we treat or correct them; Sect. 3 presents a technique to reduce the amount of data without loosing important informations regarding our monitored vessels; in Sect. 4 we present our obtained results using the proposed reduction technique on our AIS data set and Sect. 5 discusses the main conclusions obtained and planned future work.

2 AIS Data Pre-processing

AIS data is reported as ASCII data packets as a byte stream using the NMEA 0183 or NMEA 2000 data formats. AIS packets have the introducer "!AIVDM" (reports from other ships) or "!AIVDO" (reports from the "own" ship). A standard about the AIVDM/AIVDO messages is [9]. This was expanded and clarified by [10]. The ASCII format for AIVDM/AIVDO representations of AIS radio messages have been set by [11]. An example of a typical AIVDM data packet is:

!AIVDM,1,1,,B,177KQJ5000G?tO'K¿RA1wUbN0TKH,0*5C

The meaning of each field is:

- Field 1, !AIVDM, identifies this as an AIVDM packet.
- Field 2 (1 in this example) is the count of fragments in the message.
- Field 3 (1 in this example) is the fragment number of this sentence, one based.
- Field 4 (empty in this example) is a sequential message ID for multi-sentence messages.

– Field 5 (B in this example) is a radio channel code. AIS uses the high side of the duplex from two VHF radio channels: AIS Channel A is 161.975 Mhz (87B); AIS Channel B is 162.025 Mhz (88B). Codes 1 and 2 may also be encountered instead of A or B.
– Field 6 (177KQJ5000G?tO'K¿RA1wUbN0TKH in this example) is the data payload.
– Field 7 (0) is the number of fill bits requires to pad the data payload to a 6 bit boundary, ranging from 0 to 5.
– The *-separated suffix (*5C) is the NMEA 0183 checksum for the sentence, preceded by "*".

There are 27 AIS messages types, but the most used in the wild are the position reports. By regulations the frequency of messages vary and can be analyzed at http://www.itu.int/rec/R-REC-M.1371/en.

In our data reduction technique we will consider the messages 1, 2, 3 (Position Reports), 18 (Standard Class B equipment position report)and their extended data 5 (Ship static and voyage related data), 19 (Extended class B equipment position report), 24 (Static data report). Message 4 (Base Station Report) will be treated just to display the general geographic distribution of Base Stations.

The following library was used for decoding AIS stream, see https://github.com/schwehr/libais/tree/master/src/libais mC++ decoder for Automatic Identification System for tracking ships and decoding maritime information. Some extensions have been added to deploy the data in a convenient structure for this task.

Some information was eliminated as an initial cleanup:

– Coordinates greater that 180, −180 latitude and 90, −90 longitude
– The 0,0 location. As it is a real possibility that a ship can be in the 0,0 lat long spot it can be there for a specific amount of time. Usually the 0,0 messages are generated by instrumentations errors (eg. lost GPS connections).

3   AIS Data Reduction

Analyzing our AIS data set we observe that the records number can reach more than 2 million per month. Considering the increased number of vessels that install AIS devices on board, the number of records that will be stored can only increase and the only solution that we have is it to propose a reduction technique that could be applied on AIS data records in order to reduce its size without loosing any valuable informations and to be able to optimally store and query historic AIS data.

We know that all AIS broad-casted messages transmit three basic elements of information: MMSI number, message type, a repeat indicator designed for repeating messages over obstacles by relay devices.

We also know that AIS Message 1, 2, 3 (Position Report Class A) reports navigational information (longitude and latitude, time, heading, speed, ships navigation status) and AIS Message 5 contains static and voyage related data of the ship (entered by hand).

The first step is to identify the attributes within the data that can be used to reduce the number of AIS data records.

Analyzing the messages transmitted by a single vessel on a specific voyage, we can observe that the only attributes that are constantly changing are the ship location and the timestamp. We also observe that after a period on time attributes like speed and heading are also changing. Based on our observations on the AIS data set we conclude that the attributes location, speed, heading and timestamp can be used to develop our reduction algorithm. The attributes location, speed (if < 0.1 knots) and timestamp bring us informations regarding vessel's stops on it's route and their duration. The attributes location and heading let us discover if a vessel is traveling in a straight line or if it is changing it's direction and give us enough information in order to recreate the vessel's path from the starting point to the destination.

The proposed reduction technique compares all the records from the AIS data set. In this stage, we will mark redundant records as unimportant. We will set some tolerance values for our attributes and we will test the results obtain in order to better adjust them to our data set.

In the first step we will extract all the unique MMSI values. For every unique MMSI value we will extract all it's records in chronological order. The first record is considered a relevant record and it's values for attributes like long, lat, speed, heading and timestamp are used as base values for further comparisons.

