INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN: 2278 – 0211 (Online)

Ship Collision Risk Assessment Using AIS Data

Shwetabh Suman
final year (Dual Degree) M.Tech student, Department of Ocean Engineering and Naval Architecture, IIT Kharagpur, India
Vishwanath Nagarajan
Assistant Professor in the Department of Ocean Engineering and Naval Architecture at IIT Kharagpur, India
Om Prakash Sha
Head of Department ,Department of Ocean Engineering & Naval Architecture, IIT Kharagpur, India
Sahbi Khanfir
Academic promotion researcher, Graduate Schools of Maritime Sciences, Kobe University, Japan.
Eiichi Kobayashi
Professor, Graduate School of MaritimeSciences, Kobe University, Japan
Adi Maimun bin Abdul Malik
Professor , Dept. of Marine Technology, Universiti Teknologi Malaysia

Abstract:

Around 90% of world trade, by volume, is facilitated by the international shipping industry. Over the past few decades seaborne trade has expanded enormously carrying large quantities of cargo including the hazardous ones. Even a single accident can cause a serious environmental disaster and loss of human lives. This has led to increased focus on analysis of ship traffic both to prevent accidents andto make informed decisions about fairway design, traffic separation schemes and disposition of emergency services. Presently, the busy waterways are the areas with high vessel traffic density and hence pose a potential threat for collisions. Therefore, the collision risk assessment is of a significance importance for ships crossing through these narrow, shallow and busy waterways.

In this paper AIS data for Malacca Strait has been analysed for a period of one month to quantitatively assess the parameters of collision risk. The Malacca strait is divided into several zones and zone-wise analysis has been done to identify the most risky zones in which collision risk reduction measures can be prioritized. The collision risk parameters used here are index of speed dispersion and collision risk value. A procedure to quantitatively estimate the collision risk value for a particular zone has been described. . Based on these parameters it is concluded that legs 7W, 2W, 3W & 14E are most risky zones in the strait.

1.Introduction

Ships are the world's foremost means of transportation and are recognized as the most economical means of moving goods around the world. The past decade witnessed a steep increase in global trade and hence number of vessels. The ever growing number of vessels and its size is related to the possibility of accidents (collision) making ships' traffic an important factor for ship safety. Such accidents pose a great threat to environment subsequently leading to loss of life and expensive cargo. An accident in a shipping lane affects the smooth flow of traffic causing loss of both time and money. These problems concern government, industry and society thereby underlying the importance of traffic risk assessment. Earlier analysis of ship traffic has been hindered due to lack of data and the difficulty in obtaining data from all vessels passing through an area for sufficient time span to produce statistics. But with the new data acquisition equipment called Automatic Identification System (AIS) we now have abundance of data. In a very general sense, the AIS system is similar to the air traffic control system only applied to marinetraffic. Since 2004, the International Maritime Organization (IMO) has instituted carriage requirements for vessels affecting both worldwide and local shipping traffic. These carriage requirements apply to commercial vessels subject to the Safety of Life at Sea (SOLAS) convention. This mandatory regulation requires all vessels over 300 tonne on an international voyage and all domestic vessels over 500 tonne to have an AIS transponder installed. Passenger ships irrespective of size are also required to carry an AIS transponder. Local authorities may have additional requirements subject to the area's traffic. Ships are aware of eachother's position and Vessel Traffic Services (VTS) make use of AIS to increase transportation efficiency and safety by identifying, tracking and supervising the movement of these large vessels as they head into harbour, or navigate along in-land waterways or dangerous coastlines.

This paper focuses on the collision risk assessment for Singapore Strait using AIS data. The Strait of Malacca and Singapore is one of the most important shipping channels in the world providing shortest shipping route from the Indian Ocean to the Pacific Ocean. It links major economies like European Union, Middle East, India, China, Japan, South Korea etc. However, the Strait has some limitations as it is not deep enough to permit some of the largest ships (oil tankers) to pass through. At Phillips Channel close to the south of Singapore, the Strait of Malacca narrows to 2.8 km (1.5nm) wide, creating one of the world's most significant traffic choke points. Owing to vital importance of the Strait to global economy maintaining a smooth flow of traffic in the strait and ensuring navigational safety of the vessels is prime concern for the maritime authorities in Singapore.

