Exploiting the Potential of the Future "Maritime Big Data"

Bernard Garnier, Aldo Napoli

EXPLOITING THE POTENTIAL OF THE FUTURE "MARITIME BIG DATA"

Bernard GARNIER, Aldo NAPOLI***

*BlueSolutions Consulting SAS, Sophia Antipolis, France
**MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France

ABSTRACT

Today, most of the operational "abnormal behaviour" detection algorithms primarily operate kinematic rules on the vessel tracks provided by the AIS [1], [2], [3]. To be more effective, they must be associated to additional "context data", however insuring this data access is today a challenge, due to the disparity of registers and the fragmentation of actors… For example, detecting suspicious "associations of ships" requires already a complex data mining to detect indirect common ownership through the myriad of cascaded legal entities used for formal ownership and registration; as another example, maritime security actors know the importance of environmental factors (sea state, fog, clouds, moon etc) when assessing the risk of illegal passages or piracy attacks at night – but gathering the right local weather forecast data in association with abnormal behaviour detection algorithms for maritime radar systems is also a challenge.

While many more data and metadata should be browsed from all the existing maritime reporting systems and data repositories to cross-check systematically the declared versus actual behaviour of commercial, fishing and leisure ships, we shall start developing a new data access thinking to benefit from the progressive deployment of the EU-wide CISE which is approaching its pre-operational validation milestone and should be largely mature by 2020, materializing a break-through in terms of data access for the maritime security communities. In parallel, another key enabler is the capacity to collect the "local picture" gathered by genuinely cooperating shipping (sightings and nav radar) with the "big picture" (VMS, AIS, S-AIS, LRIT, satellite imaging…). The VDES will provide a very effective data uplink as an alternative to broadband maritime SatComs at the same 2020 horizon. Other planned technological gap-fillers deserve to be integrated in future data processing strategies: new space projects aim at solving the data synchronicity challenge by co-locating SAR, S-AIS and VDES payloads for specialized maritime surveillance constellations; EDRS allows downloading LEO maritime surveillance data streams in near real-time from anywhere on Earth; smart and fast embarked data processing will downsize the "rising tide" of data, extracting and tagging straight away the mere fraction of data requiring prompt human attention…

This paper will aim at delivering a sort of "wake-up call" to integrate this new data access paradigm in the current research on maritime knowledge discovery associated to the detection of safety and security threats: 2020 is tomorrow, we shall think, develop and test our toolbox at the whole scale of this "Big Data".

Index Terms— Maritime surveillance, early detection, heterogeneous correlation, CISE, weak signals analysis

1. CURRENT STATE OF PLAY

While maritime surveillance has been radically transformed by the introduction of the Automatic Identification System (AIS) in 2002 through the IMO SOLAS Agreement – suddenly populating the screen of the vessel traffic management systems (VTMS) well beyond the range of the coastal radars, the operational maritime surveillance capabilities have not much evolved since. Voluntary reporting systems (mainly AIS, VMS for fishery vessels and LRIT in distant sea lanes) remain the essential source of vessel monitoring, leaving in the shade the smaller boats… and the deliberately cheating ones. Furthermore, non-cooperative ship detection provided by maritime radars (on board ships and on the coast) is now automatically fused with AIS, no more supported by additional VHF voice contact and binoculars to confirm the vessel identity and its planned route.

The recent development of commercial satellite payloads designed to collect AIS signals from ships well beyond coastal VHF horizon (S-AIS) is a welcomed "plus" to overcome the deficient cover of coastal AIS receivers in some regions, but remains overall a moderate contribution to the VTMS operation as S-AIS largely recoups LRIT.

International cooperation has become routine, but again most if not all maritime traffic data exchange agreements relate to the data of these voluntary reporting systems; as they are often faulty and sometimes cheated, this means building the common operational maritime traffic picture on sand!

Space observation systems have been promoted as a new way to ascertain the maritime traffic data. Indeed, medium and high resolution synthetic aperture radar (SAR) on low Earth orbit (LEO) satellites provide a theoretical

capability of ship detection anywhere in high seas, but its use as a daily maritime traffic monitoring tool remains problematic (no persistence, limited refresh, significant cost, latency of several hours). It is technically impossible to acquire synchronously SAR radar and S-AIS, rendering the correlation of these data tedious and often uncertain. As a consequence, its operational use is limited to specific fishing grounds and trafficking areas (where and what to look are known ahead of data collection and processing).

