Soft Data-Augmented Risk Assessment and Automated Course of Action Generation for Maritime Situational Awareness

by

Alex Plachkov

A thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

in partial fulfillment of the requirements for the degree of

Master of Applied Science

in Electrical and Computer Engineering

Ottawa-Carleton Institute of Electrical and Computer Engineering

School of Electrical Engineering and Computer Science

Faculty of Engineering

University of Ottawa

Abstract

This thesis presents a framework capable of integrating hard (physics-based) and soft (people-generated) data for the purpose of achieving increased situational assessment (SA) and effective course of action (CoA) generation upon risk identification. The proposed methodology is realized through the extension of an existing *Risk Management Framework* (RMF). In this work, the RMF's SA capabilities are augmented via the injection of soft data features into its risk modeling; the performance of these capabilities is evaluated via a newly-proposed risk-centric information fusion effectiveness metric. The framework's CoA generation capabilities are also extended through the inclusion of people-generated data, capturing important subject matter expertise and providing mission-specific requirements. Furthermore, this work introduces a variety of CoA-related performance measures, used to assess the fitness of each individual potential CoA, as well as to quantify the overall chance of mission success improvement brought about by the inclusion of soft data. This conceptualization is validated via experimental analysis performed on a combination of real-world and synthetically-generated maritime scenarios. It is envisioned that the capabilities put forth herein will take part in a greater system, capable of ingesting and seamlessly integrating vast amounts of heterogeneous data, with the intent of providing accurate and timely situational updates, as well as assisting in operational decision making.

Acknowledgements

I would like to thank my supervisor, Dr. Voicu Groza, and co-supervisors: Dr. Rami Abielmona, Dr. Diana Inkpen, and Dr. Emil Petriu. This work would not have been possible without their encouragements, ongoing guidance, and endless insights. I would like to further extend my deepest gratitude towards my colleagues Dr. Rafael Falcon and Jamieson McCausland – thank you so much for all the valuable, thought-stimulating discussions!

Last, but certainly not least, a special thank you goes out to my family for all their immeasurable patience, support, and optimism during this time.

This work has been partially supported through software tools provided by Larus Technologies Corporation and funding provided by MITACS Accelerate programs.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACGA	Auxiliary Coast Guard Asset
ADM	Anomaly Detection Module
AIS	Automatic Identification System
AOI	Area of Interest
APGM	Asset Path Generation Module
ASM	Asset Selection Module
AU	Asset Utilization
CB	Contact-Based
CGA	Coast Guard Asset
CoA	Course of Action
CR	Common Referencing
DA	Data Association
DTG	Data Topic Group
ETURWG	Evaluation of Technologies for Uncertainty Representation Working Group
GD	Geographical Database
g-IFE	Gradual Information Fusion Effectiveness
HLIF	High-Level Information Fusion
HSIF	Hard-Soft Information Fusion
IE	Information Extraction
IF	Information Fusion
IFE	Information Fusion Effectiveness

IFE	Information Fusion Effectiveness
IG	Information Gain
i-IFE	Instantaneous Information Fusion Effectiveness
IMB	International Maritime Bureau
IMO	International Maritime Bureau
IQ	Information Quality
ISR	Intelligence, Surveillance, and Reconnaissance
LDR	Law of Diminishing Returns
LKP	Last Known Position
MDA	Maritime Domain Awareness
ME	Mission Expenses
MIP	Mean Incident Proximity
MIS	Mean Incident Severity
MMSI	Marine Mobile Service Identity
MR	Mission Requirement
MSOC	Marine Security Operations Center
MSTAR	Man-portable Surveillance and Target Acquisition Radar
MT	Mission Time
NER	Named Entity Recognition
NGA	National Geospatial-Intelligence Agency
NSGA-II	Non-dominated Sorting Genetic Algorithm II
OIT	Overall Information Timeliness
ORA	Opportunistic Response Asset
PAM	Performance Assessment Module

PCDRA	Potential Contact Detection Per Response Asset
QoI	Quality of Information
QTF	Quality Transfer Function
RAL	Risk Awareness Level
REM	Response Enactment Module
RHM	Regional Hostility Metric
RMF	Risk Management Framework
RMF	Risk Management Framework
RRDM	Response Requirements Determination Module
SA	Situational Assessment
SAM	Situation Assessment Module
SAR	Synthetic Aperture Radar
SAW	Situational Awareness
SDDRG	Soft-Data-Driven Response Generation
SME	Subject Matter Expert
URREF	Uncertainty Representation and Reasoning Evaluation Framework
USA	Unexplored Search Area
VaL	Vessel at Large
VID	Vessel in Distress
VPI	Victim Pertinence Index
WC	Weather Conditions
WWTTS	Worldwide Threats to Shipping

Chapter 1. Introduction

Maritime Domain Awareness (MDA) can be summarized as the situational knowledge of physical and environmental conditions that exist within or influence a maritime region. The intended scope of this awareness includes all behaviours that could, directly or indirectly, affect the security of the region, its economic activity or the local environment [1] [2]. MDA is achieved via persistent monitoring of the maritime environment, which allows for the identification of common trends and thus the detection of anomalous behaviour. Achieving effective MDA also entails the efficient tasking of maritime assets to counter illegal activities, as well as to assist in search and rescue (SAR) efforts.

In order to obtain accurate MDA, information derived from multiple data sources must be effectively captured, seamlessly integrated, and efficiently reasoned upon. Initial solutions attempted to resolve these challenges through the use of low-level Information Fusion (IF) modules; however, as the available data experienced exponential growth (in terms of variety, volume, velocity, and veracity), these low-level modules became inadequate [1]. To overcome this challenge, also referred to as the *Big Data Problem*, much research has gone into the development and use of High-Level IF (HLIF) techniques. HLIF is defined as Level 2 (L2) Fusion and above in the Joint Director of Laboratories (JDL)/Data Fusion Information Group (DFIG) models [3] [4], and has been successfully tasked with carrying out MDA processes (e.g., anomaly detection, trajectory prediction, intent assessment, and threat assessment) [1] [5].

Traditionally, IF systems have relied on the use of 'hard' data, that is, data provided by physical sensors, such as acoustic, radar, global positioning system (GPS), sonar, or

electro-optical sensors. This data is objective, structured, and quantitative, and maritime operators often rely on hard data generated by vessel traffic in order to identify anomalous or suspicious events at sea. Recently however, the IF community has recognized the potential of exploiting information generated by 'soft' sensors (that is, people-generated information). Conversely, this type of data is more subjective, qualitative, and transcribed in a semi-structured or unstructured format (e.g., free form natural language text). Soft data can be rich in context and provide cognitive inferences not otherwise captured by traditional sensors (e.g., judged relationships among entities and/or events) and can thus be used to further augment the situational awareness picture of the system. The exploitation of this type of data, however, comes at a cost, as its content can be exposed to bias/subjectivity, inaccuracy, imprecision, low repeatability, and conflict [6] [7]. Examples of soft data sources in the maritime domain include textual reports on vessel sightings or detailed descriptions of maritime incidents. It is therefore not surprising to see recent attempts to construct and evaluate Hard-Soft IF (HSIF) frameworks [8] [9], as well as efforts to incorporate soft information in maritime decision making like in [10], where Natural Language Processing (NLP) methods were employed to draw out meaningful information from soft data sources.

An HLIF system capable of constructing a risk-aware view of the environment (and the dynamic entities within) can enable the proactive identification of hazards, dangers, and vulnerabilities, which subsequently leads to the generation of suitable countermeasures to mitigate said risks. Such a system has already been proposed in [11] [12] and termed by the authors the *Risk Management Framework* (RMF), but considers only hard data sources.

1.1 Motivation

Maritime piracy and armed robbery at sea is estimated to cost the worldwide economy annually anywhere between $1 Billion to $16 Billion US dollars[1], whilst SAR efforts are estimated to cost the US coast guard alone approximately $680 Million USD yearly[2]. These numbers present a strong need for automated solutions in securing the maritime areas by providing accurate and timely situational updates and assisting in operational decision making. There are recent organizational initiatives to support increased collaboration and maritime information sharing among participating nations through internet networks precisely for this purpose, such as: (1) the Information Fusion Centre (IFC) hosted by the Singapore Navy, where 35 countries are exchanging data and collaborating on maritime operations, such as piracy hijacking events and SAR missions[3]; (2) the Information Sharing Centre (ISC) of the Regional Cooperation Agreement on Combating Piracy and Armed Robbery against Ships in Asia (ReCAAP), also based in Singapore, with a total of 20 currently collaborating nations[4]; and (3) the International Maritime Bureau's Piracy Reporting Centre (IMB PRC), located in Malaysia, which is responsible for disseminating piracy incident data to maritime security agencies around the world [13].

1.2 Contributions

This study focuses on extending the RMF's L2 and L3 fusion capabilities through the inclusion of soft data for the purpose of increased SA and improved CoA generation.

---

[1] http://worldmaritimenews.com/archives/134829/annual-global-cost-of-piracy-measured-in-billions/
[2] http://www.theday.com/article/20091219/INTERACT010102/912199999/0/SHANE
[3] https://www.mindef.gov.sg/imindef/press_room/official_releases/nr/2014/apr/04apr14_nr/04apr14_fs.html
[4] http://www.recaap.org/AboutReCAAPISC.aspx

It is expected that the injection of soft data into the fusion processes will yield higher mission-specific measures of performance, as this data encapsulates subject matter expertise. To the best of my knowledge, this work is the first one to apply soft data to automated CoA generation and performance evaluation in the maritime domain.

This work makes the following contributions:

1. The RMF's L2 SA capabilities are enhanced through the inclusion of soft data into its risk model. This is achieved by augmenting two of the risk features pertaining to maritime vessels by the inclusion of information derived from the processed soft data.
2. The proposed risk-centric IF Effectiveness (IFE) metric is capable of quantifying the effectiveness of the L2 process; this metric is then employed to illustrate and quantify the merit that the different soft sources provide to the L2 capabilities of the system. Furthermore, this metric is used to conduct an L2 process robustness analysis.
3. A new subsystem architecture is unveiled for the RMF's L3 capabilities, which includes a variety of modules supporting soft-data-augmented CoA generation.
4. A variety of L3 performance measures are proposed and used to assess the fitness of individual potential CoAs, as well as to quantify the improved chance of mission success brought about by inclusion of the soft data.

1.3 Organization

This thesis consists of six chapters. Chapter 2. consists of background and related work information. The next two chapters cover the two principle components of the thesis: Soft-Data Augmented Situational Assessment (Chapter 3) and Soft-Data Augmented Course of Action Generation (Chapter 4). Chapter 5 lays out the experimentation, along with the results and their analysis. Finally, Chapter 6 presents the concluding remarks and addresses future research directions.

Chapter 2. Background and Related Work

2.1 Maritime Risk Analysis

Risk can be formally defined as the effect of uncertainty on an entity's objectives [14]. A fundamental objective in MDA is the achievement of a comprehensive level of situational awareness (SAW), and within this domain, uncertainty corresponds to the lack of information to support such a level. SAW is defined as a state of knowledge of the environment, achieved and maintained through one or many SA processes [15]. Augmenting the degree of SAW in MDA can be broken down into a number of sub-problems, such as the enhancement of the collection and dissemination of data, as well as the improvement of the multitude of algorithms involved in its processing. Popular techniques employed for evaluating risk are probabilistic models, such as Bayesian Networks in [16] [17], Hidden Markov Models (HMMs) in [18] [19], and Evidence-based models in [20]. The common drawback of these approaches is that they greatly rely on *a priori* knowledge, derived from statistical analysis of large collections of historical data (e.g., event occurrence probabilities). Computational Intelligence (CI) techniques have also been employed, such as in [12] and [21] where some of the risk features of assets are quantified via Fuzzy Systems (FS). FS provide a natural, interpretable, yet effective basis for incorporating domain expert knowledge into the risk assessment process.

2.2 High-Level Information Fusion for Maritime Domain Awareness

Machine learning algorithms that are employed in the maritime domain aim to detect anomalies, or abnormal behaviors, in vessel activity data. Typically, these algorithms are trained using information pertaining to normal vessel activities in order to learn a normalcy model, and, when operational, they are used to detect any behaviour deviating from that model (i.e., anomalous behaviour). Recent HLIF techniques that have been employed for anomaly detection include neural networks [22] [23] [24], probabilistic models [25] [26] [27] [28] [29] [30], clustering approaches [24] [31] [32] [33], as well as evolutionary algorithms [34] [35]. These techniques typically fall under two categories – unsupervised vs. supervised. *Unsupervised techniques* are popular due to their self-governing nature – namely, their ability to discover hidden, meaningful structure in unlabeled data. A good argument can, however, be made for *supervised techniques*, since they incorporate domain expert knowledge into the learning process in the form of labelled training data.

The benefit of including human operators and analysts in the anomaly detection process itself (i.e., when the system is online) has been greatly underestimated in the maritime domain [36]. For this reason, the involvement of people in traditional systems has largely been only up to the point of constructing the anomaly model. There have been recent attempts to incorporate humans into the MDA processes, such as in [22], where the system can be directed for continuous online learning via operator intervention. Such evolving systems are well-suited for large, complex, dynamic environments like the maritime domain. Some recent research efforts [36] [37] are geared towards identifying

potential areas for visualization and human interaction during the anomaly detection process for improving the performance of the detectors.

Within the Joint Director of Laboratories (JDL)/Data Fusion Information Group (DFIG) models [3] [4] [38], Level 2 (L2) and Level 3 (L3) Information Fusion (IF) are respectively defined as SA and Impact Assessment (IA), with the term *High-Level IF* (HLIF) being defined as L2 and above. The role of L2 Fusion is to characterize the presently unfolding situations through the automated analysis and identification of relations existing between the entities being monitored, as well as between the entities and the environment they reside in. Upon successful characterization of situations, L3 Fusion proceeds with generating CoA recommendations and estimating their effects on the known situations. Both levels have been successfully tasked with carrying out typical MDA processes (e.g., anomaly detection, trajectory prediction, intent assessment, and threat assessment) [1] [5]. The automatic generation of suitable CoAs in the maritime domain using evolutionary multi-objective optimization (EMOO) has been addressed before in [12] and [21] but considering only hard data sources. Soft data has been employed in [10] at L2 Fusion to extract risk factors from reported maritime incidents. No studies in the maritime world appear to consider supplementing or augmenting CoA generation through the use of soft data.

2.3 Data in High-Level Information Fusion

Two broad categories of data exist in the IF domain, namely, hard and soft. Hard data is generated by physical sensors, and hence it is typically provided in a numeric format. Characteristics of this type of data are calibration, accuracy, structure, objectivity, precision, repeatability, and high frequency [6] [7]. Soft data, on the other hand, is typically

generated and transcribed by humans in intelligence reports, surveillance reports, as well as open source and social media [39] [40]. This type of data is provided in a textual format which can either be structured (e.g., restricted language such as Controlled English [41]) or unstructured (e.g., free-form natural language). The exploitation of the information provided by soft sensors comes at a cost, as its content can be exposed to bias/subjectivity, inaccuracy, imprecision, low repeatability, and conflict [6] [7]. Soft data is incorporated into HLIF systems by being processed through: Information Extraction (IE) modules, Common Referencing (CR) modules, and Data Association (DA) modules.

IE modules attempt to draw out information of interest from each of the data sources. The construction of the IE modules for soft data sources requires the elicitation of a *lexicon* – a dictionary of terms that embody factors and entities of interest in the environment [10]. This lexicon is then used for their automatic identification and extraction.

CR modules convert the content provided by IE modules into a common format via, for instance: (1) uncertainty alignment [42], as the heterogeneous data sources may express it in inconsistent forms (e.g., hard sensor uncertainty is expressed in probabilistic/quantitative format whereas soft sensor uncertainty is stated in possibilistic/fuzzy terms due to the qualitative nature of linguistic expression); or (2) temporal alignment [43], which is required when the disparate data sources present temporal features in incongruent forms (e.g., exact timestamp vs. approximate time of day vs. linguistic phrasing encompassing multiple verbal tense expressions).

The responsibility of DA modules is to effectively associate all of the available evidence (from each of the data sources) for a unique entity. The same entities in the

environment being monitored can be characterized by multiple data sources, and thus, the DA module must effectively join all of the available data for an entity into a single, cumulative data structure (also frequently referred to as 'cumulative evidence structure') [44].

2.3.1 Soft Data Categories

There are four distinct categories of soft data:

a. Observational data [40] [41] consists of reports written by human observers (e.g., intelligence data written by soldiers reporting on their patrol activities). It is typically communicated in natural language. This type of data is semantically rich, and when provided in a 'controlled language' format, it can reduce the impreciseness and uncertainty. It can be, however, ambiguous and biased, especially when not presented in a 'controlled language'. Observational data is also uncalibrated – two humans observing the same situation may provide different, even conflicting, information about the situation.

b. Contextual data [40] [42] [45] [46] contains information that can be used to characterize a situation or the surrounding environment of one; it is said to 'surround' a situation. This is information which does not characterize the entities of interest in an IF system, but rather the environment in which they reside or the situations to which they appertain. Contextual information is of two types: static (e.g., GIS database) and dynamic (e.g., weather conditions). This type of information is used to better understand the situation (e.g., by reducing the amount of uncertainty present in the information extracted from the observational data through the removal of ambiguities or addition of constraints). A large amount of

contextual information improves the situational understanding, but because it has to be merged with observational data, the data assimilation operations become time costly. Contextual data is also prone to inconsistencies and errors. No generic context representation and exploitation frameworks exist; all attempts to exploit context are tailored for the application at hand and thus the solutions are of limited scalability and poor adaptability. This issue is recognized by the authors of [46], who unveil a proposal for a middleware architecture to integrate this rich information source into IF processes.

c. Open source and social media [40] (e.g., Blogs, Twitter, Facebook) data can be very timely and typically arrives in a large volume. Arriving in a high volume, however, implies potential difficulties in its processing. This soft data type is also largely unstructured and uncontrolled.

d. Ontological data [40] [47] [48] aids in better understanding the observed situations of the world by capturing entities, processes, events, and the relationships among them in a particular domain. Its responsibility is to capture domain knowledge; when used in conjunction with the hard data, it can provide a more accurate situational understanding. Like with contextual data, augmenting the observational data with ontological information comes at a computational cost.

2.3.2 Data Imperfections

Data imperfections have been well studied in the data fusion community and can be classified in three broad categories [49] – uncertainty, granularity and imprecision. **Uncertainty** results from ignorance about the true state of the real world, and uncertain

data often entails having an associated confidence degree between zero and one [49] [50]. **Granularity** refers to the ability to differentiate between different entities present in the data. **Imprecision** can be further broken down into the following three categories: **vagueness**, described as lacking detail in such a way that the entities in the data have no direct correspondence with their referents in the world; **ambiguity**, which occurs when the entities in the data can have more than one referent in the real world; and **incompleteness**, which arises when the data possesses only partial information about the world [49] [51].

The following mathematical theories have been recently used in the HSIF community to attempt to deal with imperfect data: Dempster-Shafer Evidence Theory (DSET), Possibility Theory (PT), Random Finite Set Theory (RFST), and Markov Logic Networks (MLNs). Specific details on each of the former three theories, as well as their applicability, limitations, and use in data fusion can be found in [49]; details on MLNs can be located in [26].

DSET was used to allow for decision making under uncertainty in a hard-soft data fusion security system in [52], which is capable of assessing the threat level of situations in which objects approach and/or cross a security perimeter. It was also used in [53] to combine the uncertainties of data items in a fusion system containing only soft data. Transferable Belief Model, an extension of DSET, was used to perform reasoning under uncertainty and evaluation of the threat level caused by suspicious vessels in a hard-soft data fusion system in [45]. It was also used to combine inconsistent data and assess how much of the received data observations supported given hypotheses in a soft-data-only fusion system in [54].

PT was used to integrate imprecise qualitative data in a soft fusion framework in [42], and to model the uncertainty in soft data in an SA system in [55].

RFST was used in an HSIF framework in [56] to combine data from both hard and soft sources and enhance the performance of Kalman Evidential Filter-based target tracking.

MLNs were used in [26] to encode domain knowledge through first order logic (FOL) expressions; the latter were assigned associated uncertainty weights, and fused with real-time sensor data in order to detect maritime traffic anomalies.

2.4 Performance Evaluation in Hard-Soft Information Fusion

A natural interest in the field of HSIF is the quantification of the performance of the fusion systems. Possessing the ability to evaluate the performance of such systems is crucial for making key design and/or runtime decisions, such as which fusion algorithms are most suitable given the available data sources, or what subset of the data sources provides the highest benefit to the system (for situations in which it may be computationally intractable to process all of the available data). When it comes to evaluating performance of traditional fusion systems, the focus has largely been on computing the various *Measures of Performance* (MOPs) and *Measures of Effectiveness* (MOEs). MOPs are used to assess the ability of a fusion process to convert raw signal data into intelligible information about an entity, whereas MOEs are used to assess the ability of a fusion system as a whole to contribute to the successful completion of a mission [57] [58]. Assessing the performance of modern HSIF systems, however, where people-generated data is integrated in the fusion system, is still a burgeoning research area, and

there has been little work done in evaluating the performance of systems with individuals integrated into the fusion process [58] [49] [59].