In the past three decades several navigational solutions have been implemented in the straits to enhance navigational safety. One of such solutions includes Traffic Separation Scheme (TSS) where opposite streams of traffic have been separated by establishment of traffic lanes. A ship navigating in a traffic-lane should sail in the general direction of that lane. The body of water between two opposite lanes are no-go areas: shipping is not allowed in these areas (just like central reservation of a road), so the risks for head-on collisions are greatly reduced.The TSS rules are incorporated in the International Regulations for Preventing Collisions at Sea (under rule 10). As of now the Singapore Strait (Figure 1) has been segmented into 15 legs of voyage where 8 legs are eastbound and 7 legs are westbound (Figure 2a & 2b). The navigational rules differ from one leg to another and were formulated based on surveys from ship masters and shore based personnel having expertise in vessel traffic management.

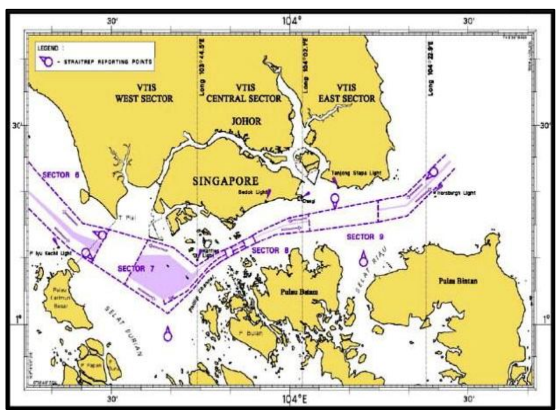

Figure 1: Traffic Separation Scheme, Singapore Strait

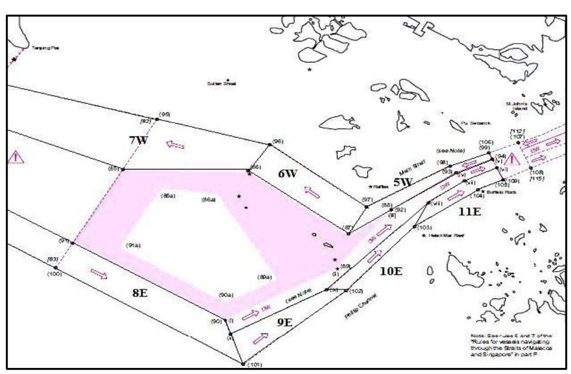

Figure 2a: Legs (Zones) of Singapore Strait

Although the solutions have been proved to be efficient in reducing collision risks these measures are, however, based on the qualitative evaluation of the experts. So far a very few studies have been done to quantitatively estimate ship collision risk in the Straits (Qu et al., 2011).

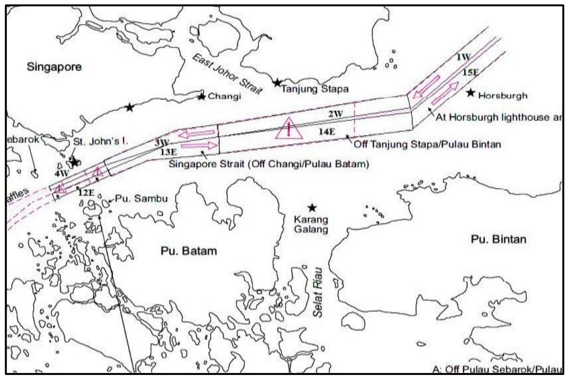

Figure 2b: 15 Legs (Zones) of Singapore Strait

Thus development of a procedure for quantitative evaluation of collision risk is important to support the expert judgements.For quantitative ship collision risk assessment for Singapore Strait, AIS data for 1month from July 1st, to July 31st, 2010 is analysed. The objective of this study is to propose a procedure for quantitative evaluation of the collision risks in individual legs of the Singapore Strait. The results are then verified based on previous studies. Riskiness of the legs is decided on the basis of two parameters: index of speed dispersion and collision risk value.

