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

R. Fablet[1], *N. Bellec*[1], *L. Chapel*[3], *C. Friguet*[3], *R. Garello*[1], *P. Gloaguen*[3], *G. Hajduch*[4], *S. Lefèvre*[3] *F. Merciol*[3], *P. Morillon*[2], *C. Morin*[5], *M. Simonin*[5], *R. Tavenard*[6], *C. Tedeschi*[2], *R. Vadaine*[4]

[1] IMT Atlantique, Lab-STICC, UBL [2] Univ. de Rennes 1, IRISA-MYRIADS [3] Univ. de Bretagne-Sud, IRISA-OBELIX
[4] CLS, Espace et Segments Sol, Brest[5] Inria, IRISA-MYRIADS[6] Univ. de Rennes 2, IRISA-OBELIX

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 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.

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

1. CONTEXT AND CHALLENGES

The surveillance of the maritime traffic is a major issue for maritime security (e.g., traffic monitoring, border surveillance, ...) as well as environmental monitoring (oil spill monitoring, ...) and marine resource management (illegal fishing monitoring, ...). Spaceborne technologies, especially satellite ship tracking from AIS messages (Automatic Identification System) and high-resolution imaging of sea surface, open new avenues to address such monitoring and surveillance objectives. Currently, operational systems cannot fully process these complete streams of satellite-derived data. For instance, the French institutional users evaluate that less than 20% of the overall AIS data (about a few tens of millions of AIS messages daily) are actually analysed for abnormal behaviour detection. Besides, the free access to Sentinel Earth Observation data streams (high-resolution Sentinel-1 SAR and Sentinel-2 optical imaging, up to a few TB daily [1]) offers novel opportunities for the analysis and detection of ship behaviours, including AIS-Sentinel data synergies.

In this context, SESAME initiative aims to develop, implement and evaluate new big-data-oriented approaches to deliver novel solutions for the management, analysis and visualisation of multi-source satellite data streams. Targeted at the automatic generation and documenting of early warnings (both in real-time and re-analysis modes), the key scientific and technological challenges cover the development of hardware and software platforms adapted to the characteristics of the data streams of interest along with the design of novel models and algorithms for AIS-Sentinel synergies and the automatic detection of abnormal behaviours. Besides the development of CLS big data infrastructure [2], the originality of the project lies in a big-data approach to jointly address these challenges based on the complementarity of the scientific expertise and knowledge of the operational needs gathered in the consortium: big-data platforms, mining strategies for time series, modeling and analysis of tracking data, Sat-AIS signal analysis, high-resolution satellite imaging.

2. RELATED WORK

Recent works highlighted the importance of new approaches for knowledge discovery and anomaly detection in maritime surveillance [3, 4, 5]. The use of AIS data becomes more and more important to improve traffic monitoring. Practical methods were developed to detect data falsification or spoofing [6] or to detect abnormal behaviours in trajectories[8, 7], both to detect dangerous movement patterns or potentially illegal behaviours. Yet, the combination of such rule-based and statistical models to big-data infrastructure and frameworks largely remains to be investigated. Besides, the emergence of deep learning techniques [9, 10] is particularly appealing to further investigate large-scale AIS data streams.

Ship detection from satellite imagery has long been an active research topic (e.g. [11]). Synergies between AIS and Sentinel data streams appear promising [5, 12] to improve

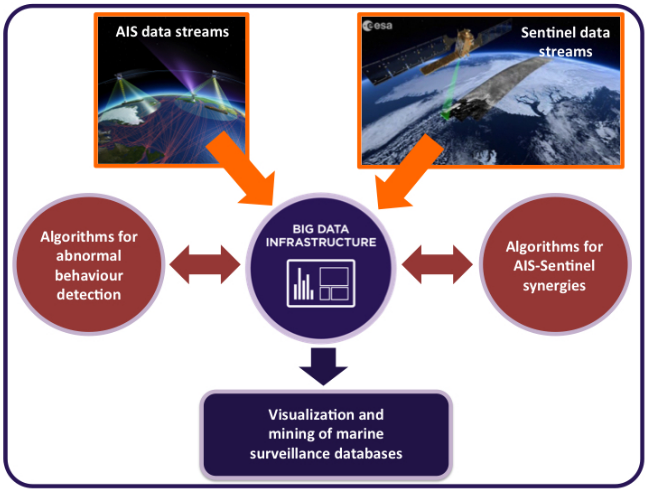

Fig. 1. **SESAME workflow for big-data-oriented maritime surveillance from multi-source AIS and Sentinel data streams.**

surveillance performance and double check the cooperative AIS data with non-cooperative sensors. We propose to unify these two approaches within a big data framework.

