Analysis of AIS-based Vessel Traffic Characteristics in the Singapore Strait

Qiang Meng*
Associate Professor
Block E1, #07-10, 1 Engineering Drive
Department of Civil and Environmental Engineering
National University of Singapore, 117576, Singapore
Phone: (65) 6516 5494
E-mail: ceemq@nus.edu.sg

Jinxian Weng
Associate Professor
MOE Key Laboratory for Urban Transportation
Complex Systems Theory and Technology
Beijing Jiaotong University
100044, Beijing, P.R. China
Phone: (86) 10 5168 4599
E-mail: wjx206@gmail.com

Suyi Li
Research Fellow
Centre for Maritime Studies
National University of Singapore
117576, Singapore

To the committee of Marine Safety and Human Factors (AW040) for re-review after revision

Paper submitted to the 93rd Annual Meeting of Transportation Research Board to be considered for presentation and publication in Transportation Research Record

Submission Date: 12/11/13

Word count:
Text: 4,713+ (8 figures + 3 tables) × 250=7,463

* Corresponding author
ABSTRACT
This study aims to analyze vessel traffic characteristics in the Singapore Strait. The real-time vessel AIS (Automatic Identification System) based data with about 4 million records are collected from the Lloyd’s Marine Intelligence Unit (Lloyd’s MIU) database for analysis. The results show that container ships have the largest proportion (36.36%) of the vessel traffic in the Singapore Strait while there is a small percentage of RORO/passenger ship (4.77%). The analysis results on the vessel characteristics suggest that special attention should be placed on the tankers, bulk carriers and LNG/LPG ships because of their bigger gross tonnage and draught. It is also found that vessels in the eastbound traffic usually sail at slightly higher speeds than that in the westbound traffic. The spatial distribution of vessel traffic flow indicates that the area between longitudes 103°48´E and 104°05´E has a larger traffic flow with the highest traffic density than the other areas. Furthermore, ship density of the westbound traffic is higher than that of the eastbound traffic. It should be pointed out that the collected AIS based data around the area between longitudes 104°10´E and 104°35´E are incomplete so that the corresponding traffic flow is underestimated. The advantages and limitations of AIS based data are discussed in detail by this study.

KEY WORDS: Ship Traffic Characteristics, Big Data Analysis, Singapore Strait, Maritime Safety
INTRODUCTION

The Singapore Strait, as shown in Figure 1, is one of the most important shipping waterways in the world (1) because it is directly connected to the Strait of Malacca. Its annually throughout is more than 100,000 vessels, carrying 80% of the oil transported to Northeast Asia as well as one third of the world’s traded goods including Chinese manufactured goods, Indonesian coffee and others (2). To enable a safer navigation, the traffic separation scheme (TSS) for eastbound and westbound traffics in the Singapore Strait was brought into force in 1981, which is indicated by the dashed line in Figure 1.

A proper understanding of vessel traffic characteristics in the Singapore Strait is necessary because it can be used to assess whether the recommended passages work as intended. On the basis of vessel traffic characteristics analysis, the maritime authorities can determine whether there is a need for new or additional buoys or new navigation routes. The vessel traffic characteristics are also fundamental for various quantitative analyses; for example, estimating occurrence frequency and consequence of navigational accidents. In addition, they form a basis to forecast future vessel traffic in the Strait.

Two types of big ship traffic data are available for the vessel traffic characteristics analysis in the Singapore Strait. The first one is the vessel traffic image data from the vessel traffic surveillance (VTS) system operated by Maritime and Port Authority (MPA) of Singapore, referred to VTS data. A VTS system is a marine traffic monitoring system established and operated by harbor or port authorities. The system uses radar, closed-circuit television (CCTV) and VHF radiotelephony to keep track of vessel movements in a limited geographical area. Wang et al. (3) claimed that coverage of a VTS is quite limited. In addition, it is difficult and also expensive to access and process VTS data because these data are protected as confidential by the maritime authorities (4). Only authorized users are allowed to use these data. It should be pointed out the VTS data merely include limited vessel traffic information such as vessel position, course and speed, excluding the other important vessel characteristics including vessel type, vessel age and draught. These missed vessel characteristics are crucial for the navigation safety studies (e.g., 5-9). For example, the number of possible ship conflicts in the Singapore Strait is dependent on traffic volume, vessel size, vessel draught, vessel speed over ground (SOG), vessel course over ground (COG) and others (10).

