Near Miss Detection for Encountering Ships in Sunda Strait

F I Prastyasari\textsuperscript{1,3} and T Shinoda\textsuperscript{2}

\textsuperscript{1}Department of Urban and Environmental, Graduate School of Engineering, Kyushu University, Japan
\textsuperscript{2}Department of Marine Systems Engineering, Faculty of Engineering, Kyushu University, Japan
\textsuperscript{3}Department of Marine Engineering, Institut Teknologi Sepuluh Nopember, Indonesia
3TE14257M@s.kyushu-u.ac.jp

Abstract. Sunda Strait has played an essential role to the Indonesian marine transportation. Safety of navigation is one issue that often leads to an accident, especially ship collision. A collision risk assessment in a waterway is critical not only to predict the probability of collision accident but also to understand its consequences to human life, shipping company, as well as to the environment. However, it could be hard to evaluate the accident due to the scarcity of real accident data. Thus, a near miss analysis can be considered as an equivalent substitute. A one-week of Automatic Identification System (AIS) data is taken as the sample to test the methodology proposed in this paper. The target area of the near miss analysis discussed in this paper is TSS area in Sunda Strait. To categorise the encounter as a safe or unsafe encounter, this paper utilises the closest point of approach (CPA) as the indicator and it contains two parameters, the distance to CPA (DCPA) and the time to CPA (TCPA). Four values of DCPA and TCPA are used as the threshold, those are introduced by Park, Fukuto and Imazu, Goodwin, and Limbach. In this time range, 135929 sample data are collected, but only 92,379 points from 276 ships are found inside the target area. The number of situations that are categorized as near miss encounters (N\textsubscript{NM}) for each threshold are 1668, 5870, 9370, and 13549, respectively. The combined threshold is proposed to adjust with the condition of Sunda Strait and produce a result with 9,946 points detected as near miss encounter.

1. Introduction
Sunda Strait has played an important role to the Indonesian marine transportation. The international passage for cargo ships and a domestic passage for passenger ferries from Merak in Java Island to Bakauheni in Sumatera Island are crossing each other in this area. Table 1 shows the collision accidents happened in Sunda Strait since September 2012 until December 2019. There are six accidents recorded in the past 10 years and all of the accidents is involving passenger ships. As we know that passenger ship carries a lot of people. Therefore, it is essential to keep the safety navigation around this area. A Traffic Separation Scheme (TSS) was proposed by the Government of Indonesia (GoI) to the International Maritime Organization (IMO) and implemented at the beginning of July 2020. TSS is a routeing measure that has a separation line to separate the opposing traffic. The purpose of TSS is to reduce the risk of collision, especially the head-on collision. To support the proposal, the GoI must give a proof that the TSS could reduce the collision risk by means of risk assessment. Many studies have been performed to help reducing the number of collisions. This include collision risk assessments that based on the ship domain [1][2], the closest point of approach (CPA) method [3] and also further developed collision avoidance systems [4][5].
A collision risk assessment in a waterway is very important not only to predict the probability of collision accident, but also to understand its consequences to human life, shipping company, as well as to the environment. It is also crucial to evaluate the collision accident to learn the factors that may cause the collision. However, it could be hard to assess the accident due to the scarcity of real accident data. One cannot expect a ship collision to happen frequently to be able to learn the background of those accidents. Thus, a near miss analysis can be considered as an equivalent substitute.

Near miss is “a sequence of events and/or conditions that could have resulted in loss. This loss was prevented only by a fortuitous break in the chain of events and/or conditions. The potential loss could be human injury, environmental damage, or negative business impact (e.g., repair or replacement costs, scheduling delays, contract violations, loss of reputation) [6]” as defined by the Maritime Safety Committee (MSC) and the Maritime Environment Protection Committee (MEPC) of the International Maritime Organization (IMO) in the MEPC. 7 Circ. 7. However, IMO does not express the limitation for the near miss collision (i.e. speed and distance). This guideline has encouraged the near miss reporting to the companies in order to increase the “just culture” and help them finding the factors contributing to the near miss or accident. Meanwhile, from the ships’ crew point of view, the safe distance of two power driven vessels is said to be 1.6 – 2.5 nm in an open sea [7]. Various studies on near miss situations, such as near miss classification by using the ship domain [8], [9], near miss classification by utilizing a Vessel Conflict Ranking Operator (VCRO) to rank the ship encounters based on their danger score [10], [11], as well as the calculation of near miss probability [12] have been carried out to support the implementation of this guideline as well as to ensure the safety navigation.

