A Scalable System for Maritime Route and Event Forecasting

Georgios Grigoropoulos
ggrigoropoulos@kpler.com
Kpler, Research Labs
Athens, Greece

Giannis Spiliopoulos
gspiliopoulos@kpler.com
Kpler, Research Labs
Athens, Greece

Ilias Chamatidis
ichamatidis@kpler.com
Kpler, Research Labs
Athens, Greece

Manolis Kaliorakis
mkaliorakis@kpler.com
Kpler, Research Labs
Athens, Greece

Alexandros Troupiotis - Kapeliaris
atroupiotis@kpler.com
Kpler, Research Labs
Athens, Greece
University of the Aegean
Ermoupolis, Syros, Greece

Marios Vodas
mvodas@kpler.com
Kpler, Research Labs
Athens, Greece

Evangelia Filippou
efilippou@kpler.com
Kpler, Research Labs
Athens, Greece

Eva Chondrodima
evachon@unipi.gr
University of Piraeus
Piraeus, Greece

Nikos Pelekis
npelekis@unipi.gr
University of Piraeus
Piraeus, Greece

Yannis Theodoridis
ytheod@unipi.gr
University of Piraeus
Piraeus, Greece

Dimitris Zissis
dzissis@aegean.gr
University of the Aegean
Ermoupolis, Syros, Greece

Konstantina Bereta
kbereta@kpler.com
Kpler, Research Labs
Athens, Greece

ABSTRACT

Digital twins serve as virtual representations of physical environments that are increasingly valuable across various sectors, including maritime operations. The complexity of monitoring vessel traffic through the Automatic Identification System (AIS), demands more sophisticated, data-driven approaches due to the extreme vessel volume and intricate vessel movement patterns. In this paper, we present a highly scalable system for maritime route and event forecasting that leverages streaming real-time AIS data for vessel route prediction, traffic state estimation, and event detection based on the actor model. The proposed solution integrates data driven models on the actor level for route forecasting that utilize vessel specific features and are adapted according to the system requirements and the limitations of AIS streaming service networks. Leveraging the forecasting functionalities, the platform is able not only to detect but also forecast events of interest for the entire global fleet accurately and consistently without any memory or system issue and continuously generate predictions for 170K vessels during performance experiments, which demonstrates the high scalability and versatility of the proposed architecture for global maritime digital twin applications.

1 INTRODUCTION

Digital twins serve as a virtual representation or a model of a physical environment mirroring its real-time state [16]. Currently, such representations find utility in multiple applications, including transportation, manufacturing, health and energy. The transfer of real-world physical information to virtual models enables users to easily simulate, facilitate data collection, analyse, perform informed decision making and improve performance of complex real world physical systems. In this context, digital twins have the potential to revolutionize maritime operations providing a realistic environment for enhanced marine traffic monitoring, vessel behavior prediction, scenario simulation, route planning and ultimately foster informed decision making, as well as safe and efficient maritime operations.

Monitoring and predicting the behavior of hundreds of thousands of vessels of the global fleet is a challenging undertaking. The Automatic Identification System (AIS), in operation for over 15 years, facilitates the collection of vessel positions worldwide [21]. At the same time, present vessel traffic management systems (VTMSs) and vessels traffic monitoring information systems (VTMIS) [19] heavily depend on basic and simplistic linear models, considering standard vessel kinematic features to predict positions and the overall vessel traffic in specific sea areas. However, these methods suffer from low accuracy, making current systems unreliable, particularly in safety-critical situations [5, 23, 33]. Given the intricate patterns of vessel movement, sophisticated techniques are essential to model and forecast vessel behavior. Recent years have seen the proposal of various architectures relying on data-driven methods [13]. Furthermore, the substantial volume of AIS messages transmitted globally on a daily basis underscores the necessity for a specialized tool capable of effectively managing big data.

Kpler uses a network of more than 7000 AIS receivers strategically placed around the world to collect information from 400K+ vessels equipped with AIS transponders. Processing streams of more that 10^9 AIS messages on a daily basis, vessel tracking data is made available publicly through the MarineTraffic website and mobile apps, allowing users to track and monitor maritime traffic globally in real-time [10]. More than 4 million users visit monthly. Given the widespread coverage and the geographical dispersion of MarineTraffic services (Figure 1) as well as the extreme volume of AIS messages that requires robust algorithms and infrastructure to manage, store, and retrieve data efficiently, scalability, efficiency, and big data processing capabilities are fundamental for successfully delivering real-time vessel tracking on a global scale.

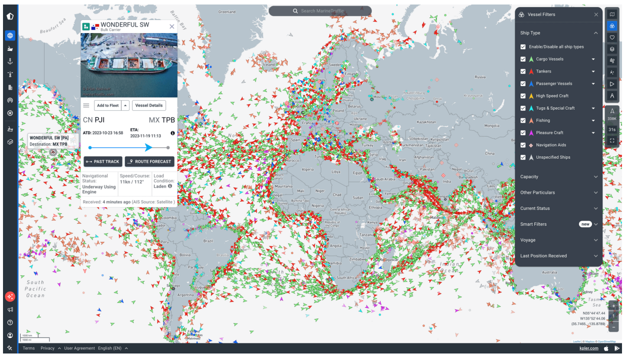

Figure 1: Global real-time vessel tracking on the MarineTraffic (Kpler) application [10].

This work proposes a pilot for maritime route and event forecasting that leverages real-time AIS data feeds and fosters data driven models for vessel route prediction, traffic state estimation and early maritime event detection aiming to further enhance the accuracy of maritime representations.

In this context, the main contributions of this work include:

(1) A distributed and highly-scalable system architecture for global vessel traffic monitoring and maritime event forecasting based on the actor model [7] that addresses key shortcomings in scalability, scope, speed and accuracy of present vessel traffic management solutions through the fusion of heterogeneous extreme-scale data associated with diverse layers of information that are specific to individual vessels.

(2) Data-driven vessel route and event forecasting models and functions integrated at the actor level that leverage vessel specific information for real-time global vessel traffic monitoring, route forecasting and precise maritime event predictions.