The second type of big ship traffic data can be obtained from the automatic identification system (AIS), which is called AIS data for short. The AIS is implemented on shore to enhance navigation safety and efficiency, and to improve situational awareness and assessment through simplified and additional information (11). AIS data overcome the limitations of VTS data due to its information completeness (12), and it can be used to analyze vessel incidents and accidents such as ship collisions and groundings in the waterways.

This study aims to analyze the vessel traffic characteristics in the Singapore Strait by means of the big AIS data. The vessel traffic characteristics are extracted from the AIS data and statistical analysis techniques are applied for the analysis. The contribution of this study is twofold. First, it makes an initiative for a comprehensive study of vessel traffic characteristics in the Singapore Strait. Maritime sectors can benefit from findings of this study; for example, enhancement of the navigational safety strategies in the Singapore Strait. Second, this study provides an insightful comparison between the AIS
data and the VTS data, which is useful for researchers to choose the appropriate data sources for the analysis of navigational safety-related issues.

FIGURE 1. The Singapore Strait
BIG VESSEL TRAFFIC DATA COLLECTION

The International Maritime Organization (IMO) International Convention for the Safety of Life at Sea (SOLAS) has required since 2002 that each larger seagoing vessel with more than 300GT and each passenger vessel should be equipped with an AIS (13). This study used the vessel traffic data originating from the Lloyd’s MIU AIS database. The big AIS data with 166,182 records covered the longitudes 103°21´E and 104°35´E of Singapore Strait, as shown in Figure 1, from the 1st of July 2009 to the 31st of July 2009. Although majority of these AIS records are accurate, some records have inaccurate ship position and speed information. We thus take the data processing method proposed by Qu et al. (14) to remedy these inaccurate data including vessel speed over ground (SOG) and position data.

The collected AIS data have the unique maritime mobile service identity (MMSI) numbers allocated to the registered vessels although these data do not include the vessel characteristics: type age, draught and gross tonnage. The vessel database in Lloyd’s MIU is used to extract these vessel characteristics by means of each MMSI number. According to the collected AIS data, 7374 vessels have been reported in the Singapore Strait in July of 2009. However, only 7102 vessels out of the 7374 vessels can be identified in the vessel database. Hence, we analyze the vessel traffic characteristics based on these 7102 vessels.

VESSEL CHARACTERISTICS

Vessel Type Distribution

The vessels passing through the Singapore Strait are classified into six categories by this study: general cargo vessels, bulk carriers, tankers, liquefied natural gas/liquefied petroleum gas (LNG/LPG), roll-on roll-off (RORO)/passenger ships and container ships. Figure 2 illustrates the percentage of each category in the Singapore Strait, which is calculated from the collected AIS data. It can be seen that container ships account for the largest proportion (36.36%) followed by bulk carriers (20.50%) and tankers (18.30%). The percentage of RORO/passenger ships (4.77%) is quite small.
FIGURE 2. Vessel type distribution in the Singapore Strait

Vessel Age Distribution

Figure 3 depicts age distribution of different vessels. According to Figure 3(a), 42% of the container ships are less than 5 years old, 62% are less than 10 years old and only 5% are older than 25 years. However, it can be found from Figure 3(b) that the general cargo ships are generally older than the container ships because their ages uniformly distribute between 5 and 35 years. Figure 3(c) shows that the bulk carriers are also older than the oil tankers as well as the container ships. The compositions of bulk carriers with different vessel ages are as follows: 26% less than 5 years old, 43% less than 10 years old and 18% older than 25 years. Figure 3(d) indicates that the tankers the similar age distribution to the container ships. More specifically, about 58% of tankers are less than 10 years old. The percentage of the old tankers (25 years old) accounts for 10%. Figures 3(e) and 3(f) also show that the LNG/LPG and RORO/passenger ships have the similar age distribution as tankers and container ships. Since new vessels usually impose high level of navigational safety while old vessels are more likely to be involved in accidents. From the safety perspective, the above distribution results might imply that the old general cargo and bulk carriers can be banned from sailing in the Singapore Strait in order to enhance the navigational safety of the Strait.
Gross Tonnage (GT) Distribution

