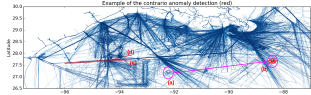


NEXT STEP FOR BIG DATA INFRASTRUCTURE AND ANALYTICS FOR THE SURVEILLANCE OF THE MARITIME TRAFFIC FROM AIS & SENTINEL SATELLITE DATA STREAMS



ABSTRACT

The surveillance of the maritime traffic is a major issue for security and monitoring issues. Spaceborne technologies, especially satellite AIS ship tracking and high-resolution imaging, open new avenues to address these issues. Current operational systems cannot fully benefit from the available and upcoming multi-source data streams. In this context, SESAME initiative aims to develop new big-data-oriented approaches to deliver novel solutions for the management, analysis and visualisation of multi-source satellite data streams going beyond the current CLS implementation. Targeted at the automatic generation and documenting of early warnings, our key originality lies in a big-data approach to jointly address these challenges based on the complementarity of the scientific and operational expertise gathered in the consortium: big-data platforms, mining strategies for time series and trajectory data, Sat-AIS signal analysis, high-resolution satellite imaging.

Keywords— Sentinel, high-resolution satellite imaging, AIS maritime traffic surveillance, big data, data mining, behaviour analysis, ship detection.

Authors

R. Fablet, L. Drumetz, N. Bellec, R. Garello
IMT Atlantique, Lab-STICC, Brest, France

L. Chapel, C. Friguet, P. Gloaguen, S. Lefèvre, F. Merciol, R. Tavenard
Univ. de Bretagne-Sud/Univ. de Rennes 2, IRISA-OBELIX, Vannes, France

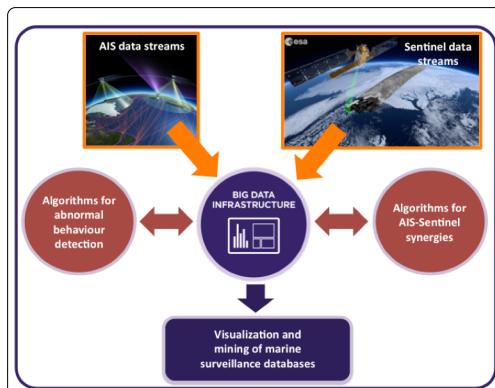
P. Morillon, C. Morin, M. Simonin, C. Tedeschi
INRIA/Univ. de Rennes 1, IRISA-MYRIADS, Rennes, France

G. Hajduch, R. Vadaine
CLS, Espace et Segments Sol, Brest, France

Acknowledgments



SESAME Workflow

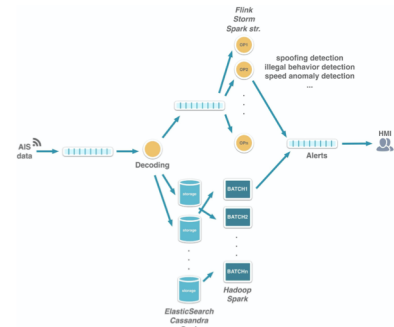


SESAME embeds the implementation of the proposed solutions for dual case-studies for the automatic generation and documenting of early warnings:

- the real-time analysis of AIS-Sentinel data streams
 - the re-analysis of large-scale AIS-Sentinel datasets.
- These case-studies comprise both the evaluation of algorithms and models as well as big-data-oriented infrastructures and frameworks. Grid'5000 platform [4] will provide an initial flexible and scalable testbed, whereas CLS big data infrastructure will be the targeted platform to validate how the proposed solutions scale-up to realistic large-scale datasets.

Targeted SESAME Big Data infrastructure

The envisioned pipeline of computation and storage for the AIS data will exploit and combine state-of-the-art big-data-oriented framework such as for instance Cassandra, Hadoop, Spark.



The considered pipeline involves two main parts:

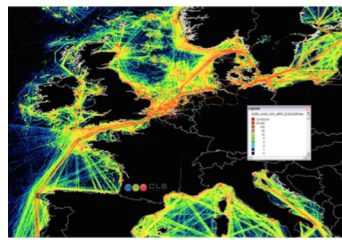
- the real-time stage where data streams are analyzed in near real-time;
- the batch stage where larger history of data can be processed at once.

