An Intelligent Framework for Vessel Traffic Monitoring using AIS Data

Nicos Evmides
Cyprus University of Technology
Limassol, Cyprus
nicos.evmides@cut.ac.cy

Lambros Odysseos
Cyprus University of Technology
Limassol, Cyprus
lambros.odysseos@cut.ac.cy

Michalis P. Michaelides
Cyprus University of Technology
Limassol, Cyprus
michalis.michaelides@cut.ac.cy

Herodotos Herodotou
Cyprus University of Technology
Limassol, Cyprus
herodotos.herodotou@cut.ac.cy

***Abstract*—Automatic identification system (AIS) data provides a wealth of information regarding vessel traffic and is used for a variety of applications such as collision detection and avoidance, route prediction and optimization, search and rescue operations, etc. However, several challenges exist when working with AIS data including huge volume and velocity (as AIS signals are sent by vessels every few seconds), message duplication, various types of data irregularities, as well as the need for real-time processing and analysis. This paper presents a new framework for collecting, processing, storing, and analyzing AIS data in real time plus a set of algorithms for doing so in an efficient and scalable way. At the same time, a set of intelligent services are provided as building blocks for improving and creating new AIS data driven applications. This framework has been operational for the past few years in Cyprus, and has collected and processed around one billion AIS messages from the Eastern Mediterranean Sea.**

***Index Terms*—AIS data, intelligent vessel traffic monitoring**

I. Introduction

The automatic identification system (AIS) is an automated tracking system used on ships and by vessel traffic services for monitoring vessel movements in real time. Signals are sent using VHF signals at frequent time intervals (ranging from seconds to minutes) containing encoded information regarding a ship's attributes at the given time the signal was sent. These attributes include position coordinates in latitude and longitude, speed over ground, course over ground, the vessel's unique identification number (MMSI), and many more.

Despite its simplicity, AIS has become an important source of information for both the maritime industry and researchers. It provides basic information in open air which is readily available to anyone using an AIS receiver; hence, information provided by AIS can be further utilized for various purposes. Existing research work analyzes traffic patterns, performs route prediction, collision detection, ETA prediction, performance optimization, etc. Moreover, AIS has found many practical applications over the years. For example, it has been used for collision avoidance by keeping vessels informed of the positions of nearby vessels as well as for monitoring fishing fleets, navigation aid, maritime security, search and rescue operations, accident investigation, etc. [1]. Maritime Informatics, a discipline that has evolved in recent years, is concerned with digital technology in the shipping industry and makes frequent use of AIS data [2]–[4].

Several challenges exist when working with AIS data. AIS signals are transmitted every few seconds by each vessel. This creates a challenge to manage the data, as the volume and velocity of arriving data is very high [5]. Data needs to be collected and managed effectively to produce meaningful results. Multiple shore-based stations located at distant geographic locations could be receiving the same signal. This creates duplicated messages that need to be dealt with to keep a single version of the message. Some messages may arrive delayed due to station processing and/or network delays, and hence, consideration for delayed messages should also be made. While most AIS information is generated automatically via onboard instruments, some fields such as the destination port is manually entered by the ship's personnel. Currently, there is no agreed standard of how the destination field should be reported [6], [7]. As a result, manually entered data is often polluted with misspellings, truncations, inserted punctuation marks, unexpected abbreviations, and other irregularities. Past studies estimate that 49% of the reported AIS destination entries are polluted with similar errors [8].

The aforementioned issues create several challenges for performing real-time analysis to support decision making procedures as well as creating intelligent services and applications. This paper presents a new framework for collecting, processing, storing, and analyzing AIS data in real time plus a set of algorithms for doing so in an efficient and scalable way. At the same time, a set of intelligent services are provided as building blocks for creating new AIS data driven applications. More specifically, the key contributions of this paper include:

- An end-to-end framework design and architecture for processing AIS data in real time, preparing them for easy use by high level applications;

This work was co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (STEAM Project: INTEGRATED/0916/0063 and MARI-Sense Project: INTEGRATED/0918/0032).

- A set of intelligent processing algorithms to enable deduplication of messages received from multiple sources, cleaning dirty fields of manually entered data, and online data aggregation to support continuous data analysis; and
- A set of intelligent services that utilize the processed AIS data efficiently, including online data analytics, real-time monitoring of areas of interest, and collision detection and avoidance.