2.Data Collection And Processing

2.1.AIS As Data Source

In simple terms, AIS is a technology to make ships visible to each other. It involves exchanging data withother nearby ships and AIS Base stations electronically. AIS data is sent every few seconds over two dedicated digital marine VHF channels. The transmission interval varies based on speed and for those changing course; faster and turning vessels are updated more frequently. The AIS transponder works in continuous mode regardless of whether the vessel is off-shore, within coastal or inland waters, or at anchor. Moored and anchored vessels broadcast their position less frequently. Table 1 shows the details of rate of data transmission under various manoeuvring conditions. AIS information includes static, dynamic, and voyage related elements. Static information: IMO Number, MMSI, Call sign & Name, Length and Beam, Vessel type and location of antenna. These are programmed into the unit at commissioning. Dynamic information: Vessel's position (latitude & longitude), Time (in UTC), Course over ground (COG), Speed over ground (SOG), Heading and Rate of turn (ROT). These are derived from interfaces with the ship's GPS and other sensors. Voyage-related details are entered manually by the master through a password-protected routine and include Vessel's draft, hazardous cargo (type), destination & ETA. A MMSI (Maritime Mobile Service Identity) is a series of nine digits which uniquely identifies ship stations, shipearth stations, coast stations, coast earth stations, and group calls.

It is important that the structure of MMSI number is understood prior to data analysis since we are concerned with data of individual ships only. The first three digit of a MMSI number is known as Maritime Identification Digits (MID) which represents country code. This 3 digit code is country specific and varies from 200-799. Table 2 represents various MMSI categories used in present day navigation.

Manoeuvring Situation	Sample Rate
At Anchor	3 min
Speed 0-14 knots	12 s
Speed 0-14 knots and changing course	4 s
Speed 14-23 knots	6 s
Speed 14-23 knots and changing course	2 s
Speed > 23 knots	3 s
Speed > 23 knots and changing course	2 s

Table 1: Sample rate of variable data from AIS

0	Ship group, coast station, or group of coast stations
1	Not used (the 7digit identity beginning with "1" is used by Inmarsat A)
2-3-4-5-6-7	MMSI's used by individual ships
8	Assigned for regional Use
9	Assigned for national Use

Table 2:MMSI categories (First digit of MID or MMSI)

2.1.Data Pre-Processing

The AIS data for analysis was available on hourly basis for the month of July. One hour record consisted of around 90,000 data sets and hence huge amount of data

for span of 1 month. In order to carry out zone (Leg) wise risk assessment it is important that meaningful data are extracted from the available lot for individual legs. Sometimes due to poor transmission or malfunctioning of sensors, vague data are broadcasted and recorded. Such meaningless data are termed as 'Noise' which has to be filtered before further analysis. It should be pointed out that it is the vessels in the traffic waterways which contribute to the collision risk in the shipping lanes and not the motionless vessels which areeither anchored, moored or lie at fairways in the ports. For our analysis all vessels with speed (SOG) less than 0.2 knots were not included. The records were grouped based on the location and MMSI number of the ships. The average speed limit for the vessels transiting the strait is 12 knots (Qu et al., 2011).

3.Collision Risk Parameters

3.1.Speed Dispersion Index

This is basically a macroscopic index to represent collision risk. Greater differences in the speed amongvessels are related to higher collision risk and therefore, greater variance in speed represents higher likelihood of vessel collision. Vessels are at risk of collisionwhen they come within a certain distance of their closest pointof approach (CPA). Evidently, higher degree of speed dispersionwould result in greater variance of headways between consecutivevessels, and thus indicates higher possibility of coming within theCPA. Accordingly, the possibility of collision risks would be higher.To calculate speed dispersion first all the records are categorised into 15 groups based on the position data. This is done by identifying the leg in which the ship is navigating. The vertices and edges of each leg are identified using SHIP's Routeing published by IMO. For example, Fig 2 shows the vertices and edges of Leg 1W whose geographical coordinates are given in Table 3. Let (x, y) denotes the position of any ship obtained from the record. The ship will lie in Leg 1W if x and y satisfies following Eqs. 1, 2, 3 & 4:

$$y - y1 \geq \frac{y2-y1}{x2-x1}(x - x1) \quad (1)$$

$$y - y2 \geq \frac{y3-y2}{x3-x2}(x - x2) \quad (2)$$

$$y - y3 \geq \frac{y4-y3}{x4-x3}(x - x3) \quad (3)$$

$$y - y4 \geq \frac{y1-y4}{x1-x4}(x - x4) \quad (4)$$

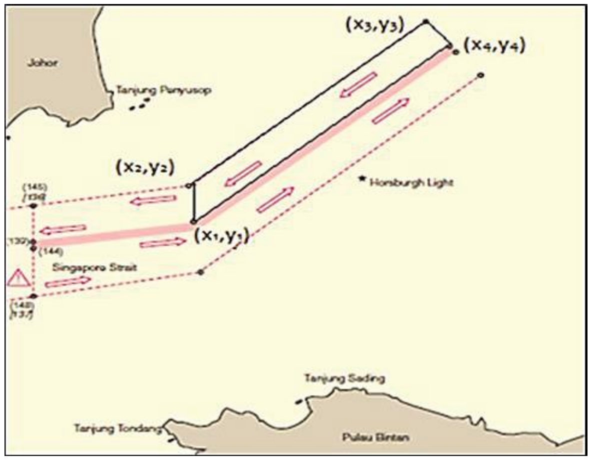

Figure 3: Leg 1W of Singapore Strait

For each leg the data sets are further grouped on the basis of MMSI number. Thus all the information of a particularship from the time it enters the leg till it exits is

Vertices	**Longitude (x_i)**	**Latitude (y_i)**
(x_1 ,y_1)	104° 19.70'E	1° 18.00'N
(x_2 ,y_2)	104° 19.50'E	1° 19.40'N
(x_3 ,y_3)	104° 26.32'E	1° 25.40'N
(x_4 ,y_4)	104° 27.05'E	1° 24.55'N

Table 3: Latitude & Longitude coordinates of vertices of Leg 1W

recorded. Universal Transverse Mercator (UTM) coordinate system has been used for calculations. A time series of each ship, from the time it enters the leg till it exits, is generated and standard deviation (S) of the speed of each ship is calculated. Thus if there are N number of ships crossing through the leg, a distribution of the N standard deviations can be plotted. Figure 4 shows the similar plot for zone 1W for the month of July, 2010. The Speed Dispersion Index (SDI) for the leg is defined as maximumof standard deviationsfor which N > 10, i.e. no. of ships greater than 10, as given in Eqs. 5 & 6:

$$SDI = \max (s_i) ; \qquad (5)$$

s_i= mean Standard Deviation of that distribution.

It is logical that a large deviation of few ships cannot be reflective of the nature of whole leg and hence a lower limit of 10 ships has been used here.

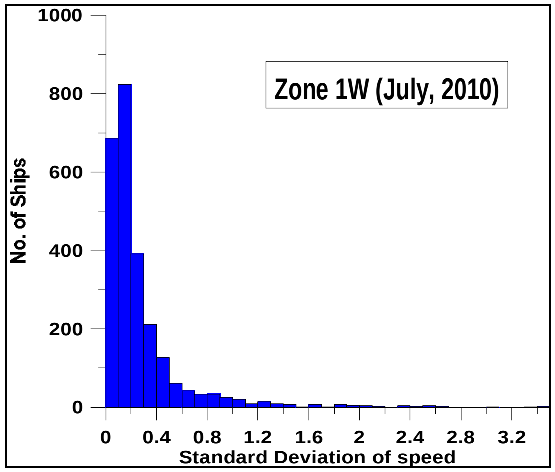

Figure 4: Distribution of standard deviation of ship speeds

3.2.Collision Risk Value

In this section a procedure is explained for quantitative estimation of the collision risk value in a particular leg of the Strait. This can be treated as a microscopic index for ship collision risk assessment. The method uses a probabilistic risk assessment model derived from orderedprobit regression model (Chin et al., 2010). Since the calibration of the regression model incorporates risks perceived by Singapore port pilots, the method is well-suited for risk assessment for Singapore Strait. It also differentiates in the collision risk value based on visibility conditions (day/night) and vessel gross tonnage (GT) which makes the analysis more accurate for the Strait is under operation throughout day and night.

