Dr. Rami Abielmona is the vice president of Research & Engineering at Larus Technologies Corporation (Rami.Abielmona@larus.com).

TACKLING BIG DATA
IN MARITIME DOMAIN AWARENESS

Photo: RCN

In a world where more than 40 percent of the population lives within 100 kilometers of a coast and where traditional and asymmetric threats to physical and cyber infrastructures and borders continue to rise each year, countries are becoming increasingly aware of the gaps that exist in their ability to achieve persistent surveillance and continuous awareness of their maritime domains.

Persistent surveillance is an essential component in a global system to ensure Maritime Domain Awareness (MDA). The latter is defined as the situational understanding of activities that impact maritime security, safety, economy or environment. MDA involves a system of people, processes and technological tools that discover, sense, analyse and react to events, and perform physical and virtual defence of the country's borders. It includes the capture and storage of domain knowledge obtained along with the actions, effects and outcomes for use in planning future surveillance operations.

The outcome expected from MDA is the effective tasking of joint and interagency assets to respond to offensive/illegal activities, disasters and rescue scenarios in the maritime domain. In Canada, MDA requires the surveying of 10 million square kilometers across the Pacific, Atlantic and Arctic oceans, over 200 thousand kilometers of coastline and five million square kilometres of Arctic landmass, and the inherent challenge of monitoring and controlling the vast amount of data and information that will be generated.

This activity falls within the jurisdiction of the Marine Security Operations Centres (MSOCs) and the Canadian Forces' (CF) Regional Joint Operations Centres (RJOCs). These organizations are responsible for detecting and assessing Canadian marine security threats and providing support to responders. Threats include individuals, vessels, cargo and infrastructure involved in any activity that could pose a risk to the safety, security, environment or economy of Canada.
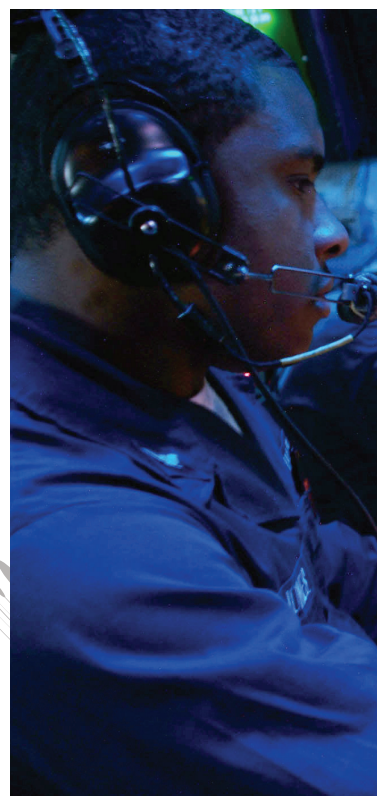
Using the many existing loosely connected surveillance and exploitation systems, operators and analysts have been overwhelmed by the tide of incoming data, including sensor outputs, databases, reports and other sources of information. This situation has led to operator/analyst fatigue, overload, stress and inattention which, in turn, have led to human errors.

We have seen that on a limited basis, surveillance solutions have been effective, particularly where the regions of interest were well delineated, data sources structured and precise, events-of-interest few and far between, and response requirements neither time-critical nor calculated.

However, this level of performance is not sustainable over time and on a global scale. Any proposed solution to these challenges will need to feature continuous awareness of the environment unconstrained by data parameters or geographical boundaries, i.e., persistent surveillance.

Persistent surveillance

To enable effective continuous awareness, threat mitigation and re-

sponse to territorial breaches, persistent surveillance is needed and must be instituted in a systematic way. Persistent surveillance systems incorporate multiple collection, exploitation and dissemination capabilities that cooperatively detect, classify, identify, track, corroborate and assess situations within maritime areas.

This cooperative approach has two significant, positive effects: it permits the creation of fused information and intelligence products for use by decision and policymakers, and it results in effectiveness and efficiency benefits due to the systems being coordinated, widely dispersed, remotely controlled and intelligent.

Additionally, there are many potential data sources that can be inputted into these systems. These sources fall in two categories: structured vs. unstructured (sometimes referred to as hard vs. soft).

Structured, or "hard", indicates data that has a high observational sampling rate, is easily repeatable, and is calibrated and precise, such as data from radar-based, tracking-based and imagery-based sensors. Unstructured, or "soft", indicates data that provides relations between discovered entities; typically it has a low observational sampling rate, is easily repeatable, is less precise and is uncalibrated, such as human observation-based (e.g., field reports), web-based (e.g., websites, pages, forums) and map-based (e.g., navigational charts, climate maps).

Information fusion

To accurately and effectively monitor a maritime area, the vast depth and breadth of incoming data must be interpreted and managed. Often referred to as the "Big Data Problem," this state is best handled through the creation and maintenance of a real-time representative model of the world. Early solutions at-

tempted to resolve this challenge through low level Information Fusion (IF) modules that used complex mathematical formulations or brute force number crunching. However, these solutions were inadequate because the complexity created by the four-dimensional vector (variety, volume, velocity and veracity) quickly increased to the point where low level IF modules were overwhelmed. Low level IF was only capable of performing fusion when the data itself was limited in volume, involved few types (low variety), did not frequently change in mission-critical applications (low speed) and was somewhat trustworthy (high veracity). As data complexity continued to grow exponentially, researchers realized that at some point a new computational paradigm was required.

To address the challenges of Big Data, High-Level Information Fusion (HLIF), which in the Joint Director of Laboratories (JDL) model is defined as Fusion Level 2 and above, has become the focus of research and development efforts. HLIF uses a mixture of numeric and symbolic reasoning techniques running in a distributed fashion while presenting internal functionality through an efficient user interface. HLIF allows the system to learn from experience, capture human expertise and guidance, automatically adapt to changing threats and situations, and display inferential chains and fusion processes graphically.

Instead of attempting to keep up with the ever increasing complexity of the four-dimensional data streams, HLIF, aided by Computational Intelligence (CI), allows one to model, and therefore, better understand the data stream sources and better adapt to the dynamic structures that exist within the data. CI-based algorithms furnish an HLIF system with its reasoning, inference and learning capabilities and involve the design of computational architectures, methodologies and processes to address complex real-world problems using nature-inspired approaches.

HLIF capabilities are continuing to evolve to alleviate the challenges presented by Big Data, including:

- anomaly detection, a process by which patterns are detected in a given dataset that do not conform to a pre-defined typical behavior (e.g., outliers);
- trajectory prediction, a process by which future positions (i.e., states) and motions (i.e., trajectories) of an object are estimated;
- intent assessment, a process by which object behaviors are characterized based on their purpose of action; and
- threat assessment, a process by which object behaviors are characterized based on the object's capability, opportunity and intent.

Hence, an HLIF- and CI-based continuous MDA solution improves on existing persistent surveillance methods by generating an understanding of the objects, actions and intentions. It adds automation to the surveillance process by fusing a multitude of structured and unstructured data sources through computational intelligence algorithms and behavior analysis into a decision support system. The solution needs to learn and continuously improve upon itself in real-time to provide true and timely information on maritime activities, reduce operator workload, provide an accurate and reliable world model and enable interoperability and knowledge sharing. 🍁