The paper is organised as follows: First, Section 2 presents an overview of existing solutions for maritime situational awareness and vessel traffic management. In Section 3 the system architecture is described. Subsequently, the major components related to the forecasting (Section 4) models and the event detection and forecasting features (Section 5) of the application are presented. These sections focus on describing the methodology followed in developing and integrating the respective components to the digital twin platform. Section 6 presents the evaluation of the integrated models and functions as well as the results on the system scalability performance. Finally, Section 7 presents the main conclusions and future outlook of this work.

2 RELATED WORK

Applications for maritime situational awareness are constantly developed and researched by multiple parties around the world. Most commercially and non-commercially available solutions focus on specific aspects in the fields of maritime operations and maritime monitoring. These include primarily use cases for maritime security, safety [15, 24, 27, 29], operations [1, 5, 12, 16, 18, 25] and autonomous vessels [18, 22]. In this context, such solutions are mostly deployed on a local or regional level depending on the limitations and requirements of the specific use case. These include the monitoring of maritime activities and vessel traffic in prespecified sea areas (e.g. coastal areas, ports) for vessel traffic management and operations. Other use cases, focus primarily on maritime security and compliance such as smuggling, trafficking activities and illegal fishing in specific sea areas, through the fusion of multiple sensor views, while leveraging anomaly detection algorithms for agile decision making. Typical competences of these platforms include data fusion capabilities, fusing views of global data such as the AIS with local views from sensors that, depending on the use case, include video cameras, radar, remote sensing or vessel integrated sensors [1, 5, 12] focusing on monitoring and anomaly detection. Finally, the enhanced situational awareness views are delivered to the end users through virtual or augmented reality modules.

Therefore, maritime situational awareness platforms are severely restricted in their scalability capabilities as they are tailored to provide enhanced views in specific sea areas limited by their use case requirements and affiliated sensor ranges. Additionally, their components, focused on monitoring and event detection, have limited event and vessel behavior forecasting capabilities. Thus, they do not foster proactive decision making, as they concentrate on delivering present situational information rather than exploiting historical and real-time streaming data flows to automate event detection procedures as well as early event forecasting and action planning. Situation and behavior forecasting is supported only through the combined views from different data sources and analytics that are presented to the end users in real time, while situational interpretation, forecasting and decision making is limited by the end user experience and competence. Thus, the fusion of historical information and insights together with real time event detection and monitoring coupled with data-driven forecasting components has the potential to further expand the capabilities of maritime situational awareness platforms. This can be achieved through automated early event detection and forecasting functions that foster proactive action planning, improve interpretability and reduce task complexity for human operators.

3 SYSTEM ARCHITECTURE

Figure 2 presents the high level system architecture of the pilot platform. Based on the state of the art findings and the research needs assessment, the proposed system aims to diversify its application scope in maritime operations and vessel traffic management by addressing the two main shortcomings of maritime situational awareness systems, namely their scalability and forecasting capabilities. This is achieved through the adoption of a scalable and distributed system architecture for the fusion of heterogeneous extreme-scale data associated with diverse types of information specific to vessels, sea areas and routes. Additionally, data driven models, trained over historical mobility vessel data are integrated into the platform and and coupled with the real-time streaming feed of AIS data, the real-time event detection functions and the visualization of historical aggregated vessel mobility metrics creating a a multi-layer information system for efficient action planning and enhanced decision making.

For addressing the scalability, real-time processing and forecasting requirements, the proposed multi-layered system is based on the actor model [7]. In particular, the actor model implementation of the Akka framework [8] is selected for its dynamic scaling capabilities that allows the development of concurrent and distributed systems with a focus on asynchronous message-passing communication. This can be advantageous for building complex system logics and orchestrating distributed concurrent

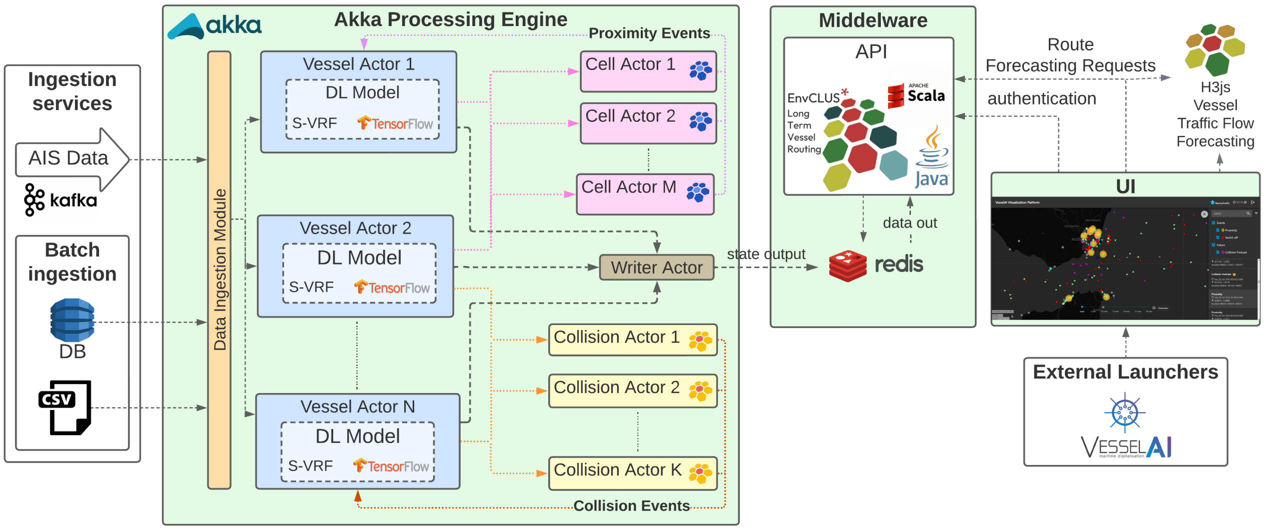

Figure 2: The proposed system architecture based on the actor model.

workflows. Additionally, the lightweight, isolated actors facilitate the development of a scalable and responsive system with built-in mechanisms for messaging between actors and for handling state and failures. Its real-time processing capability allows the processing data-streams coming from multiple Kafka connections [11]. In this context, the data ingestion services of the processing engine consume streaming real-time positional AIS data transmitted from the global fleet which is captured by satellite services, the Kpler MarineTraffic AIS terrestrial network and third-party providers of AIS real-time data operating globally. Additionally, at the initialization phase, any static information required to be fused with the streaming information is provided, either by direct requests to databases or in the form of static files located in the file-system. As soon as the information is retrieved, it is cached in memory, available for fast retrieval from all actors in each node.