<u>3.2.1Proximity Parameters</u>

The two proximity indicators used in the risk calculation are DCPA (Distance at Closest Point of Approach) and TCPA (Time to Closest Point of Approach) which represents spatial and temporal closeness between a pair of vessels at any instant of time. DCPA and TCPA are respectively the probable distance between a vessel pair at their closest point of approach (CPA) and time required to reach CPA. Figure 5 represents a typical interaction between two vessels i and j, regarded as own ship and target ship respectively, approaching from their current positions (x_i, y_i)and (x_j, y_j)with speeds V_o and V_T. If they maintain their speeds and courses they will reach at CPA after a time period equal to TCPA.From this condition, DCPA and TCPA are derived as Eqs. 6, 7, 8, 9 & 10:

$$DCPA = D^{*}Cos\,(C_{rt} - B_{rt}) \qquad (6)$$

TCPA = D* Sin (C_{rt}–B_{rt}) / V_{rt} (7)

$$V_{rt} = \sqrt{V_o{}^2 + V_T{}^2 - 2V_oV_T\cos(C_T - C_O)} \quad (8)$$

$$C_{rt} = \frac{\cos^{-1}(V_O - V_T\cos(C_T - C_O))}{V_{rt}} \quad (9)$$

B_{rt} = True Bearing (T) – Ship's heading (C_O) (10)

Where D, Crt, Brt and Vrt respectively are absolute distance, relative course, relative bearing and relative velocity.True bearing is measured relative to true north in clockwise direction. DCPA, TCPA and Vrt are measured in cable lengths, minutes and m/s respectively.

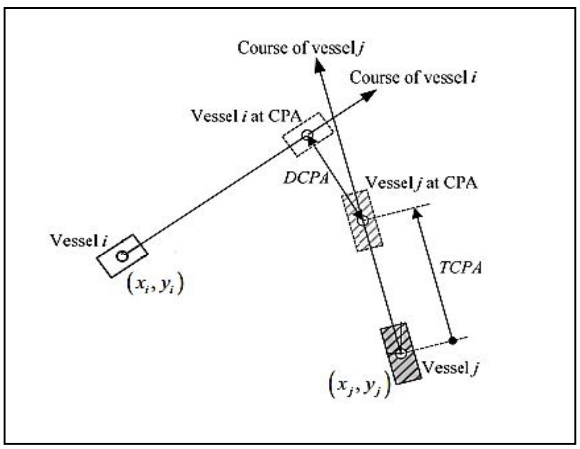

Figure 5: Vessel interaction showing indicators DCPA & TCPA

3.2.2.Risk Calculation Using Proximity Parameters

The ordered probit model is formulated as Eq. 11:

$Y_i^* =$ □$\mathbf{X_i}$+ □$_i$ (11)

Where Y_i^* is a continuous latent variable measuring collision risk, □□is a vector containing coefficients□_□&□□□and **X$_i$**is a vector containing DCPA and TCPA for ith data set.The procedure explained by Chin et al. (2010) to calculate the collision risk for a given Xi in an interaction is used here. The measurement model, in which the latent variable Yi* is mapped on to an observed ordinal variable y, is specified as Eq. 12:

$y_i = m$ if $□_{m-1} \leq Y_i^* \leq □_m$; for m =1 to J(12)

Where J is the number of ordinal categories in y and □ are unknownparameters to be estimated. Here m = 1 implies very high risk and m = 5 is safe.Based on the normality assumption of the error term, the probability of risk level m for given X_i can be predicted as Eq. 13:

$$\hat{P}r(y = m|\boldsymbol{X_i}) = F(\hat{\tau}_m - \hat{\boldsymbol{\beta}}\boldsymbol{X_i}) - F(\hat{\tau}_{m-1} - \hat{\boldsymbol{\beta}}\boldsymbol{X_i}) \quad (13)$$

Where F is cumulative distribution function of □. The collision risk (C) for given X_i can be computed as Eq. 14:

$$C(\mathbf{X_i}) = \sum_{m=1}^{J} RS_m \times \hat{P}r(y = m|\boldsymbol{X_i}) \; ; 0 \leq P_c \leq 1 \quad (14)$$

RS_m being the risk score values are assigned to each risk level based on the thresholds. The RS_m representsthe probability of collision for risk level m. The risk scores for VHR and Safe levels are assigned values of 1 and 0, respectively. For our study the regression coefficients are taken from the study done by Chin et al. (2010). Table 4 a and 4 b shows the estimates of the regression coefficients under different visibility condition for vessels categorized on the basis of Gross Tonnage (GT).