The GT distributions for different vessel types are shown in Figure 4. It can be seen from Figure 4(a) that the GTs of container ships are mainly between 10,000 and 100,000 tons. There are less than 10% of the container ships with a GT larger than 100,000 tons. Compared with the container ships, the general cargo ships have much less gross tonnage, as shown in Figure 4(b). The largest GT for the general cargo ships are found not to exceed 50,000 tons and majority of the general cargo ships (73%) have a GT less than 10,000 tons. Figure 4(c) provides a multi-modal distribution on the GT associated with
bulk carriers. It has a similar distribution to the GT of container ships when the GT is less than 50,000 tons. However, there is a very small percentage of the bulk carrier with a GT between 50,000 and 80,000 tons. According to Figure 4(d), the GT of tankers also shows a multi-modal distribution. It can be found that there are about 25% of tankers with a GT larger than 150,000 tons. This implies that more and more large-sized tankers are used by shipping liners because of the scale advantages. Figure 4(e) shows that the LNG/LPG ships have the same GT distribution as the tankers. Since the large-sized tankers and LNG/LPG ships may lead to catastrophic consequences once they are involved in a navigational accident, special focus should be placed on these vessels in the Singapore Strait. Figure 4(f) shows that 85% of RORO/Pas-senger ships are less than 60,000 tons and 97% are less than 70,000 tons.

![FIGURE 4. Gross tonnage distributions of vessels in the Singapore Strait](image-url)
**Vessel Draught Distribution**

Figure 5 shows the age distribution of different vessels in the Singapore Strait. According to Figure 5(a), it can be seen that about 70% of the container ships have a draught between 7 and 9 meters. General cargo ships have a similar vessel draught distribution to container ships. Moreover, 70% of general cargo ships are found to have a draught less than 8 meters, as shown in Figure 5(b). Different from the container ships, there are a bigger percentage of general cargo ships with draught of more than 10 meters. From Figure 5(c), it can be found that draughts of the bulk carriers are larger draught than that of the container ships and the general cargo ships. More specifically, more than 85% of bulk carriers have a draught larger than 15 meters, as shown in Figure 5(d). Compared with the bulk carriers, tankers have much larger draughts, as shown in Figure 5(e). It can be seen that the draught of tankers ranges from 10 to 25 meters. It should be pointed out that the very large vessels (e.g., large-sized bulk carriers and tankers) have to use the deep water routes.
TRAFFIC CHARACTERISTICS IN THE SINGAPORE STRAIT

Traffic Volume
Table 1 gives the monthly vessel traffic volumes for various origin-destination (O-D) pairs based on the collected AIS records, where O1 is the east entrance of Singapore Strait, O2 is the west entrance of Singapore Strait, D1 is the area of Brani Keppel and Tanjong Pagar container terminals, D2 is the Pasir Panjang container terminal, D3 is the Jurong Island area and D4 is the Sembawang port area. According to the table, it can be seen that the majority of vessels visiting container terminals (D1 and D2) are the container terminals. Among the vessels visiting container terminals, majority of them are container ships, and most chemical tankers call the Jurong Island.
TABLE 1 Monthly Vessel Volumes for Various O-D Pairs in the Straits