The data stream is partitioned in multiple ways to take advantage of the actor model. The core partitioning functionality generates multiple actors N, with each one corresponding to a specific vessel as it is defined by its unique Maritime Mobile Service Identity (MMSI). For improving the forecasting capabilities of maritime situational awareness platforms, novel short-term and long-term route forecasting models, as well as event detection and forecasting methods are developed and integrated with the proposed architecture. Forecasting models are loaded and applied at an actor level to process AIS data and generate vessel specific predictions. The models are mapped on an 1-to-1 basis for each respective *vessel actor*. Specifically, the short-term vessel route forecasting model is mounted only once in memory, serving simultaneously the requirements of each of *vessel actor* defined by the processing engine and generating in this way dedicated predictions per *actor* (real vessel) using vessel specific information provided by MarineTraffic according to digital twin concept. In this way, at any point in time, the system provides the most updated predictions for the entire global fleet.

Two additional actor classes are defined on the spatial level utilizing the H3 spatial index [26], a class for proximity event detection with variable size M and a class for collision forecasting with variable size K. These actors consume the combined output of all *vessel actors* N and determine the state of their respective event class. Specifically, AIS positional data are sent to the *cell actors* for proximity event detection, while short-term position forecasts, are forwarded to *collision actors*. In both cases, the information organised in messages, is sent to the corresponding actor with respect to the type of the operation and the H3 index of each message. Based on the final state status, they communicate their state back to the respective affected subset of *vessel actors*. In this context, it is possible that some *cell* or *collision actors* receive more messages than others. This can be the case in sea areas with high vessel traffic. However, based on initial experimental results, it is found that these cases do not slow down the system. The respective methodologies and model architectures for forecasting are described in more detail in subsequent sections.

Finally, the actor states are stored by the *writer actor* in a Redis [14] database in order to be visualized by the User Interface (UI) through a dedicated API responsible to interface the frontend with the backend systems. Depending on system and application requirements, multiple *writer actors* may exist and be supported by Akka concurrently. In such cases, each other type of actor may be assigned to send its output to a specific *writer actor* and maintain the connection to the Redis database. In the context of this work, a single *writer actor* has been defined to write all actor outputs to the Redis database. The Redis and the API belong to the Middleware component. The end user is able to interact with the system by exploring the visualized route an event states through the UI.

4 FORECASTING

The system aims to expand functionalities and applications for the end users through the integration of data-driven models for vessel behavior prediction and planning. The accurate prediction of vessel positions is significant for applications in maritime operations and safety. Since, vessel routing is one of the most significant operational tasks for vessel crews and vessel control center operators. In this context, the proposed platform integrates two distinct Vessel Route Forecasting (VRF) models, one for long-term route forecasting and planning and one for short-term route forecasting. Both models consider vessel specific attributes (type,

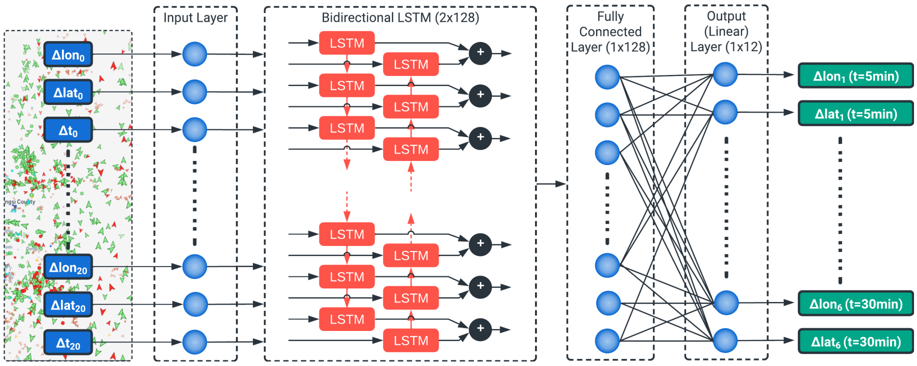

Figure 3: The S-VRF model architecture integrated with the ingestion services on the actor level. It consists of one input layer, one BiLSTM layer, one fully connected layer, and an output layer for 6 transitions (Δlon,Δlat) in 5-minute intervals and up to a 30-minute time horizon.

dimensions, draught, etc.) as well as past motion behavior features for facilitating predictions.

4.1 Long-term Vessel Route Forecasting

For the long-term route forecasting (L-VRF) needs, the system integrates an extended version of EnvClus [34], namely the EnvClus* model [28, 35], through API calls that forecasts the vessel path towards a destination port, given a specific port of origin (Figure 4a).

The method trains a dedicated model for each distinct pair of origin-destination ports by exploiting the collective intelligence generated by hundred of thousands of vessel routing decisions that are extracted from historical AIS data by MarineTraffic (Kpler). The positional AIS data is clustered in order to extract common pathways of vessel movements. In turn, these pathways are translated into a weighted transitions graph, representing the patterns of movement found in the historical data. Using the resulting graph we are able to generate a prediction of the path the vessel is going to follow towards its destination port. Segments from historical AIS data are being used to generate these predictions, allowing for the extraction of typical realistic paths that avoid dangerous routes or crossing over land. Vessel-specific information is utilized to generate the best-suited forecasts for each query, by enhancing the graph with classification models in significant graph nodes (route junctions). Features may include the vessel type, length, draught, deadweight tonnage (DWT) or trip related information (time of day, month).