Iterating through all the records of the MMSI we will compare the selected attributes values. If the values of lat, long and timestamp are equal the record is considered duplicate and is marked as unimportant. If the values are different we will compare the speed and heading. If those values are higher than our tolerance values compared to the base attribute values, then the record will be considered important and marked accordingly. The values used for further comparisons will be updated with the ones of the latest record marked as important. The process will continue until all the records are compared and marked accordingly.

4   Results

The AIS data set that we used for our experiments contains informations gathered from Black Sea area and contains 136 008 000 records - no preprocessing applied on the initial data set. In order to store the data we used a PostgreSQL database with PostGIS extension.

In the following subsections we will present the results obtained after we detect and correct the errors on our AIS data set and applied the reduction technique presented in Sect. 3 and finally we will compare the data visualization of our data set before (Fig. 1(a)) and after the reduction was applied (Fig. 1(b) and (c)).

4.1   Data Reduction Applied on AIS Data Set

Our initial dataset contained 136 008 000 records (area Constanta port, Romania) and some informations were excluded as an initial cleanup (mentioned in Sect. 2), and after this correction we reduced the amount of records with 20%.

For this set we followed the algorithm described in Sect. 3 using different parameters for speed and heading of the vessels. The results obtained are presented in Table 1.

**Table 1.** Initial records vs. reduced records

Initial number of records	Unique MMSI	Speed difference param	No. of records after reduction
752 552	458	Less than 0.1	248 743
752 552	458	Less than 0.15	204 338
752 552	458	Less than 0.2	202 248
752 552	458	Less than 0.5	187 881
752 552	458	Less than 1	177 845

In the following section we will show some visualization of our dataset on the map, in order to easily compare the results obtained after applying our data reduction algorithm.

We observed that the reduced data still preserve unaltered information about the position, speed, heading and path of a vessel and the number of records are substantial reduced.

4.2   Data Visualization

In this section we present the visualization of the initial dataset, the visualization of the reduced dataset using different parameters for the speed difference and in the last image we will have a representation of the initial dataset compared with the reduced dataset resulted after we applied our reduction algorithm.
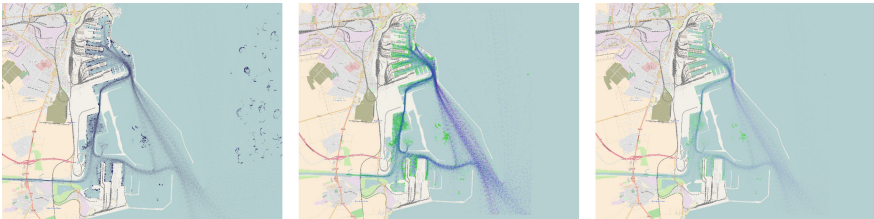
(a) Initial dataset - density map   (b) Speed difference less than 0.2   (c) Speed difference less than 0.5

Fig. 1. Visualization of the reduced dataset

5   Conclusions

As a conclusion for our experiment we consider that our reduction algorithm can be successfully used on AIS datasets (we preserve unaltered information for speed, heading, position and path of vessels) and the reduced information can be easily managed by applications that can be used in ports for the organization and planning of maritime traffic especially within ports or other dense traffic areas.

We also plan to improve our reduction algorithm by reducing data based on timestamp differences and we also want adapt it for real-time data streams.

References

1. What is the Automatic Identification System (AIS)? https://help.marinetraffic.com/hc/en-us/articles/204581828-What-is-the-Automatic-Identification-System-AIS-
2. Lanitis, A., Taylor, C.J., Cootes, T.F.: Automatic face identification system using flexible appearance models. Image Vision Comput. **13**(5), 393–401 (1995)
3. Harati-Mokhtari, A., et al.: Automatic Identification System (AIS): data reliability and human error implications. J. Navig. **60**(3), 373–389 (2007)
4. Automatic identification system. https://en.wikipedia.org/wiki/Automatic_identification_system
5. Wang, J., et al.: A new automatic identification system of insect images at the order level. Knowl.-Based Syst. **33**, 102–110 (2012)
6. http://www.navcen.uscg.gov/
7. Greene, M.: Radio frequency automatic identification system. U.S. Patent No. 5,204,681, 20 April 1993
8. https://github.com/trendmicro/ais
9. ITU Recommendation M.1371, Technical Characteristics for a Universal Shipborne Automatic Identification System Using Time Division Multiple Access [ITU1371]
10. IALA Technical Clarifications on Recommendation ITU-R M.1371-1
11. IEC-PAS 61162–100, "Maritime navigation and radiocommunication equipment and systems" [IEC-PAS]