A demand for doing the near miss analysis in Sunda Strait is considered to be essential, because the collision accident report is very limited. This paper discusses about the near miss detection method for encountering ships in Sunda Strait by utilizing the Automatic Identification System (AIS) data from October 2018 until October 2019 and then classify the encounters as safe or near miss encounter based on the thresholds set by previous researches. Finally, the results from each limit are compared and adjusted with the condition at the Sunda Strait. The following chapters are organized as follows. Section 2 describes about the AIS and its data contents. The next is Section 3 that outlines the methodology used to classify the encountering ships to near miss or no near miss situation. Discussion and results are given in the Section 4 followed by Section 5 with the conclusions.

| No. | Date       | Ship(s) Involved                                      | Location       | Severity               |
|-----|------------|------------------------------------------------------|----------------|------------------------|
| 1   | Dec 3, 2019| KMP Amarisa (Pssgr ship, 126 m)                      | Bakauheni Port |                        |
|     |            | KMP Nusa Agung (Pssgr ship, 115 m)                   |                |                        |
|     |            | KMP Suki 2 (Pssgr ship, 99 m)                        |                |                        |
| 2   | April 22, 2019 | KMP Windu Karsa (Pssgr ship, 85 m)               | 1 fatalities & 2 major injuries |
|     |            | Virgo 18 (Pssgr ship, 134.6 m)                       |                |                        |
| 3   | July 5, 2017| Mutiara Persada I (Pssgr ship, 132.4 m)             | Bakauheni Port |                        |
|     |            | Port Link III (pssgr. ship, 243 m)                  |                |                        |
|     |            | In contact                                           |                |                        |
| 4   | May 3, 2014| KMP Marisa Nusantara (pssgr. ship, 85 m)            | 3nm from Bakauheni Port | - |
|     |            | Qihang (cargo ship, 144 m)                           |                |                        |
| 5   | Jan 28, 2014| KMP Jatra III (pssgr. ship, 89.95 m)                 | Not mentioned  |                        |
|     |            | MT Soechi Chemical VII (cargo ship, 107 m)           |                |                        |
| 6   | Sept 26, 2012| KMP Bahuga Jaya (pssgr. ship, 92 m)                | 4nm east of Rimau Balak Island | 7 fatalities |
|     |            | MT Norgas Cathinka (cargo ship, 110 m)               |                |                        |
2. Automated Identification System

2.1. AIS Transponder

Automated Identification System is a broadcasting system that has the capability to provide information from one ship to other ships or to the coastal authorities autonomously [13]. The AIS uses VHF maritime band to send and receive the data. As regulated by the International Convention for the Safety of Life at Sea (SOLAS) Chapter V Regulation 19 that AIS shall be installed onboard ships with 300 GT upwards that serve international voyages, 500 GT upwards for ships that do not serve international voyage, and all passenger ships [14]. AIS can store up to 63 messages, however only 27 of them have been utilized up to this date. Those messages are divided into two categories of data, static and dynamic. Static data is related with the identity and voyage of the vessel. It includes theIMO number, ship name, ship type, ship particulars, etc. Meanwhile, the dynamic data is related with the data that keep changing along with the movement of the vessel.

There are two types of AIS, Class A and Class B. The AIS Class A meets all standards required by the IMO and able to transmit the position-related data automatically every 2-10 seconds depends on the speed and the course of the ship. However, if the ship is at anchor or moored the data will be transmitted every 3 minutes. Class B type has a lower specification and transmit the data of ship’s position less frequently than the previous one. Both classes can also send the static data of a vessel every 6 minutes.