EnvClus* scales at a global level for any origin-destination port pair and is able to generalize and scale on trip data for unseen origin-destination pairs and vessel types [28, 34, 35]. Finally, aggregated mobility statistics regarding the vessel traffic at the selected area are also generated and visualized for the user. These statistics, called Patterns of Life [32], are extracted from historical data from relevant trips and provide a more complete overview of the historical traffic in the area (Figure 4b). The fusion of the present vessel position, with the route forecasts and the aggregated vessel mobility insights, allows the user to assess the efficiency of the current vessel route, evaluate possible rerouting strategies and alternative routes towards the destination port and detect possible deviations from common vessel traffic patterns.

4.2 Short-term Vessel Route Forecasting

In [4] a Short-term Vessel Route Forecasting (S-VRF) Long Short Term Memory (LSTM) model is proposed that was trained on real world AIS datasets covering specific sea regions. It showcased significant performance improvements in comparison to state-of-the-art methods for vessel location forecasting [4], and baseline techniques for vessel trajectory forecasting [3]. The model considers special characteristics of the AIS databases such as the irregularity of AIS transmissions. The model is able to generate vessel location predictions given variable input sequences k of past vessel latitudinal displacements $[(p_{lat_0}, p_{lon_0}, t_0), ..., (p_{lat_k}, p_{lon_k}, t_k)]$ and for a variable number of future transitions r up to a preset prediction horizon Δt at (fixed) timestamps with a sampling rate equal to $\Delta t/r$. Thus, the original problem formulation is generic, with flexible input and output requirements as defined by the end user and depending on the quality and availability of data sources. Initial integration of the LSTM model presented in [4] in the system architecture resulted in poor performance. This is attributed to the special requirements of the real-time Akka based system, which include memory restrictions, scalability, transferability, stable on-stream system performance as well as the fact that the model has to generate predictions using on stream AIS position transmissions, the retraining, modification and investigation of a new model architecture was necessary.

Specifically, a new S-VRF model architecture has to be defined that fullfills the requirements of the system architecture. More specifically, the tensor input size is reduced to 20 past vessel spatiotemporal displacements in comparison to the initial tensor with a maximum size for 1000 displacements and variable filling. In this way the restrictions regarding system memory allocation are met, while the complexity for the training task by the model is simplified as the model is trained with a fixed tensor size input. The prediction output is fixed to six spatial transitions with 5-minute intervals and up to a 30-minute time horizon. Finally, a 30 second downsampling rate is set as a minimum limit for AIS aggregated sequential transmission intervals and is validated after additional experimentations. As AIS transmission frequencies show extreme irregularities due to various factors that include but are not limited to, environmental factors, vessel motion state (speed, turning), equipment quality and reliability and AIS network coverage, the preprocessing component of the model is modified to be able to consider aggregated AIS message transmissions up to a 30-second downsampling rate.

The new model architecture is presented in Figure 3. The architecture is also modified, compared to [4], with updated layer sizes and the use of a bidirectional LSTM layer (BiLSTM), instead of the standard LSTM layer. BiLSTM adds one more LSTM layer, which

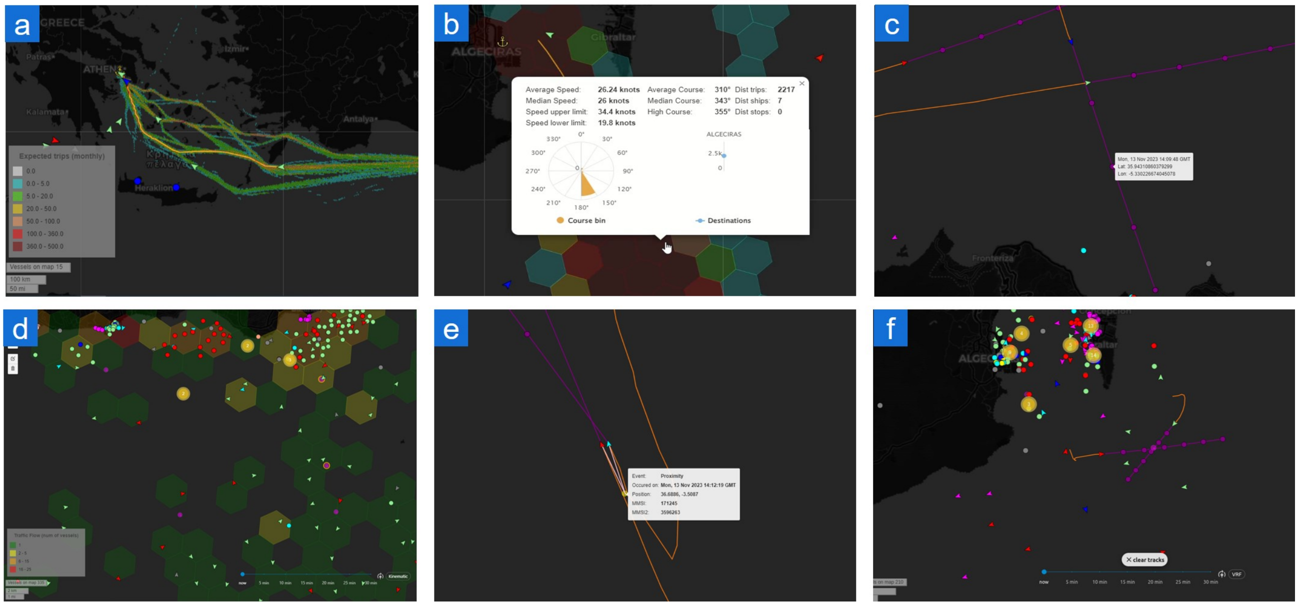

Figure 4: Visualization of the key features and forecasting components on the user interface of the Maritime Situational Awareness Platform. a) L-VRF, b) Aggregated vessel mobility statistics inspection in long-term routing, c) S-VRF, d) Vessel Traffic Flow Forecasting, e) Proximity Event Detection, f) Vessel Collision Forecasting using the S-VRF.

reverses the direction of information flow. BiLSTMs address the probability that the state of an element in the sequence does not only depend on past element states but on future element states. Concatenation is used for combining the bidirectional LSTM-layer outputs. Finally, the BiLSTM is coupled with L1 in-layer regularization for reducing overfitting.

