# KNOWLEDGE DISCOVERY FROM MARITIME MOVING OBJECTS - APPLICATION TO AEGEAN SEA –

**Cyril Ray, Naval Academy Research Lab, December 2010**

## 1. Introduction

Maritime environment has a huge impact on the world economy and our everyday lives. Beyond, being a space where numerous marine species live, the sea is also a place where human activities evolve: sailing, cruising, fishing, goods transportation... Nowadays, the vast majority of world goods is carried by sea by huge container-carriers. The disasters and damages caused in the event of major sea collisions can pose serious threats to the environment and human lives. Such disasters and damages also often lead to highly negative effects on maritime ecosystems and are threats not only for the important populations of marine protected and endangered species, but also for economic, scientific, and cultural sectors.

The lack of efficient mechanisms able to manage, monitor and regulate ship traffic conditions is exacerbated by the fact that collaborations and actions between countries to deal with potential maritime accidents, piracy which could result in an ecological or socio-economical disaster are still limited [1]. In Europe, safety and security has therefore become a major concern of maritime navigation, especially when considering the increasing growth of maritime traffic, and constant decrease of crews on decks. In most of the densely trafficked, enclosed or semi-enclosed seas of the world, measures are being applied to better regulate and monitor boat traffic, so as to reduce the possibilities of maritime accidents and ensure pollution prevention from boats. Such measures include ship traffic lanes and separation schemes, routing and monitoring systems, reporting systems, decision-aid systems, and the definition of exclusion areas and *Particularly Sensitive Sea Areas* (PSSA).

This objective stresses the need for the development of maritime navigation systems whose objective will be to contribute to a safer sea by efficient monitoring and analysis of mobilities at sea. Advances in the last decade in telecommunication, positioning systems, distributed architectures favoured emergence of platforms and European programs (e.g. MARNIS, INTERRISK) related to maritime information systems. In order to facilitate navigation decisions and warn about possible collisions, Automatic Identification Systems (AIS) have been progressively implemented on commercial ships. This system, whose objective is to identify and locate vessels at distance, automatically broadcasts location-based information through wireless communications. The International Maritime Organization (IMO) has made the AIS a mandatory standard for the Safety of Life at Sea (SOLAS).

The Automatic Identification System usually integrates a transceiver system, a GPS receiver and other navigational sensors on board such as a gyrocompass and a rate of turn indicator. An AIS transponder runs in an autonomous and continuous mode, and broadcasts a position report regularly according to the ship behaviour. The information is broadcasted within a range of 35 nautical miles to surrounding ships and maritime Vessel Traffic Systems (i.e. maritime authorities) on the ground. AIS messages include data on the ship characteristics like position, route, speed, and estimated arrival time to a given place.

Such location-based data include partial information where trajectories are difficult to identify and classify, as well as detailed data where contextual information helps to understand ships behaviours

[1] A. Miliou, M. Demetriou, C. Caprio, T. Tsimpidis, S. Barnicoat, Boat traffic characteristics in three passages of the Aegean Sea: evidence, risk of maritime accidents, strategy for protection, 2010

according to space, time, destination and ships' types. Both are based on geographic positions and require error-detection, filtering process and trajectory reconstruction.

**2. Research issues and context**

One of the main interesting challenges still opened for the successful development of maritime information systems relies in the management and analysis of the large and increasing amount maritime spatio-temporal datasets coming from different sources (automatic identification system, databases, and proprietary systems). These typical moving objects provide research opportunities where to apply and evolve existing algorithms and techniques for knowledge discovery on movement behaviours at sea. Research issues on such information still include:

- Integration of heterogeneous databases: sensors, normalisation (hardware, exchange languages...), development of federated and distributed spatio-temporal databases, pre-processing that includes errors, filtering and reconstruction algorithms, semantic issues.
- Patterns discovery and analysis: modelling of moving objects and trajectories, trajectory querying and simplification, automatic annotation, similarity functions, classification and clustering algorithms, knowledge discovery (trends, unusual behaviours, and event detection).
- Physical-based analysis: correlation between environmental data (currents, tides, winds) and moving objects: optimal routes, exclusion areas (pollutions...), port management...
- Simulation and decision-aid system: interpolation and extrapolation of data for simulation and replay of trajectories, 2D and 3D geo-visualisation, tracking and context-aware monitoring systems, cross-based analysis taking into account the human factor, maritime rules and transportation features, training-based simulation systems.

The work did in this short term scientific mission is part of an innovative and inter-disciplinary network initiated within MOVE network in September 2010. The objective of this network is to explore some of these issues for knowledge discovery from maritime data. This network is inter-disciplinary as it associates MOVE research labs with complementary activities and expertise on the four aforementioned research issues but also associates final end-users that have expressed concrete needs in terms of knowledge discovery and trajectory understanding at sea: French maritime control centers and marine officers onboard, Netherland coast guards and Greek institute for marine conservation (Archipelagos). This research is based on a very large datasets of locations coming from the Automatic Identification System and provided by IREnav (France); Infolab (Greece); MARIN (Netherlands) [2]. The objective of this STM is to the study

The research relies, by the end of 2010, on three main events in which I collaborate:

- Short term scientific mission by Fabio Pinelli (KDD Lab) at Naval Academy Research Lab (September, 2010). The work focused on temporal distribution of movements, application of a clustering algorithm using different distance measure in order to identify similar behaviours in space and in time and, temporal analysis of emerging clusters (French AIS data).