2.4.1 Performance Metrics

Innate differences between physical sensors and humans make it difficult or impossible to use these MOPs and MOEs when people are integrated into the fusion process; physical sensors can be calibrated for consistent performance, and although people can be trained, the differences in internal factors (such as cognitive abilities, biases, and stress) can cause unknown performance variation [42] [58] [59].

When it comes to evaluating performance, two broad classes of metrics have been previously identified: **quality-based metrics** and **runtime-based metrics**; the optimization of both classes typically involves conflicting objectives [8]. There currently does not exist much in the area of runtime-based metrics – only algorithm computational times are given, as in [44] [8]. Overall, more emphasis has been put into developing quality-based metrics; the latter allow for the quantification of the performance of an IF system. Gaining insight into the system's performance is crucial for unlocking answers to questions such as which combinations of hard and soft data sources yield the highest value, as well as which fusion algorithms are most suitable for these data sources. There are three levels on which the quality of a fusion system can be judged: the **input information level**, the **fusion algorithm level**, and the **system level**.

2.4.1.1 Input Information Quality Evaluation

The theory behind information quality evaluation is thorough, as it largely borrows from existing fields (e.g., information theory). There is a myriad of quality dimensions which have been identified (e.g., accuracy, timeliness, integrity, relevance, among others);

however, input information quality evaluation appears to have gained little traction in the IF domain. A few relevant studies done in this area are [7] [39] [60] [61]; each of these presents an ontology on the *Quality of Information (QoI)*, as it pertains to IF. A prevailing theme in these studies is that no universal methodology exists to assess QoI in a general IF system. The dimensions used to evaluate the QoI are context-dependent (i.e., based on specific user goals and objectives) and thus the manner in which the QoI is evaluated will vary from one fusion system to another.

The quality level of a specific dimension can be evaluated using one of several methods (not all are possible in every situation): *a priori* knowledge (e.g., training level of a human observer), judgement of human experts, supervised learning from an existing dataset, or conflict level of information between different sources, among others [39]. Once the different information quality dimensions are evaluated, they can be integrated into a single, concise, unified quality score (UQS). Different ways of achieving an UQS include: the **combination rule** used in a mathematical framework in which the quality dimensions are represented, if a mathematical framework is used; a **weighted average** of the individual quality scores, normalized to unity; or the training of a **neural network** on a dataset labeled with subjective UQS labels [39].

Guidelines for confidence levels that can be assigned to the estimated reliability levels of information and information sources can be found in the US Military Standard 640 (MIL-STD-640) [62]; this standard provides quantitative labels for the otherwise qualitative descriptions of the confidence levels found in the NATO Standardized Agreement for Intelligence Reports - STANAG 2022.

2.4.1.2 Fusion Algorithm Quality Evaluation

Quality evaluation at the algorithmic level is typically treated like a classification problem. Ground truth values of the expected output of fusion algorithms are generated by human analysts, and Precision, Recall, and F-measure of the results are reported [42] [44] [8]. For processing algorithms, performance is typically assessed based on the number of detections or lack of detections of entities and relationships between entities. For DA, evaluators are concerned with the number of correctly associated (true positives – TPs), correctly not associated (true negatives – TNs), incorrectly associated (false positives – FPs) and incorrectly not associated (false negatives – FNs) entities or relationships. For state estimation, evaluation is done based on the performance of inexact graph matching algorithms; i.e., the number of correctly identified, correctly not identified, incorrectly identified, and incorrectly not identified instances of the template data graph within the cumulative data graph. The performance of the uncertainty alignment process is difficult to assess "due to the uncertain nature of the inferences" made by it [42] and it is thus evaluated indirectly through process refinement (by observed improvements to results of downstream fusion processes).

In order to assess the “true” performance of a fusion algorithm, any errors generated by upstream processes must be accounted for, as was demonstrated in [8]. The system evaluated in this research was an HSIF prototype ingesting the SYNCOIN (hard-soft) dataset [58]. The output of both the hard and soft data processing algorithms was an attributed data graph. It was the task of the DA algorithms to merge the distinct attributed graphs into a cumulative evidence graph. The DA process was evaluated at two levels in order to demonstrate the adverse effect upstream processes can have on downstream ones.

The evaluation at the first level disregards any errors that have been generated by upstream processes, and thus simply calculates the aforementioned performance metrics (i.e., Precision, Recall, F-measure) with a solution key. The evaluation at the second level aimed to determine the "true" performance of the DA process, and thus take into account upstream errors (e.g., missing or incorrect entity typing generated by the upstream natural language processing system), and determined the FPs and FNs as a result of these errors. The identified entity pairs are then simply disregarded from the calculation of the performance metrics. This approach was tested on three DA algorithms in [8], and each of them, unsurprisingly, reported higher performance when the upstream errors were being accounted for.

The limitation in this approach is that the performance of the algorithms is judged only in an 'offline' mode; such methods do not allow for an understanding of the instantaneous fusion quality of the results provided by the fusion system in a deployed environment.

2.4.1.3 System-Level Quality Evaluation

This level of evaluation is concerned with the integrative performance of the different components of an IF system – the input data, the fusion algorithms, as well as the humans themselves, in the case of a human-in-the-loop (HITL) IF systems.

Some recent efforts to model fusion performance include the research performed in [63], which is concerned with judging the quality of HITL fusion tools and aids (e.g., advanced human interaction displays, as well as social network analysis and collaboration tools). The research was conducted on a cyber-infrastructure that was set up in order to simulate real-world scenarios. The performance measures used were based on:

containment (number of events considered to be out of control), resource allocation efficiency (by determining cumulative event intensities), and team information sharing (e.g., frequency of communication interface use). The metrics were evaluated at different time slots, thereby evaluating the impact of a situation on the performance, as the situation was unfolding. Questionnaires given to the participants were also used to evaluate their SA level.

More recently, in [60], the authors put forward the idea that, if a fusion system is broken down to its most elementary modules, and if a quality transfer function (QTF) can be determined for every module, then given the input quality of the data into the module, the output quality can be determined via this function. With such a model, the data and its quality are propagated together through the fusion processing pipeline, and therefore, one can attain detailed insights into which modules are adversely affecting the QoI at any time instant. The approach solves the issue of aggregating the data qualities from the different modules of the HSIF system, because the last module in the pipeline will be providing a global performance measure. If a complete behaviour model of each module cannot be obtained, then experimentation can be performed to estimate the QTF. This experimentation would require varying the input data quality of a module, determining the output quality, and recording this input-output pair. Subsequently, the input-output pairs could be used to determine the QTF via statistical methods (e.g., regression).

It is worth noting here that an IF system's performance can be multifaceted and with conflicting objectives (e.g., not just concerned with meeting fusion goals, but also minimizing the expenditure of resources that are required to achieve these goals) [42].

2.4.2 Uncertainty Handling Capabilities

Evaluating the manner in which uncertainty is dealt with in a fusion system has been recently recognized as being distinct from evaluating how the overall system is performing [64]. The Evaluation of Technologies for Uncertainty Representation Working Group (ETURWG) is developing a comprehensive framework, namely the Uncertainty Representation and Reasoning Evaluation Framework (URREF), which is concerned with evaluating how uncertainty is managed (represented and reasoned with) in fusion systems [64] [61] [65]. As part of this work, a comprehensive URREF ontology has been constructed which specifies the concepts that are related to this evaluation. This ontology is meant to be used as a guideline for selecting the actual criteria to be used as part of the uncertainty evaluation in a particular fusion system.

2.5 Maritime Domain Basics

A maritime region is encompassed by a set of hard and soft data sources. This region consists of an AOI selected for monitoring; vessels are traversing this area and are tracked via active or passive sensing modalities, such as *Synthetic Aperture Radar* and *Automatic Identification System* (AIS) transceivers. These physical (hard) sensors produce periodical data about each of the vessels (speed, position, vessel type, course, etc.), which is captured and processed at Marine Security Operations Centres (MSOCs). Textual reports detailing specific sea incidents are submitted to maritime organizations, and can also be used for processing within the MSOCs.

2.5.1 Incident and Response Reports

There are four openly available soft incident and response data sources in the maritime domain:

a. International Maritime Bureau (IMB) Incident Reports[5]

b. National Geospatial-Intelligence Agency's (NGA) Worldwide Threats to Shipping (WWTTS) Incident Reports[6]

c. International Maritime Organization (IMO)[7]

d. Regional Cooperation Agreement on Combating Piracy and Armed Robbery against Ships in Asia (ReCAAP)[8]

These unstructured and semi-structured data sources provide details on incidents, such as: (1) the name of the vessel involved in the incident, (2) the type (category) of the vessel, (3) the position and location of the incident, and (4) a brief transcription of the events that occurred during the incident (e.g., armed pirates boarding the vessel). Additionally, information on the type of actions taken by the crew, the consequences to the crew, as well as the actions taken by the coastal agency is sometimes provided, though not consistently. Coastal agency actions are also typically vague. Figure 1 and Figure 2 present examples of ReCAAP and WWTTS reports, respectively.

---

[5] https://icc-ccs.org/piracy-reporting-centre/live-piracy-report

[6] http://msi.nga.mil/NGAPortal/MSI.portal?_nfpb=true&_pageLabel=msi_portal_page_64

[7] www.imo.org/en/OurWork/Security/PiracyArmedRobbery/Reports

[8] www.recaap.org/AlertsReports/IncidentReports

N°	Ship Name Type of Ship Flag Gross Tonnage IMO Number	Date Time	Position of the incident*	Details of the incident	Consequences for crew, ship, cargo	Action taken by the master and the crew	Was the incident reported to the coastal authority? Which one?	Reporting State or international organization	MSC.4/Circ.181 ANNEX 2 Coastal State Action Taken
1	2	3	4	5	6	7	8	9	10
IN INTERNATIONAL WATERS									
1	DELFA Bulk carrier Marshall Islands 31261 9330094	04/01/2012 03:00 UTC	EAST AFRICA Gulf of Aden Somalia 13° 10.00' N 049° 12.00' E	About five pirates armed with guns in a skiff chased and fired upon a bulk carrier underway. The onboard security team returned fire and the skiff turned away and aborted the attack.	Pirates fired upon the vessel	The onboard security team returned fire at the pirates	Yes CSO	Marshall Islands ICC-IMB Piracy Reporting Centre Kuala Lumpur, UKMTO, NATO Northwood, Royal Oman Police, Yemen Coast Guard	-
2	SENANUR CEBI Bulk carrier Turkey 31763 9491367	04/01/2012 07:35	EAST AFRICA Gulf of Aden Yemen 12° 14.60' N 044° 11.80' E	Pirates in a skiff chased and attempted to board the ship underway. The ship enforced anti-piracy measures, increased speed and made evasive manoeuvres, resulting in the pirates moving away.	-	Ship enforced anti-piracy measures, increased speed and made evasive manoeuvres	Yes UKMTO	ICC-IMB Piracy Reporting Centre Kuala Lumpur, UKMTO, NATO Northwood, Yemen Coast Guard	A warship and a helicopter came to the location

Figure 1: An example ReCAAP Incident and Response Report[9]

> 3. (U) PHILIPPINES: On 10 October, seven robbers boarded MV CECILIA near 05:20 N - 125:31 E, 4.5 nm southeast of Olanivan Island. Two speed boats with seven persons armed with guns approached the vessel. Suspecting a boarding attempt, the Master instructed the crew to close and lock all access to the vessel's accommodation area. The armed robbers boarded the vessel, but departed when they could not enter the ship. The Master raised the alarm, informed the local authorities and notified the owners. A coast guard vessel was dispatched to investigate.

Figure 2: An example WWTTS report[10]

[9] www.recaap.org/AlertsReports/IncidentReports
[10] http://msi.nga.mil/NGAPortal/MSI.portal?_nfpb=true&_pageLabel=msi_portal_page_64

2.5.2 Search Patterns

In SAR operations, there are four popular types of search patterns with which response assets can be tasked to execute. These are the Track Crawl, Parallel Track Line, Outwards Expanding Square, and Inwards Expanding Square search patterns [66] [67].

The *Track Crawl* is a search pattern assigned to a vessel or an aircraft and tasks them to follow the track of the vessel at large (VaL). This search pattern is used when the VaL will most likely be close to its intended track. A visualization of this search pattern is presented in Figure 3.

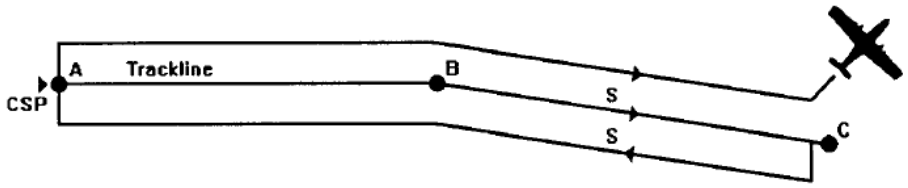


Figure 3: Track crawl search pattern, as presented in [67]

The *Parallel Track Line* is a search pattern providing uniform area coverage. This pattern can be executed by one or more vessels and/or one or more aircraft. The vessels or aircraft follow (in an outward fashion) parallel tracks along the expected drift direction of the VaL. This search pattern is useful when the search area is large and the VaL last known position (LKP) is not known with a good precision. A visual representation of this search pattern is presented in Figure 4.

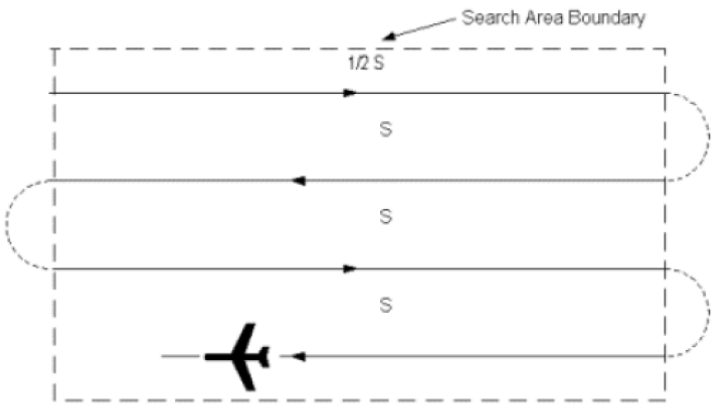


Figure 4: Parallel track line search pattern, as presented in [66]

The *Outwards Expanding Square* is a search pattern which starts at the VaL's LKP and expands outward in concentric squares. In addition to vessels, the search pattern can be used by an aircraft. Turns are 90 degrees. It is used when the VaL is thought to be within a small area. A depiction of this search pattern is presented in Figure 5.

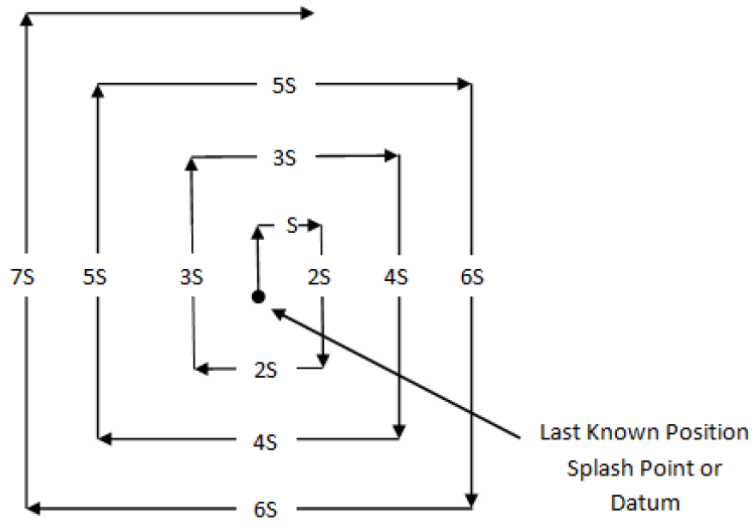


Figure 5: Outwards expanding square search pattern, as presented in [66]

Inwards Collapsing Square is a search pattern that is similar to the *Outwards Expanding Square*; however, instead of starting off at the last known VaL position, this location is visited last. It is used in the same situations as the outwards expanding square. The depiction of this pattern is the same as the one presented in Figure 5, except with the directionality of the arrows being reversed.

Chapter 3. Soft-Data Augmented Situational Assessment

SA capabilities of an IF system aim to characterize presently unfolding situations through the automated analysis of input data. This section illustrates the soft-data augmented methodology used for proactive identification of vessels which may fall under distress. Along with this methodology, a new L2 performance metric is proposed. Furthermore, it is demonstrated how this metric can be used to quantify the merit of the different soft data sources, as well as how it can be applied to perform a robustness analysis on the system.

3.1 Environment and Data Sources

A thorough description of the maritime environment can be found in Section 2.5. The data sources used in this work to achieve SA in the maritime domain are:

a. **AIS** [Hard] – A passive data source providing vital ship information, such as the vessel category it belongs to, its unique nine-digit Marine Mobile Service Identity (MMSI) number and its spatio-temporal features (e.g., speed, position, vessel type, course). The AIS feed was provided by exactEarth[11].

b. **Sea State Reports** [Hard] – A data source describing the motion of sea waves. The information extracted from this data set is used in conjunction with the Douglas Sea Scale[12]. This data source was simulated, as access to sea state reports, in the regions which were studied as part of this research was not available.

[11] http://www.exactearth.com
[12] http://en.wikipedia.org/wiki/Douglas_sea_scale

c. **WWTTS** [Soft] – An incident and response report data source, previously discussed in Section 2.5.1. The reports span the entire world and are provided on a weekly basis.

d. **IMB** [Soft] – A second incident and response report data source (refer to Section 2.5.1.). Reports from this source can be obtained in real-time, or in batch mode (on a quarterly or annual basis). In addition to providing maritime crime information similar in nature to WWTTS, this semi-structured data source also declares whether an incident was successful or merely attempted (e.g., vessel boarding by pirates was attempted but unsuccessful).

e. **GeoNames Geographical Database** [Soft] – A data source containing tuples of location names and their geographical coordinates[13]. It is used to supplement the WWTTS and the IMB incident reports for incidents which report only general locations (e.g., 50 nautical miles northeast of Jakarta), as they will be translated into specific latitude and longitude values, via a distance and heading from a point calculation.

3.2 Soft-Data Extension of the Risk Management Framework

Vessels in the maritime environment are characterized by the set of risk features previously defined in the RMF [12]: (1) the *Collision Factor*, indicating the probability of a vessel colliding with another maritime object; (2) the *Sea State*, capturing the risk posed by the surrounding weather conditions at sea; (3) the *Regional Hostility Metric*, indicating the degree of hostility of the region a vessel is navigating through; and (4) the *Degree of*

[13] http://www.geonames.org/

Distress, a multifarious feature encompassing different distress factors such as the environmental impact of a potential catastrophe involving the vessel, the threat to the human lives aboard the vessel, etc. As they have been previously defined, all four risk features are fully characterized by hard data sources.

This study extends the hard-source-based characterization methodology with two real-world soft maritime incident soft sources (i.e., WWTTS and IMB). Figure 6 presents the RMF's components pertaining to L2 fusion processes.

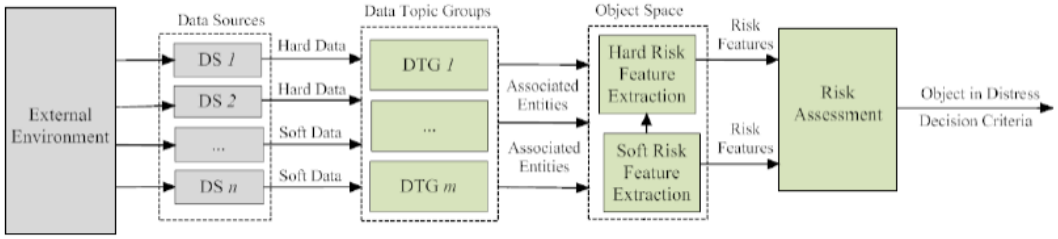


Figure 6: The RMF's SA components, adapted from [12]

3.2.1 Data Topic Groups

This study proposes a grouping of data sources in the system according to the type of information they provide. For instance, WWTTS and IMB provide incident reports of very similar nature, and thus belong to the same group, namely, the Maritime Incidents group. A detailed diagram of the DTG subcomponent is presented in Figure 7. This subcomponent takes multiple data sources as input and is responsible for providing all the information available on unique entities described by the input data sources.