2.2. AIS Data

2.2.1. Raw AIS Message. The AIS message, that is received by the Vessel Traffic Services (VTS), is not readable by the human because it is encrypted by using National Marine Electronics Association (NMEA) 0183 sentences. Hence, it must be decoded first to be a useful source for the purpose of ship tracking system. An example of the raw AIS message is like “!’AIVDM,1,1,,A,14eG:o@034o8sd<L9i:a;WF>062D,0*7D”. The first part is “!’AIVDM” which has a meaning of this message is the message from another vessel. The real encoded AIS data that contains the ship’s information is “14eG:o@034o8sd<L9i:a;WF>062D”. This message needs to be decoded using at least 4 steps:

a. Decode the message by using American Standard Code for Information Exchange (ASCII) Payload Armoring to be 6-bit binary number
b. Concatenate all 6-bit numbers to be a binary payload message
c. Convert from sixbit to ASCII characters
d. From there, we can extract the useful information such as the MMSI number, which started from bit 8 for 30 bits; heading, which is located from bit 128 for 9 bits; course over ground, bit 116 for 12 bits (and divide by 10); speed over ground, bit 50 for 10 bits (and divide by 10); latitude, bits 89 for 27 bits (a signed binary number, divide by 600000); and longitude, bit 61 for 28 bits (a signed binary number, divide by 600000).

2.2.2. Static Data. The static data can be described as the data that do not change with the movements of the ship. It is related with the identity of the ship as well as the information about the voyage taken by the ship. The data included in this category are IMO number, call sign, ship name, ship type, ship principal dimension, draught, type of positioning system and its antenna location, destination, and estimated time of arrival (ETA). The ship is updating these data in every five minutes or less.

2.2.3. Dynamic Data. Different from the static data, the dynamic data are updated more frequently (i.e. every 2 to 10 seconds) depend on the ship’s movement. The dynamic data consist of Maritime Mobile Service Identity (MMSI) number which is a unique 9-digit number that identify the ships’ station; AIS navigational status such as underway using engine, at anchor, moored, etc; rate of turn in degree per minute; speed over ground in knots; coordinate of the ship that consists of latitude and longitude; course
over ground, relative to the true north; ship heading; ship bearing at own position; and the timestamp indicating when the data is sent.

2.2.4. Data Gap. Although AIS data can be compelling to help analyzing the near miss situations, but sometimes the data can contain error messages, or the station does not even receive the data correctly. This problem could lead to a gap of data and would make the analysis a little more difficult, because the amount of information is not sufficient that makes the interpolation is needed from the available data to fill the missing data. For the purpose of this paper, the AIS data of Sunda Strait between October 2018 until October 2019 was collected. The time data is recorded in epoch time, which is the number of second since January 1, 1970 in midnight GMT (Greenwich Mean Time). The start time of the whole data is 1538961377, that equals to October 8, 2018 1:16:17 GMT and the end is on 1571697861 or October 21, 2019 22:44:21 GMT. Table 2 shows the proportion of each dataset that is separated by the data gap that is interpreted as missing data with unit of second.

2.2.5. Data sample. Considering the datasets of the missing data in the previous sub chapter, a set of data is taken as the sample to test the methodology proposed in this paper. A one-week data is taken from 1538961377 (October 8, 2018 1:16:17 GMT) until 1539492288 (October 14, 2018 4:44:48 GMT) as the data sample, because fewer missing data is found in that time range compared to other weeks. In this time range, the total of 135929 AIS data is found, however not all the data is located in the TSS area.

3. Near Miss Detection Methodology

3.1. Target Area
The near miss analysis is proposed for the TSS area in Sunda Strait, which is bounded by a polygon that connects four points that have coordinates of (106.041, -5.876), (105.883, -6.073), (105.718, -5.908), and (105.797, -5.796) as shown in Figure 1. Only the ship positions that located inside this polygon are going to be analysed in this paper, because the authors want to focus on the number of near miss that happened in the TSS area before this routeing measure is implemented in June 2020.