An example of the visualization of the prediction for end-user through the user interface is presented in Figure 4b.

5 MARITIME SITUATIONAL AWARENESS

The system provides users with typical functions for maritime event detection. These include composite events regarding vessel movement and AIS transmission status that can be detected and logged for inspection. Such events include the close proximity between vessels (see Figure 4e) and the switch-off of the AIS transmitter on a vessel [9]. The system additionally leverages the S-VRF model integrated on the actor level for maritime event forecasting. Here, two operations have been integrated into the platform, one for vessel traffic flow forecasting and one for vessel collision forecasting.

5.1 Vessel Traffic Flow Forecasting

The objective of Vessel Traffic Flow Forecasting (VTFF) is to predict the evolution of traffic (number of requests) in a specific region at a future time step/window, drawing insights from historical traffic flow data [31]. Typically, vessel traffic flow data adopt the form of spatiotemporal raster data [6, 30]. Raster data are grid-based data referring to observations of a continuous spatiotemporal field represented at fixed locations or regions in space and time. More specifically, in this work, we focus on maritime traffic data taking the form of raster data categorized as grid-based. Grid-based approaches organize raw traffic data within a set of grids, simplifying the problem scale [30].

In the literature, the most promising methods used in predicting vessel traffic flow mainly employing grid-based representation analysis [20], approach the VTFF problem from two different perspectives [17]: a) indirectly - as a route forecasting application by estimating future traffic based on vessel locations produced by VRF algorithms, and b) directly - as a flow sequence forecasting problem by predicting future traffic through sequence analysis of historical traffic flow. In [17] a comparative analysis between the indirect and direct VTFF strategies was presented, concluding that the indirect paradigm generally demonstrates superior prediction accuracy, often exceeding 1.5 times the accuracy of the direct VTFF alternative. The indirect VTFF not only predicts more accurately than the direct strategy but is also less computationally demanding, especially when the underlying VRF has already been implemented within the same system.

The VTFF approach based on the S-VRF model (cf. Section 4.2) has been integrated into the actor model based system. The process involves feeding the framework with streaming AIS messages, employing the S-VRF model to produce future trajectories composed of 6 transitions at a sampling rate of 5 minutes up to 30 minutes. The predicted locations by the S-VRF model are allocated into a spatiotemporal grid formulated by the H3 grid [26]. The resulting vessel counts represent the volume of vessels, i.e., the vessel traffic flow, in each grid cell and time window.

An illustration of the forecasted traffic flow through the user interface is depicted in Figure 4d. Specifically, vessels are displayed within grid cells, with only the active cells (those containing vessels) being visible. Furthermore, cells indicating low traffic are depicted in dark green, cells with medium traffic are represented in light green, and cells with high traffic are displayed in red.

5.2 Vessel Collision Forecasting

The vessel collision forecasting algorithm is integrated in the system on the actor level and utilizes the selected S-VRF model for detecting probable imminent collisions between two vessels.

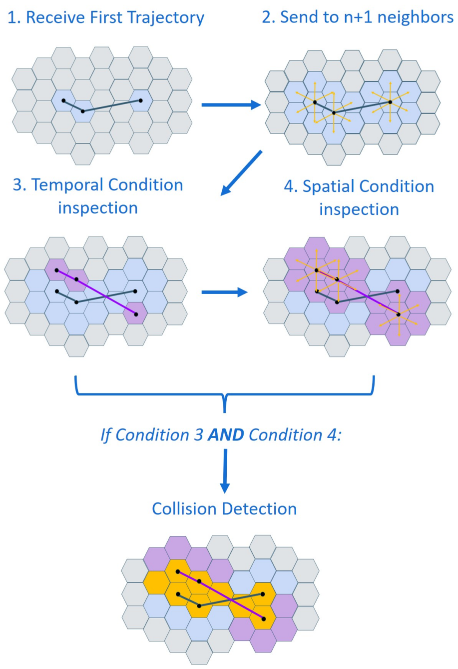

Figure 5: Example of information sharing for collision prediction among actors, for two vessel trajectories. In this example the trajectories contain 3 predicted points (instead of 6 that are produced by the S-VRF model for simplification).

First, streaming AIS messages are ingested. Subsequently, each AIS message from a vessel is redirected to an actor and generates a route forecasting prediction with the S-VRF model. The output of each trajectory prediction per vessel/actor and for every received AIS message consists of 7 positions (1 present position and 6 position predictions). The output positions are assigned to the respective cell of the H3 grid [26] n and each $n+1$ nearest cell. Each cell corresponds to an actor that is responsible for all the messages of the cell. To assess if two vessels are into a collision course, the algorithm first checks the temporal intersection using the 30 minute maximum prediction window and according to a system defined time interval threshold that accounts for close proximity vessel passes. Then, the spatial intersection of the forecasted trajectories is assessed. If both conditions are true, then a potential collision is detected and logged for the end-user. Figure 5 presents an example of the information sharing between the actors for resolving the collision prediction problem. An example of the visualization of the prediction for end-users through the UI is presented in Figure 4f. The user is able, using the list of events, to be notified regarding forecasted collisions. The events automatically appear in an event list, which can be used for quick navigation to the location of the forecasted collision. In addition, the user can check the estimated time of the collision, as well as the MMSIs of the involved vessels.

6 EVALUATION

6.1 S-VRF Evaluation

The S-VRF model was trained and tested using archived AIS stream data from MarineTraffic (Kpler). The dataset includes AIS transmissions from 24 hours on the 02.11.2021 covering the entire European continent, the North Atlantic Ocean, the Barrents Sea, the Kaspian Sea, the Red Sea and the Persian Gulf.