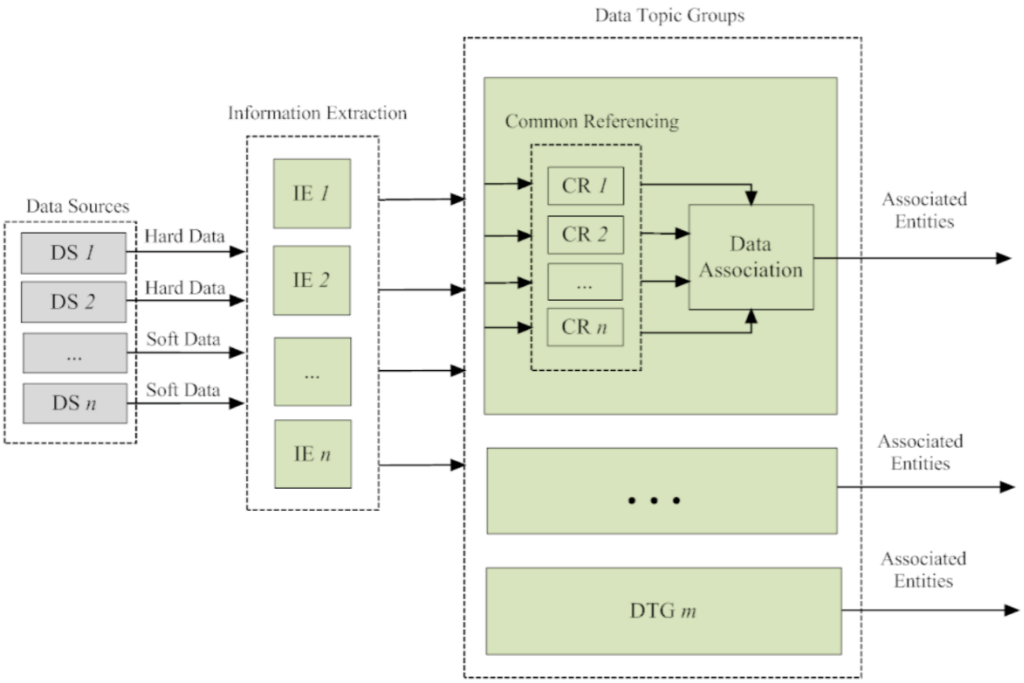


Figure 7: DTG Module

The DTGs and their respective data sources currently present in the system are shown in Table 1. Note that it is also possible for a single data source to belong to multiple DTGs, if that source provides relevant information to each of those groups.

Table 1: Data topic groups

Data Topic Group	**Data Sources**
Contact-Based (CB)	AIS
Weather Conditions (WC)	Sea State Reports
Geographical Databases (GD)	GeoNames Geographical Database
Maritime Incidents (MI)	WWTTS, IMB

The CB DTG could involve hard data sources such as AIS, radar, and electro-optical, but could also contain soft data sources, such as vessel report sightings provided by people. For the purposes of this study, the CB DTG was constrained to the AIS source. The IE submodule for the CB DTG is responsible for processing the raw dynamic vessel information, and extracting the vessel type, and its LKP.

The WC DTG could involve hard data sources such as those generated by buoys deployed around the world, measuring and quantifying local environmental characteristics (e.g., wave height), or could even include qualitative descriptions of the physical terrain (provided by human reporters). For the purposes of this study, the data source used in the WC DTG was simulated, as it was not possible to obtain access to a real-world one. The IE submodule for this DTG is responsible for extracting the quantitative or qualitative descriptions of the physical terrain.

The GD DTG contains data sources that provide geographical location information about places. The GD data source used as part of this study was the GeoNames Geographical Database. The IE submodule for this DTG is tasked with extracting the geographical name and place coordinate pairs.

The MI DTG is responsible for grouping together soft data sources containing maritime incident information. The IE submodule for this DTG is responsible for processing these raw textual incident and response reports, and extracting information pertaining to the different categories of interest (e.g., vessel type, incident keywords, incident time, incident location). More formally, this DTG processes a textual response

report, and produces the set, *R*, containing relevant response information extracted from the report:

$$\boldsymbol{R} = \{\boldsymbol{vesselType}, \boldsymbol{K}, \boldsymbol{incidentTime}, \boldsymbol{incidentLocation}\} \quad \textbf{(1)}$$

where K is the set of incident keywords discovered in the report.

The MI IE submodule makes use of the Natural Language Processing (NLP) technique called Named-Entity Recognition (NER) in order to extract information from the textual incident report and construct the set *R*. There is a lexicon constructed for each of the elements in the set *R*.

3.2.2 Risk Features

Two of the above risk features, namely the *Degree of Distress* and the *Regional Hostility Metric* from [12], are augmented to include risk factors derived from the data sources belonging in the MI DTG. The conceptual definitions of the two risk features remain as previously stated in Section 3.2.

3.2.2.1 Regional Hostility Metric

The *Regional Hostility Metric* of vessel *x*, *RHM(x)*, is formally defined as:

$$RHM(x) = w_{MIP}\, MIP(x) + w_{MIS}\, MIS(x) + w_{VPI}\, VPI(x) \quad \textbf{(2)}$$

where w_{MIP}, w_{MIS}, and w_{VPI} are used to respectively weigh: (a) the significance of vessel *x*'s proximity to known maritime incidents, (b) the importance of the severity of the incidents around the vicinity of the vessel, and (c) the pertinence of the nearby incidents to the vessel.

A suggested weighting is w_{MIP} = 0.4, w_{MIS} = 0.3, and w_{VPI} = 0.3; however, these weights could vary according to the maritime operator's preferences.

The *Mean Incident Proximity*, *MIP(x)*, computes the total risk posed by the proximity of the vessel to the reported incidents in its spatial surroundings; it is formally defined as the average of the risk values associated with the proximity of the *n* closest incidents:

$$MIP(x) = \frac{1}{n}\sum_{i=1}^{n}\theta(dist(x.loc, i.loc)) \quad \textbf{(3)}$$

where *x.loc* represents the geographical location of vessel *x*, *i.loc* represents the geographical location of incident *i*, *dist()* is a function which calculates the geographical distance (in km) between two locations, and Θ represents the *incident proximity risk*, which is a fuzzy set over the domain of distance values. The trapezoidal fuzzy membership function that models this set (according to the vessel's speed) adopts the following parametric configurations: (A = 0, B = 0, C = 5, D = 10) for slow-speed vessels; (A = 0, B = 0, C = 21.6, D = 43.2) for medium-speed vessels; and (A = 0, B = 0, C = 40, D = 80) for fast-speed vessels. These values emerged after consultations with a subject matter expert and may change depending on the region of interest.

The *Mean Incident Severity*, *MIS(x)*, computes the risk posed by the seriousness of the surrounding maritime incidents; it is formalized as the average of the risks induced by the severity of each of the *n* closest incidents:

$$MIS(x) = \frac{1}{n}\sum_{i=1}^{n}\psi(i.type) \quad \textbf{(4)}$$

where *i.type* denotes the type of incident *i*, and $\psi()$ is a function named *incident severity risk*, which maps an incident type to a numeric severity level (from 0 to 1). An example severity mapping is shown in Table 2.

Table 2: Reported maritime incidents and their severity values

Incident Category	Incident Keywords	Severity Values
Bomb Threat	bombed	1.00
Terrorism	terrorist, terrorism	0.90
Hostage Scenario	hijacked, abducted, kidnapped, hostage, kidnapping	0.80
Damage to the Crew	fired, tied up with rope	0.70
Theft	robbed, attacked, robbers, criminals, robbery, theft, stole equipment	0.60
Invasion	boarded, clashed with, boarded the vessel, knives, invaded, trespasser	0.50
Near Invasion	attempted boarding, forced, crime, threat, surrender	0.40
Threatened	chased, threatened, threat, suspect, escape, blocking, risk	0.30
Approach	suspicious approach, suspiciously approached, approached	0.20
Crew Error	crashed, negligence	0.10
Unknown	other risks	0.05

The *Victim Pertinence Index*, *VPI(x)*, quantifies how relevant the surrounding *n* incidents are to vessel *x*; it is formally defined as the maximum similarity between the type of vessel *x* and the type of vessels involved in each of the surrounding *n* incidents.

$$VPI(x) = \max_i \{\delta(x.type, i.vesseltype)\} \quad \textbf{(5)}$$

where *x.type* denotes the type of vessel *x*, *i.vesseltype* denotes the type of vessel that was involved in incident *i*, and *δ(x.type, i.vesseltype)* is computed as follows:

$$\delta(x.type, i.vesseltype) = \begin{cases} 1 & \text{if } x.\text{type is the same as } i.\text{vesseltype} \\ 0.5 & \text{if } x.\text{type is similar to } i.\text{vesseltype} \\ 0 & \text{if } x.\text{type is unrelated to } i.\text{vesseltype} \end{cases} \quad \textbf{(6)}$$

Table 3 defines the similarity matrix for vessel categories while Table 4 maps a vessel type to its category.

Table 3: Similarity matrix for vessel categories

	Cargo Transport	**Tanker/ Industrial**	**Warship**	**Small Military Vessel**	**Small Transport/ Utility**
Cargo Transport	1.0	0.5	0.0	0.0	0.5
Tanker/Industrial	0.5	1.0	0.0	0.0	0.0
Warship	0.0	0.0	1.0	0.5	0.0
Small Military Vessel	0.0	0.0	0.5	1.0	0.5
Small Transport/ Utility	0.5	0.0	0.0	0.5	1.0

Table 4: Vessel types and their corresponding category

Cargo Transport	Tanker/Industrial	Warship	Small Military Vessel	Small Transport/Utility
Bulk carrier	Chemical tanker	warship	Coast guard boat	Fishing trawler
Car carrier	Heavy lift vessel		Naval patrol vessel	Japanese harpoonists
Cargo ship	Lpg tanker			Militant anti-whaling group
Carrier ship	Mother vessel			Skiff
Container ship	Oil tanker			Speed boat
Dry cargo vessel	Product tanker			Tug boat
General cargo ship	Tanker			Dredger
Livestock carrier				Fishing vessel
Lng carrier				vessel
Refrigerated cargo ship				
Merchant ship				

3.2.2.2 Degree of Distress

The *Degree of Distress* of vessel *x*, *DoD(x)*, is augmented from [12] with another distress factor, the *Risk of Attack*, that will be derived from textual information mined directly from the semi-structured maritime incident reports; it is formally defined as:

$$\mu_{DD}(X) = 0.3\mu_{RP}(x) + 0.2\mu_{RE}(x) + 0.2\mu_{RF}(x) + 0.3\mu_{RA}(x) \quad \textbf{(7)}$$

Where the integer values present a suggested weighing, and $\mu_{RP}(x)$, $\mu_{RE}(x)$, and $\mu_{RF}(x)$ remain as previously defined as in [12] and $\mu_{RA}(x)$ is the probability that vessel x will be attacked due to the category it belongs to (according to its vessel type). More formally, $\mu_{RA}(x)$is the square root of the conditional probability that *x.category* would be the subject of a maritime incident *I*:

$$\mu_{RA}(x) = \sqrt{P(x.\text{category}|I)} \quad \textbf{(8)}$$

3.3 Performance Evaluation

As was discussed in Section 2.4, quantifying the effectiveness of an IF system is still a burgeoning area in the fusion community. The prevalent theme in the IF quality evaluation literature is that there is no commonly agreed-upon method and set of quality measures that can be used to evaluate a general fusion system; this is largely because quality is dependent on user objectives in a specific context [7] [60] [39] [61].

Recent attempts to quantify the IFE were [62] and [68], where the proposed metric consists of three aspects: (1) the **information gain**, quantifying the aggregation level of content provided by the data sources; (2) the **information quality**, which is a measure of data characteristics such as timeliness and information confidence; and (3) the **robustness** defined as the ability of the system to cope with real-world variation [68].

This research proposes a generic risk-aware IFE metric, capable of quantifying the effectiveness of the HSIF process. Furthermore, an IFE evaluation from two perspectives is formulated – from an **instantaneous** and from a **gradual** perspective. The typical evaluation done in terms of comparing against ground truth values (computing Precision, Recall, and F-measure) provides an overall ("averaged") performance assessment;

however, it does not allow for an understanding of the instantaneous fusion quality of a system. In a complex and dynamic environment, where decision makers need to act promptly and decisively, gauging the instantaneous effectiveness of the fusion system becomes of vital importance; as such, the metric is formulated in a way which does not require knowledge of ground truth.

The goal of an *Instantaneous IFE* (i-IFE) metric is to permit for the evaluation of the system effectiveness at any given point in time. This is an imperative feature for an IF system to possess, as it provides real-time insights into the effect an unfolding scenario has on the IF system.

A *Gradual IFE* (g-IFE) metric, on the other hand, is meant to efficiently and accurately express the IFE of the system over a period of time. The gradual metric, then, is an amalgamation of the effectiveness of the system over the set of scenarios which will unfold in the region the IF system is monitoring. Thus, a g-IFE evaluator is a crucial system component, as it can unlock the answer to questions such as: what group of data sources yields the highest merit for a particular region, or for a particular period of time. An advanced IF system will track its fusion effectiveness over the duration of its lifetime, and will be capable of automatically determining the answers to such questions, and reconfigure itself accordingly.

3.3.1 Instantaneous Information Fusion Effectiveness

As previously discussed, this research expands on the domain-agnostic *Information Gain* (IG) and *Information Quality* (IQ) measures proposed in [62] and [68], to formulate a risk-aware fusion effectiveness metric, whose building blocks are presented in Figure 8.

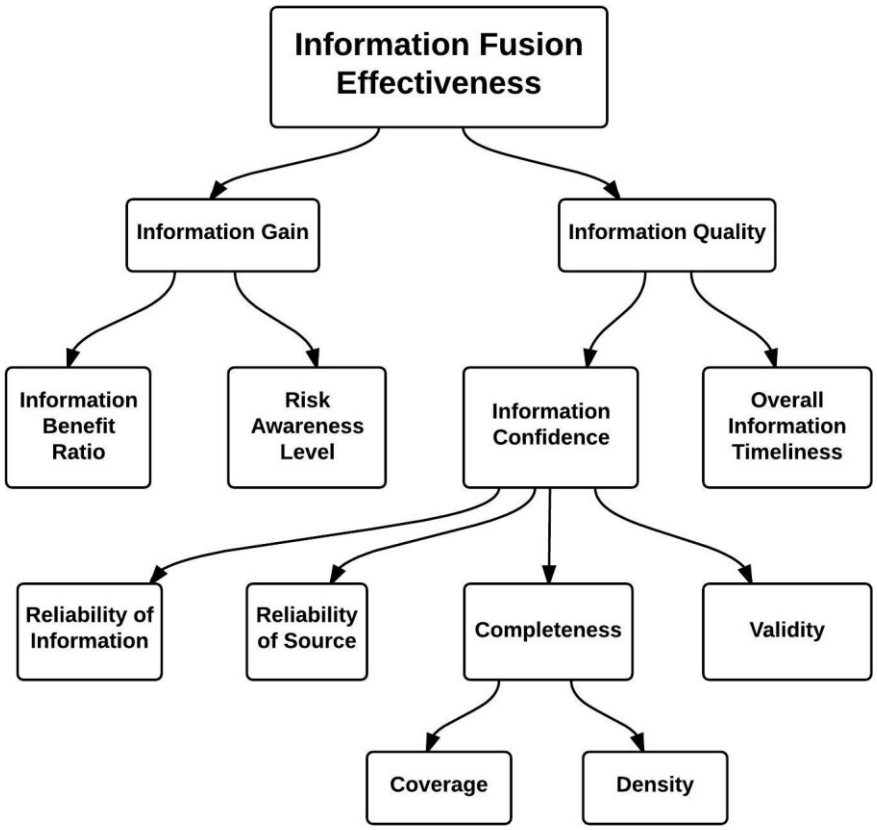


Figure 8: Instantaneous information fusion effectiveness (i-IFE) hierarchy

From a top-level perspective, the *i-IFE* metric is calculated as the product of IG and IQ; furthermore, $i\text{-}IFE \in [0, 1]$.

$$i\text{-}IFE = IG * IQ \tag{9}$$

Table 5 presents commonly used symbols in the equations from the subsequent sections.

Table 5: Notations for i-IFE

Symbol	Definition
S	a maritime scenario; a static snapshot (of configurable time length) of the assets present in a particular region
$S_{endtime}$	denotes a scenario's end time
x	a maritime vessel
i	a maritime incident
$NR_n(x.loc)$	the set of nearest n incident reports to vessel x
$NR_r(x.loc)$	the set of nearest relevant incident reports to vessel x
$NW_n(x.loc)$	the set of nearest weather condition measurements to vessel x
$I_{NG}(x)$	the set of nearest n incidents which report only general locations, or more formally: $I_{NG}(x) = \{ i \in NR_n(x.loc) : i.locSpecific = null\}$
A	the set of attributes across all data sources belonging to a particular data topic group that are contributing to the fusion process
O	the set of distinct objects present in the data sources which belong to a particular data topic group
W	the set of all possible objects (in the environment being monitored)
$DTGR$	data topic group relevance – a weight meant to quantify the relevance of the DTG to the fusion objectives.

3.3.1.1 Information Gain

IG is defined as a measure of the level to which the system is benefitting from the data sources in all of the DTGs, and the extent to which the system is aware of the risk present in the scenario. IG is thus formalized as the weighted sum of the Information Benefit Ratio (IBR) and the Risk Awareness Level (RAL):

$$IG = w_{IBR} IBR + w_{RAL} RAL \tag{10}$$

The sum of the weights should be unity. In this study, the weights were chosen as wIBR = wRAL = 0.5; however, these weights are parameterizable, in case the Subject Matter Expert (SME) decides that one of the two quality measures carries a greater influence over the system effectiveness.

3.3.1.1.1 Information Benefit Ratio

The IBR is a measure of the number of system units (vessels) that are making use of the information present in a particular DTG; IBR is calculated for each of the DTGs. The IBR of the system is the weighted sum of the IBR values for each of the DTGs. The semantics behind the weights, called DTGRs, are defined in Table 5.

$$IBR_{System} = \sum_{i \in DTG} DTGR_i IBR_i \tag{11}$$

IBR for the **Contact-Based** DTG is defined as the ratio of the number of vessels that have received contact information from one of the data sources belonging to that group to the total number of vessels in the scenario S. More formally:

$$IBR_{CB} = \frac{1}{|S|}|\{x \in S : x.CB \neq \emptyset\}| \quad \textbf{(12)}$$

where *x.CB* holds the set of contacts reported by vessel *x* during the scenario, and $|S|$ represents the cardinality of the set *S*.

IBR for the **Maritime Incident** DTG is governed by the Law of Diminishing Returns[14] (LDR). A vessel is considered to have benefitted from an incident only if that incident is within the proximity of the vessel, and if the type of the vessel involved in the incident has a similarity greater than zero (as defined in Table 3). Furthermore, a vessel has a higher overall information value benefits from being aware of more relevant incidents in its proximity; however, the difference in information benefit (from a risk perspective) of being aware between zero and one incident is greater than the difference between being aware of, for instance, 49 and 50 incidents (LDR being a driving force). To successfully model these characteristics, a series having the base form of $f(n) = \sum_{i=1}^{n} \left(\frac{1}{2}\right)^i$ was selected. It is important to note that this series converges: $\lim_{n\to\infty} f(n) = 1$. IBR_{MI} is furthermore averaged over all the vessels in the scenario, and is formally defined as:

$$IBR_{MI} = \frac{1}{|S|}\sum_{x \in S} B(x) \quad \textbf{(13)}$$

[14] http://www.britannica.com/topic/diminishing-returns

where:

$$B(x) = \begin{cases} \sum_{i=1}^{|NR_r(x.loc)|} (\frac{1}{2})^i & \text{if } |NR_r(x.loc)| > 0 \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(14)}$$

IBR for the **Weather Conditions** DTG is the ratio of the number of vessels that had access to WC information in their proximity to all the vessels in the scenario. More formally:

$$IBR_{WC} = \frac{1}{|S|} |\{x \in S : NW(x.loc) \neq \emptyset\}| \quad \textbf{(15)}$$

IBR for the **Geographical Databases** DTG is a measure of the number of vessels in the scenario with maritime incidents in their proximity that only contained general locations, and for which specific coordinate locations were successfully extracted from the data sources belonging to the GD DTG. More formally:

$$\textbf{(16)} \quad IBR_{GD} = \frac{1}{|S|} \sum_{x \in S} \frac{1}{|I_{NG}(x)|} \sum_{i \in I_{NG}} \Omega(i)$$

where:

$$\Omega(i) = \begin{cases} 1 & \text{if } |\{p \in GD : p.name = i.locGeneral.name\}| \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(17)}$$

3.3.1.1.2 Risk Awareness Level

The RAL metric represents how aware scenario-based assets are of the overall risk in terms of the multiple risk features (e.g., collision factor, degree of distress) each asset is able to compute.

$$RAL = \frac{1}{|S|} \sum_{x \in S} \frac{|\{r \in R : x.r \neq null\}}{|R|} \quad \textbf{(18)}$$

where R is the set of risk features.