The rest of the paper is organized as follows. Related work is presented in Section II. Section III describes the AIS framework architecture for an end-to-end handling of AIS messages. The detailed description of the processing functions and the services are presented in Sections IV and V, respectively. Next, Section VI presents a case study for the Eastern Mediterranean region, where the framework is currently deployed, while Section VII concludes the paper.

II. Related Work

AIS data is widely used in the maritime sector for improving the efficiency of shipping operations, enabling situation awareness, performing route analysis and destination prediction, and detecting potential ship collisions, among others [9]. Recently, Fu et al. [10] have reviewed maritime traffic surveillance systems for spatiotemporal data collection and then presented a computational framework to efficiently compress, transfer, and acquire the necessary information for the further analysis of large-scale AIS data; unlike our approach that focuses on real time analytics and value extraction from AIS data.

Maritime Situational Awareness (MSA) [11] is another framework for detecting and forecasting maritime events of interest, such as illegal fishing, deviation from common routes, or vessels entering shallow waters. MSA also includes provisions for operating on data synopses (e.g., samples, frequency moments) to boost scalability when approximate values are acceptable. The datAcron system [12] supports processing streaming surveillance data, including AIS data, to perform online data cleaning, trajectory reconstruction and compression, as well as trajectory enrichment with other data sources, such as weather or contextual data. Another work also performs link discovery to calculate relations between vessels and areas, which can then be used to detect various types of dangerous or suspicious vessel activity [13].

AIS data has found many applications over a spectrum of maritime research areas. Ship *route analysis* is one that attempts to create a normalcy model using historical AIS data, which will later be compared to new incoming AIS data to *predict the future route* of a ship or *detect any route anomalies*. Pallotta et al. [14] presented a fully unsupervised learning methodology that clusters AIS data into waypoints to construct routes, whereas Sheng et al. [15] proposed an AIS-data-based trajectory clustering model to analyze routes. Using AIS data, elements such as distance, heading, speed, historical routes, etc. can be analyzed to *identify potential collisions* that may occur in order to take early action to avoid such a situation [16], [17]. Finally, *risk assessment*, and *situation awareness* in general, refer to the broader spectrum of risks related to the vessel itself, its surrounding, and the environment, such as narrow navigation paths, shallow waters, high vessel traffic, oil spillage, etc. Various works have utilized AIS data to perform such tasks [18], [19].

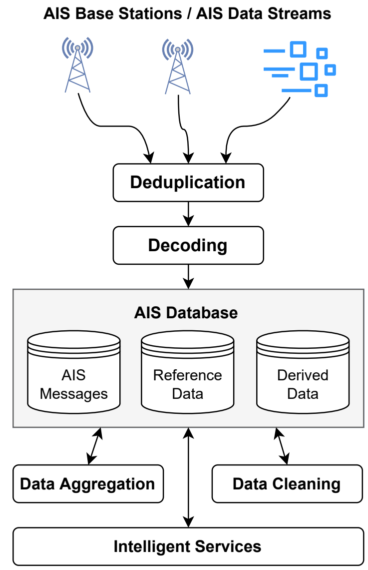

Fig. 1. AIS framework architecture.

III. AIS Vessel Monitoring Framework

The end-to-end architecture of our AIS framework is shown in Figure 1. In order to have the data readily available to be used by higher level applications, the data needs to be collected, pre-processed, organized, and stored in a flexible way. The process starts with the collection of AIS data originating from multiple sources, including live AIS base stations or other AIS data streams. Since data is received from multiple sources, some of which may be located in close geographical proximity, the same AIS message may be received more than once. Only one version of the data must be kept as duplicate messages increase storage requirements, reduce performance, and may skew statistics. Hence, the first process run is deduplication (detailed in Section IV-A), which keeps only a single version of a message. Deduplication can take place without decoding the message by comparing encoded parts of the message, thereby minimizing decoding overheads.

Further processing requires the message to be decoded to get the individual message fields, such as the unique vessel ID, position coordinates, speed and course over ground, destination port, etc. The messages are decoded using libais 0.17 [20] and then are stored in a relational database. According to the AIS standard, there are 27 different kinds of messages, each with its own set of fields and purpose. Thus, different messages are stored in different tables in the database. In addition, the database stores some reference data collected from other sources. For example, it stores a port data table, which lists over 17 thousand worldwide ports from the World Port Index, as well as a vessel registry containing over 23 thousand vessels

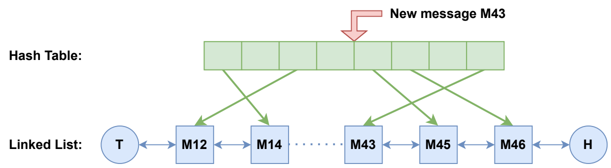

Fig. 2. Custom data structure used during deduplication.