With these limitations, the blue borders remain largely porous to all sorts of trafficking (drugs, arms, migrants…) while endangered fishing species are still poached at high scale (IUU fishing remains as high as 25 to 30% of all catches).

In the field of Defence, Intelligence Agencies are actively seeking for illegal arms trade, intrusions in territorial waters and security threats of all sorts; however cross-border information exchanges (e.g. under the auspices of NATO) are most often limited to share a "list of usual suspects" designating about 2000 vessels of specific interest to be jointly monitored. In parallel, anti-drug operations are also most of the time driven by human intelligence (HUMINT), leaving 90% of the traffic undetected. In short, the challenge relate to the large number of "unknown unknown" threats.

To overcome the incapacity of the current state-of-play to anticipate the creativity and flexibility of criminal schemes, there is a clear need for changing the maritime surveillance paradigm: monitoring ship tracks on big screens is not enough!

2. LOOKING FOR WEAK SIGNALS

Time is long gone where seas appeared as the ultimate area of freedom, where the Master was invested of every power… after God. All maritime activities are today explicitly regulated, even in High Seas, as a result of international treaties and agreements (IMO…), regional agreements (Baltic…), national and local regulations. The respective authority of Flag States, Port States, Coastal States etc. is internationally agreed and results into massive data collection from all operators: every ship voyage generates dozens of massive files on the ship itself, the voyage, the cargo and the people on board.

Each of these files are directed to a particular Maritime Authority which screens the documents, clears the corresponding ship operation and triggers possible controls.

Criminal gangs are thus used to provide "clean" documents to avoid controls, e.g. cargo or fish catch declarations that will look "business as usual" for the custom officer or fishery inspector respectively. In the same time, these data are not today available in parallel to the serious crime investigators that could detect inconsistencies or possible correlations with their own investigations.

This "fragmentation" of the State controls of maritime activities is inherited from the absence of an holistic vision of the maritime economy common to almost all States: a comprehensive survey undertaken by DG Mare confirmed a split of the maritime authority prerogatives between more than 10 different administrations all across the 21 EU maritime nations, with many different Ministries involved (Transport, Energy, Environment, Agriculture, Interior, Economy, Defence…). Furthermore, this organizational mix differs significantly form a country to the next, with hardly a cross-border match between mandates and legal prerogatives, hence making inter-administration cooperation furthermore complex. This results into what is usually called "data silos", each of them only exploited under a single angle.

In essence, maritime surveillance operations are not distinct from any business, and the general approach of "Strategic Early Warning Systems" theorized from 1975 ref [4, 5] to provide on-time strategic reaction capabilities is perfectly applicable. The central element is that disruptions do not emerge without warning, however these warnings remain most often undetected as they don't come from the expected channels of business information. These warning signs are described as "weak signals" [4], a concept aimed at early detection of those signals which could lead to strategic surprises -- events which have the potential to jeopardise an organization's strategy. Brison and Wybo [6] represent the life cycle of weak signals as four successive steps associated with barriers that the weak signal has to overcome. These four steps are: Detection, Interpretation, Transmission and Priority setting. The extraction of such weak signals from the massive data and meta-data collected on Internet by the GAFA turns to be potentially extremely profitable – currently turning as the 21th century gold rush… Everyone has experienced already how effectively e-advertising can be targeted by processing the heterogeneous navigation data and metadata of your internet browser.

There are already demonstrative contributions of Open Source data mining and weak signals analysis in the maritime security domain, such as the ConTraffic web-service of the JRC able to alert Custom Authorities on particular containers associated with "abnormal" voyage histories (https://contraffic.jrc.ec.europa.eu/).

The largest potential relate to the application of weak signals detection over widely heterogeneous data possibly correlated. Ref [7] proposes an interesting application of this approach to the detection of cyber-intrusions, which is not dissimilar to our own preoccupation. Proper "Features Detection" comes as a cornerstone of efficient

heterogeneous correlation. (http://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0013-4)..