Note that a risk factor is *null* if its associated data sources did not provide relevant information within the scenario of interest. RAL is not calculated for every one of the DTGs; instead, there is a single RAL per scenario.

3.3.1.2 Information Quality

IQ is judged in the system based on the Information Confidence (IC) and Overall Information Timeliness (OIT) quality measures; it is formally defined as their weighted sum:

$$IQ = w_{IC} IC + w_{OIT} OIT \quad \textbf{(19)}$$

The sum of all the weights must be unity; for this study, these weights were set as $w_{IC} = w_{OIT} = 0.5$, but they are left parameterizable to be set according to the operator's preference.

3.3.1.2.1 Information Confidence

IC is based on several information quality measures, namely *Reliability of Information* (ROI), *Reliability of Source* (ROS), *Validity* (V), and *Completeness* (C); it is formally defined as their weighted sum:

$$IC = w_C C + w_V V + w_{ROI} ROI + w_{ROS} ROS \quad \textbf{(20)}$$

The sum of all the weights must be unity. For this research, the weights were all equal: $w_C = w_V = w_{ROI} = w_{ROS} = ¼$, however, as with Equations (10) and (19), this is kept parameterizable according to the operator's preferences.

A. Reliability

ROS and ROI values have been defined according to the US Military Standard 640 (MIL-STD-640) [62], which provides quantitative values for the otherwise qualitative descriptions of the reliability levels of information and source of information found in the NATO Standardized Agreement for Intelligence Reports – STANAG 2022. The need for assessing ROS and ROI separately stems from the fact that information's quality is instantaneous, whereas a source's quality is enduring (i.e., based on previous experience with that information source) [7]. An unreliable source may provide a reliable piece of information at a certain point in time; the converse is also true.

B. Validity

Validity is defined as the "degree of truthfulness" in information [38]. In the RMF, this degree will be judged by adherence to the expected format of the attributes of the data in a particular DTG:

$$V_{DTG} = \frac{1}{|A|} \sum_{a \in A} \frac{\sum_{o \in O, o[a] \neq null} FC(o[a])}{|\{o \in O : o[a] \neq null\}|} \quad \textbf{(21)}$$

where the attributes for the CB and MI DTGs are $A_{CB} = A_{MI} = \{location, time, vessel\ type\}$, the attributes for WC are $A_{WC} = \{location, time\}$, and the attribute for the GD DTG is $A_{GD} = \{location\}$. FC refers to format check, which returns unity if the attribute is in the correct format, or zero otherwise. The format check of the coordinate returns unity if they are over water, or zero if they are over land.

The overall Validity measure of the system is the weighted sum of the individual Validity measures:

$$V_{System} = \sum_{i \in DTG} DTGR_i V_i \tag{22}$$

C. Completeness

The Completeness of a data source is broadly defined as the "proportion of missing values" [60]. The authors in [69] expand on this notion by taking into account the total information present in the real world. Completeness of a source then becomes a construct which describes the amount of information provided by the source as compared to the total amount of information present in the world (domain) which the source is describing. The formalization is accomplished by introducing two dimensions of Completeness, namely Coverage and Density.

The Completeness of a DTG (C_{DTG}) is formalized as the product of the Coverage of the DTG (Cv_{DTG}) and the Density of the DTG (d_{DTG}):

$$C_{DTG} = d_{DTG} * Cv_{DTG} \tag{23}$$

The overall Completeness measure of the system is the weighted sum of the individual DTG Completeness measures:

$$C_{System} = \sum_{i \in DTG} DTGR_i C_i \tag{24}$$

a. Coverage

The Coverage of an information source is defined as the ratio of the number of objects present in the source to the total number of possible objects in the world. This notion is expanded in the system by applying the Coverage to a DTG (i.e., the Coverage is a measure applied to all the data sources belonging to the DTG):

$$Cv_{DTG} = \frac{|O|}{|W|} \tag{25}$$

where *O* and *W* remain as previously defined in Table 5.

The authors in [69] point out that it is not always possible to have explicit knowledge of *W*, and therefore, in such a case, the coverage of a source can be only estimated by a domain expert (an SME).

b. Density

The Density of an information source quantifies how much data the source provides for each of the objects described within it; it is thus a measure of the number of non-null attributes the source provides. The notion of Density is expanded within this research by applying it to a DTG. More formally:

$$d_{DTG} = \frac{1}{|A|} \sum_{a \in A} \frac{|\{o \in O : o[a] \neq null\}|}{|O|} \tag{26}$$

For the data sources belonging to the CB DTG, the system is only concerned with location, time stamp, vessel type, and its speed: A_{CB} = *{location, time, vessel type, speed}*. For the data sources belonging to the MI DTG, the system ingests the location (can be general or specific), the time, the type of vessel, and the incident type: A_{MI} = *{location, time, vessel type, incident type}*. For the data sources belonging to the GD DTG, the attributes used are the coordinate location and the place names: A_{GD} = *{location, place name}*. Finally, for the data sources belonging to the WC DTG, the attributes of importance are the coordinate location and the time stamp: A_{WC} = *{location, time stamp}*.

3.3.1.2.2 Overall Information Timeliness

Overall Information Timeliness (OIT) is modeled using a Mamdani Fuzzy Inference System (FIS) and is a function of the Information Timeliness (IT) of the different DTGs.

The IT of the CB DTG, IT_{CB}, can be expressed as the average, across all vessels, of the average delay (in seconds) between consecutive contact reports, and between the latest contact report and the end time of the scenario, from the data sources in the DTG for each vessel:

$$IT_{CB} = \begin{cases} \frac{1}{|S|} \sum_{x \in S} \min(IT(x), \tau_{CB}) & \text{if } |S| > 0 \\ \infty & \text{otherwise} \end{cases} \quad \textbf{(27)}$$

where τ_{CB} is the CB's *temporal staleness point* and represents the time past which the timeliness (from a relevance perspective) of the information provided by a single vessel is considered to be fully outdated, and *IT(x)* is defined as:

$$\textbf{(28)} \quad IT(x) = \begin{cases} \frac{1}{n}(S_{endtime} - x.t[1]) & \text{if } n > 0 \\ \infty & \text{otherwise} \end{cases}$$

where t is the sorted collection of time stamps reported by vessel x in the scenario, $x.t[1]$ represents the first time stamp, and n is the number of time stamps reported by that same vessel.

FIT_{CB} is the fuzzy variable whose crisp input is IT_{CB}, and has the linguistic terms CB_{Recent} and CB_{Old}.

The IT of the MI DTG, IT_{MI}, is the average, across all vessels, of the average age (in days) of each vessel's n closest maritime incident reports to the scenario's end time, drawn from data sources belonging to the DTG:

$$IT_{MI} = \begin{cases} \frac{1}{|S|} \sum_{x \in S} \min(IT_{NR_n}(x), \tau_{MI}) & \text{if } |S| > 0 \\ \infty & \text{otherwise} \end{cases} \quad \textbf{(29)}$$

where τ_{MI} is the MI's temporal staleness point, and $IT_{NRn}(x)$ represents the timeliness of the n nearest incident reports to vessel x, from the scenario's end time, and is formally defined as:

$$IT_{NR_n}(x) = \begin{cases} \frac{1}{|NR_n(x.loc)|} \sum_{i \in NR_n(x.loc)} (S_{endtime} - i.t) & \text{if } |NR_n(x.loc)| > 0 \\ \infty & \text{otherwise} \end{cases} \quad \textbf{(30)}$$

where $x.t$ represents the last reported time by vessel x, and $i.t$ represents the reported time of incident i.

FIT_{MI} is the fuzzy variable whose crisp input is IT_{MI}, and has the linguistic terms MI_{Recent} and MI_{Old}.

The IT of the WC DTG, IT_{WC}, is the average, across all vessels, of the age (in hours) of each vessel's closest weather condition reports drawn from the data sources in the DTG:

$$IT_{WC} = \begin{cases} \frac{1}{|S|} \sum_{x \in S} \min(IT_{NW_n}(x), \tau_{WC}) & \text{if } |S| > 0 \\ \infty & \text{otherwise} \end{cases} \quad \textbf{(31)}$$

where τ_{WC} is the WC's *temporal staleness point*, and $IT_{NWn}(x)$ represents the timeliness of the n nearest weather reports to vessel x and is formally defined as:

$$IT_{NW_n}(x) = \begin{cases} \frac{1}{|NW_n(x.loc)|} \sum_{w \in NW_n(x.loc)} (S_{endtime} - w.t) & \text{if } |NW_n(x.loc)| > 0 \\ \infty & \text{otherwise} \end{cases} \quad \textbf{(32)}$$

where $w.t$ represents the last reported time of the weather station w.

FIT_{WC} is the fuzzy variable whose crisp input is IT_{WC}, and has the linguistic terms WC_{Recent} and WC_{Old}.

OIT is then calculated as the center of gravity of the fuzzified output of the Mamdani FIS:

- if FIT_{CB} is CB_{Recent} and FIT_{MI} is MI_{Recent} and FIT_{WC} is WC_{Recent}

  then OIT is OIT_{Recent}
- if FIT_{CB} is CB_{Recent} and (FIT_{MI} is MI_{Old} or FIT_{WC} is WC_{Old}) then

  OIT is $OIT_{Acceptable}$
- if FIT_{CB} is CB_{Old} then OIT is OIT_{Old}

The set of membership functions associated with the OIT fuzzy variable are OIT_{Recent}, $OIT_{Acceptable}$, and OIT_{Old}.

Note that the GD DTG does not contribute to the OIT metric because the group only contains data sources with static (i.e., time-invariant) information. Also, FIT_{CB} is chosen to play a predominant role in the above rules due the fact that the CB DTG contains the information most vital to the system; the latter's capability to accurately detect risk vastly deteriorates when there is only very scarce and untimely CB-related data.

3.3.2 Gradual Information Fusion Effectiveness

The g-IFE metric is calculated as the trimmed, or truncated mean of the set of i-IFE values collected for each of the scenarios present in a region over a user-specified time period. The trimmed mean is chosen due to its robustness in relation to outliers [70]. Outlier i-IFE values (i.e., ones that are unusually high or low) are possible when the system first starts due to the data transmission rates (e.g., a scenario has very little CB data due to the CB transmission rates, but in reality, the region has many maritime entities); however, as the system remains active over a period of time, the knowledge of the entities which are present grows, and outlier i-IFEs are no longer expected.

$$E = \{i\text{-}IFE \in \mathcal{R} : i\text{-}IFE_n \leq i\text{-}IFE_{n+1}\} \tag{33}$$

$$g\text{-}IFE = \overline{E}\{0.20\} \tag{34}$$

where $\mathcal{R}$ is the region , E is the set of ordered *i-IFEs*, and $\bar{E}\{0.20\}$ is the 20% trimmed mean of E.

Since *i-IFEs* ϵ [0,1], from Equation (34) it follows that *g-IFEs* ϵ [0,1].

3.3.3 Robustness Analysis

As previously discussed, robustness is defined as the ability of a system to cope with real-word variations [68]. Within the proposed HSIF system, robustness is judged via *Information Loss Tolerance* (ILT) experiments and *Use Case Coverage* (UCC) analysis.

3.3.3.1 Information Loss Tolerance

An ILT analysis is done on the system by randomly dropping a defined percentage of entities (e.g., incident information, weather condition measurements) and measuring the new *g-IFE* values. These *g-IFE* values are then used to determine the effect this disturbance has on the performance of the system.

In order to determine the ILT level, it is necessary to determine the disturbance percentage the information loss has caused on the *g-IFE* metric:

$$\alpha = \frac{g\text{-}IFE(InfoLoss)}{g\text{-}IFE} - 1 \quad \textbf{(35)}$$

where α represents the percentage gain (+) or loss (-) relative to *g-IFE*. The optimal case is that α is equal to 0%, which occurs when the disturbance inserted into the system had no effect on the *g-IFE* metric. Note that as previously discussed, *g-IFE* and *g-IFE(InfoLoss)* are always positive values in the interval [0, 1].

The *ILT* can then be defined as follows:

$$ILT = \begin{cases} 1 & \text{if } |\alpha| \leq \beta \\ 1+\alpha & \text{if } \alpha < -\beta \text{ (if it was a loss under } \beta \text{ percent)} \\ 1-\alpha & \text{if } \beta < \alpha < 1 \text{ (if it was a gain between } \beta \text{ and } 100\%) \\ 0 & \text{otherwise} \end{cases} \quad \textbf{(36)}$$

where β represents a threshold percentage of allowed *g-IFE* metric variation. If the percentage disturbance caused to the *g-IFE* metric is under the threshold, then the metric is said to be tolerant of the information loss [suggested $\beta = 0.05$ (i.e., 5%)].

3.3.3.2 Use Case Coverage

The second aspect of robustness is UCC – the ratio of the total number of use cases in which the system can be run whilst providing useful L2 fusion capabilities (i.e., the core functionality is still available) to the total number of possible ways to run the system. If there are X information sources, then there are 2^X use cases (i.e., all possible combinations of the sources). The system's UCC then can be determined as the total number of cases (data source combinations) which would render the system useful divided by the total number of possible combinations:

$$UCC = \frac{\delta(2^D)}{|2^D|} \quad \textbf{(37)}$$

where D is the set of data sources used by the system, 2^D is the powerset of D, δ() is a function that identifies the number of useful elements in the powerset (in the RMF context – number of useful use cases).

3.3.4 Profitability Analysis

Profitability analysis can be conducted on the *g-IFE* metric by starting with a baseline configuration (e.g., system running with AIS data only) and calculating the *g-IFE* percent gain for another configuration (e.g., adding a new data source). This analysis will demonstrate the underlying merit provided by the additional sources.

$$\Pi_B(A) = \begin{cases} \frac{g\text{-}IFE(B,A)}{g\text{-}IFE(B)} - 1 & \text{if } g\text{-}IFE(B,A) \neq 0 \\ 0 & otherwise \end{cases} \quad \textbf{(38)}$$

where $\prod_B(A)$ denotes the profitability – *g-IFE* percent gain or loss – achieved by adding a set of sources *A* into the system, given a set, *B*, of baseline sources. Note that it is possible to have infinite profit, but not infinite loss, as *g-IFE* can never be negative.

3.4 Chapter Summary

This chapter proposed a new L2 SA soft-data augmented extension to a recently proposed RMF [12], for the purpose of proactively identifying vessels which may fall in distress. A new L2-based IFE metric is proposed for the purpose of evaluating instantaneous and gradual performance of the fusion process of the system. Further presented were two types of analyses, Robustness and Profitability, which can be performed on the basis of the IFE metric, for the purpose of: (1) gauging the system's ability to cope with real-world variation and (2) determining the merit provided by different data sources, respectively.

Chapter 4. Soft-Data Augmented Course of Action Generation

An integral part of L3 fusion is the automatic generation of suitable CoAs to tend (respond) to situations having been identified by the L2 modules of the system, typically with the intention to lower a risk or a hazard present in the environment being monitored by the system. This chapter proposes an architecture for a system capable of generating automatic CoAs for multiple, concurrently unfolding situations, through the guidance of soft data. This system integrates soft data, in the form of specific maritime incident and response reports, into its CoA generation process with the expectation that it will yield a higher chance of mission success, as judged by a mission-specific measure of performance, PCDRA, as defined in Section 4.2.5. This expectation arises from the fact that such data captures operational (e.g., incident response) mission knowledge, typically locked within the minds of SMEs. The ability of handling concurrent situations is expected to yield lower average asset utilization in missions (AU, also defined in Section 4.2.5) than when the same situations are being processed independently by the system, due to a more optimal asset-to-situation assignment.

Note that the term '*Response*' is used to signify the act of tending to one or many unfolding situations.

4.1 Environment and Data Sources

The maritime environment and available soft data sources remain as previously described in Sections 3.1 and 2.5.1, respectively. Whilst the aforementioned data sources supply sufficient L2-related information, they fail short in providing an adequate level of L3-related information. Details on mission operations, such as what actions were taken by coastal agencies to respond to the incident, are typically missing or undisclosed, and when

these details are present, they are quite vague in content. For instance, consider the incident response report captured in Figure 1. The coastal agency action taken is listed as "a warship and a helicopter came to the location". The relevant pieces of information which can be extracted from this data are that two assets of types *warship* and *helicopter* were dispatched; however, there is no mention about other key details, such as what search patterns, if any, these assets might have been executing in an attempt to locate the perpetrators (the pirates in this particular case), nor is there any way to gage what sort of characteristics these assets possess (in terms of their associated operational costs or speed capabilities, or what kind of onboard equipment each has that can be used to aid in the mission). These data sources also do not provide any description of the physical landscape (e.g., cloud coverage, rain intensity).

For this reason, it was deemed necessary to construct synthetic incident reports, which *augment* the original incident reports, by including more detailed operational information. These synthetic reports do not deviate from the real-world ones in any aspect, aside from the addition of information about: (1) what search patterns (as presented in Section 2.5.2) if any, the assets were instructed to execute; (2) unique asset identifiers, so that the system can be aware of the asset's characteristics (what specific onboard detection equipment the asset had, as well as the asset's associated speed and cost to move); and lastly, (3) a brief mention of the physical landscape of the response. More information on synthetic incident response reports can be found in Section 4.2.1.1.

4.2 System Description and Blueprint

The *Soft-Data-Driven Response Generation* (SDDRG) system is capable of performing its L3 duties through the aid of seven modules: (1) the *Anomaly Detection Module* (ADM), which is responsible for determining the confidence levels for different anomaly types (e.g., piracy events, vessels in distress) for each of the assets being monitored; (2) the *Situation Assessment Module* (SAM), which determines the most pressing situations the system will tend to; (3) the *Response Requirements Determination Module* (RRDM), which uses historical incident data to infer the response requirements based on the type of unfolding situations and the manner in which similar, previous (historical) situations were dealt with; (4) the *Asset Selection Module* (ASM), which is responsible for selecting the assets that will tend to the unfolding situations of interest; (5) the *Asset Path Generation Module* (APGM), which generates tracks for all the assets, based on their designated search areas and assigned search patterns; (6) the *Response Enactment Module* (REM), which is responsible for carrying out the response simulation; and lastly, (7) the *L3 Performance Assessment Module* (PAM), which tracks and calculates six performance metrics, according to which each potential response is judged. The architectural blueprint of the system is presented in Figure 9. This research focuses on the development of *L3 modules*, which are all shaded in *grey* within the blueprint.

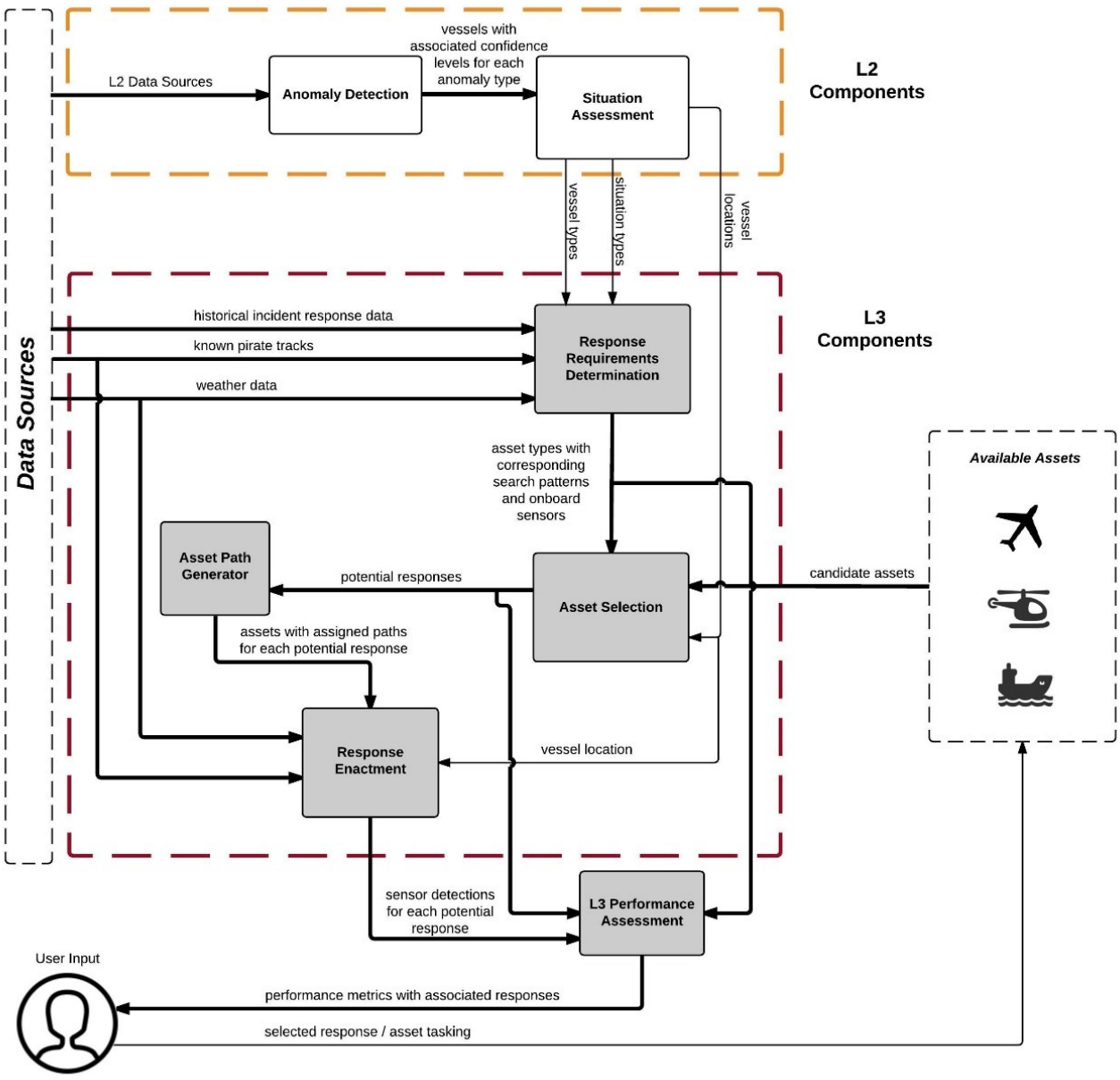


Figure 9: Soft-data-driven response generation system

Table 6 presents the inputs and outputs (IOs) for all the different modules of the SDDRG system, as related to L3 fusion.