Coefficients	**VC1**		**VC2**	
	Day	***Night***	***Day***	***Night***
***DCPA(	)***	0.266	0.2179	0.5611	0.6502
TCPA(□□)	0.1168	0.0902	0.3278	0.2637
Thresholds	***Day***	***Night***	***Day***	***Night***
□□	0.2716	0.3271	0.7505	1.3021
□□	1.0468	1.2946	2.5342	3.3943
□□	2.1088	1.9947	4.6098	5.9758
□□	3.1519	3.0112	6.9348	8.5806
Risk Scores	***Day***	***Night***	***Day***	***Night***
$\mathbf{RS_{VHR}}$	1	1	1	1
$\mathbf{RS_{HR}}$	0.9138	0.8913	0.8917	0.8482
$\mathbf{RS_{MR}}$	0.6678	0.5700	0.6345	0.6044
$\mathbf{RS_{LR}}$	0.3309	0.3375	0.3352	0.3035
$\mathbf{RS_{safe}}$	0		0	

Table 4a: Regression Coefficients (Chin et al., 2010)

It is very important to note that the above method provides collision risk value on one to one basis where one vessel is termed as own vessel and other as targetvessel. Hence, by utilizing the input data, the proximity indicators and collision risk (C) values are calculated for all possible vessel pairs in the leg (fairway) in consideration. To form the pairs, the first vessel in the first update-cycle segment of the database is kept as own vessel, while the rest are considered as target vessel one after another. Thus a collision risk matrix (CR) is generated with elements C_{ij} representing collision riskvalue between own vessel i and target vessel j. Supposethere are n vessels in the leg, the collision risk matrix at any instant of time, t, is given as Eqs. 15 & 16:

$$\text{CR}(t) = \begin{pmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{pmatrix} \quad (15)$$

Where,

$$C_{ij} = \begin{cases} 0; \text{ for } i = j \text{ or TCPA} < 0 & (16) \\ C; \text{ otherwise}; \quad 0 < C < 1. \end{cases}$$

Coefficients	**VC3**		**VC4**	
	Day	***Night***	***Day***	***Night***
DCPA(	0.2641	0.271	0.2641	0.271
TCPA(□□)	0.1151	0.1181	0.1151	0.1181
Thresholds	***Day***	***Night***	***Day***	***Night***
□□	0.3212	0.5363	0.3212	0.5363
□□	1.5432	1.8126	1.5432	1.8126
□□	2.3581	2.7565	2.3581	2.7565
□□	3.4408	3.9437	3.4408	3.9437
Risk Scores	***Day***	***Night***	***Day***	***Night***
$\mathbf{RS_{VHR}}$	1	1	1	1
$\mathbf{RS_{HR}}$	0.9066	0.8640	0.9066	0.8640
$\mathbf{RS_{MR}}$	0.5515	0.5403	0.5515	0.5403
$\mathbf{RS_{LR}}$	0.3146	0.3010	0.3146	0.3010
$\mathbf{RS_{safe}}$	0		0	

Table 4b: Regression Coefficients (Chin et al., 2010)

A negative TCPA implies the two vessels have already crossed their closest point of approach (CPA) and hence no risk of collision between two vessels at that instant. It is required to measure the overall severity of collision or conflict in the zone (leg) at time, *t*, given that the risks of individual interactions have been calculated. For our analysis, this has been taken as maximum of all the elements of collision risk matrix, CR, i.e. $C_{max} = max (C_{ij})$; i, j varies from 1 to *n*. Hence a time series of collision risk value is generated on hourly basis. From these set of time series risk assessment can be done ondaily basis where mean collision risk per hour is considered to be the indicative factor. Similarly by taking the mean of this data, daily average of collision risk is calculated. Finally the quantitative risk valuefor each zone for a period of 1 month is given by averaging the daily mean values of CR. In case of zone 1W it comes out to be 0.38.