The exact area of coverage defined in the WGS84 projection system is ((24.0000°,-41.99983°), (24.0000°,68.9986°), (78.9862°,-41.99983°), (78.9862°,68.9986°)). The total dataset corresponds to 1,4617,382 AIS messages (16.93GB) that were transmitted by 14,895 distinct vessels (MMSIs) over the 24-hour period. As the AIS dataset originated from the streaming service of MarineTraffic with AIS messages coming from both the MarineTraffic (Kpler) AIS terrestrial network and third-party AIS satellite services, the sampling rate was dynamic and irregular for all vessels. Considering the 30 second set minimum downsampling rate, after aggregating the AIS messages for each vessel during preprocessing, the final average sampling rate is 78.6 seconds with a standard deviation of 418.3 seconds for the entire dataset.

In order to transform the AIS vessel positions into the fix-sized input and output tensors, the distinct vessel trajectories are segmented into distinct partitions and shuffled building a final total dataset of 232,852 trajectories. The resulting ground-truth trajectory segments include the 20 spatiotemporal transitions used as the input tensor and the spatial displacement in the target 30-minute temporal prediction window which is transformed into the output tensor by interpolating and transforming the respective transitions into six 5-minute segments. 50% of the dataset was used for training, while the remaining 50% for validation (25%) and for testing (25%). Finally, the test results were compared to a simple linear kinematic model which utilizes the last reported AIS position, reported AIS speed (knots) and course (°) to predict future vessel positions in the same time horizons.

Table 1: S-VRF performance results on the AIS MarineTraffic Stream Dataset. Comparison of the Average Displacement Error (ADE) in meters. over all six prediction horizons (t=5min,...t=30min).

ADE per prediction horizon	Linear Kinematic Model	S-VRF	% Difference
t = 5min	97.7	91.7	-6.1
t = 10min	256.6	232.0	-9.6
t = 15min	457.0	408.7	-10.6
t = 20min	688.2	609.5	-11.4
t = 25min	943.5	828.9	-12.1
t = 30min	1216.3	1060.2	-12.8
Mean ADE	609.9	538.5	-11.7

Table 1 presents the performance results for based on the average displacement error metric for each prediction horizon. The new model manages to outperform the linear kinematic model for all predictions horizons. Results are also comparable to the original generic architecture [4], given the significantly wider area of coverage as well as the limitations and challenges of the integration with streaming AIS data.

6.2 Vessel Collision Forecasting Evaluation

For the needs of the evaluation of the vessel collision forecasting algorithm, a synthetic dataset of vessel proximity events was utilized. The dataset consists of 4,658 samples/AIS messages originating from 213 unique vessels in the Aegean Sea generating 237 vessel proximity events [2]. Two additional sub-datasets were created, one by taking only the vessels that will come into close proximity in less than 2 minutes (Sub dataset A) and one with vessels that will come into close proximity in less than 5 minutes (Sub dataset B). Sub dataset A has 61 events and Sub dataset B consists of 152 events. Eight different sets of experiments were tested using the simple linear kinematic model and the S-VRF model for the three different vessel event specifications (Original Dataset, Sub dataset A (Proximity), Sub dataset B (Proximity)). For each experiment the accuracy, precision, and recall based on the measured True Positives (TP), False Positives (FP) and False Negatives (FN) was assessed.

Evaluation results are presented in Table 2. The collision prediction using the data driven S-VRF model offers higher accuracy than using the outcome of the linear kinematic model. In all experiments, collision prediction with the S-VRF model achieves more than 90% in terms of accuracy, precision, and recall, compared to the ground truth data. Collision prediction using the S-VRF model generates more false positives, while the linear kinematic more false negatives leading to small differences in the F1-score metrics. Eventually, as the investigated method for collision forecasting is a safety-critical application for the detection and mitigation of vessel collisions with significant consequences in case of non-detections, recall is the decisive performance metric as it indicates that the system is effectively capturing and identifying all events that actually require attention or intervention. Under this assumption, the S-VRF model excels over the linear kinematic model in both critical performance metrics, the accuracy and the recall.

6.3 System Scalability Evaluation

The system scalability is evaluated using the global real-time AIS data stream of MarineTraffic AIS Service by Kpler. In the experimental evaluation the system is deployed in a single virtual machine (VM) with 12 cores and 128 GB RAM, the S-VRF is selected as a typical use case for generating predictions. During the evaluation, the system was operational for 72 hours without any memory or system issue and managed to continuously generate predictions for 170K real vessels. This number corresponds to all vessels that were tracked during this 72-hour period globally by the MarineTraffic AIS Network Service. Thus, the proposed system is highly scalable. As depicted in Figure 6 the system is able to perform complex calculations on live data at a global scale reporting very low processing times. It averages less than a few milliseconds calculated using moving window of 100 actors (vessels). The system during the initialisation phase (up to 5K actors (vessels)) peaks on processing time due to excessive needs of computational resources and massive introduction of new actors. After this phase the rate of unseen vessels decreases significantly and thus the system passes onto a stable state, where given an increasing the number of actors it is able to continue ingesting AIS messages in real time.

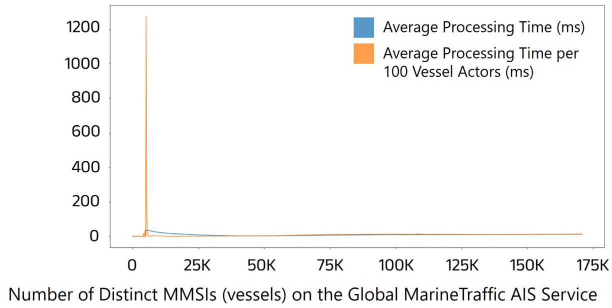

Figure 6: The average processing time with respect to the total number of actors active on the system, along with a moving window average of 100 actors.

7 CONCLUSIONS AND FUTURE WORK

In this work we showcase how the adoption of extreme-scale solutions can address shortcomings in scalability & scope, speed and accuracy of vessel traffic management applications. In this context, we presented a highly scalable platform for maritime situational awareness serving as a digital twin that focuses on both the detection and the forecasting of vessel behavior and maritime events for efficient and safe maritime operations. The proposed solution that is based on the actor model is able to perform short- and long-term route forecasting utilizing vessel specific features, using AI models that are integrated on the actor level. Leveraging the forecasting functionalities, the platform is able to detect and forecast events of interest for the entire global fleet accurately, consistently and efficiently.