Table 6: Inputs and outputs of the SDDRG system

IO Name	IO Description
Vessel locations	IO coming from the L2 SA module. It specifies the last known positions of the vessels that are involved in the most pressing (as judge by a threshold value) situations that the system has to deal with.
Vessel types	IO coming from the L2 SA module. It specifies the type of vessels that are involved in the most pressing situations that the system has to deal with.
Situation types	IO coming from the L2 SA module. It specifies the type of situations the system shall deal with.
Asset types with corresponding search patterns and onboard sensors	IO determined based on the historical response data. The most similar scenarios to the currently unfolding situations are located in the response data; similarity is based on: (1) the currently unfolding situation types, (2) the vessels involved, and (3) the prevailing (current) weather conditions. The RRDM module then takes these historical scenarios, and outputs the asset types, assigned search patterns, and onboard sensors that partook in the historical missions.
Candidate response assets	Assets that are candidates to be used within a response. This includes coast guard assets, auxiliary coast guard assets, as well as opportunistic platforms (i.e., ships or aircraft which are in the vicinity and able to provide assistance, based on their capabilities).
Potential responses	A collection of potential responses, which can be enacted to tend to the current situations. Each response has a collection of assets that have been selected to participate. Each asset has a designated search area to cover, with a specific search pattern to be executed within the search area. Each asset also has an assigned sensor with a particular quality level.
Assets with assigned paths for each potential response	A search path (a collection of sequential latitude, longitude, and altitude values) is generated for each asset participating in a response.

Sensor detections for each potential response	A statistical measure provided for each response. It quantifies the likelihood of the assets detecting the VaL.

4.2.1 Response Requirements Determination Module

The RRDM uses characteristics of the currently unfolding situations and analyses the soft, historical incident response data to determine the optimal asset requirements for the mission to be carried out. The RRDM contains two submodules – the *Soft IE Submodule* and the *Case-Based Reasoning* (CBR) *Submodule*, each of which are discussed in the subsequent sections.

4.2.1.1 Soft IE Submodule

As with Section 3.2.1, the soft IE submodule is responsible for processing the raw textual incident and response reports, and extracting information pertaining to the different categories of interest through the NER technique. The difference from the previous formulation in Equation (1) is that there are two additional categories of interest, capturing L3 information:

$$R' = \{R, W, A\} \tag{39}$$

where R remains as previously defined in Equation (1), repeated here for convenience:

$$R = \{vesselType, K, incidentTime, incidentLocation\}$$

A is the superset containing the sets of tuples of assets along with their assigned search patterns, sensor types, and sensor qualities; and W is the set containing the weather-related information surrounding the response:

$$A = \{\{asset_1, searchpattern_1, sensortype_1, sensorquality_1\},$$
$$\{asset_2, searchpattern_2, sensortype_2, sensorquality_2\}, \ldots,$$
$$\{asset_n, searchpattern_n, sensortype_n, sensorquality_n\}\} \quad (40)$$

$$W = \{cloudDensity, rainDensity\} \quad (41)$$

The elements in set A that are not immediately present in the soft report (i.e., the sensor type and sensor quality) are retrieved based on the unique asset identified as being present in the report.

The elements of the set W are all numerical values between 0 and 1, inclusively. The intensity and density values selected to quantify the qualitative descriptions are configurable by the human operator. Table 7 presents one possible configuration.

Table 7: Weather conditions and their associated intensity and density values

Weather Category	Keywords	Intensity/Density Value
Rain	Extreme rain	1.00
	Heavy rain	0.75
	Moderate rain	0.50
	Light rain	0.25
	No rain, clear skies	0.00
Clouds	Extremely dense clouds	1.00
	Dense clouds	0.75
	Moderate clouds	0.50
	Light clouds	0.25

	No clouds, clear skies	0.00

Figure 10 presents an example incident response report with the different elements belonging to R' highlighted.

While underway a bulk carrier ceased transmitting AIS information at approximately 100 nautical miles southeast of St John's, Canada. After an unsuccessful attempt to contact the vessel crew, a response mission was launched. Helicopter-3 was assigned a square in search pattern. FastUAV-2 was assigned a parallel track line search pattern. SlowUAV-3 was assigned a square in search pattern. Aircraft-3 was assigned a parallel track line search pattern. Tugboat-3-A was assigned a square out pattern. Tugboat-3-B was assigned a square out pattern. Tugboat-3-C was assigned a square out pattern. Lastly, Speedboat-3 was assigned a parallel track line search pattern. At the time of the incident, moderate rain present with extremely dense clouds were in the response region.

Figure 10: Example synthetic incident report with L3 information

After the soft IE submodule processes this report, it will have the following form:

R = *{bulk carrier, ceased transmitting AIS, 100 NM south east St John's Canada, incident time not found}*

R' = *{R, {{Helicopter-3, square in, SAR sensor, very high quality}, ... {Speedboat-3, parallel track line, MSTAR sensor, very high quality}}, {moderate rain, extremely dense clouds}}*

4.2.1.2 Case-Based Reasoning Submodule

The RRDM employs a case-based reasoning technique, Nearest Neighbor (NN), in order to locate the most similar scenario in the historical response data. The dissimilarity between the currently unfolding situation, *S*, and a response report, *R*, is calculated using a modified Euclidean Distance function that takes categorical data into account:

$$D(S,R) = \sqrt{\delta s^2(S,R.s) + \delta v^2(S.v,R.v) + \sum_{w \in W}(w_S - w_R)^2} \quad (42)$$

where *W* remains as previously defined; *R.s* represents the situation being described in report *R*; *S.v* and *R.v* represent the vessel type in situation *S* and report *R,* respectively.

Furthermore, *δs()* calculates the dissimilarity between two different situations, and *δv()* calculates the dissimilarity between two different vessel types; more formally:

$$\delta s(S,R) = 1 - \Omega s(S,R) \quad (43)$$

$$\delta v(S,R) = 1 - \Omega v(S,R) \quad (44)$$

where Ω*s()* calculates the similarity degree between two different situations, and is derived from Table 8. Ω*v* calculates the similarity between two different vessel types, and is derived from Table 3 and Table 4.

Table 8: Similarity matrix for situations

	Piracy	**VID**
Piracy	1.0	0.5
VID	0.5	1.0

4.2.2 Asset Selection Module

The ASM is responsible for selecting which response assets will be carrying out the mission, based on the mission requirements (which define what specific types of assets are required for each situation, as well as their associated onboard sensors). The module provides a designated search area for each asset. Asset search area designation is optimized with the popular Non-dominated Sorting Genetic Algorithm II (NSGA-II) [71]. This algorithm is used for EMOO to produce a set of spread non-dominated candidate solutions with varying degrees of mission latency, cost, response area gap sizes, and levels to which they meet mission requirements.

4.2.2.1 Response Encoding

Whenever an incident is detected, a response grid pertaining to that incident is generated and broken down into a square grid. Each cell is a square, which is entirely enclosed within the sweep area of the smallest-sweeping sensor of any of the selected assets. Figure 11 presents an example response grid, with three (coloured) subgrids, each of which is assigned to a response asset. The grid also shows two continuous gaps in coverage (in white).

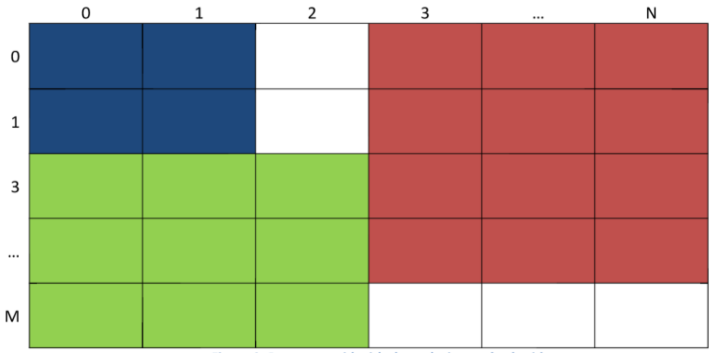


Figure 11: Example response grid with three designated subgrids

Responses are encoded as four-layer chromosomes. Each gene represents an asset, which is available to engage in the response. The first layer codifies an inclusion/exclusion bit on the gene. The second layer encodes which response grid the asset will be exploring, as there can be multiple response grids (in the case where multiple, concurrent situations are unfolding). The third layer encodes the designated subgrid (within the selected response grid in the second layer) for that asset, by encoding the row and column indices of the top left corner location of the response grid, as well as the length and width of its designated subgrid. The fourth layer encodes the type of search pattern the asset will have to execute.

The types of search patterns that an asset can execute are derived from the historical incident response data. Assets already have specific sensors (specific types and qualities) that are mounted on them; thus, the sensor details are *not* part of the encoding (assets cannot swap equipment between one another), but are instead used as part of the calculation of the *Mission Requirements* objective function presented in Section 4.2.2.2. The chromosome encoding of a candidate response (mission) is presented in Table 9.

Table 9: Chromosome encoding of a candidate response

Asset	**A_1**	**A_2**	**A_3**	**A_4**	**…**	**A_N**
Inclusion / Exclusion	Include	Exclude	Include	Include	…	Include
Designated Search Area	1	1	2	1	…	3
Designated Subgrid	<0,0,2,3>	<5,1,4,8>	<3,0,5,4>	<8,8,1,3>	…	<6,2,3,3>
Designated Search Pattern	Parallel Track Line	Parallel Track Line	Track Crawl	Outwards Expanding Square	…	Track Crawl

In the case that no soft data is available, the designated search pattern defaults to an adhoc pattern. This pattern is generated according to the algorithm presented in Figure 12. The adhoc search pattern starts off by selecting the upper left corner cell of the response subgrid, and proceeds by selecting a random neighbour cell. The selection process will give precedence to unexplored neighbouring cells; however, if no unexplored ones exist, it will choose one that has already been visited.

```
cell = designatedAssetSubgrid.upperLeftCornerCell
for i = 1 to asset.numCellsInSubArea
  pattern.add(cell)
  newCell = getRndUnexploredNeighbour(cell)
  if newCell = null
    newCell = getRndNeighbour(cell)
  end if
  cell = newCell
end for
return pattern
```

Figure 12: Adhoc search pattern generation algorithm

4.2.2.2 Response Objectives

The NSGA-II optimizer makes use of the four objective functions. The first three, which it aims to minimize, are: (1) the mission time (MT), which is determined by calculating how long it would take each asset to get to and traverse its designated search area; (2) the mission expenses (ME), which is calculated by how much cost is incurred by each asset as it travels to and subsequently traverses its designated search area; and (3), the unexplored search area (USA), which quantifies the percentage of the search area which remains unexplored. Lastly, the final objective function, mission requirements (MR), is

attempted to be maximized and determined by calculating the level to which the requirements are met by comparing the number and type of required assets (as extracted from the soft data) to the number and type of ones that have been selected to participate in the potential response.

More formally, the four objective functions are defined as:

$$MT = \sum_{a \in A} (T_{a.loc,a.subgrid.start} + T_{a.subgrid.start,a.subgrid.end}) \tag{45}$$

$$ME = \sum_{a \in A} (E_{a.loc,a.subgrid.start} + E_{a.subgrid.start,a.subgrid.end}) \tag{46}$$

$$USA = 1 - \frac{|\cup_{a \in A} a.gridCells|}{SearchGrid.numGridCells} \tag{47}$$

$$MR = \sum_{a \in A} \beta(\alpha) \tag{48}$$

where A is the set of selected response assets; $E_{a.loc,a.subgrid.start}$ and $T_{a.loc,a.subgrid.start}$ represent respectively the expenses that would accumulate and the time it would take asset a to move from its current location to the starting point of its assigned response subgrid; $E_{a.subgrid.start,a.subgrid.end}$ and $T_{a.subgrid.start,a.subgrid.end}$ represent respectively the expenses that would accumulate and the time it would take asset a to move from the starting location to the ending location of its subgrid;

a.gridCells is the set of grid cells belonging to the search grid that asset *a* is responsible for visiting, ∪ is the union of such sets (for all selected assets); *SearchGrid.numGridCells* returns the total number of grid cells that the search grid contains; and $\beta(\alpha)$ is a function which returns 1 when the selected asset satisfies a selection requirement, and 0 otherwise. Satisfying a selection requirement entails the asset: (1) being of a required type (e.g., aircraft), and (2) having an onboard sensor of equivalent or higher quality than the required quality.

If the system is running with historical response reports, NSGA-II will use all of the above objective functions. However, if the system is running without any historical response reports, the genetic optimizer ignores *MR*.

4.2.2.3 Evolving Operators

There are two types of evolving operators used in this study – a custom mutation operator and a standard cross-over operator. The standard cross-over operator works by randomly selecting a cross-over point and swapping the information in the four layers of each parent chromosome before and after the point, thereby generating two offspring. The custom mutation operator ensures that each layer of the gene is mutated based on an input probability, as presented in Figure 13. The mutation operator is also responsible for ensuring that if the asset's subgrid boundaries extend beyond those of the full search grid (as a result of the mutation), it will trim the subgrid to fall entirely within the search grid.

comments: `rnd.nextDouble()` returns a uniformly distributed real number in the range of 0 to 1

`rnd.nextInt(N)` returns a uniformly distributed integer ranging from 0 to N-1

`getPatternsForAsset(assetType, subgrid)` returns a list of search patterns (if any) which were used for this asset type in the closest historical soft incident report (for the situation described in the provided subgrid)

`checkGridBoundaries(gene, subgrid)` checks whether the designated subgrid of the asset, as designated by the row, column, and their respective offsets falls fully within the boundaries of the search grid; if it does not, the offsets are trimmed to guarantee that the asset's subgrid does not extend beyond the search grid

---

inputs:

chromosome	`c`,
random number generator	`rnd`,
mutation probability	`p`,
search grids	`grids`

```
for each gene g in c
      if (rnd.nextDouble() <= p)
            g.flipInclusionBit()
      end if
      if (rnd.nextDouble() <= p)
            s = grids.get(rnd.nextInt(grids.length))
      end if
      if (rnd.nextDouble() <= p)
            g.rowIndex = rnd.nextInt(s.numberOfRows)
      end if
      if (rnd.nextDouble() <= p)
            g.colIndex = rnd.nextInt(s.numberOfCols)
      end if
      if (rnd.nextDouble() <= p)
            rangeToMaxRow = s.numberOfRows = g.rowIndex
            g.rowOffset = rnd.nextInt(1 + rangeToMaxRow)
      end if
      if (rnd.nextDouble() <= p)
            rangeToMaxCol = s.numberOfCols = g.colIndex
            g.colOffset = rnd.nextInt(1 + rangeToMaxCol)
      end if
      if (rnd.nextDouble() <= p)
            patterns = getPatternsForAsset(g.assetType, s)
            g.searchPattern = patterns.get(rnd.nextInt(patterns.length))
      end if
      checkGridBoundaries(g, s)
end for
```

Figure 13: Custom Mutation Operator

4.2.3 Asset Path Generation Module

The APGM is a small module whose sole responsibility is to create actual search paths for each asset. It has knowledge of the real-world geographical area that the NSGA-II search grid covers; based on the collection of grid cells that each response asset has to visit in this search grid, it generates an ordered collection of waypoints for every asset.

4.2.4 Response Enactment Module

The REM is responsible for carrying out a simulation of each potential response, and providing high-fidelity sensor detection information from the sensors of each of the assets participating in the response. In order to achieve this functionality, the REM's underpinning simulation engine of choice is a proprietary Intelligence, Surveillance, and Reconnaissance (ISR) system developed by *Larus Technologies*[15]. This system contains models for Man-portable Surveillance and Target Acquisition Radar (MSTAR) and Synthetic Aperture Radar (SAR) sensors. The ISR tool also accepts weather information, which influences the probability of detection of the different sensors. The qualities of the sensors (the power levels) that are aboard the assets also affect the probability of detection of targets (the VaLs). More information related to this tool and its sensor models can be found in Section 5.2.1.1.

[15] https://www.larus.com/

4.2.5 Performance Assessment Module

There are six different metrics that are used to evaluate responses; these, along with the details of each response are presented to the human operator, who proceeds to select which response, if any, should be carried out, given his or her training, expertise and intuition.

The first four metrics in the PAM are the previously defined objective functions used by NSGA-II, and thus remain as previously defined in Section 4.2.2.2.

The fifth metric, termed Potential Contact Detection Per Response Asset (PCDRA), quantifies the number of potential VaL contacts that are detected by each response asset during the simulations. The higher this value is, the greater the probability of detecting the VaL during the execution of the real-life mission. More formally, the PCDRA is defined as:

$$PCDRA = \frac{1}{|A|}\sum_{a\varepsilon A}\sum_{v\varepsilon V}\lambda_a(v.C) \tag{49}$$

where A is the set of response assets; V is the set of potential paths that the VaL could have taken; $v.C$ is the set containing the sorted collection of contacts in potential path v; and $\lambda_a()$ is a function that takes in a set of potential contacts, and returns the number of those that have been detected by asset a.

The sixth metric, termed Asset Utilization (AU), quantifies the percentage of assets partaking in a response. More formally, AU is defined as:

$$AU = \frac{\{a\ \varepsilon\ A : a.participates = true\}}{|A|} \tag{50}$$

where A represents the set of assets.

4.3 Chapter Summary

This chapter proposed a system capable of generating automatic CoAs for multiple, concurrently unfolding situations, through the guidance of soft data. It provides an overview of the different system modules, required to achieve these desired L3 functionalities. Furthermore, a set of performance measures that can be used to assess specific potential mission responses is proposed.

Chapter 5. Experimentation

This section presents the experiments conducted for the purpose of empirical evaluation on the L2 and L3 methodologies described in Chapter 3 and Chapter 4.

5.1 Situational Assessment Experiments

Two regions are selected for the purpose of evaluating the risk-aware HSIF model, and the practicability of the IFE metric. The first experiment is situated in the Malaysia/Singapore region, whereas the second is set in the Bangladesh region. The former region is a hotspot for maritime piracy, with an ever-increasing number of incidents[16], as reported by the IMB. In contrast, the latter region was selected due to its relatively placid nature.

5.1.1 Experimental Configuration

This section lays out the configuration of the system that was used for the maritime experiments.

5.1.1.1 Coverage

The Coverage measures for CB, MI, and GD for these experiments are non-computable, since it is not possible to be cognizant of all the entities that are present in the maritime environment being described by the data sources (there is no access to 'ground truth'); the actual Coverage values are thus provided by domain experts, as is typically done in such a case [69]. The values with which the system was run are as follows: $Cv_{CB} = Cv_{MI} = 0.6$. These values were selected through SME feedback from Larus Technologies

[16] http://maritime-executive.com/article/pirates-hijack-tanker-off-malaysian-coast

and IMB, respectively. Since the only data source present in the WC DTG was a simulated one, this DTG's Coverage value was set to be Cv_{WC} = 1.0. Furthermore, Cv_{GD} was estimated to be 1.0, since the only data source currently present in the group is GeoNames DB, a comprehensive database with over 10 million geographical names.