4.Results

A summary of the two collision risk assessment parameters for each leg of the Singapore Strait is shown in Table 5. It can be seen that the leg with higher speed dispersion index has greater collision risk value. Thus SDI as macroscopic index justifies the risk of ship collision in an area. It is clear from the table that legs 7W, 2W, 3W & 14E are most vulnerable legs as far asship collision is considered. In west side leg 7W and in east side leg 8E are the most risky. These legs being the entrance and exit of the canal respectively have high density of traffic (taking into account all sorts of vessels). From the analysis it was found that the number of ships in the outer most legs like 7W and 8E are quite higher than the narrower legs 4W and 12E. Therefore, probability of ships posing threat to other ships is relatively higher in crowded zones than the zones (4W & 12E) with fewer ships and where strict navigational restrictions are implemented. This, however, does not mean that legs 4W and 12E are safe. This can be seen in the Table 5 where CR value for zone 4W is 0.44 making it riskier than few legs. It may be noted that for analysis all type of vessels are taken into account. So, for example, risk of fishing vessels or speeding boats to a cargo ship is an important consideration for our study.

Zone (Leg)	Speed Dispersion Index	Collision Risk Value
1W	1.50	0.38
2W	3.00	0.70
3W	2.97	0.56
4W	2.00	0.44
5W	1.48	0.37
6W	2.44	0.48
7W	4.00	0.80
8E	2.39	0.45
9E	1.24	0.24
10E	1.35	0.31
11E	1.28	0.28
12E	1.44	0.35
13E	2.44	0.47
14E	2.45	0.49
15E	1.29	0.28

Table 5: SDI and CR values for each Zone (Leg)

Figure 6a, 6b, 7a & 7b shows the comparative risk plots for west and east zones respectively. The missing plot from 13th -16th July is due to the unavailability of AIS data for this duration. Hence no mean CR value could be computed for these three days.

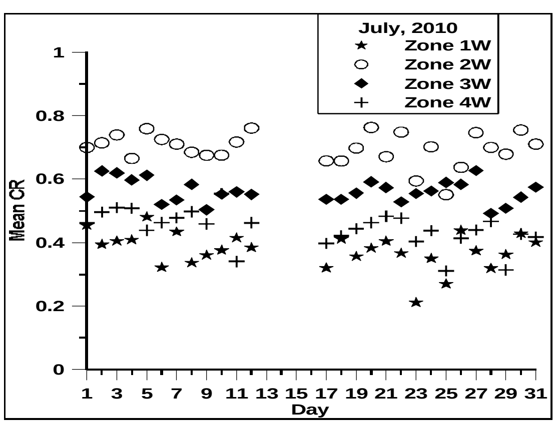

Figure 6a: Mean CR v/s Day (West Zone)

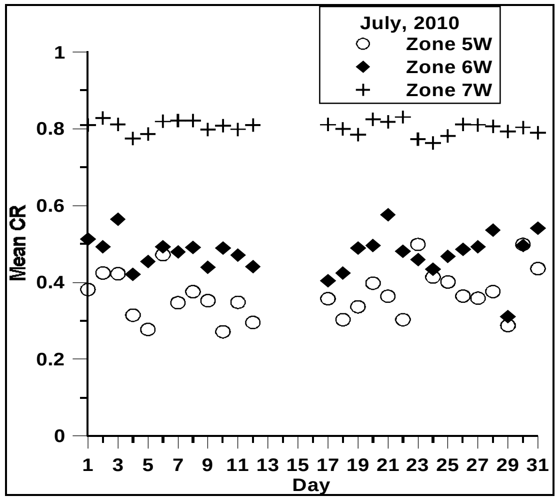

Figure 6b: Mean CR v/s Day (West Zone)

5.Conclusion

In this paper two ship collision risk parameters; speed dispersion index and collision risk value has been used. The data cleansing and analysis was done using MATLAB. Based on these parameters legs 7W, 2W & 3W on west side and legs 14E, 13E & 8E on east side are identified as the risky legs in the strait.Hence, the risk reduction measures should be prioritized in these legs. Although the traffic separation scheme has been effective in reducing the collision risks in the strait, the ever increasing number of ships and global marine trade