As another inspiring example, heterogeneous correlation will soon offer a totally secure substitute to traditional passwords: a very clear signal has been given by Google at its annual I/O conference 2016, announcing the availability of a "Authentication APS developer kit" by the end of 2016 based upon "behavioural biometrics" (Abacus project). This component of Chrome will allow elaborating a "trust score" to enable the user login, achieving its authentication through a comprehensive pattern of behavioural features currently captured by the smartphone own sensors: how you type, where you are, how fast you move, your voice intonations etc. - none of them "mediated" by an explicit identification request (https://www.newscientist.com/article/2091203-google-plans-to-replace-smartphone-passwords-with-trust-scores/).

The transposition of these advanced IT concepts to cross-correlate the data respectively collected by Port Authorities, Customs, VTMS, fishery control agencies, Defence, Law enforcement agencies etc. seems straightforward: this shall allow directing the operator's focus on few "abnormal/suspicious" seafarers trying to hide in the global maritime traffic while clearing without any further investigation most of the tracked vessels.

3. THE REVOLUTION TO COME: EASING DATA ACCESS

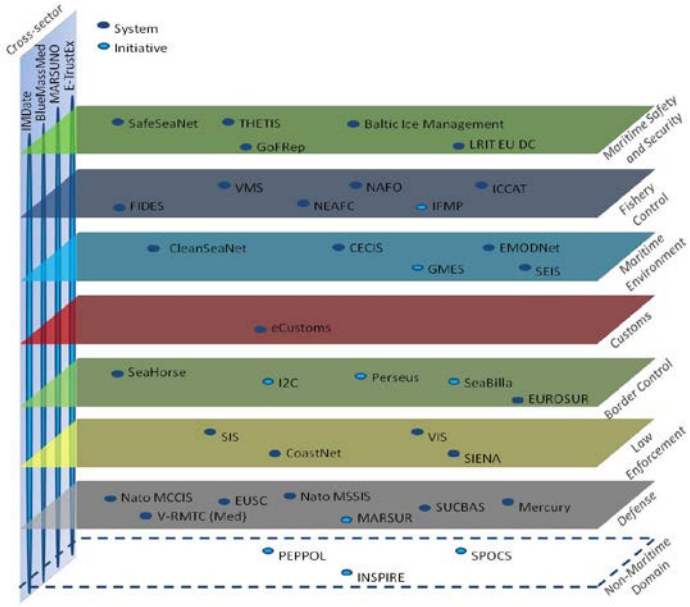
The key point to enable weak signal analysis is to access as many data as possible without any prior selection: searching the "unknown unknown" means that the data owner has no clues to pre-select what is worth being shared with its partnering maritime authorities – so it is a totally distinct approach from current "need to know" (and even the emerging "dare to share") the data he has already identified as suspicious.

The second characteristic of this activity is to process all sorts of data collected in the global framework of "maritime management", including the associated meta-data. This is totally distinct from building a "common maritime picture" by fusing all the ship tracks collected across the community of maritime administrations: instead of aggregating every possible data of the same nature (ship tracks), the purpose is to browse with pre-determined search strategies (e.g. to build a confidence index) all possible layers of data (e.g. ship owner, previous ports of call, container numbers, crew list, average speed, mix of cargo, berthing records, occurrence of encountering with ship Y etc etc)

The conjunction of the total dematerialization of all the shipping documents (e-maritime, single national window…) and of the go-ahead for the Common Information Sharing Environment (CISE) is creating the framework of a massive "maritime Big Data" which is the pre-requisite for deploying advanced data mining tools underlying the weak signal analysis approach.

This requires however building effectively a CISE with all the features of the original vision of DG Mare in terms of seamless access to all relevant national/sectoral data repositories.

Fig.1 is a familiar conceptual view of CISE (excerpt from the Deloitte report ref [8]) showing the various "User Community Layers" expected to rally the common exchange environment. On this graphic, developing heterogeneous correlations would come as achieving a seamless permeability between any layers, to conduct the correlation by picking data of all 7 colours and detect



anomalies that no layer would ever suspect.