[21]. Both tables are used by the data cleaning procedure as reference. Finally, the database contains several tables containing data derived automatically by the other modules, such as a mapping from dirty destination port entries into clean, standardized port names or aggregated statistics.

Most fields in an AIS message are automatically generated by onboard equipment, but some, such as the destination port field, are manually entered by the vessel's crew. These fields are often polluted with misspellings, wrong abbreviations, or other irregularities. In addition, several messages may arrive with some corrupt data, or missing information (e.g., missing vessel name or dimensions). The data cleaning module, described in Section IV-B, is responsible for cleaning and correcting the data as they arrive in the system. For example, in the case of the destination port, the data cleaning process tries to match the dirty entry with a standardized entry in our database and update the dirty field with a standardized value.

Generating insightful statistics and analytical reports requires high level descriptive metrics of the data. The process to acquire these metrics is called data aggregation (presented in Section IV-C), which is the calculation of descriptive metrics for the data such as counts, max, min and average across different dimensions (e.g., per time, per area). AIS data can very quickly grow in size, making data storage and access of AIS data a challenge. By implementing live data aggregation as data come in, we reduce the access and calculation time needed when information is on demand.

Finally, the framework boasts a set of intelligent services that utilize the processed AIS data collected to produce meaningful results and services to higher level applications end users. These services currently include online data analytics and reporting, real-time monitoring of vessels entering/exiting various areas of interest, and collision detection and avoidance, presented in more detail in Section V.

IV. AIS Data Intelligent Processing

This section presents the key processing tasks, namely deduplication, data cleaning, and online data aggregation, that perform pre- and post-processing of the data to shape it in a more organized and efficient manner.

A. Deduplication

AIS messages are collected by the framework from multiple receiving stations or data streams. Hence, it is frequent to receive duplicate messages within a very short period of time, typically within a few seconds. In addition, it is also possible to receive messages out of order, that is, receive a message that

Algorithm 1 Deduplication Process

1: **procedure** ProcessNewMessage($message$)
2:   **if not** $buffer$.**contains**($message.payload$) **then**
3:     $node = buffer.head$
4:     **while** $node.message.time > message.time$ **do**
5:       $node = node.previous$
6:     **end while**
7:     $buffer$.**insertAfter**($node, message$)
8:     $decoder$.**process**($message$)
9:   **end if**
10:   $node = buffer.tail$
11:   **while** $node.message.time < currTime - interval$ **do**
12:     $node = node.next$
13:     $buffer$.**delete**($node$)
14:   **end while**
15: **end procedure**

refers to an event that took place before an already received message. The deduplication module is able to handle these cases efficiently by utilizing a custom data structure that is used to buffer the messages received in a certain time window (e.g., the last 60 seconds). This data structure maintains a doubly linked list of time-ordered messages along with a hash table that maps a message's payload to the list's node that stores the message, as shown in Figure 2. The presence of the hash table makes checking for an existing duplicate message very efficient, while the linked list enables traversing the messages from both ends.

Algorithm 1 shows the overall deduplication process. When a new AIS message is received, the custom buffer is checked to determine whether the same message has arrived during the configured time window (line #2). This check happens only based on the encoded payload to avoid unnecessary decoding of duplicate messages. If the message is distinct from the existing messages, then the list is traversed from its head backwards until a node is found with a message containing an earlier (or equal) timestamp compared to the new message (lines #3-6). Note that this check is typically very fast because most messages arrive in the correct order. Next, the message is inserted at the correct position in the list and sent to the decoder (lines #7-8). Finally, the list is traversed from its tail forward in order to delete from the buffer all messages that are older than the current time minus the configured time window (lines #10-14).

B. Data Cleaning

As mentioned earlier, some of the AIS fields are entered manually by the vessel crew and may be polluted with misspellings, truncations, unexpected abbreviations, or other irregularities. This data needs to be converted into a standardized form to be usable by high level applications [6]. To achieve this goal, we employ the use of *fuzzy matching*, a general technique for finding strings that match a pattern approximately rather than exactly. In simple terms, given an input string, fuzzy matching will find the most similar string from a list of strings based on a given similarity function.