<table>
<thead>
<tr>
<th>Vessel type</th>
<th>O1-D1</th>
<th>O1-D2</th>
<th>O1-D3</th>
<th>O1-D4</th>
<th>O1-O2</th>
<th>O2-D1</th>
<th>O2-D2</th>
<th>O2-D3</th>
<th>O2-D4</th>
<th>O2-O1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Container ship</td>
<td>256</td>
<td>104</td>
<td>85</td>
<td>0</td>
<td>498</td>
<td>166</td>
<td>44</td>
<td>81</td>
<td>0</td>
<td>345</td>
</tr>
<tr>
<td>General cargo</td>
<td>5</td>
<td>6</td>
<td>25</td>
<td>1</td>
<td>56</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>81</td>
</tr>
<tr>
<td>Bulk carrier</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>431</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>197</td>
</tr>
<tr>
<td>Tanker</td>
<td>1</td>
<td>6</td>
<td>46</td>
<td>4</td>
<td>334</td>
<td>0</td>
<td>1</td>
<td>24</td>
<td>0</td>
<td>145</td>
</tr>
<tr>
<td>LNG/LPG</td>
<td>0</td>
<td>3</td>
<td>56</td>
<td>6</td>
<td>248</td>
<td>0</td>
<td>3</td>
<td>26</td>
<td>0</td>
<td>116</td>
</tr>
<tr>
<td>RORO/Passenger</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>431</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>198</td>
</tr>
<tr>
<td>TOTAL</td>
<td>262</td>
<td>119</td>
<td>228</td>
<td>13</td>
<td>1998</td>
<td>169</td>
<td>49</td>
<td>150</td>
<td>5</td>
<td>1082</td>
</tr>
</tbody>
</table>

Traffic Flow
The spatial distribution of monthly vessel traffic flow in the Singapore Strait is shown in Figure 6. It can be seen that the Area 1 (between longitudes 103°48’E and 104°05’E) has a larger traffic flow than the other areas. However, it can be inferred from the figure that the collected AIS data in the Area 2 (between longitudes 103°10’E and 104°35’E) are incomplete so that the AIS-based traffic flow in this area may be underestimated. This is because traffic counts between the Areas 1 and 2 are significantly distinct. In principle, the traffic flow in the Area 1 should be slightly larger than the Area 2 because only a small portion of vessels outside the Singapore Strait could enter the Area 1. Likewise, we can claim that the AIS data in Area 3 are also incomplete according to Figure 6. It should be pointed out that this claim could be supported by the study of Wu et al. (15) which analyzed the VTS data in the Singapore Strait.

FIGURE 6. Vessel traffic flow distribution

Traffic Density
The vessel spatial distribution in the Singapore Strait on the July 1 of 2009 is presented in Figure 7. The pink dot represents a westbound vessel while the blue dot represents an eastbound vessel. It can be seen that majority of the vessels locate within the area between longitudes 103°48’E and 104°05’E because of their high traffic density.
Moreover, most westbound vessels locate in the north of the Singapore Strait while the eastbound vessels mainly distribute in the south of Singapore Strait. The westbound vessels concentrate within the area between longitudes 104°05´E and 104°25´E while the distribution of eastbound vessels has been rather dispersed within this area. Figure 7 also shows that the number of eastbound vessels is larger than the number of westbound vessels in the area between longitudes 103°48´E and 104°05´E. In addition, it can be found from Figure 7 that some of AIS data are missed within the area between longitudes 103°10´E and 104°35´E. This finding is consistent with the findings from Figure 6.

Table 2 gives the vessel traffic density at different segments of the Singapore Strait. It can be clearly seen that the traffic density varies with different water areas. The unequal traffic density is mainly due to the differences in the segment width and traffic flow. Since the water area between longitudes 103°48´E and 104°00´E is the narrowest among all the water areas, the corresponding traffic density is obviously the largest. As the vessel traffic density could reflect the navigational safety level to some extent, this result suggests that it is more likely to cause navigational accidents in this water area, as compared with the other water areas. Moreover, it can be found from Table 2 that the westbound traffic has a higher density than the eastbound traffic in this water area. This is because the lane width for the westbound traffic is narrower than the eastbound traffic although traffic volumes are almost the same for both courses. It should be pointed out that the collected AIS-based traffic density in the area between 104°06´E and 104°16´E is underestimated, compared with the VTS data collected by Wu et al. (15). This may be due to the loss of AIS data in this area.