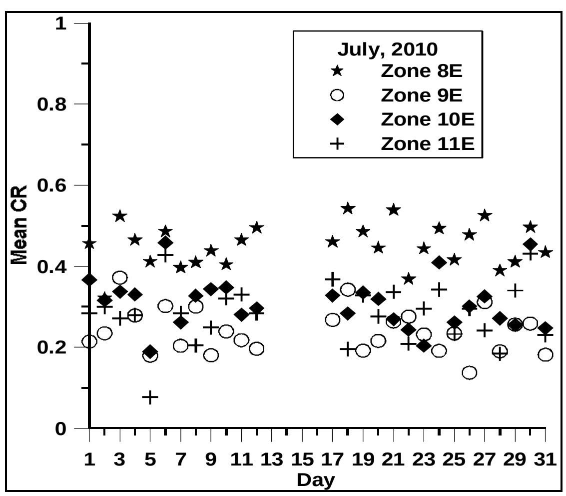

Figure 7a: Mean CR v/s Day (East zone)

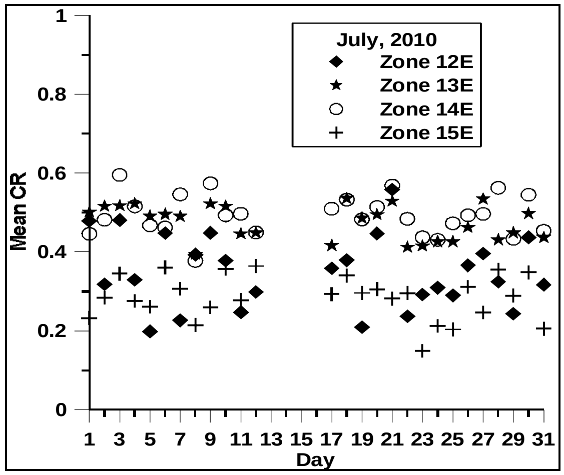

Figure 7b: Mean CR v/s Day (East zone)

pushes the strait to its limit in terms of ship handling and passage.Thus time to time risk assessment of the strait has to be done and risk mitigation measures need to be implemented to ensure safe navigation of ships. Since the average speed limit in the various legs of Singapore Strait is 12 knots; it was found out that the mean speed of vessels overshooting this limit, for all zones, was between 17-18 knots.

6.Acknowledgement

This paper is supported as a part of the research project 'Research on ship maneuvering and propulsion performance using data from Voyage Data Recorder (VDR) and Automatic Identification System (AIS)'sponsored by Ministry of Shipping, Govt. of India. AIS data for Malacca strait was provided by Professor Eiichi Kobayashi, Kobe University, which made this study possible. We acknowledge him for his support. The authors thank all the members involved in the project for their assistance and support in evaluation of the data and completion of this study.

7.Reference

1.Chin, H.C. & Debnath, A.K., 2009. 'Modelling perceived collision risk in port water navigation'. Safety Science, Elsevier 47, 1420-1416.

2. Meng, X.Y., Bai, Y.M. & Han, X.J., 2009. 'Study on shipping safety strategy based on accident forecast model'. Taylor & Francis Group,London,UK.

3. Liu, Y.H., & Liu, H.X., 2006. 'Case learning based on evaluation system for vesselcollision avoidance'. Proceedings of the fifth international conference on machinelearning and cybernetics, Aug 13-16, Dalian.

4. Qu, X., Meng, Q. & Suyi, L., 2011. 'Ship collision risk assessment for the SingaporeStrait.' Accident Analysis and Prevention, Elsevier 43, 2030-2036.

5. Aarsaether, K.G. & Moan, T., 2009. 'Estimation of Navigation Patterns from AIS'. TheJournal of Navigation 62, 587-607.

6. Kao, S.L., Lee, K.T., Chang, K.Y. & Ko, M.D. 2007. 'A fuzzy logic method for collisionavoidance in vessel traffic service'. The Journal of Navigation 60, 17-31.

7. Tasseda, E.H. 2008. 'A study on dynamic traffic management system based on virtualand synthetic AtoN AIS system'. Thesis, Tokyo University of MarineScience andTechnology.

8. Puszcz, A., Gucma, L. &Gucma, M., 2008.'Development of a model for simulation of vessel traffic streams.'Maritime Traffic Engineering Institute.

9.SHIPS' Routeing, 2008. International Maritime Organization (IMO).