We have developed a fuzzy matching algorithm that receives dirty port destinations entries from AIS messages and finds the corresponding clean port names based on matching similarities with clean, standardized port records from a reference ports table (recall Section III). Fuzzy Lookup builds and utilizes an Error-Tolerant Index (ETI) for finding matching rows in the reference table. Each record in the reference table is broken up into words, known as tokens, and the ETI keeps track of all the places in the reference table where a particular token occurs. Moreover, the algorithm uses a set of domain-specific rules as well as a custom distance function that takes into account the edit distance, the number of tokens (e.g., port codes, country codes, port names), token order, and relative frequencies. Hence, the matching process is resilient to a variety of errors that are present in the input records, while the cleaned records are associated with a score that indicates the quality of the match. More details can be found in [6].

The output of the Fuzzy Matching algorithm is a mapping table that maps the dirty destination data to the clean data from the reference table. The mapping table can then be used by applications for converting dirty entries to clean ones very efficiently. In addition, the mapping table gradually grows over time as more data is cleaned. Hence, as time goes by, new dirty entries are more likely to already exist in the mapping table and do not need to go through the Fuzzy Matching algorithm.

C. Online Data Aggregation

Data aggregation concerns the procedure of calculating accumulating metrics like counts, max, min and average on AIS messages in order to provide insightful statistics. Since AIS messages are received at a rapid data income rate, we needed a way of optimizing the aggregation process in order to report various statistics with a decreased latency time. To achieve this, we have configured events that are scheduled to execute within our database at frequent time intervals. These events calculate and store various statistics that concern the previous day across various dimensions, such as the number of messages and unique ship visits per day, per port area, per ship status (e.g., moored, anchored), and more.

Whenever data statistics are requested for reporting, we dynamically calculate the metrics of the current day and aggregate them with the previously pre-calculated data. This enables our framework to be fast and report live aggregate results back to the user with very low latency time (i.e., within milliseconds). The reason is because we avoid re-calculating past metrics that will not change in the future. More specifically, since AIS messages are received at real time, it is not expected to receive past messages and thus, previously calculated aggregations will not change. This is a key feature that significantly boosts performance.

V. Intelligent Services

The efficient processing of AIS data described above have enabled us to develop a set of intelligent services for providing online data analytics, real-time monitoring of areas of interest, and collision detection and avoidance, described next.

Algorithm 2 Monitoring areas of interest in real time

procedure NOTIFYAREASOFINTEREST($area$, $vessel$)
  $pastState$ = **getPastState**($area$, $vessel$)
  if $pastState$ == `outside` **then**
    if isInArea($area$, $vessel$) **then**
      notifyEntering($area$, $vessel$)
    else if isInArea($area$, **project**($vessel$)) **then**
      notifyApproaching($area$, $vessel$)
    end if
  else if $pastState$ == `approaching` **then**
    if isInArea($area$, $vessel$) **then**
      notifyEntering($area$, $vessel$)
    end if
  else if $pastState$ == `inside` **then**
    if not isInArea($area$, $vessel$) **then**
      notifyExiting($area$, $vessel$)
    end if
  end if
  setState($area$, $vessel$)
end procedure

A. Online Data Analytics

Our AIS framework provides a wide range of insightful statistics and intelligent data analytics for a particular period of time (e.g., last day, last week), that we also visualize through a web dashboard. First, we numerically report the total messages received, discriminated by message type (e.g., Class A, Class B, Base Station, etc.). Then, we provide the number of unique ships seen grouped by country of registration using a pie chart. Furthermore, we have also utilized line graphs to visually report metrics like number of signals received per day, number of ships seen per day, number of ships moored at specific ports per day, and number of ships at anchoring area per day. We have also integrated a few multiline graphs. The first one, again visualizes the number of ships moored at a particular port per day grouped by the different berth areas at the port. The other multiline graph reports the minimum, maximum, and average waiting time of the ships that are located at the anchoring area per day. All of these data analytics and visualizations provide the user and higher-level applications with important information and meaningful representations that can be further analyzed as needed.