- MARIN Showcase (October, 2010). The workshop focused on pattern identification and geo-visualization of maritime traffic (Netherland AIS data). Attendees were Fraunhofer Institute, IAIS; Information Systems Department, University of Minho; Faculty of Mathematics and Computer Science, University of Hagen; KDD Lab, University of Pisa; Naval Academy Research Lab; University of Venice; Technical University Eindhoven; MARIN.

[2] Describing respectively traffic from Brest bay (3 years, 6GB) - Aegean sea (3 years, 3 TB) – North sea (2 weeks, 310 MB)

- This short term scientific mission (Cyril Ray, Naval Academy Research Lab) at Infolab (December, 2010). The work focused on AIS parsing, database modelling, data analysis for Archipelagos and trajectory modelling (Greek AIS data).

Amongst many scientific, technical and methodological issues to address by MOVE research labs, there is still a need for spatio-temporal data mining techniques that will favour, from raw data, added-value information, understanding and suitable results for these aforementioned “clients”. This has motivated a pragmatic approach, for this STSM, that try first to answer to their needs focusing on quantitative analysis, event discovery and trajectory understanding.

**3. Objectives, resources and results**

The objective of this STSM is partly based on some requirements and queries set by Archipelagos, to investigate factors which contribute most to the risk of a shipping accident and identification of dangerous areas. This has been done by traffic database processing through semi-quantitative risk analysis of shipping traffic in the Aegean Sea. Typical questions expressed are:

- For each ship and for each day
  - Extract / draw the ship tracks (detailed and simplified, e.g. using Douglas-Peucker)
  - Calculate the number of sharp changes in direction (e.g. over 10 deg.) Calculate average and minimum distance from shore; where and when
  - Calculate the maximum number of ships in the vicinity of the ship (e.g. 10 nautical mile radius)
  - Find whether (and how many times) the ship go through narrow passages or biodiversity boxes

- For the full population of ships
  - Find typical routes vs. Outliers (full or small part of its track)
  - Recognize template movement (according to IMO specifications), e.g. drifting

The available elements to achieve this task are:

- Shapefile of the Greek mainland and islands
- Narrow passages and biodiversity boxes
- Database systems : MySQL or PostgreSQL with PostGIS
- Moving object database (MOD) engine : Hermes
- LOCOSS platform and a first AIS data model
- Data with full activity of ships in eastern Mediterranean for a period of 3 years

The original AIS data were provided by IMIS Hellas (http://www.imishellas.gr), which owns a global network of AIS receivers. The data are stored in MySQL MyISAM tables. Existing database have several drawbacks: Large number of files, unstructured MySQL database, redundancy, modified frames that do not belong to AIS specification anymore, inconsistent data, and no indexes. As a consequence, any processing of the data directly from these MySQL tables is possible without further processing at the application level. The solution identified, and the first step of this STSM, was to parse the original MySQL database in order to extract original AIS frames within text files that can be parsed again and integrated in a structured spatial database.

The second step of the work focused on AIS specification and the production of an adapted data model. Automatic identification system provides several meaningful information that can be considered in a purpose of movement discovery and understanding. The data model proposed relies on the following classification of AIS messages: static, dynamic and route-based. The database has been implemented using the open-source object-relational DBMS PostgreSQL and PostGIS that is a spatial database extension for PostgreSQL database. The AIS tables contain roughly the following information:

- **Static**: MMSI number, IMO code, call sign, ship name and type, dimension
- **Dynamic**: Position (Longitude, Latitude in WGS84), time, speed, heading, course over ground (COG), rate of turn (ROT), navigational status
- **Trajectory-based**: Destination, estimated time of arrival (ETA), draught

A benefit of this methodology has been to provide a structured database with spatial types, functions and indexes. It also reduces the size of original database by 3 without indexes (by 2 with indexes: 1,8 GB for a each IMIS Hellas file → 950 MB).

The third step of the work relied on a generic and scalable information system that have been designed to integrate, store, model, analyze and visualize spatially related data in time. The LOCOSS framework aims to provide an integrated system for the real-time monitoring and tracking of different types of mobile objects. This framework has been already experimented in the context of maritime navigation in Brest coastal area. The platform, developed so far, has been designed with four tiers client-server architecture and organized through a distributed data and processing model. This information system is a distributed platform based on three functions: (1) integration of positioning information, (2) data management and processing and (3) web-based visualisation. The platform is a computing system that includes a Java-based platform for mobile objects and PostGIS spatial database for data manipulation and storage.

A large part of this STSM has been devoted to the update, bug tracking and improvements with new facilities and parsing capabilities of this LOCOSS platform. The result is a more robust software able to parse all AIS frames and to insert them in a spatial database after error-checking and point-based distance filtering.

In addition a new web-based KML generator has been designed for fast and efficient visualisation of database content. This application generates on-demand KML files that can display data in any compatible software (ie. Almost all GIS and viewers like Google Earth).

Finally, the last part of the STSM has been devoted to queries and answers for Archipelagos. A technical report (currently 28 pages) explains the methodology applied to extract, parse data and fill the spatial database with AIS information and geographic features. It shows how to exploit PostGIS database, types and functions to queries the data and obtain answer to Archipelagos point-based questions. It shows the limits of point-based querying and proposes improvements using trajectory-based analysis to identify spatial relationships.

In conclusion, this joint inter-disciplinary STSM has been already the opportunity of fruitful exchanges on maritime data mining. This is still an ongoing research we would like to formalise with several papers. Current results include: a data model, a parsing software, a web-based visualisation application, a methodology for the integration and querying to Aegean Sea AIS data. Next steps concern a new clustering algorithm, the connexion of the database with Hermes, route detection and advanced pattern detection.