5.1.1.2 Overall Information Timeliness

The fuzzy linguistic terms of the OIT sub-component of the *i-IFE* metric were defined according to maritime subject matter expertise (as provided by Larus Technologies), and were constructed as follows:

- CB_{Recent}(trapezoidal, A = 0, B = 0, C = 600, D = 1200), and
  CB_{Old}(trapezoidal, A = 600, B = 1200, C = ∞, D = ∞); units are in seconds.
- MI_{Recent} (trapezoidal, A = 0, B = 0, C = 500, D = 732), and
  MI_{Old} (trapezoidal, A = 500, B = 732, C = ∞, D = ∞); units are in days.
- WC_{Recent} (trapezoidal, A = 0, B = 0, C = 1, D = 2), and
  WC_{Old} (trapezoidal, A = 1, B = 2, C = ∞, D = ∞); units are in hours.
- OIT_{Recent} (triangular, A = 0.5, B = 1, C = 1.5),
  $OIT_{Acceptable}$(triangular, A = 0, B = 0.5, C = 1),
  and OIT_{Old} (triangular, A = 0.5, B = 0, C = 1.5)

The reader may notice that the OIT_{Recent}'s parameter C is set to 1.5, and that the OIT_{Old}'s parameter A is set to -0.5; however, the range of values that the crisp OIT output can take is in the region [0,1], as determined by the respective centroids of the OIT_{Recent} and OIT_{Old} fuzzy linguistic terms.

The temporal staleness points represent the point in time past which the timeliness of the information provided by the DTGs is considered to be outdated, and this concept is represented by the B parameter of their respective *Old* linguistic terms. Therefore, these parameters were configured as follows: τ_{CB} = CB_{Old}'s B parameter; τ_{MI} = MI_{Old}'s B parameter; and τ_{WC} = WC_{Old}'s B parameter.

5.1.1.3 Reliability

The system was run with the reliability values presented in Table 10. The CB DTG contains only the AIS data with the provider being exactEarth[17], which is the foremost leader in satellite-based AIS data dissemination; thus, its ROS level was deemed to be high; however, the ROI value was slightly reduced from the maximum as AIS data can be intentionally spoofed. The MI DTG contains data provided by two governmental entities – namely, NGA and IMB – both of which exhibit high degrees of reliability (ROS and ROI) since their incident reports are peer-reviewed by maritime operators prior to their online distribution. The WC DTG only contained simulated data, and was thus assigned the highest reliability levels. The GD DTG contains a database of geo-locations which are constantly revised (by users of the dataset through a wiki interface), and was thus assigned high degrees of reliability as well.

[17] http://www.exactearth.com/products/exactais#overview

Table 10: Reliability values

DTG	***ROI***	***ROS***
CB	0.8 (minimal doubt)	0.75 (fairly reliable)
MI	0.8 (minimal doubt)	0.85 (very reliable)
WC	1.0 (trusted)	0.95 (highly/consistently reliable)
GD	0.8 (minimal doubt)	0.95 (highly/consistently reliable)

5.1.1.4 Data Topic Group Relevance

For the following maritime experiments, the system was run with the following DTG Relevance (DTGR) values: $DTGR_{CB} = 0.5$, $DTGR_{MI} = 0.3$, $DTGR_{WC} = 0.1$, and $DTGR_{GD} = 0.1$. The CB DTG was given the highest relevance, because it contains data sources that are crucial for the system's ability to administer fundamental L2 fusion capabilities. The MI DTG was given the next highest relevance, as it contains data sources providing pivotal information for the proactive identification of vessels that may come into distress. The remaining two DTGs do not carry as high of significance, and were henceforth assigned lower relevance values.

5.1.2 Incident and Response Reports

A two-year time period (2011 and 2012) was selected for conducting experiments for each of the maritime regions. Figure 14 unveils the worldwide maritime incident distribution in this time period per incident category for three configurations: (1) only WWTTS incident and response reports, (2) only IMB incident and response reports, and (3) combined WWTTS and IMB incident and response reports.

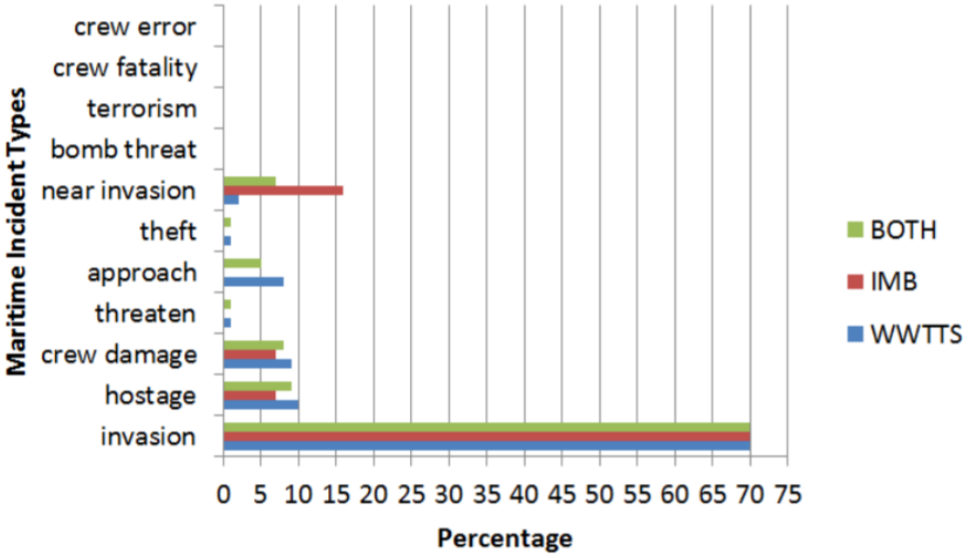


Figure 14: Maritime incident distribution per incident category

5.1.3 Scenario 1: Malaysia/Singapore

A series of scenarios that stretch over a period of 17 hours during the year of 2012 were selected; these scenarios were mostly in the Malaysia/Singapore region. The chosen time period was further separated into six consecutive scenarios (five three-hour long scenarios followed by a single two-hour one). The motivation behind the chosen time lengths was to ensure that some scenarios will contain an ample amount, some will contain

a scarce amount, and some will not contain any CB-related data. As previously discussed, CB-related data is pivotal for the RMF's L2 capabilities, and having such a data-diverse set of scenarios provides a good basis for experimental validation of the IFE metric. The risks, as well as the *i-IFE* metric, were then evaluated for each of the scenarios; lastly, the *g-IFE* metric was evaluated for the region during that 17-hour time period. The region along with the vessels and incident reports during the sixth scenario are displayed in Figure 15. The scenario contains 49 oil tankers, 2 tugboats, and 1 fishing vessel along with WWTTS (red) and IMB (black) incidents.

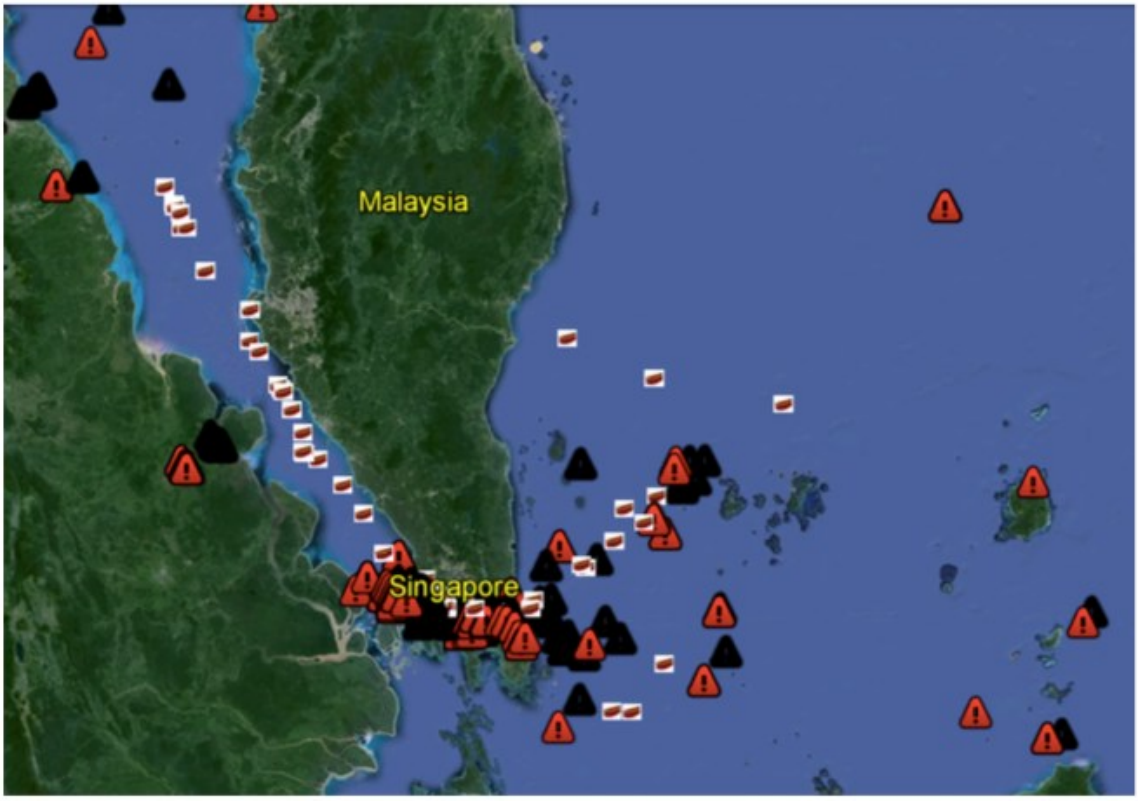


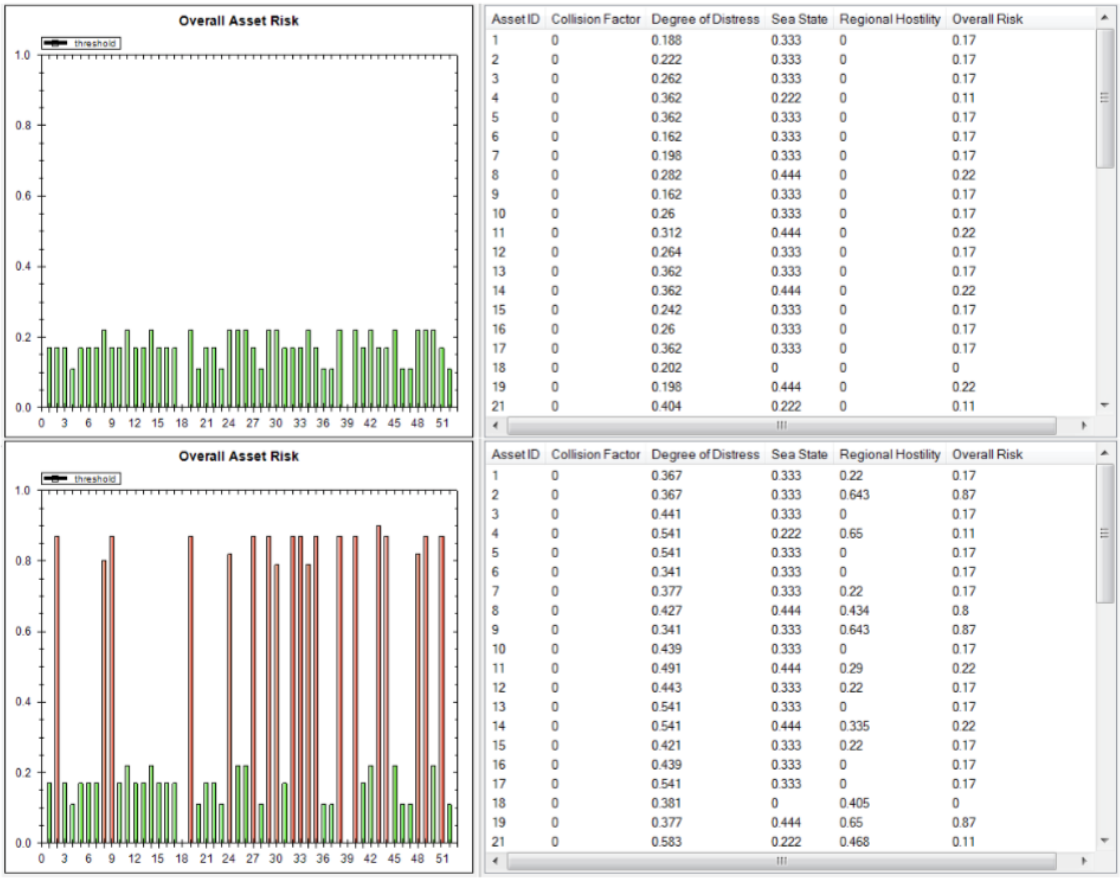
Figure 15: Vessels and incident reports in the Malaysia/Singapore region

The WWTTS source provided 98 incident reports for the Malaysia region, whereas the IMB data only provided 85; however, it is important to note that the IMB data source was observed to be the one providing more relevant incidents with respect to the assets in this region. The region was dominated by the presence of oil tankers – out of 52 vessels, 49 were oil tankers. The IMB source contained 17 incidents that were not captured by the WWTTSS source, but all belonged to the *Cargo Transport* and *Tanker/Industrial* categories (both of which are relevant to oil tankers). Whilst the latter data source provided much more information about *Small Transport/Utility*-related incidents (30 incidents that were not covered by IMB), they were not as relevant to the assets present in the region. Since more relevant incidents were brought in by IMB than WWTTS, it is expected that a system running with the former data will have a higher IBR than one running with only the latter. Similarly, if a system is already running with WWTTS data, and IMB data is added alongside the WWTTS, the IBR is not expected to increase by much.

5.1.3.1 Regional Risk Assessment

Figure 16 demonstrates a risk view of the Malaysia/Singapore region during the sixth scenario. The top portion presents the risk picture without any of the soft maritime incident data sources being ingested by the system (i.e., the MI DTG was deactivated) for the 51 assets present in the scenario. The bottom portion depicts the computed asset risk with both of the WWTTS and IMB data sources for the same scenario (with the same assets). The vertical axis presents the computed risk level for each asset. As can be seen in the bottom portion, the overall system-level risk has increased due to the relevant information provided by the incident sources; specifically, the maritime incident

information increased the values associated with the Regional Hostility and the Degree of Distress risk features.



Asset ID	Collision Factor	Degree of Distress	Sea State	Regional Hostility	Overall Risk
1	0	0.188	0.333	0	0.17
2	0	0.222	0.333	0	0.17
3	0	0.262	0.333	0	0.17
4	0	0.362	0.222	0	0.11
5	0	0.362	0.333	0	0.17
6	0	0.162	0.333	0	0.17
7	0	0.198	0.333	0	0.17
8	0	0.282	0.444	0	0.22
9	0	0.162	0.333	0	0.17
10	0	0.26	0.333	0	0.17
11	0	0.312	0.444	0	0.22
12	0	0.264	0.333	0	0.17
13	0	0.362	0.333	0	0.17
14	0	0.362	0.444	0	0.22
15	0	0.242	0.333	0	0.17
16	0	0.26	0.333	0	0.17
17	0	0.362	0.333	0	0.17
18	0	0.202	0	0	0
19	0	0.198	0.444	0	0.22
21	0	0.404	0.222	0	0.11

Asset ID	Collision Factor	Degree of Distress	Sea State	Regional Hostility	Overall Risk
1	0	0.367	0.333	0.22	0.17
2	0	0.367	0.333	0.643	0.87
3	0	0.441	0.333	0	0.17
4	0	0.541	0.222	0.65	0.11
5	0	0.541	0.333	0	0.17
6	0	0.341	0.333	0	0.17
7	0	0.377	0.333	0.22	0.17
8	0	0.427	0.444	0.434	0.8
9	0	0.341	0.333	0.643	0.87
10	0	0.439	0.333	0	0.17
11	0	0.491	0.444	0.29	0.22
12	0	0.443	0.333	0.22	0.17
13	0	0.541	0.333	0	0.17
14	0	0.541	0.444	0.335	0.22
15	0	0.421	0.333	0.22	0.17
16	0	0.439	0.333	0	0.17
17	0	0.541	0.333	0	0.17
18	0	0.381	0	0.405	0
19	0	0.377	0.444	0.65	0.87
21	0	0.583	0.222	0.468	0.11

Figure 16: Malaysia/Singapore region risk assessment

5.1.3.2 Performance Evaluation and Empirical Analysis

5.1.3.2.1 Instantaneous IFE

Figure 17 visualizes the *i-IFE* results for the Malaysia/Singapore region for each of the six scenarios. Figure 18 unveils the results for the IBR, RAL, OIT, and IC subcomponents of the *i-IFE* metric.

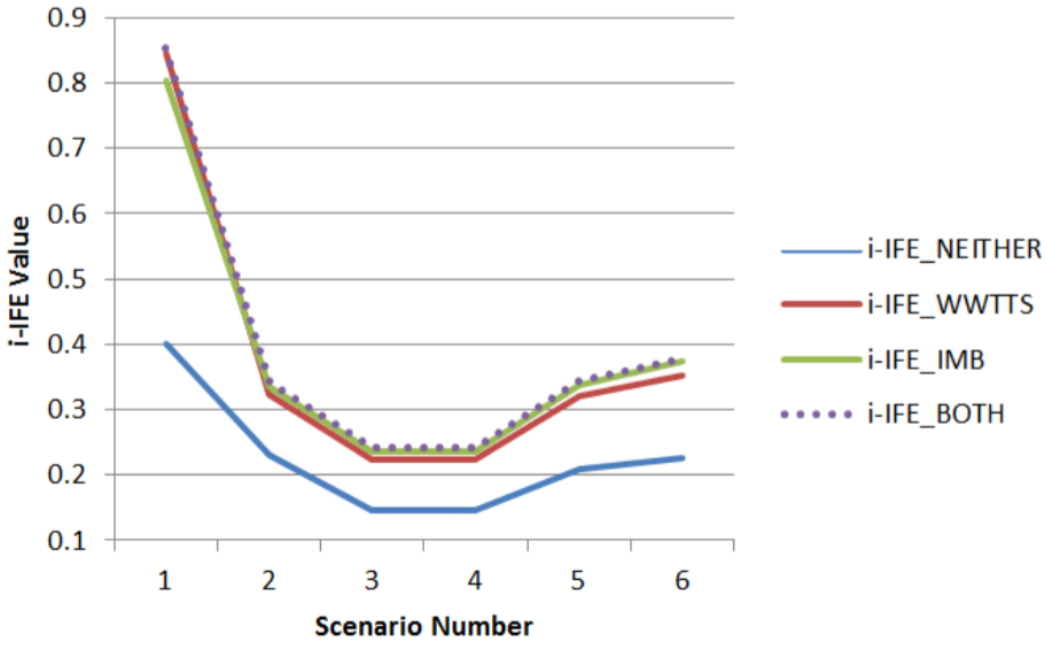


Figure 17: *i-IFE* results in the Malaysia/Singapore region

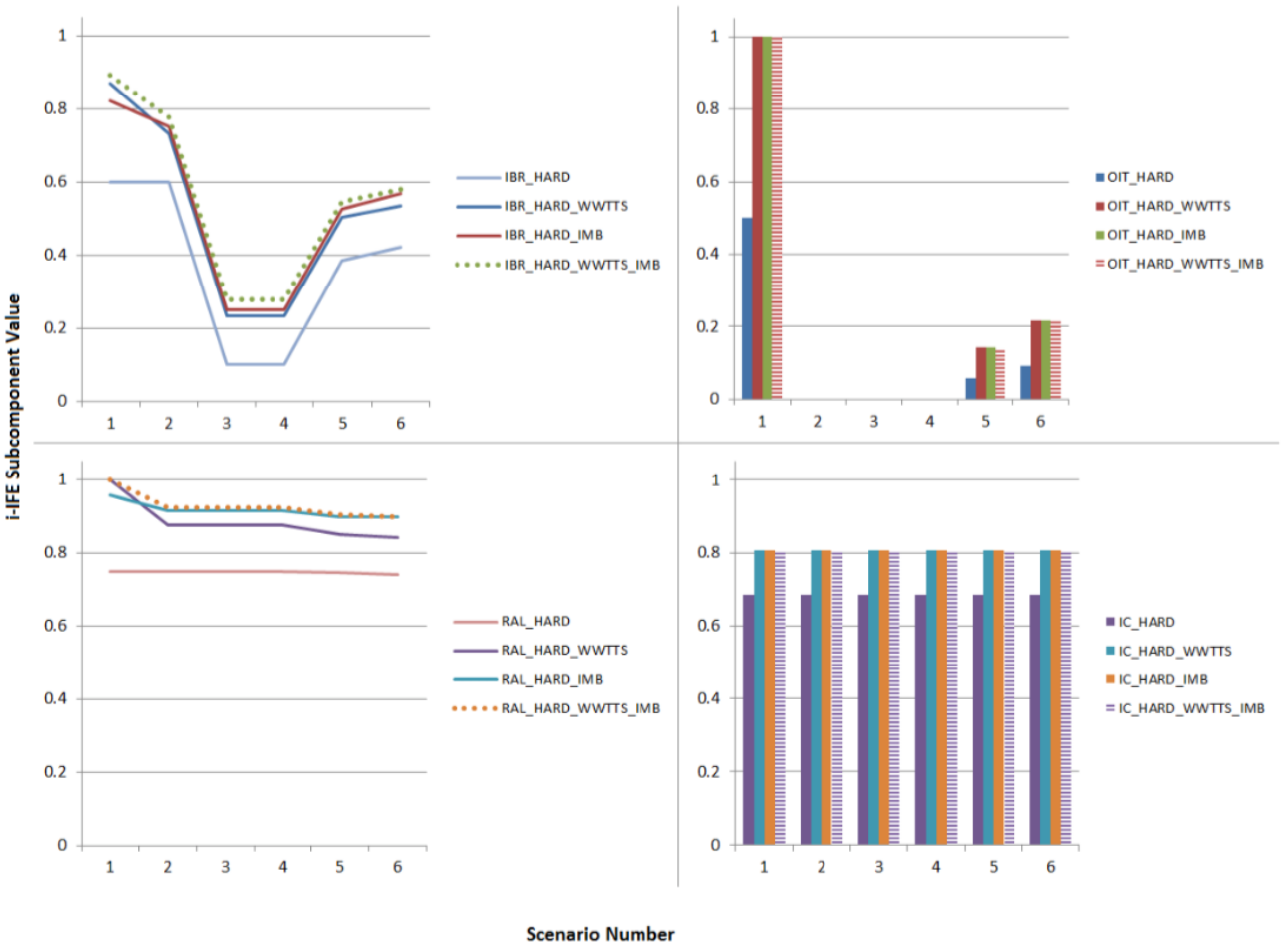


Figure 18: *i-IFE* subcomponents in the Malaysia/Singapore region

One can observe that adding any of the two soft sources to a system running with only hard data sources, consistently yielded a higher IC, IBR, RAL, and consequently higher i-IFE values between the six different scenarios. Furthermore, having both of the soft sources ingested by the system caused the IBR and RAL to be consistently higher (as opposed to the system being run with only a single one of the soft sources), due to the fact that each of the incident sources brought in some relevant incident data. This observed rise in IBR and RAL, and the steady nature of OIT and IC (between running the system with

both soft data sources as opposed to a single one) consequently caused the observed rise in i-IFE results (as presented in Figure 17). The OIT was the lowest when the system was run with only hard data. This was due to the fact that when there is no MI data available, the timeliness value of the MI DTG going into the FIS (IT_{MI}) is equal to infinity as previously defined. OIT also remained constant between three system configurations (running the RMF with only WWTTS, only IMB, and both soft sources) because both of the WWTTS and IMB sources exhibited similar timeliness characteristics. Scenario 2 contained very untimely CB-related data, whilst scenarios 3 and 4 contained no CB-related data at all; although all of these scenarios contained MI- and WC-related data, the OIT was 0 due to the predominant role FIT_{CB} plays in the fuzzy rules.