B. Real-time Monitoring of Areas of Interest

Using AIS, it is possible to track vessels in real time that approach, enter, or exit various areas of interest such as port areas, marinas, fishing areas, or marine protected areas. The areas of interest are defined using coordinate points, which connect and form a location polygon. Then, we can check if a vessel is inside an area of interest by checking whether the coordinates of the vessel received in an AIS message are located within the corresponding enclosed area of the polygon. Given that messages are received every few seconds, the key challenge lies in ensuring that only distinct notifications are sent when a vessel is approaching, entering, or exiting an area the first time. Algorithm 2 shows (a simplified version of) the logic used to generate these notifications to interested applications. The input corresponds to one area and information about one vessel received in an AIS message, including its current geolocation, speed, and course. First, the past state of the vessel is retrieved from the database, representing the state reported by the last seen message and corresponds to whether the vessel was outside, approaching, or inside the area (line #1). If the vessel was outside the area and it is now inside the area, a notification is sent for the vessel *entering* this area (lines #3-5). If the vessel is not inside the area, its future location is projected to see if the vessel will enter the area in the near future; if so, a notification is sent for the vessel *approaching* this area (lines #6-8). Currently, the projected location is computed based on the current speed and course of the vessel but more advanced (e.g., machine learning-based) techniques can be used. Similarly, if the vessel was approaching the area and it is now inside the area, a notification is sent for the vessel *entering* the area (lines #9-12). Otherwise, if the vessel was inside the area and now is outside, then a notification is sent for the vessel *exiting* the area (lines #13-16). Finally, the new state is saved in the database.

A variety of optimizations are also implemented to ensure that the notification process is both efficient and robust. First, we use micro-batching to collect all messages received during a small time interval (e.g., last 1 minute) and then invoke Algorithm 2 to avoid flooding the database with requests per single message. Given that the majority of messages received correspond to vessels being outside the various areas of interest, we do not store the 'outside' state in the database, but rather remove any previous states associated with the corresponding vessel and area. Hence, only the states 'inside' and 'approaching' are stored. Finally, given the ship dimensions and heading, we compute four geolocations for each vessel (corresponding to the four corners of the vessel's bounding rectangle). A vessel is considered to be inside an area only when all four geolocations are within that area.

C. Collision Detection and Avoidance

Another beneficial use of AIS data involves making a prediction whether a collision between two vessels is likely to occur. Given the speed and course of two vessels (from AIS messages), we can calculate the positions of the two vessels in the near future (e.g., in 20 minutes), assuming the vessels maintain their course and speed. These calculations create two line segments, one for each vessel, from their current to their future position, as seen in Figure 3. If the two segments intersect, then the two vessels will pass from the same location in the near future. However, for a collision to occur, the two vessel paths must intersect at approximately the same time. Hence, we next compute the time it will take for each vessel to reach the point of intersection, given their current speed. If the two times are within a small delta from each other (e.g., 4 minutes), then a warning is issued that a collision is likely to occur. More advanced closest point of approach (CPA) algorithms are also investigated. Finally, the list of warnings is maintained in the database to avoid issuing repeated collision warnings, unless the warning parameters change.

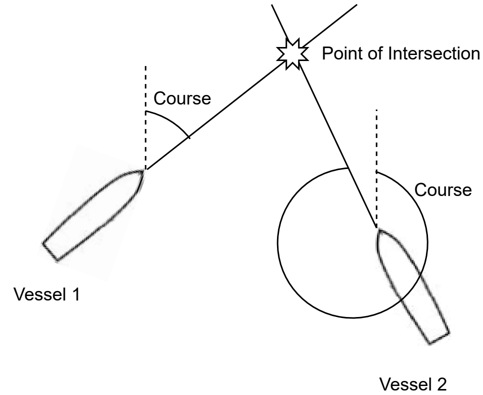

Fig. 3. Vessel course intersection.

TABLE I
AIS MESSAGE TYPES AND COUNTS COLLECTED TO DATE.

Category	**Message Type**	**Count**
Class A Position Report	1, 2, 3	754,206,078
Base Station Report	4	88,780,049
Class A Static and Voyage Data	5	40,883,397
Class B Position Report	18	41,986,520
Other Messages	6–17, 19–27	48,933,056

The collision detection procedure must be checked frequently between pairs of vessels navigating at a close distance. To optimize this process, we collect the most recent AIS messages in frequent intervals and sort them based on latitude and longitude. Next, we check for collision only between vessels whose distance is smaller than a predefined threshold (e.g., 50 Km) instead of generating all possible pairs of vessels. We also have a slightly modified version of the collision detection procedure for the scenario where one vessel is moving and the other vessel is stationary. In particular, we create a line segment for the stationary vessel based on its length (plus some buffer length) and set it perpendicular to the course of the moving vessel.