Table 2. Dataset of missing AIS data

| Dataset | Date and time start | Date and time end | Duration of missing data (s) |
|---------|---------------------|-------------------|-----------------------------|
| 1       | 2018/10/08 1:17:57  | 2018/10/08 13:29:00| 43863                       |
| 2       | 2018/10/08 13:32:56 | 2018/10/08 14:12:03| 2347                        |
| 3       | 2018/10/09 7:00:56  | 2018/10/09 13:06:40| 21944                       |
| 4       | 2018/10/11 3:45:14  | 2018/10/11 11:07:42| 26548                       |
| 5       | 2018/10/13 11:24:35 | 2018/10/14 1:46:23 | 51708                       |
| 6       | 2018/10/14 4:44:48  | 2018/10/25 7:07:22 | 958954                      |
| 7       | 2018/12/07 19:27:09 | 2019/02/18 2:21:12 | 6245643                     |
| 8       | 2019/02/18 2:45:26  | 2019/02/27 7:56:26 | 796260                      |
| 9       | 2019/04/19 9:35:30  | 2019/05/08 2:48:55 | 1617205                     |
| 10      | 2019/05/08 3:59:05  | 2019/07/01 9:31:06 | 4685521                     |
| 11      | 2019/07/02 3:30:45  | 2019/08/19 20:10:15| 4207170                     |
| 12      | 2019/08/20 7:22:04  | 2019/10/21 15:07:06| 5384702                     |
| Total   |                     |                   | 27781663                    |
3.2. Near Miss Analysis

The methodology used in the near miss classification is described in the Figure 2. First, the AIS data that has been collected is containing some important data that is used in the calculation such as the time of the data was received, MMSI number, ship coordinate (longitude and latitude), ship speed, and ship heading.

![Figure 1. The target area of the study](image1)

![Figure 2. Flowchart of near miss classification analysis](image2)
Second, check the location of the ship coordinate whether it is inside the target area or not. The target area is marked by the polygon that has been defined in the previous subchapter. If the ship coordinate is outside the target area then the data is passed, and the program takes the next coordinate to be checked.

Third, the ship coordinate that located inside the polygon is taken as the own ship and the next ship coordinate inside the polygon is modelled as the target ship. The calculation is performed to determine whether the ship positions are categorised as a safe encounter or an unsafe encounter based on the previous researches. To categorise the encounter as a safe or unsafe encounter, this paper utilises the closest point of approach as the indicator. The CPA consists of two parameters, spatial or the distance to CPA (DCPA) and temporal or the time to CPA (TCPA). There are several values of DCPA and TCPA that is used as the threshold. A research conducted by Goodwin stated that the safe DCPA is 5.6 nm, research by Fukuto and Imazu added the value of safe DCPA as 1.0 nm and the safe TCPA is 5 minutes [16], Limbach also mentioned in their research that the safe DCPA is 5.6 nm, and Park in his research found that the DCPA is 0.15 nm and the TCPA is 3 minutes [17]. Those are used to see how many encounters are categorised as the safe encounter.

The last step is to determine the status of the encounter of the own ship and the target ship. If the ship coordinate that located inside the polygon is taken as the own ship and the next ship coordinate to be checked.

4. Discussion and Results

4.1. TCPA and DCPA Calculation

The number of data used in this analysis has been outlined in Chapter 2 that is 135,930 from a one-week time range which received from 333 ships. Figure 3 shows the heatmap of the area where all the AIS data is located for this period of time. The black-lined polygon is the target area of this paper or the location of the TSS of Sunda Strait.

However, only 92,379 points are inside the target area. In order to determine the number of ship route, the total number of MMSI is used and it found that in this period, there were 276 ships passing the Sunda Strait. From these number of ships, the own ship and the target ship are modelled. The program automatically takes the first point that lies inside the polygon as the own ship and takes the next ship that detected to be inside the polygon to be the target ship. The calculation of TCPA and DCPA is based on these following equations:

\[ TCPA(t) = -\frac{[(y_j-y_i)(y_j-y_i)+(x_j-x_i)(x_j-x_i)]}{(y_j-y_i)^2(x_j-x_i)^2} \]  
\[ DCPA(t) = \sqrt{[(y_j-y_i) + (y_j-y_i) \times TCPA]^2 + [(x_j-x_i) + (x_j-x_i) \times TCPA]^2} \]  

where,

- \(x_i, y_i\) : position of own ship in x and y-axis
- \(x_j, y_j\) : position of target ship in x and y-axis
- \(\dot{x}_i, \dot{y}_i\) : speed component of own ship in x and y-axis
- \(\dot{x}_j, \dot{y}_j\) : speed component of target ship in x and y-axis