In future work, we aim to leverage Kafka topics to produce streams of dedicated system, model and actor-based outputs and to develop API endpoints for facilitating external user interaction. We also intend to incorporate the detection and forecasting of even more complex maritime activities, leverage new data sources to improve model prediction performance (e.g. weather data, vessel sensors) and fuse heterogeneous extreme-scale data associated with diverse related types of information. This also includes the enrichment and fusion of the H3 spatially indexed AIS mobility data with weather related features and forecasts that will further improve the situational interpretability and comprehensiveness of maritime operations for the end user. Additionally, we aim to further integrate additional pattern extraction and forecasting models specifically targeted to common maritime operations and industry needs. New assets may include the monitoring and prediction of berth and port congestion, the automated rerouting for vessel collision avoidance and the consideration of weather related features in vessel routing. In this context, we aim to preserve and expand the highly scalable and adaptable architecture of the platform and ultimately create a detailed representation for vessel mobility that serves as a digital twin of global maritime situational awareness.

ACKNOWLEDGMENTS

The presented digital twin, the system architecture, the Short-term (S-VRF) and Long-term (EnvClus*) Vessel Route Forecasting model versions, the version of the indirect Vessel Traffic Flow Forecasting methodology adopted from [17] as implemented within the proposed system architecture in this paper and the event detection and forecasting functions, were developed within this work and are owned by Kpler.

This research is partially supported by European Union's Horizon 2020 research and innovation programme under grant agreement No 957237, project Enabling Maritime Digitalization by Extreme-scale Analytics, AI and Digital Twins (VesselAI) and by European Union's Horizon 2020 research and innovation programme under grant agreement No. 101092749, project "Critical Action Planning over Extreme-Scale Data (CREXDATA)".

REFERENCES

[1] Nur Assani, Petar Matić, and Marko Katalinić. 2022. Ship's Digital Twin—A Review of Modelling Challenges and Applications. Issue 12. https://doi.org/10.3390/app12126039

Table 2: Evaluation of vessel collision forecasting

Dataset	Model	Temporal Difference Threshold (min)	Total Proximity Events	TP	FP	FN	Precision	Recall	F1-Score	Accuracy
All Events	Linear Kinematic	2	237	203	3	34	0.98	0.85	0.91	0.85
	S-VRF		237	214	11	23	0.95	0.9	0.92	0.9
	Linear Kinematic	5	237	222	4	15	0.98	0.93	0.95	0.93
	S-VRF		237	231	16	6	0.93	0.97	0.95	0.97
Sub dataset A	Linear Kinematic	2	61	60	1	1	0.98	0.98	0.98	0.98
	S-VRF		61	60	5	1	0.9	0.98	0.94	0.98
Sub dataset B	Linear Kinematic	5	152	147	4	5	0.97	0.96	0.96	0.96
	S-VRF		152	149	16	3	0.9	0.98	0.94	0.98

[2] Ilias Chamatidis, Giannis Spiliopoulos, Manolis Kaliorakis, Georgios Grigoropoulos, and Konstantina Bereta. 2023. *Synthetic AIS Dataset of Vessel Proximity Events*. https://doi.org/10.5281/zenodo.8358665

[3] Eva Chondrodima, Petros Mandalis, Nikos Pelekis, and Yannis Theodoridis. 2022. Machine Learning Models for Vessel Route Forecasting: An Experimental Comparison. In *2022 23rd IEEE International Conference on Mobile Data Management (MDM)*. 262–269. https://doi.org/10.1109/MDM55031.2022.00056

[4] Eva Chondrodima, Nikos Pelekis, Aggelos Pikrakis, and Yannis Theodoridis. 2023. An Efficient LSTM Neural Network-Based Framework for Vessel Location Forecasting. *IEEE Transactions on Intelligent Transportation Systems* 24, 5 (2023), 4872–4888. https://doi.org/10.1109/TITS.2023.3247993

[5] Samantha Cope, Ellen Hines, Roger Bland, Jerry D. Davis, Brendan Tougher, and Virgil Zetterlind. 2020. Application of a New Shore-Based Vessel Traffic Monitoring System Within San Francisco Bay. *Frontiers in Marine Science* 7 (2 2020). https://doi.org/10.3389/fmars.2020.00086

[6] Shengnan Guo, Youfang Lin, Shijie Li, Zhaoming Chen, and Huaiyu Wan. 2019. Deep Spatial–Temporal 3D Convolutional Neural Networks for Traffic Data Forecasting. *IEEE Transactions on Intelligent Transportation Systems* 20, 10 (2019), 3913–3926. https://doi.org/10.1109/TITS.2019.2906365

[7] Carl Hewitt, Peter Bishop, and Richard Steiger. 1973. A universal modular actor formalism for artificial intelligence. In *Proceedings of the 3rd international joint conference on Artificial intelligence*. 235–245. https://doi.org/10.5555/1624775.1624804

[8] Lightbend Inc. 2022. Akka. http://akka.io/

[9] Ioannis Kontopoulos, Konstantinos Chatzikokolakis, Dimitris Zissis, Konstantinos Tserpes, and Giannis Spiliopoulos. 2020. Real-time maritime anomaly detection: detecting intentional AIS switch-off. *International Journal of Big Data Intelligence* 7, 2 (2020), 85–96. https://doi.org/10.1504/IJBDI.2020.107375

[10] Kpler. 2023. Unlock Maritime Insights with MarineTraffic. https://www.kpler.com/product/maritime

[11] Jay Kreps. 2011. Kafka : a Distributed Messaging System for Log Processing. https://api.semanticscholar.org/CorpusID:18534081

[12] Jae-Hoon Lee, Yoon-Seo Nam, Yonghwan Kim, Yuming Liu, Jaehak Lee, and Heesuk Yang. 2022. Real-time digital twin for ship operation in waves. *Ocean Engineering* 266 (2022), 112867. https://doi.org/10.1016/j.oceaneng.2022.112867

[13] Huanhuan Li, Hang Jiao, and Zaili Yang. 2023. Ship trajectory prediction based on machine learning and deep learning: A systematic review and methods analysis. *Engineering Applications of Artificial Intelligence* 126 (2023), 107062. https://doi.org/10.1016/j.engappai.2023.107062

[14] Redis Ltd. 2023. Redis. https://redis.io/

[15] ST Engineering Advanced Networks & Sensors Pte Ltd. 2021. Maritime Situational Awareness System.