3. PROPOSED WORKFLOW

As sketched in Fig.1, the proposed workflow relies on the implementation and evaluation of a big-data-oriented framework for the management, mining and visualisation of the considered multi-source data streams. It combines the development of hardware and software big data platforms with novel models and algorithms for the detection of abnormal behaviours and AIS-Sentinel synergies. 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 and the re-analysis of large-scale AIS-Sentinel datasets. They will comprise both the evaluation of algorithms and models as well as big-data-oriented infrastructures and frameworks. Grid'5000 platform[14] 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.

The algorithms and models will be evaluated on two case studies representative of the challenges of global maritime surveillance. The first case study will focus on the monitoring of IUU (Illegal, Unreported, and Unregulated) fishing activities occuring inside the economic exclusive zones (EEZ). A more general approach will be taken for the second case study where the maritime traffic and illegal activities will be analyzed on a global scale.

4. PRELIMINARY RESULTS

4.1. Big data framework for AIS data streams

CLS operational rule-based architecture for the mining of AIS data streams will provide a baseline architecture for the development and evaluation of big-data-oriented infrastructure. Preliminary analyses point out the requirement for the exploitation of big-data-oriented frameworks to scale up to large-scale AIS datasets. Grid'5000 is used as a pre-production testbed to evaluate the framework.

Fig. 2 depicts the envisioned pipeline of computation and storage for the AIS data. It is split into two parts: first, the real-time stage where data streams are analyzed in near real-time, and second, 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 can be alerts passed to a human analyst 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).

Our initial investigations focused on the characterization of the AIS data stream. In particular, based on one month of AIS data on a global scale, we started investigating the unsortedness in AIS message arrivals. We compared the order in which AIS messages has been stored in the storage system versus its sending timestamp. This will in turn help us replay the global AIS stream realistically for simulation and validation purposes. The initial platform version was based on a Spark cluster[1] for the batch processing system and Kafka[2] for the message brokering and real-time processing system. The platform deployment is fully automated and run on the Grid'5000 platform [14].

4.2. Synergy between Spaceborne SAR data and AIS data

Regarding the joint usage of spaceborne SAR data and AIS information, CLS has demonstrated its relevance for the characterisation of SAR vessel detection performances using interpolated/extrapolated AIS tracks as ground truth [11]. It is operationally used for the monitoring of oil spills at sea and the identification of potential polluter sources [5] in the framework of the European Maritime Safety Agency (EMSA) CleanSeaNet Service [13].

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

[1] urlhttps://spark.apache.org/
[2] `https://kafka.apache.org`

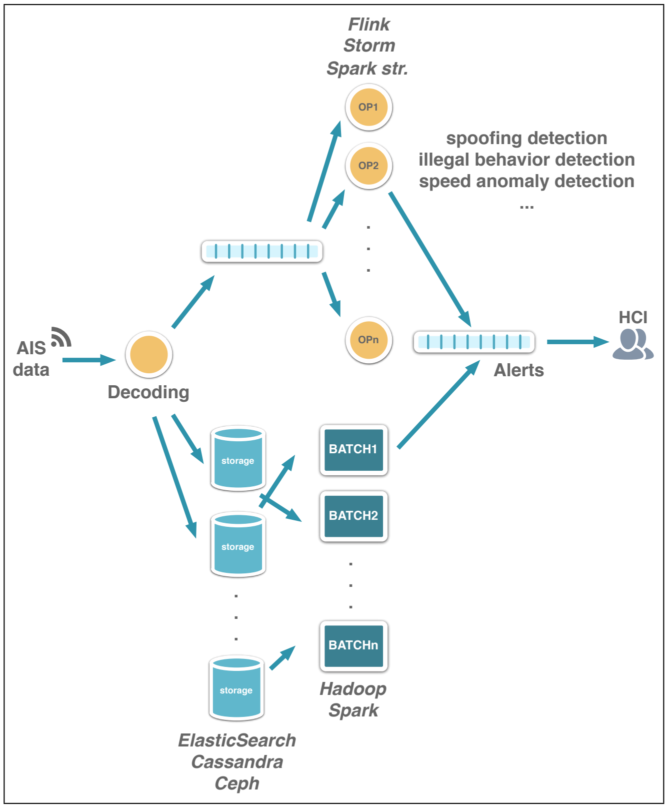

Fig. 2. **Targeted SESAME logical infrastructure:** it aims to generate alerts for abnormal behaviours from the AIS data stream. As illustrated, the infrastructure will exploit and combine state-of-the-art big-data-oriented framework such as for instance Cassandra, Hadoop, Spark, Fink,...