4.2. Results

The calculation produced 18754 of total ship encounters (\(N_{TE}\)) as depicted by Figure 3. According to this result, 4900 encounters have negative CPA (\(N_{NE}\)) and it can be inferred that such encounters have passed the crossing point and can be categorized as safe encounters. The colored lines represent each threshold set by previous researches. The black line is a threshold set by Fukuto and Imazu set the value for both safe DCPA and TCPA as 1 nm and 5 minutes. The blue line as proposed by Park, has also said
that a safe encounter is an encounter that has DCPA and TCPA more than 0.15 nm and 3 minutes, respectively. However, the other two lines, pink and green, set a safe encounter based on the DCPA only. The pink line is a DCPA threshold set proposed by Goodwin and the value is 2.35 nm. The last one, the green line, which set by Limbach has the DCPA value as 5.6 nm. An encounter which located in the right-hand and upper side of each threshold is classified as a safe encounter, because the value of the DCPA and TCPA are greater than the threshold. The number of safe encounter (N_{SE}) is found to be 12186, 7984, 4484, and 305 encounters for Park, Fukuto and Imazu, Goodwin, and Limbach, respectively. Hence, the number of situations that are categorized as near miss encounters (N_{NM}) for each threshold are 1668, 5870, 9370, and 13549, respectively. The result of this calculation is summarized in Table 2. The near miss encounters found in the calculation result involved 61, 65, 66, and 67 ships for each threshold.

4.3. Combined Threshold
The results of DCPA and TCPA calculation above show how many encounters that classified as near miss encounter and safe encounter. We do not want to focus to the negative encounter, instead we want to emphasize the positive side where the near miss encounter happens in this area. Looking at the traffic characteristics of Sunda Strait that is dominated by passenger ships and cargo ships, then we need to consider the opinion of the ships’ crew about the safe DCPA that lies within the range of 1.6 – 2.5 nm and also by looking at the research conducted by Fukuto and Imazu [16] that was carried out on the Bungo Strait in Japan, which has a relatively similar traffic pattern with many cross routes, set the TCPA value as 5 minutes. Hence, a combined threshold is proposed by adjusting the previous thresholds with the current condition of Sunda Strait. The value of DCPA and TCPA of the combined threshold is set to be 2.5 nm and 5 minutes, respectively. The CPA calculation is performed again for the proposed threshold and the result is illustrated by Figure 5. This threshold classified 9,946 points as the near miss encounter, which located somewhere between Goodwin and Limbach. The summary of all CPA calculation results is summed up in Table 3.

5. Conclusions
This research analyses the near miss encounter in the Sunda Strait and the area of TSS Sunda Strait is chosen as the target water area. A sample of AIS data with the time period of one week started from October 8 until October 14, 2018. In this time range, 135929 sample data are collected, and the near miss analysis is carried on for this data. There are some points that can be drawn from this study. First, current AIS data that has been collected contains too many gaps. The more data that can be collected can make the analysis more precise. Second, although the number of collisions in Sunda Strait is low, the number of near misses is still high. Third, the combined threshold is used to roughly estimate the number of near miss cases in the target area by utilizing both ships’ crew opinion and a study which analyzed area has a similar characteristic with Sunda Strait. The last point us that the input from the ships’ crew in Sunda Strait is necessary to get the exact safer CPA to be analyzed in this study.

Table 3. Calculation result

| No. | Type of Encounter   | No. of encounters for each threshold per week |
|-----|---------------------|---------------------------------------------|
|     |                     | Park       | Fukuto and Imazu | Goodwin | Limbach | Combined |
| 1   | Negative encounter (N_{NE}) | 4,900        | 4,900             | 4,900    | 4,900   | 4,900    |
| 2   | Safe encounter (N_{SE})    | 12,186      | 7,984             | 4,484    | 305     | 3,908    |
| 3   | Near miss encounter (N_{NM}) | 1,668      | 5,870             | 9,370    | 13,549  | 9,946    |
| 4   | Total encounter (N_{TE})   | 18,754      | 18,754            | 18,754   | 18,754  | 18,754   |
| 5   | Ships involved in N_{NM}   | 61          | 65                | 66       | 67      | 66       |
Figure 3. Heatmap and the target area for October 8-14, 2018