Outputs of these two stages can either be stored persistently, pushed to other message queues for future reuse or generate alerts passed to analysts for further analysis. The different processing components in this architecture rely on a message brokering system taking care of messages buffering and distribution between the various actors. This architecture is extensible in the sense that new components (e.g corresponding to new types of alerts) can be added as well as new storage backends (e.g for indexing purpose).

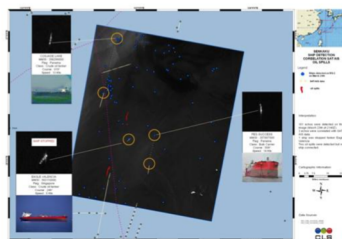
Case-studies for the Synergy between AIS & Sentinel data streams

References and prior work

- [1] Maritime Knowledge Discovery and Anomaly Detection Workshop, Ispra, Italy, July 2016, European Commission, pp. 24–27 – ISBN 978–92–79–61301–2.
- [2] R. Pelich et al., "Performance evaluation of sentinel-1 data in SAR ship detection," in *IEEE IGARSS*, 2015.
- [3] European Maritime Safety Agency, "Clean sea net service," <https://portal.emsa.europa.eu/web/csn>.
- [4] F. Cappello et al., "Grid'5000: a large scale and highly reconfigurable grid experimental testbed," in *IJHPCA*, 20(4): 481–494, 2006.
- [5] V.D. Nguyen, R. Vadaine, G. Hajduch, R. Garello, R. Fablet. A multi-task deep learning model for vessel monitoring using AIS streams. Big Data Maritime Workshop, NATO STO CMRE, La Spezia, May 2018.
- [6] P. Gloaguen, L. Chapel, C. Friguet, R.Tavenard. Scalable clustering of trajectories within a continuous time framework. Submitted



AIS and Sat-AIS data
(Millions of GPS positions daily)



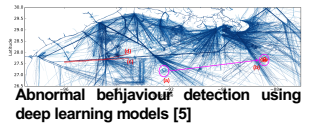
Sentinel-1 SAR images
(4 TB daily)

Preliminary big-data-oriented analysis of AIS data streams

Our initial investigations focused on the characterization of the AIS data streams from a 1-month AIS dataset on a global scale. We started investigating the unsortedness in AIS message arrivals for simulation and validation purposes. We compared the order in which AIS messages has been stored in the storage system versus their sending timestamp. The initial platform version was based on a Spark cluster¹ for the batch processing system and Kafka for the message brokering and real-time processing system. The platform deployment is fully automated and run on the Grid'5000 platform [4].

Preliminary results on abnormal behaviour detection

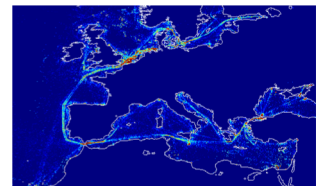
We develop statistical models [5,6] for the analysis of AIS data streams with a view to model vessel trajectory behaviours. These models trained using non-supervised schemes can then be exploited for multiple tasks including vessel track interpolation, abnormal behaviour detection....



Abnormal behaviour detection using deep learning models [5]

Preliminary analysis of AIS-Sentinel synergies for vessel detection

Preliminary results highlight the relevance of the synergy between SAR and AIS information in terms of geographic coverage and of detection and characterisation of abnormal activities at sea. We also demonstrated the feasibility of the construction of large-scale datasets of SAR echoes acquired in various configurations corresponding to a subset of known vessels with a view to apply machine-learning-based detection and classification strategies. Considering a four-month dataset of Sentinel-1A satellite data over Europe from March to June 2017, we collected 5,414 SAR images.



Mapping of the cumulated number of AIS messages matched with SAR echoes from March to June 2017.

This preliminary analysis pinpoints the potential added-value of the synergy between AIS and Sentinel data for: i) the automated documentation of abnormal behaviour (e.g., transshipment activities), ii) the automated creation of groundtruthed SAR datasets for vessel detection and recognition.