Fig 1: "CISE Landscape", from ref [8] p.189

There is currently a risk that a number of User Communities (UCs) still consider the horizontal permeability as the principal scope for CISE, materializing into a collection of "common sectoral operational picture" built from the data considered by each data owner as "interesting to share" within the same UC for improving cross-border cooperation. At a time where the Pre-Operational Validation project CISE-2020 is launching the procurement of critical IT software bricks of the future CISE, opening this discussion seems critical.

4. THE VDES OPPORTUNITY

The AIS is currently under revision at international level (IALA, IMO) to incorporate the capability of broadband data transfer (Very high frequency Data Exchange System, VDES) while improving as well the capture of AIS signals by satellites (S-AIS).

Planned to enter into service by 2020, as for CISE, the VDES will provide all reporting vessels with a capability to contribute to the global maritime surveillance picture. It shall be seen having the potential of a Copernican revolution: moving from the era of unilateral reporting toward VTMS operators with little if no feedback, ship masters will be able to exchange all sorts of maritime safety and security notices with the neighbouring ships (via VHF) and with the whole community (via S-AIS) at no cost (compared to the SatCom broadband links currently needed). EU Maritime Authorities should monitor more closely and possibly influence the current phase of standard VDES messaging definition to secure the effective contribution of every cooperative ship to report its local environment as a contribution to the grand picture currently compiled by National Authorities, Regional Commissions, EMSA and NATO. Early reporting of the sighting by a cargo or a ferry of “strange” ships around her (fishing vessels out of fishing grounds, old cargo much too low over the water, towed pateras or RHIBs, unusual routes, apparent rendez-vous at sea, low flying plane etc) could help directing patrols well before incidents might trigger alerts. VDES has the potential to turn every cooperative ship as an “in-situ” maritime surveillance sensor, able to transmit messages, pictures, radar screenshots, open comments etc.

5. RECOMMENDATIONS

Maritime authorities will not transform their data sharing and data access policies overnight. In the same time, major enablers of a new way to think the maritime surveillance data analysis are at stake now depending on short term decisions at EU level (CISE) or UN/IMO (VDES) to freeze the technical requirements of these new systems.

Our R&D community has the duty to launch now very imaginative and convincing “vertical” data mining experiments (wrt Fig.1) to demonstrate the power of weak signal analysis, the way to build unprecedented “trust scores” and the consequence of this transformation of the notion of “abnormal behaviour detection” on the requirements (including at some stage underlying legal agreements) of both CISE and VDES. Without offering attractive use-cases, we might miss this challenging 2020 milestone, with the risk of not meeting again before long such opportunity.

6. REFERENCES

[1] Pallotta G., Vespe M. and Bryan K. (2013), “Vessel pattern knowledge discovery from AIS data: a framework for anomaly detection and route prediction”, Entropy 2013, 15, 22218-2245.

[2] Vandecasteele A., Devillers R. and Napoli A. (2014), “From Movement Data to Objects Behavior Using Semantic Trajectory and Semantic Events”, Marine Geodesy, Taylor & Francis, 2014, 37 (2 - Special Issue: Special Issue on Coastal and Marine Geographic Information Systems), pp.126-144.

[3] Idiri B. and Napoli A. (2012), “The automatic identification system of maritime accident risk using rule-based reasoning”, 7th International Conference on System Of Systems Engineering - IEEE SOSE 2012, Jul 2012, Genoa, Italy. pp.125-130.

[4] Ansoff, H. I. (1975), “Managing Strategic Surprise by Response to Weak Signals”, California Management Review, vol. XVIII no. 2, pp. 21–33.

[5] Gilad, B. (2003), “Early Warning: Using Competitive Intelligence to Anticipate Market Shifts, Control Risk, and Create Powerful Strategies", AMACOM.

[6] Brizon, A. and Wybo, J-L. (2009), “The life cycle of weak signals related to safety”, Int. J. Emergency Management, Vol. 6, No. 2, pp.117–135

[7] Zuech R., Taghi M Khoshgoftaar and Randall Wald (2015), “Intrusion detection and Big Heterogeneous Data: a Survey”, Journal of Big Data, Springer 2015.

[8] Deloitte Consulting (2012), “Study on the current surveillance IT landscape and the resulting options for the Common Information Sharing Environment for Surveillance in the Maritime Domain”, ISA/DG Mare/DIGIT.