VI. Case Study: Eastern Mediterranean

Our AIS framework is currently collecting AIS data from one in-house AIS base station and two AIS streams gathering data from 15 other base stations located along the coast of Cyprus and covering most of Eastern Mediterranean sea. Over the past few years, we have collected, processed, and stored close to 1 billion AIS messages, the majority of which (78%) belong to the category Class A Position Report, reporting detailed location information for commercial vessels. Table I shows the number of messages collected per AIS message category to date. Moreover, Figure 4 shows the AIS messages across all categories collected over the past two years, grouped

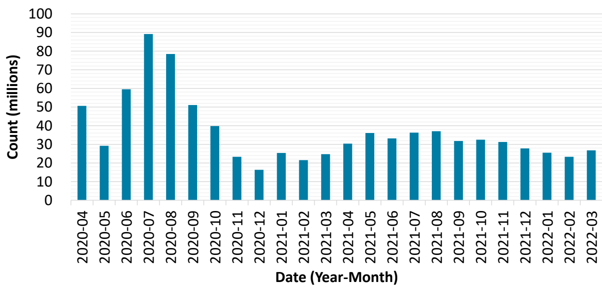

Fig. 4. Collected AIS messaged over the past two years.

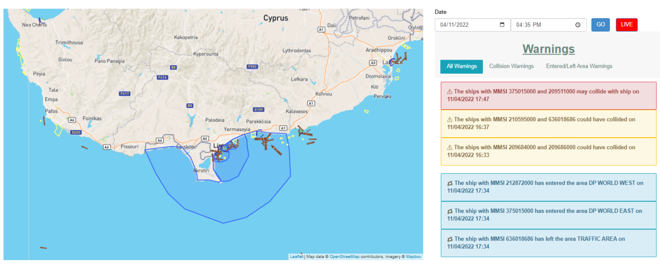

Fig. 5. Collision warnings and notifications sent for vessels entering or exiting areas near Cyprus, as visualized in our framework.

by year and month. On average, 13 messages per second and over 1.1 million messages per day are received and stored in the database. Notice the seasonal pattern that reveals that more messages are collected over the summer months instead of the winter months as VHF signals are impacted by weather.

Another interesting observation from Figure 4 is that prior to October 2020, more messages were collected and stored in the database as that was the time before deploying our deduplication process. Regarding the data cleaning process, 91.6% of the Class A Static and Voyage Data messages containing destination ports were successfully cleaned, while 6.6% contained empty or invalid fields, and only 1.9% of the fields remained unmatched; highlighting the high accuracy of our fuzzy matching approach. Finally, Figure 5 shows a screenshot from a web interface visualizing collision warnings and notifications regarding vessels entering or exiting various areas of interest near Cyprus. In total, 291K notifications were sent over the past 12 months for areas within and around the Port of Limassol, Cyprus, showcasing the fine-grained vessel traffic monitoring that our framework can achieve.

VII. Conclusions

AIS data have a wide spectrum of applications such as route prediction, ETA estimation, collision detection, risk assessment, etc. However, their use is surrounded with various challenges including their huge volume and velocity, duplicate messages, dirty fields, etc. This paper presents a new framework boasting intelligent processing approaches and services that cover the full cycle of AIS data collection and management in an efficient and effective way, that enables its easier use by current and future higher level applications. In the future, we plan to incorporate into the framework more open data provided by other services such as Copernicus and Galileo to complement the collected AIS information.

Acknowledgment

We would like to thank Theofilos Onisiforou, Theofanis Christou, Elizabeth Pieri, Constantinos Christofi, and Paul Constantinou for their contributions in implementing various parts of the framework as part of their undergraduate theses.