[16] Zhihan Lv, Haibin Lv, and Mikael Fridenfalk. 2023. Digital Twins in the Marine Industry. *Electronics* 12 (2023). Issue 9. https://doi.org/10.3390/electronics12092025

[17] Petros Mandalis, Eva Chondrodima, Yannis Kontoulis, Nikos Pelekis, and Yannis Theodoridis. 2022. Machine Learning Models for Vessel Traffic Flow Forecasting: An Experimental Comparison. In *2022 23rd IEEE International Conference on Mobile Data Management (MDM)*. 431–436. https://doi.org/10.1109/MDM55031.2022.00094

[18] Kongsberg Maritime. 2023. Situational Awareness - Enhancing safety and efficiency.

[19] Saeed Mehri, Ali Asghar Alesheikh, and Anahid Basiri. 2021. A Contextual Hybrid Model for Vessel Movement Prediction. *IEEE Access* 9 (2021), 45600–45613. https://doi.org/10.1109/ACCESS.2021.3066463

[20] Duc-Duy Nguyen, Chan Le Van, and Muhammad Intizar Ali. 2018. Vessel Trajectory Prediction Using Sequence-to-Sequence Models over Spatial Grid. In *Proceedings of the 12th ACM International Conference on Distributed and Event-Based Systems (DEBS '18)*. Association for Computing Machinery, New York, NY, USA, 258–261. https://doi.org/10.1145/3210284.3219775

[21] International Maritime Organization. 1980. SOLAS Chapter V, Regulation 19.2. https://www.navcen.uscg.gov/sites/default/files/pdf/AIS/SOLAS.V.19.2.1-5.pdf [Accessed 16-10-2023].

[22] Myron Papadimitrakis, Marios Stogiannos, Haralambos Sarimveis, and Alex Alexandridis. 2021. Multi-Ship Control and Collision Avoidance Using MPC and RBF-Based Trajectory Predictions. *Sensors* 21, 21 (2021). https://doi.org/10.3390/s21216959

[23] Gesa Praetorius. 2014. Vessel Traffic Service (VTS) : a maritime information service or traffic control system? : understanding everyday performance and resilience in a socio-technical system under change. (2014). Issue 3729.

[24] Science, Technology Organisation Centre for Maritime Research, and Experimentation (STO CMRE). 2023. Maritime Security - Maritime Situational Awareness (MSA).

[25] Panagiotis Tampakis, Eva Chondrodima, Aggelos Pikrakis, Yannis Theodoridis, Kostis Pristouris, Harry Nakos, Eleni Petra, Theodore Dalamagas, Andreas Kandiros, Georgios Markakis, Irida Maina, and Stefanos Kavadas. 2020. Sea Area Monitoring and Analysis of Fishing Vessels Activity: The i4sea Big Data Platform. (2020), 275–280. https://doi.org/10.1109/MDM48529.2020.00063

[26] Uber Technologies. 2019. Uber H3: Uber's Hexagonal Hierarchical Spatial Index.

[27] Terma. 2023. Complete Situational Awareness.

[28] Alexandros Troupiotis-Kapeliaris, Nicolas Zygouras, Manolis Kaliorakis, Spiros Mouzakitis, Giannis Tsapelas, Alexander Artikis, Eva Chondrodima, Yannis Theodoridis, and Dimitris Zissis. 2022. Data Driven Digital Twins for the Maritime Domain. In *Technology and Science for the Ships of the Future*. IOS Press, 744–751. https://doi.org/10.3233/PMST220087

[29] A. C. van den Broek, Martin Neef, Patrick Hanckmann, Sicco Pier van Gosliga, and D. van Halsema. 2011. Improving maritime situational awareness by fusing sensor information and intelligence. *14th International Conference on Information Fusion* (2011), 1–8. https://api.semanticscholar.org/CorpusID:2371319

[30] Zhe Xiao, Xiuju Fu, Liye Zhang, and Rick Siow Mong Goh. 2020. Traffic Pattern Mining and Forecasting Technologies in Maritime Traffic Service Networks: A Comprehensive Survey. *IEEE Transactions on Intelligent Transportation Systems* 21, 5 (2020), 1796–1825. https://doi.org/10.1109/TITS.2019.2908191

[31] Xueyan Yin, Genze Wu, Jinze Wei, Yanming Shen, Heng Qi, and Baocai Yin. 2022. Deep Learning on Traffic Prediction: Methods, Analysis, and Future Directions. *IEEE Transactions on Intelligent Transportation Systems* 23, 6 (2022), 4927–4943. https://doi.org/10.1109/TITS.2021.3054840

[32] Dimitris Zissis, Konstantinos Chatzikokolakis, Giannis Spiliopoulos, and Marios Vodas. 2020. A Distributed Spatial Method for Modeling Maritime Routes. *IEEE Access* 8 (2020), 47556–47568. https://doi.org/10.1109/ACCESS.2020.2979612

[33] Dimitrios Zissis, Elias K Xidias, and Dimitrios Lekkas. 2016. Real-time vessel behavior prediction. *Evolving Systems* 7 (2016), 29–40. Issue 1. https://doi.org/10.1007/s12530-015-9133-5

[34] Nikolas Zygouras, Giannis Spiliopoulos, and Dimitris Zissis. 2021. Detecting representative trajectories from global AIS datasets. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. 2278–2285. https://doi.org/10.1109/ITSC48978.2021.9564657

[35] Nikolas Zygouras, Alexandros Troupiotis-Kapeliaris, and Dimitris Zissis. 2024. EnvClus*: Extracting Common Pathways for Effective Vessel Trajectory Forecasting. *IEEE Access* 12 (2024), 3860–3873. https://doi.org/10.1109/ACCESS.2023.3349081