5.1.3.2.2 Gradual IFE

Table 11 presents the *g-IFE* results for running the system with four different configurations –hard sources only, hard sources and WWTTS, hard sources and IMB, hard as well as both soft sources. The *g-IFE* was higher whenever any of the soft sources were turned on, as compared to when the system was running with only hard data. The metric reported the highest value whenever both of the soft sources were being processed. The small *g-IFE* increase seen when the WWTTS data source is added to the system when it is already processing IMB data is mainly attributed to the observed increase in the *IBR* value in each of the six scenarios; this was due to the fact that the WWTTS data source did bring in some relevant incident information (incidents belonging to the *Cargo Transport* and *Tanker/Industrial* categories) not already captured by IMB.

Table 11: Gradual IFE results in the Malaysia/Singapore region

	Hard sources only	***Hard and WWTTS***	***Hard and IMB***	***Hard, WWTTS, and IMB***
g-IFE	0.20325	0.305	0.32075	0.326

5.1.3.2.3 Profitability Analysis

The profitability of adding the IMB data source when the system is already ingesting the WWTTS data source is calculated to be $\Pi_B(\{IMB\})$ = 0.068852459, or approximately 6.89%; where $B = \{AIS, SS, GeoNames, WWTTS\}$.

The profitability of adding the WWTTS data source when the system is already ingesting the IMB data source is calculated to be Π_B ({WWTTS}) = 0.016367888, or approximately 1.64%; where $B = \{AIS, SS, GeoNames, IMB\}$.

The profitability of adding the WWTTS and IMB data sources when the system is running with only hard data is calculated to be Π_B ({WWTTS, IMB}) = 0.603936, or approximately 60.40%; where $B = \{AIS, SS, GeoNames\}$.

Hence, the IMB data source was calculated to be approximately 4.21 times more profitable for this region. This is largely attributed to the more *relevant* information brought into the system by IMB over WWTTS.

5.1.3.2.4 Robustness Analysis

As was defined in Section 3.3.3.1, *ILT* is concerned with randomly discarding entities from data sources and observing the effect this disturbance has on the system via

the *g-IFE* calculation. The experiments performed consisted of randomly discarding 40%, 80%, and 95% of incidents from the soft data sources, and the weather measurements from the weather conditions source. As can be seen from the results presented in Table 12, there was a consistent and expected decline when more data was being discarded. The system was judged to be approximately 83.51% tolerant of a disturbance that would cause it to lose 95% of the input incident and weather data.

Table 12: ILT results in the Malaysia/Singapore region

Discarded Entities Percentage	*g-IFE*	ILT
40%	0.31375	1.000000000
80%	0.30150	0.924846626
95%	0.27225	0.835122699

From Section 3.3.3.2, UCC is a measure of the ratio of the total number of use cases in which the system can be run whilst providing useful fusion capabilities. The core functionalities within the RMF L2 fusion (SA) are risk modeling, risk assessment, as well as the proactive identification of vessels in distress. The RMF currently cannot achieve any of these functionalities without the AIS data source, which implies that any 'useful' use case is one that includes AIS data. There are currently five data sources used by the system, which implies 32 possible ways of running the system. From the 32, 16 of those are ones that include the AIS data source (and are therefore 'useful'). Thus, for this set of experiments, UCC = 16/32 = 0.5.

5.1.4 Scenario 2: Bangladesh

The next set of experimental scenarios is situated in the Bangladesh region and spans over a period of 16 hours during the year of 2012. The 16-hour time period was separated into six consecutive scenarios, each of a three-hour length, except for the last one, which was a single hour. The reasoning behind this partitioning was idem to the Malaysia experiment – having a data-diverse set of scenarios for good experimental validation. The risks as well as the *i-IFE* metric were evaluated for each of the scenarios, and lastly, the g-IFE metric was evaluated for the region. The region, along with the vessels and the maritime incidents during the sixth scenario, is displayed in Figure 19.

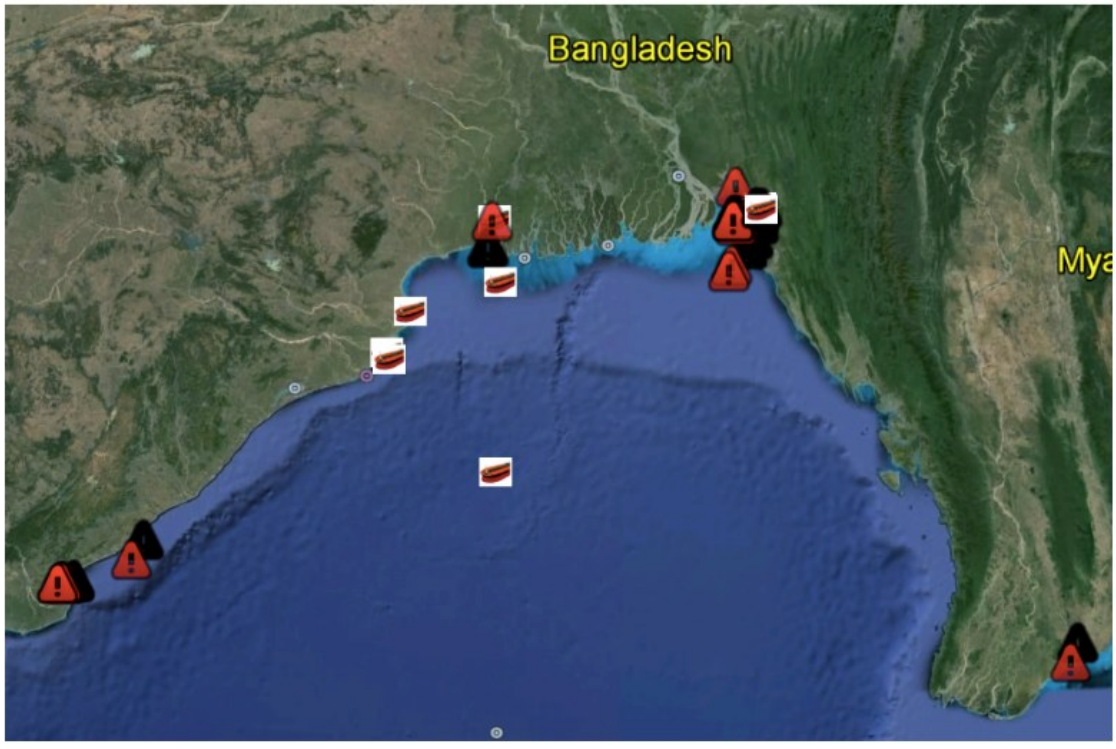
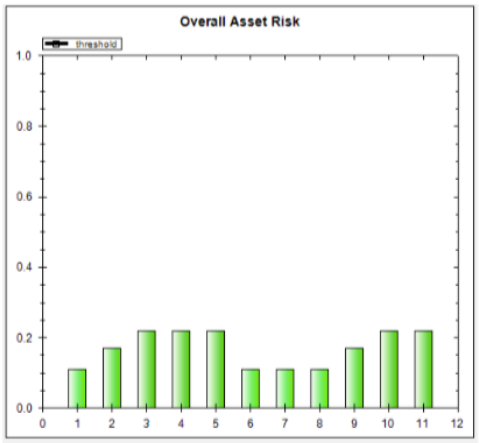


Figure 19: Vessels and incident reports in the Bangladesh region

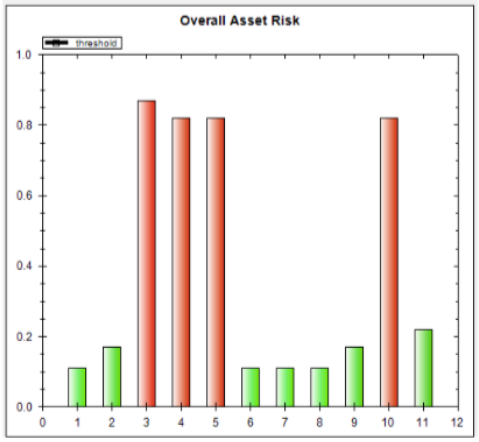
In this experiment, WWTTS and IMB contained roughly an equal number of maritime incidents (23 and 20 incidents, respectively). As in Malaysia, the Bangladesh region was dominated by oil tankers (8 out of the 13 vessels), and IMB provided information about three incident reports belonging to the *Tanker/Industrial* category, none of which were captured by WWTTS. Both of the data sources brought in an equal amount of *Cargo Transport* incident data. Thus, the IMB source provided more relevant information to the region's assets; it is therefore expected that a system running with this data will have a slightly higher IBR than one running with only the WWTTS. Similarly, if a system is already running with IMB data, and WWTTS data is added alongside the IMB, the IBR is not expected to improve by much.

5.1.4.1 Regional Risk Assessment

Figure 20 demonstrates a risk view of the Bangladesh region during the sixth scenario. The top portion presents the risk picture without any of the soft maritime incident data sources being ingested by the system (i.e., the MI DTG was deactivated), whereas the bottom portion depicts the computed asset risk with both of WWTTS and IMB data sources for the same scenario, with the same 11 assets. As before, the vertical axis presents the computer risk level for each asset. As can be seen in the bottom portion of the figure, the information brought from the incident sources has increased the risk level (due to heightened *Regional Hostility* and *Degree of Distress* values) for four of the assets.



Asset ID	Collision Factor	Degree of Distress	Sea State	Regional Hostility	Overall Risk
0	0	0.188	0.222	0	0.11
1	0	0.322	0.333	0	0.17
2	0	0.25	0.444	0	0.22
3	0	0.278	0.444	0	0.22
4	0	0.372	0.444	0	0.22
5	0	0.31	0.222	0	0.11
6	0	0.286	0.222	0	0.11
7	0	0.186	0.222	0	0.11
8	0	0.162	0.333	0	0.17
9	0	0.246	0.444	0	0.22
10	0	0.26	0.444	0	0.22



Asset ID	Collision Factor	Degree of Distress	Sea State	Regional Hostility	Overall Risk
0	0	0.367	0.222	0	0.11
1	0	0.501	0.333	0	0.17
2	0	0.395	0.444	0.65	0.87
3	0	0.457	0.444	0.5	0.82
4	0	0.551	0.444	0.5	0.82
5	0	0.489	0.222	0.335	0.11
6	0	0.431	0.222	0	0.11
7	0	0.331	0.222	0	0.11
8	0	0.341	0.333	0	0.17
9	0	0.425	0.444	0.5	0.82
10	0	0.439	0.444	0.335	0.22

Figure 20: Bangladesh region risk assessment

5.1.4.2 Performance Evaluation

5.1.4.2.1 Instantaneous IFE

Figure 21 visualizes the *i-IFE* results for the Bangladesh region for each of the six scenarios. Figure 22 displays the results for the IBR, RAL, OIT, and IC subcomponents of the *i-IFE* metric.

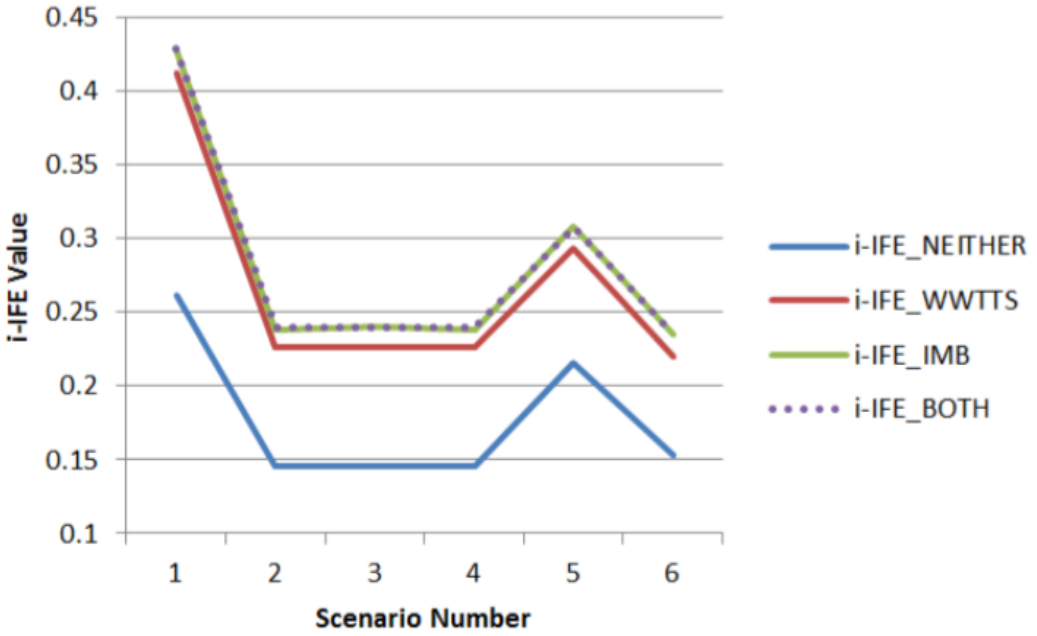


Figure 21: *i-IFE* results in the Bangladesh region

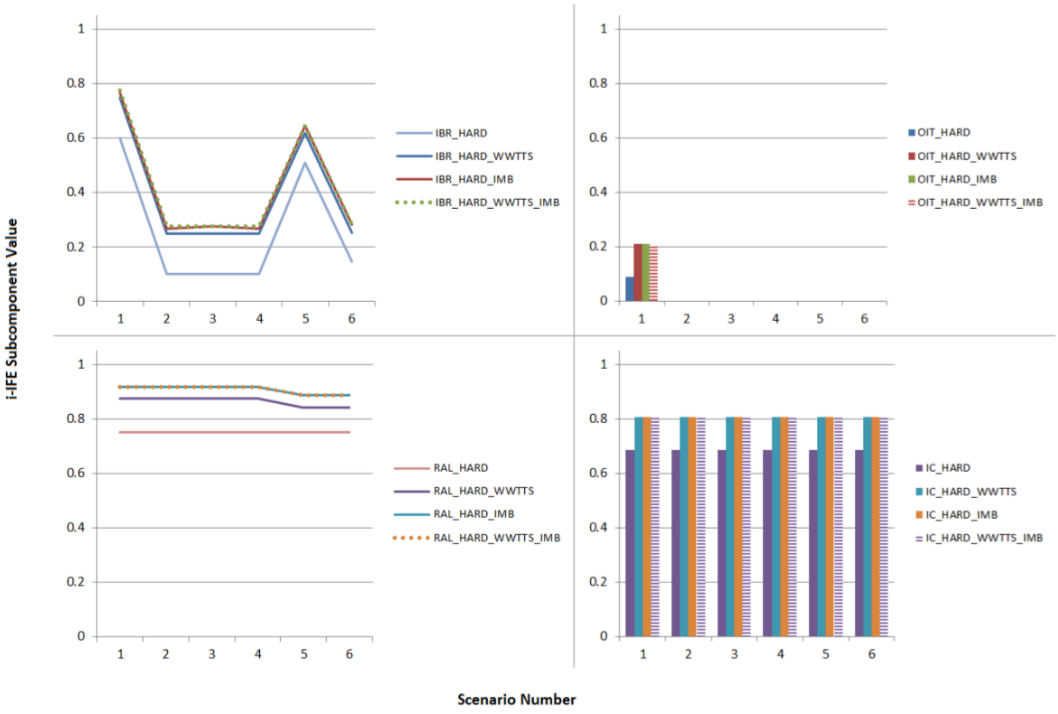


Figure 22: *i-IFE* subcomponents in the Bangladesh region

From the experimental results, one can observe that adding any of the two soft sources to a system running with only hard data sources consistently yielded a higher IC, IBR, RAL, and consequently higher *i-IFE* values between the six different scenarios. The OIT was the lowest when the system was run with only hard data; the reasoning is idem to the Malaysia experiment. OIT also remained constant between three system configurations (running the RMF with only WWTTS, only IMB, and both soft sources), because both of the WWTTS and IMB sources exhibited similar timeliness characteristics. Scenarios 5 and 6 contained very untimely CB-related data, whilst scenarios 2 through 4 contained no CB-related data at all; although all of these scenarios contained MI- and WC- related data, the OIT for each of them was 0 due to the predominant role FIT_{CB} plays in the fuzzy rules.

The *i-IFE* and all of its subcomponents were virtually identical when the system was run only with IMB data and run with both WWTTS and IMB data; this seems to suggest that the IMB data source may be sufficient for the selected region (i.e., there is no need to be processing any WWTTS data, as it seems to provide little or no merit to a system already processing IMB data within this region).

5.1.4.2.2 Gradual IFE

Table 13 presents the *g-IFE* results for running the system with the same four different dataset configurations as the experiment in the Malaysia region. The g-IFE was higher whenever any of the soft sources were turned on, as compared to when the system was running with only hard data. The *g-IFE* was the highest whenever both of the soft sources were being ingested; however, the value was not much higher than when only the

IMB data source was being processed. This was indeed the expected behaviour, since WWTTS contained fewer relevant incidents for this region than IMB did.

Table 13: Gradual IFE results in the Bangladesh region

	Hard sources only	**Hard and WWTTS**	**Hard and IMB**	**Hard, WWTTS, and IMB**
g-IFE	0.16525	0.24275	0.256	0.257

5.1.4.2.3 Profitability Analysis

The profitability of adding the IMB data source when the system is already ingesting the WWTTS data source is calculated to be Π_B ({IMB}) = 0.058702369, or approximately 5.87%; where $B = \{AIS, SS, GeoNames, WWTTS\}$.

The profitability of adding the WWTTS data source when the system is already ingesting the IMB data source is calculated to be Π_B ({WWTTS}) = 0.00390625, or approximately 0.39%; where $B = \{AIS, SS, GeoNames, IMB\}$. This low value is consistent with the previous observation that adding a WWTTS data source when the system is already ingesting IMB data seems to provide little or no merit.

The profitability of adding the WWTTS and IMB data sources when the system is running with only hard data is calculated to be Π_B ({WWTTS, IMB}) = 0.5552194, or approximately 55.52%; where $B = \{AIS, SS, GeoNames\}$.

Thus, the profitability brought into the system by the IMB data source was calculated to be roughly 15 times greater than that of the WWTTS data source. This again

is attributed to the more relevant information brought into the system by IMB over WWTTS.