References

[1] E. Batty, "Data Analytics Enables Advanced AIS Applications," in *International Workshop on Mobility Analytics for Spatio-temporal and Social Data*. Springer, 2017, pp. 22–35.
[2] M. Lind, M. P. Michaelides, R. Ward, and R. T. Watson, *Maritime Informatics*, 1st ed., ser. Progress in IS. Springer, Cham, 2020.
[3] ——, *Maritime Informatics: Additional Perspectives and Applications*, 1st ed., ser. Progress in IS. Springer, Cham, 2021.
[4] A. Artikis and D. Zissis, *Guide to Maritime Informatics*. Springer, 2021.
[5] H. Herodotou, S. Aslam, H. Holm, and S. Theodossiou, "Big Maritime Data Management," in *Maritime Informatics*, M. Lind, M. Michaelides, R. Ward, and R. T. Watson, Eds. Springer, Cham, 2021, pp. 313–334.
[6] N. Evmides, S. Aslam, A. Televantos, A. Karagiannis, A. Paraskeva, M. Michaelides, and H. Herodotou, "Employing Fuzzy Matching for Cleaning Manual AIS Entries," in *Proc. of World of Shipping Portugal - An International Research Conference on Maritime Affairs*, 2021.
[7] N. B. Abdallah, C. Iphar, G. Arcieri, and A.-L. Jousselme, "Fixing Errors in the AIS Destination Field," in *Oceans 2019-Marseille*. IEEE, 2019.
[8] A. Harati-Mokhtari, A. Wall, P. Brooks, and J. Wang, "Automatic Identification System (AIS): Data Reliability and Human Error Implications," *The Journal of Navigation*, vol. 60, no. 3, pp. 373–389, 2007.
[9] S. Aslam, M. P. Michaelides, and H. Herodotou, "Internet of Ships: A Survey on Architectures, Emerging Applications, and Challenges," *IEEE Internet of Things journal*, vol. 7, no. 10, pp. 9714–9727, 2020.
[10] X. Fu, Z. Xiao, H. Xu, V. Jayaraman, N. B. Othman, C. P. Chua, and M. Lind, "AIS Data Analytics for Intelligent Maritime Surveillance Systems," in *Maritime Informatics*. Springer, 2021, pp. 393–411.
[11] M. Vodas, K. Bereta, D. Kladis, D. Zissis, E. Alevizos, E. Ntoulias *et al.*, "Online Distributed Maritime Event Detection & Forecasting over Big Vessel Tracking Data," in *Proc. of the IEEE International Conference on Big Data (Big Data)*. IEEE, 2021, pp. 2052–2057.
[12] G. A. Vouros, A. Vlachou, G. M. Santipantakis, C. Doulkeridis, N. Pelekis, H. V. Georgiou, Y. Theodoridis, K. Patroumpas *et al.*, "Big Data Analytics for Time Critical Mobility Forecasting: Recent Progress and Research Challenges," in *EDBT*, 2018, pp. 612–623.
[13] G. M. Santipantakis, A. Vlachou, C. Doulkeridis, A. Artikis, I. Kontopoulos, and G. A. Vouros, "A Stream Reasoning System for Maritime Monitoring," in *25th Intl. Symp. on Temporal Representation and Reasoning (TIME 2018)*, vol. 120, 2018, pp. 20:1–20:17.
[14] G. Pallotta, M. Vespe, and K. Bryan, "Vessel Pattern Knowledge Discovery from AIS Data: A Framework for Anomaly Detection and Route Prediction," *Entropy*, vol. 15, no. 6, pp. 2218–2245, 2013.
[15] P. Sheng and J. Yin, "Extracting Shipping Route Patterns by Trajectory Clustering Model based on Automatic Identification System Data," *Sustainability*, vol. 10, no. 7, p. 2327, 2018.
[16] P. Silveira, A. Teixeira, and C. G. Soares, "Use of AIS Data to Characterise Marine Traffic Patterns and Ship Collision Risk off the Coast of Portugal," *The Journal of Navigation*, vol. 66, no. 6, pp. 879–898, 2013.
[17] S. Hornauer and A. Hahn, "Towards Marine Collision Avoidance based on Automatic Route Exchange," *IFAC Proceedings Volumes*, vol. 46, no. 33, pp. 103–107, 2013.
[18] J.-F. Balmat, F. Lafont, R. Maifret, and N. Pessel, "A Decision-making System to Maritime Risk Assessment," *Ocean Engineering*, vol. 38, no. 1, pp. 171–176, 2011.
[19] H. Zheng-wei, Y. Fan, and L. Li-rong, "Ship Safe Navigation Depth Reference Map based on AIS Data," *Journal of Traffic and Transportation Engineering*, vol. 18, no. 4, pp. 171–181, 2018.
[20] "Libais." [Online]. Available: https://pypi.org/project/libais/
[21] "World Port Index," 2019, National Geospatial-Intelligence Agency. Twenty-Seventh Edition Pub 150. [Online]. Available: https://msi.nga.mil/Publications/WPI