**TABLE 2 Vessel Traffic Density in the Singapore Strait (vessel/nm²)**
<table>
<thead>
<tr>
<th>Area</th>
<th>Eastbound</th>
<th>Westbound</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>103°48´E -103°54´E</td>
<td>0.28</td>
<td>0.27</td>
<td>0.55</td>
</tr>
<tr>
<td>103°54´E -104°00´E</td>
<td>0.19</td>
<td>0.36</td>
<td>0.55</td>
</tr>
<tr>
<td>104°00´E -104°06´E</td>
<td>0.08</td>
<td>0.39</td>
<td>0.47</td>
</tr>
<tr>
<td>104°06´E -104°12´E</td>
<td>0.03</td>
<td>0.13</td>
<td>0.16</td>
</tr>
<tr>
<td>104°12´E -104°18´E</td>
<td>0.02</td>
<td>0.13</td>
<td>0.15</td>
</tr>
</tbody>
</table>

**Traffic Speed**

The @Risk software is used to fit the vessel traffic speed distributions. This software makes use of chi-squared fit statistics to measure how well a distribution fits the collected data. Those which fit best (with the lowest chi-squared statistics) were selected. Figure 8 includes two graphs which describe the best-fitted distributions of vessel traffic speed for different sailing directions. Figure 8(a) shows that the beta general distribution is able to capture the speed of vessels in the eastbound traffic. The majority of vessel speeds (90% of vessels) range from 5 to 19 knots, and the average traffic speed is around 12 knots. Figure 8(b) indicates that the average vessel speed for the westbound traffic is about 11 knots. The comparison of two graphs in Figure 8 demonstrates that the vessels in the westbound traffic generally sail at a slightly lower speed than the eastbound traffic. The lower traffic speed in the westbound lanes shows that the speed limit strategy for the large-sized vessels in the narrow segments of the westbound lanes is able to reduce the overall traffic speed, which further enhances the navigational safety.
Figure 9 shows that the Weibull distribution could describe the speed distribution of different vessel types. It can be found that the container ship has the largest average sailing speed (11 knots), followed by the LNG/LPG (10 knots) in the Singapore Strait. Since the higher speed could cause more severe damages, it can be inferred that the container ship should be placed much attention because of its relatively big speed.
A one-way ANOVA analysis is also carried out to test whether the eastbound traffic speed is statistically higher than the westbound traffic speed. Table 3 gives the
ANOVA test results, showing that the eastbound traffic speed is statistically higher than the westbound traffic, at a significance level of 0.001. The slightly lower westbound traffic speed may be explained by the fact that the lane width for the westbound traffic is generally narrower than the eastbound traffic.

**TABLE 3 Statistical Results of Traffic Speed for Different Sailing Directions and Time Periods**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Traffic Mean</th>
<th>Std.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastbound</td>
<td>12.13</td>
<td>4.47</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Westbound</td>
<td>10.75</td>
<td>4.06</td>
<td></td>
</tr>
<tr>
<td>Time period</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day</td>
<td>11.50</td>
<td>4.59</td>
<td>0.003</td>
</tr>
<tr>
<td>Night</td>
<td>11.00</td>
<td>4.03</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 also demonstrates that the time of a day could affect the vessel traffic speed in the Singapore Strait. More specifically, the vessel traffic speed during the day is found to be statistically larger than the traffic speed at night at a significance level of 0.01. The larger vessel traffic speed during the day may be explained by the fact that the visibility during the day is much better than that at night. In general, good visibility could encourage vessels to sail at high speeds.

**ADVANTAGES & DISADVANTAGES OF AIS DATA**

Although both VTS and AIS data can be used for analyzing vessel traffic characteristics, AIS data has the following advantages over the VTS data. First, the AIS data can be easily obtained because it is possible for any interested parties to get access to AIS data from the shore-based AIS stations. We can directly buy AIS data from companies or set up relatively low cost receivers to receive AIS data under the acquiescence of maritime authorities. Second, the AIS is able to detect the equipped vessels in situations where the radar detection is limited such as around bends, behind hills and in conditions of restricted visibility by fog, rain, and other adverse weather conditions. Third, AIS data could provide more information than the VTS data. In general, the VTS can only display the position and, possibly, a calculated course and speed of the vessels. Compared with the VTS data, the AIS data contain more information (16), as shown below:

1. Static information. These data are entered into the AIS on installation and need to be changed only if the vessel name or call sign changes. The static information consists of: (i) MMSI number; (ii) Call sign and name; (iii) Vessel characteristics (e.g., type, length, gross tonnage (GT), age and others) and (iv) Location of position fixing antenna such as GPS/DGPS.
2. Dynamic information. Most dynamic data will automatically be updated through the AIS-connected vessel sensors. Each vessel must dynamically provide the following information: (i) Vessel’s position with accuracy indication (for better or worse than 10 m); (ii) Time in UTC (coordinated universal time); (iii) Course over ground (COG); (iv) Speed over ground (SOG); (v) Heading; (vi) Navigational status (e.g., not under command, constrained by draught and others).
3. Voyage related information. The voyage related data are entered manually during each voyage. They comprise the vessel’s draught, type of cargo, and the number of persons on board if it is requested.
Although the AIS data could compensate for the weak points of the VTS data, they have the following limitations. First, the AIS data do not cover records of all vessels. In reality, only vessels equipped with the AIS transmitters can be seen. However, the SOLAS only requires that all vessels above 300 gross tonnages should be equipped with the AIS transmitters. All fishing vessels less than 15 meters are not mandated to install the AIS transmitters. In other words, the AIS records on the fishing vessels are missing. In addition, there are missing AIS records for the vessels leaving a port and appearing in another port. The second limitation is that the quality of AIS data is not good as the VTS data. This is because the AIS signals are transmitted from other vessels. The information typed in can be incorrect or there can be defects that may give an incorrect impression of the vessel’s speed or position. Bailey (17) claimed that 80% of AIS records contain some error or inaccuracies. The third limitation is that the range of AIS data coverage is a little small. In reality, AIS receivers cannot receive AIS messages outside 30 nautical miles. This may be one possible reason that the AIS data are lost within the Areas 1 and 3 as shown in Figure 6.

Therefore, there is a critical need to improve the AIS data and completeness. One effective method is to minimize the data entry errors. In addition, we can use other data sources (e.g., the VTS data) to refine the AIS data. Since the AIS has a small coverage range, another method is to set up more shore-based AIS stations, which could greatly improve the AIS data completeness. The microscopic traffic simulation approach can be also applied to generate the dynamic vessel information (i.e., position, SOG, COG) in order to compensate for the missing AIS records.

CONCLUSIONS
This study conducted a comprehensive analysis of the vessel traffic characteristics in the Singapore Strait using the big AIS data. According to the analysis, it can be seen that container ships account for the largest proportion (36.36%) in the ship traffic in the Singapore Strait while RORO/Passenger ships have a small percentage of (4.77%). In addition, attention should be paid to the management of tankers, bulk carriers and LNG/LPG ships when they sail in the Singapore Strait because of their bigger gross tonnage and draught. It was found that the vessels in the eastbound traffic usually sail at slightly higher speeds than the westbound traffic. The analysis results also showed that the area between longitudes 103°48´E and 104°05´E has the largest traffic flow. This area is also found to have the highest traffic density. Moreover, the westbound traffic in this area has a higher density than the eastbound traffic. However, it was found that the collected AIS data in the area between longitudes 104°10´E and 104°35´E are incomplete so that the traffic flow is underestimated within this area.

Finally, the advantages and disadvantages of AIS data were discussed. Compared with the VTS data, the AIS data has the advantages of easy access and more information provided. However, the AIS data have the limitations of data missing, and a small range of data coverage. Therefore, some possible measures were proposed to improve the AIS data quality, such as minimizing the data entry errors, setting up more shore-based AIS stations and introducing other data sources to refine the AIS data.

ACKNOWLEDGEMENTS
This study is supported by the research project “Impact Analysis of Large Ships on Ship Traffic in the Straits of Malacca and Singapore” funded by the Simulation and Modeling R&D Programme of Maritime Research Institute of Singapore.

REFERENCES