Figure 4. DCPA and TCPA calculation results
Figure 5. DCPA and TCPA calculation results of combined threshold

References
[1] L. Zhang and Q. Meng, “Probabilistic ship domain with applications to ship collision risk assessment,” Ocean Engineering, vol. 186, p. 106130, Aug. 2019, doi: 10.1016/j.oceaneng.2019.106130.
[2] X. Qu, Q. Meng, and L. Suyi, “Ship collision risk assessment for the Singapore Strait,” Accident Analysis & Prevention, vol. 43, no. 6, pp. 2030–2036, Nov. 2011, doi: 10.1016/j.aap.2011.05.022.
[3] Y. Yoo and J.-S. Lee, “Evaluation of ship collision risk assessments using environmental stress and collision risk models,” Ocean Engineering, vol. 191, p. 106527, Nov. 2019, doi: 10.1016/j.oceaneng.2019.106527.
[4] Y. Sato and H. Ishii, “Study of a collision-avoidance system for ships,” Control Engineering Practice, vol. 6, no. 9, pp. 1141–1149, Sep. 1998, doi: 10.1016/S0967-0661(98)00107-5.
[5] J. M. Mou, C. van der Tak, and H. Lijtse, “Study on collision avoidance in busy waterways by using AIS data,” Ocean Engineering, vol. 37, no. 5, pp. 483–490, Apr. 2010, doi: 10.1016/j.oceaneng.2010.01.012.
[6] IMO, “Guidance on Near-Miss Reporting.” Oct. 10, 2008.
[7] S. Vujčić, D. Mohović, and R. Mohović, “A Model of Determining the Closest Point of Approach Between Ships on the Open Sea,” PROMET, vol. 29, no. 2, pp. 225–232, Apr. 2017, doi: 10.7307/ptt.v29i2.2197.
[8] E. van Iperen, “DETECTION OF HAZARDOUS ENCOUNTERS AT THE NORTH SEA FROM AIS DATA.” p. 12, 2012.
[9] F. Goerlandt, J. Montewka, H. Lammi, and P. Kujala, “Analysis of Near Collisions in the Gulf of Finland,” Advances in Safety, Reliability and Risk Management – Bérenguer, Grall & Guedes Soares (eds), 2012.
[10] W. Zhang, F. Goerlandt, J. Montewka, and P. Kujala, “A method for detecting possible near miss ship collisions from AIS data,” Ocean Engineering, vol. 107, pp. 60–69, Oct. 2015, doi: 10.1016/j.oceaneng.2015.07.046.
[11] W. Zhang, F. Goerlandt, P. Kujala, and Y. Wang, “An advanced method for detecting possible near miss ship collisions from AIS data,” Ocean Engineering, vol. 124, pp. 141–156, Sep. 2016, doi: 10.1016/j.oceaneng.2016.07.059.
[12] K.-I. Kim, J. S. Jeong, B.-G. Lee, International Maritime Transportation Science, Mokpo National Maritime University, and Electronics and Telecommunications Research Institute, “Study on the
Analysis of Near-Miss Ship Collisions Using Logistic Regression,” *J. Adv. Comput. Intell. Intell. Inform.*, vol. 21, no. 3, pp. 467–473, May 2017, doi: 10.20965/jaciii.2017.p0467.

[13]IMO, “AIS transponders.” http://www.imo.org/en/OurWork/Safety/Navigation/Pages/AIS.aspx (accessed Nov. 27, 2019).

[14]IMO, *International Convention for the Safety of Life at Sea (SOLAS) 1974* 1980.

[15]E. M. Goodwin, “A Statistical Study of Ship Domains,” *J. Navigation*, vol. 28, no. 3, pp. 328–344, Jul. 1975, doi: 10.1017/S0373463300041230.

[16]J. Fukuto and H. Imazu, “New collision alarm algorithm using obstacle zone by target (OZT),” *IFAC Proceedings Volumes*, vol. 46, no. 33, pp. 91–96, 2013, doi: 10.3182/20130918-4-JP-3022.00044.

[17]J. W. Park, “A Study on the Estimation of a Safe Speed for Prevention of Ship’s Collision in VTS Area,” Mokpo Maritime University, Korea, 2014.