the construction of large-scale datasets of SAR echoes acquired in various configuration corresponding to a subset of known vessels with a view to applying machine-learning-based detection and classification strategies. Considering a four-month dataset of Sentinel-1 A satellite data over Europe from March to June 2017, we collected 5414 SAR images. They were systematically processed using CLS vessel detection algorithm. The detected SAR echoes were then matched with interpolated/extrapolated AIS data. For illustration purposes, we depict in Fig. 3 the spatial mapping of the cumulated number of SAR-AIS matches over the considered 4-month period. These results highlight shipping routes and areas where the coverage of the AIS network is significantly lower than in densely-monitored areas such as in the English channel. The Bay of Biscay as well the northern coast of Tunisia are typical examples of areas with a high-traffic but a relatively low coverage in terms of coastal AIS receiver networks. In such areas, the main source of AIS data is issued from satellite AIS data, but it cannot provide a space-time coverage similar to dense coastal AIS networks. Our future work will further explore the potential of AIS-SAR synergies

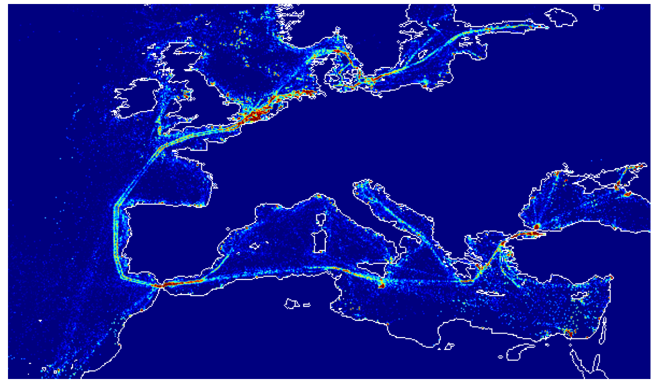

Fig. 3. **Maps of the cumulated number of AIS messages matched with SAR echoes from March to June 2017:** the redder the color, the higher the density of AIS-SAR matches. For this analysis, we processed 5414 Sentinel-1 A satellite images using CLS operational vessel detection system.

for the creation of groundtruthed SAR image datasets for the developement of novel learning-based ship vessel detection and identification strategies.

5. CONCLUSION

The expected impacts of the project include both dissemination actions to the scientific community, including a maritime surveillance benchmark suite, and technological transfers to CLS with respect to future national and international calls on operational systems and services for maritime traffic surveillance and high-resolution environment monitoring.

6. ACKNOWLEDGEMENTS

The authors acknowledge the support of DGA and ANR under reference ANR-16-ASTR-0026 (SESAME initiative), the labex Cominlabs, the GIS BRETEL (CPER/FEDER framework) the Booster MoreSpace and CNES for access to the PEPS cluster.

7. REFERENCES

[1] A.G. Castriotta, "Sentinel data access report 2016.," Tech. Rep.

[2] E. Lambert et al, "Hadoop platform for georeferenced mobiles and gridded data," in *Submitted to Big Data From Space 2017*, 2017.

[3] FP7 DOLPHIN Consortium, "FP7 DOLPHIN Project," `http://gmes-dolphin.eu/`,

[4] "Maritime Knowledge Discovery and Anomaly Detection Workshop," Ispra, Italy, July 2016, European Commission, pp. 24–27 – ISBN 978–92–79–61301–2.

[5] N. Longépé et al, "Polluter identification with spaceborne radar imagery, ais and forward drift modeling," *Marine Pollution Bulletin*, vol. 101, no. Issue 2, pp. 826–833, 2015.

[6] C. Ray et al, "Methodology for Real-Time Detection of AIS Falsification," in *Maritime Knowledge Discovery and Anomaly Detection Workshop*, Michele Vespe and Fabio Mazzarella, Eds., Ispra, Italy, July 2016, pp. 74–77.

[7] G. Pallotta et al, "Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction," *Entropy*, 15(6):2218–2245, 2013.

[8] W. Hu, et al, "A system for learning statistical motion patterns," *IEEE PAMI*, 28(9):1450–1464, 2006.

[9] Y. Le Cun, et al, "Deep learning," *Nature*, 521(7553):436-444, 2015.

[10] X. Jiang et al, "Trajectorynet: An embedded gps trajectory representation for point-based classification using recurrent neural networks," *arXiv preprint arXiv:1705.02636*, 2017.

[11] R. Pelich et al, "Performance evaluation of sentinel-1 data in SAR ship detection," in *IEEE IGARSS*, 2015.

[12] F. Mazzarella et al, "Sar ship detection and self-reporting data fusion based on traffic knowledge," *IEEE GRSL*, 12(8):1685–1689, 2015.

[13] European Maritime Safety Agency, "Clean sea net service," `https://portal.emsa.europa.eu/web/csn`.

[14] F. Cappello et al, "Grid'5000: a large scale and highly reconfigurable grid experimental testbed," in *IJHPCA*, 20(4): 481-494, 2006.