5.1.4.2.4 Robustness Analysis

As was the case in the Malaysia/Singapore experiment, there were three configurations used for running the *ILT* experiments – 40% of entities from the MI and WC data discarded, 80% of entities discarded, and lastly, 95% of entities discarded. As can be seen from the results presented in Table 14, there was a consistent and expected decline when more data was being dropped. The system was judged to be approximately 73.64% tolerant of a disturbance that would cause it to lose 95% of the input incident and weather data. This *ILT* value is lower than the one of the Malaysia/Singapore experiment, due to the fact that the Bangladesh region had much less available incident data (since there are many less incidents). Conversely, the Malaysia/Singapore region contained an overabundance of incident data, due to its hostile nature.

Table 14: ILT results in the Bangladesh region

Discarded Entities Percentage	*g-IFE*	ILT
40%	0.25625	1.000000000
80%	0.23150	0.900778210
95%	0.18925	0.736381323

UCC remains the same (*UCC* = 0.5) as it was in the Malaysia/Singapore experiments due to the fact that the same data sources were employed for both of the experiments.

5.2 Course of Action Generation Experiments

There are two types of experiments run through the SDDRG system. The first one focuses on exploring the benefit provided by the soft incident response data in a single situation occurring off the East Coast of Canada. In addition to exploring the value provided by soft data, the second experiment explores a pair of events occurring off of the East Coast of Africa in order to demonstrate the merit of considering multiple situations concurrently.

5.2.1 Experimental Setup

In both of these experiments, NSGA-II was configured to run for 10,000 generations or 5 minutes, whichever came first. The study considers two evolving operators (mutation and cross-over), as previously defined in Section 4.2.2.3. The crossover probability was 0.9 and the mutation probability was 0.3. Due to restrictions of the available multi-criteria decision analysis (MCDA) framework (under which the NSGA-II implementation resides), no constraints were able to be imposed on the chromosomes. The epsilon value for the Epsilon-Box Dominance Archive was 0.01, and the selection operator used was a binary tournament.

5.2.1.1 Response Simulation Environment

As previously discussed, the sensor simulation engine used as part of this research was *Larus Technologies*' proprietary ISR tool. This simulator provided access to two radar models: MSTAR and SAR. The latter sensors are ones that are mounted on air and space platforms; they look down on targets located on the Earth's surface. By contrast, the former sensors are located near the surface of the Earth, often mounted on tall masts, lifting them high above the ground. The height of the mast determines the distance to the horizon, past

which targets cannot be detected. The probability of target detection for both of these sensor types is directly related to their corresponding power levels, as well as being directly affected by the prevailing weather conditions in the region they are monitoring.

For the purposes of this work, it was necessary to augment the ISR tool's data processing capabilities. By default, this simulation platform processes data in real-time; however, for the purposes of this research, it was necessary to obtain simulations for many potential responses, spanning over long durations. Thus, the data processing capability was augmented to be 1000 times faster than real-time. The simulations were performed on an Intel i7 quad-core processor with 16 GB of RAM.

5.2.2 Response Asset Characteristics

Table 15 presents the maximum speed, the moving cost, and the sensor type and its associated quality mounted on each of the different types of assets present in the VID experiment. Within the system, MSTAR sensors are mounted on vessels, whereas the SAR sensors are mounted on aerial platforms. Higher quality sensors are mounted on platforms with higher operational costs. The quality of the sensor is determined by its associated power level (i.e., higher quality means a higher-power sensor).

Table 15: Response asset characteristics

Asset Type	Max Speed (km/h)	Moving Cost ($/km)	Sensor Quality	Sensor Type
Fast UAV	300	120	Low	SAR
		140	Medium	
		160	High	
		180	Very high	
Slow UAV	180	70	Low	SAR
		90	Medium	
		110	High	
		130	Very high	
Helicopter	200	75	Low	SAR

		95	Medium	
		115	High	
		135	Very high	
Aircraft	170	40	Low	SAR
		60	Medium	
		80	High	
		100	Very high	
Tug boat	60	175	Low	MSTAR
		195	Medium	
		215	High	
		235	Very high	
Speed boat	100	40	Low	MSTAR
		60	Medium	
		80	High	
		100	Very high	
General cargo boat	60	175	Low	MSTAR
		195	Medium	
		215	High	
		225	Very high	

5.2.3 Experiment 1: Single Situation Handling – Canadian East Coast

This scenario comprises a VID which needs to be located in the worst of weather conditions – dense clouds with extreme rain. The incident is situated in the East Coast of Canada, approximately 330 km from St John's, Newfoundland. The LKP of the VID is at latitude of 46.0, and longitude of -49.0. A visualization of the response region is presented in Figure 23; there were 22 CGAs grounded at St John's (not visualized), two non-grounded CGAs (one helicopter and one fast UAV), and five ORAs (aircraft)

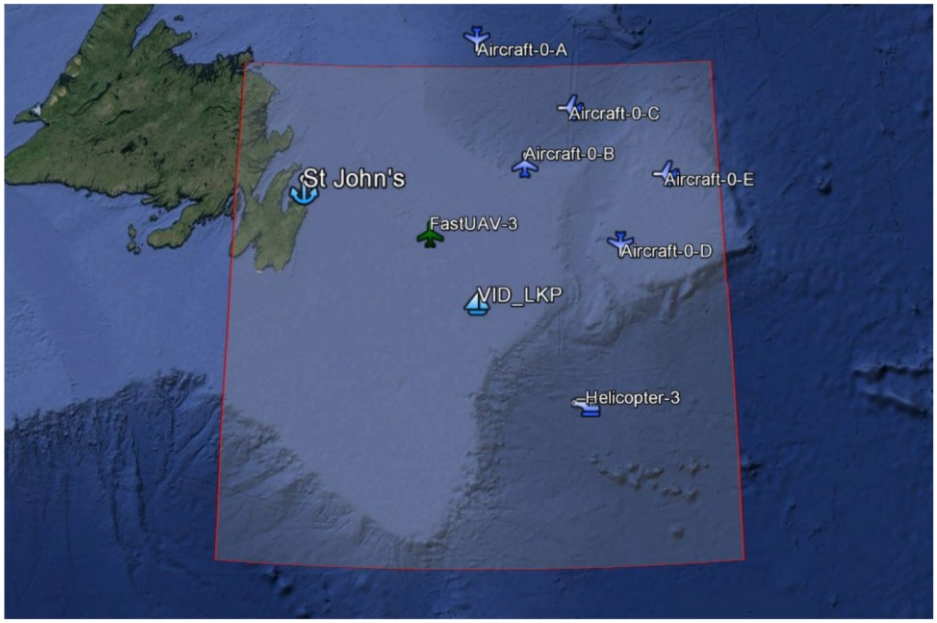


Figure 23: VID Scenario - East Coast of Canada

The historical incident response report which was deemed by the system to be the most pertinent to the current situation was previously presented in Figure 10. The report describes a VID situation taking place during bad weather. As a result, the coastal agency decided to dispatch assets with higher quality sensors (all of the assets had 'very high' quality sensors, with the exception of *FastUAV-2*, which had a 'high' quality onboard sensor). This report tells the system to explore solutions comprising of assets with high quality sensors, via the MR objective function used within NSGA-II.

Thus, for this experiment, there are three trends which are expected to be observed in the experimental results. The first trend is examining relations between running the system with vs. without soft data; and the remaining two trends pertain to relations which are expected to be observed only within the soft data results:

Expected Trend 1. It is anticipated that whenever simulations are run with the soft data enabled (vs. disabled), there will be a higher chance of mission success, as judged by the average PCDRA value obtained by running simulations on each set of generated CoAs.

Expected Trend 2. Furthermore, it is expected that within the results generated by running the system with soft data present, there will be a correlation between the degree to which the MR objective is met, and the calculated PCDRA (i.e. the better the MR objective is satisfied, the higher the probability of VID detection should be). The reverse, however, may not be necessarily true, since the system may conceive solutions, having not previously been identified or explored in the soft data.

Expected Trend 3. Similarly, there is an expected correlation between the MR and ME objectives, as higher-quality sensors are attached to more expensive platforms.

5.2.3.1 Scenario Assets

In this scenario, most of the CGAs are docked at St John's (Newfoundland, Canada) which is located approximately at latitude of 47.555, and longitude of -52.7067. More specifically, 22 of the CGAs are docked (eight speed boats, eight tug boats, one general cargo boat, two slow UAVs, one fast UAV, and two helicopters), and the remaining two CGAs are travelling in the surrounding area (one fast UAV and one helicopter). There are also five ORAs in the vicinity, which can partake in the response (five aircraft). The details of all these assets are presented in Table 16 and Table 17.

Table 16: VID experiment coast guard assets

Asset ID	Asset Type	Sensor Quality	Asset Location (Name or Lat, Lon)	Status
Speedboat-0-A	Speed boat	Low	St John's	Docked
Speedboat-0-B	Speed boat	Low	St John's	Docked
Speedboat-0-C	Speed boat	Low	St John's	Docked
Speedboat-0-D	Speed boat	Low	St John's	Docked
Speedboat-0-E	Speed boat	Low	St John's	Docked
Speedboat-0-F	Speed boat	Low	St John's	Docked
Speedboat-3-A	Speed boat	Very high	St John's	Docked
Speedboat-3-B	Speed boat	Very high	St John's	Docked
Tugboat-0-A	Tug boat	Low	St John's	Docked
Tugboat-0-B	Tug boat	Low	St John's	Docked
Tugboat-0-C	Tug boat	Low	St John's	Docked
Tugboat-0-D	Tug boat	Low	St John's	Docked
Tugboat-0-E	Tug boat	Low	St John's	Docked
Tugboat-1	Tug boat	Medium	St John's	Docked
Tugboat-3-A	Tug boat	Very high	St John's	Docked
Tugboat-3-B	Tug boat	Very high	St John's	Docked
Generalcargo-0	General cargo	Low	St John's	Docked
SlowUAV-0	Slow UAV	Low	St John's	Docked
SlowUAV-2	Slow UAV	High	St John's	Docked
FastUAV-1	Fast UAV	Medium	St John's	Docked
Helicopter-0	Helicopter	Low	St John's	Docked
Helicopter-2	Helicopter	High	St John's	Docked
FastUAV-3	Fast UAV	Very high	<47.0, -50.0>	Not docked
Helicopter-3	Helicopter	Very high	<44.5, -46.853>	Not docked

Table 17: VID experiment opportunistic response assets

Asset ID	Asset Type	Sensor Quality	Asset Location (Name or Lat,Lon)
Aircraft-0-A	Aircraft	Low	<50.0, -49.0>
Aircraft-0-B	Aircraft	Low	<48.0, -48.0>
Aircraft-0-C	Aircraft	Low	<49.0, -47.0>
Aircraft-0-D	Aircraft	Low	<47.0, -46.0>
Aircraft-0-E	Aircraft	Low	<48.0, -45.0>

5.2.3.2 Experimental Results

This section presents the experimental results gathered by running the system on the same VID scenario with and without the historical incident response data. The performance metrics gathered from each of these two configurations are laid out in Table 18 and Table 19, respectively.

A human operator looking at the solution set presented in Table 18, who is primarily concerned with overall response time, may consider *Response 12* as viable, in that it has a low response time, yet a decently high chance the VID will be detected (as judged by the response's PCDRA value). If this operator, however, is willing to allocate a bit of extra time (for a 52.84% longer mission), he or she could opt in to use *Response 8* and drastically bring down the overall mission cost (by about 58%), whilst greatly increasing the overall *PCDRA* level (roughly 2.5 times higher), and hence increasing the probability of detecting the VID.

Table 18: Experiment results with historical incident response data

Response Number	ME ($)	MR	MT (min)	USA (%)	PCDRA
1	1,285,485.98	4	7,716.28	0.00	38.30
2	1,129,650.55	4	8,037.73	0.00	55.95
3	1,207,408.25	4	7,809.23	0.00	38.85
4	1,190,460.33	4	7,841.85	0.00	27.95
5	3,023,906.55	4	5,549.43	0.00	39.26
6	963,590.19	3	7,020.97	0.00	23.97
7	33,795.41	0	298.19	94.81	0.00
8	72,400.29	1	845.35	98.96	54.78
9	78,438.53	1	278.58	95.50	0.00
10	80,671.99	0	258.35	98.96	0.00
11	206,503.21	1	537.52	93.43	1.91
12	173,736.18	2	553.06	97.23	21.98

Figure 24 presents the normalized values for *PCDRA*, *ME*, and *MR* from Table 18. An interesting observation one can make from the graph is that there is a good correlation between the level to which the *MRs* are met and the *PCDRA* – as was previously anticipated in *Expected Trend 2*. Responses which met the MR objective with a level of 4 had an average of 2.73 times higher *PCDRA* versus the remainder of the responses (meeting the *MR* with levels ranging from 0 to 3). Furthermore, responses which met the *MRs* with a level of 4 were also on average 6.82 times more expensive than the remainder of the responses, due the use of higher-quality, but more expensive platforms – as anticipated in *Expected Trend 3*.

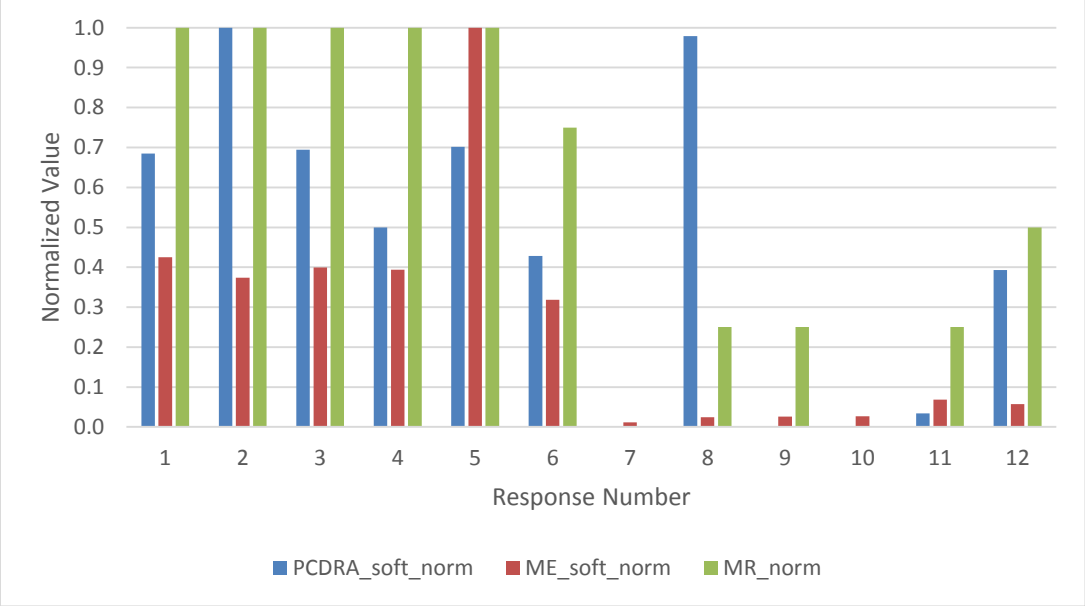


Figure 24: Normalized soft data PCDRA, ME, and MR

Response 9 presents an interesting data point, adhering to *Expected Trend 2*; the level to which *MR's* were met was low, yet after simulating the response, its associated *PCDRA* value was calculated to be quite high. This response may present a viable solution,

having not been previously considered in (historical) incident responses. In real-world exercises, such a response should be considered as a prime candidate for a detailed analysis, as it may provide valuable insights into previously-unconsidered CoA generation characteristics, which may further lead to an overall response process improvement.

A human operator, who does not have access to historical incident response data, would find him or herself with the results presented in Table 19. He/she would conduct a similar analysis, as presented above, and observe trade-offs in the different responses in terms of the various performance metrics. For instance, if the operator wants to maximize the chance of detecting the VID, but is concerned with neither the overall time nor cost of the mission, he or she could opt in for *Response 2*. If the operator is more sensitive to cost, but wishes to maintain the probability of detection reasonably high, he or she could select *Response 4* and save approximately 27.74% in mission expenses whilst roughly halving the number of potential VID contacts that each response asset can detect (i.e. the *PCDRA*).

Table 19: Experiment results without historical incident response data

Response Number	ME ($)	MR	MT (min)	USA (%)	PCDRA
1	6,209,923.14	N/A	44,954.88	0.00	19.72
2	6,701,877.41	N/A	35,368.16	0.00	26.72
3	7,106,243.64	N/A	38,450.67	0.00	2.47
4	4,842,652.20	N/A	38,929.73	0.00	12.80
5	148,505.91	N/A	1,172.78	84.78	0.00
6	281,949.99	N/A	2,698.34	85.12	2.87
7	291,217.74	N/A	2,707.28	79.93	5.95
8	336,652.87	N/A	2,239.92	79.93	3.78
9	375,789.14	N/A	1,293.78	84.08	3.69
10	674,632.95	N/A	1,612.65	71.97	0.00
11	522,222.34	N/A	2,003.35	80.28	3.81
12	1,027,456.31	N/A	2,238.14	67.13	7.94

Figure 25 presents the PCDRA values obtained by running the system with and without soft incident response data. The average *PCDRA* obtained by running the SDDRG was calculated to be roughly 3.38 times higher when historical incident response data was available. This observation is consistent with *Expected Trend 1*. This significant increase represents a tangibly higher chance of VID detection in the real world, and can be attributed to the fact that the historical data pointed the system towards the use of higher-quality sensors (due to the prevailing weather conditions).

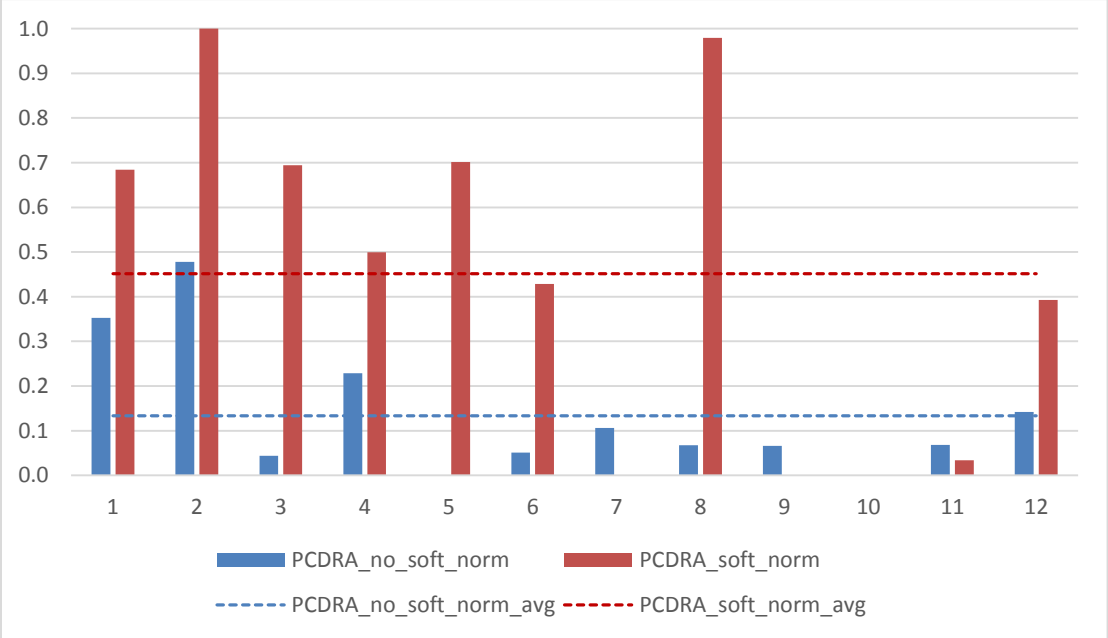


Figure 25: Normalized soft vs no soft PCDRA comparison

5.2.4 Experiment 2: Concurrent Situation Handling – Somalian East Coast

This experiment comprises two situations – a VID and a piracy event, both occurring at roughly the same time. To gather experimental data, the system was run with three different configurations:

Configuration 1. Generating experimental results with soft data enabled by considering the VID and piracy event independently (in a sequential fashion).

Configuration 2. Generating experimental results with soft data by concurrently assigning assets to both the VID and piracy event.

Configuration 3. Generating experimental results in same way as in *Configuration 2*, but without considering soft data.

The three different configurations are all executed with the same set of assets. The scenario occurs in the North-East Coast of Africa. The region has a Somalian coast guard station located at latitude of 11.776586, and longitude of 51.242778; a Somalian auxiliary coast guard station located at latitude of 4.407503, and longitude of 47.784330; and a Yemeni auxiliary coast guard station located at latitude of 15.769778, and longitude of 52.043575. The pirate attack under consideration occurs approximately 500 km off of the Somalian coast line – at latitude of 7.525418, and longitude of 54.950306; at the time of the incident the weather was relatively calm. The VID is approximately 530 km off of Oman's coast at latitude of 14.06291, and longitude of 59.720956; as before, the VID situation is occurring in bad weather conditions. Both of these incidents are depicted in Figure 26.

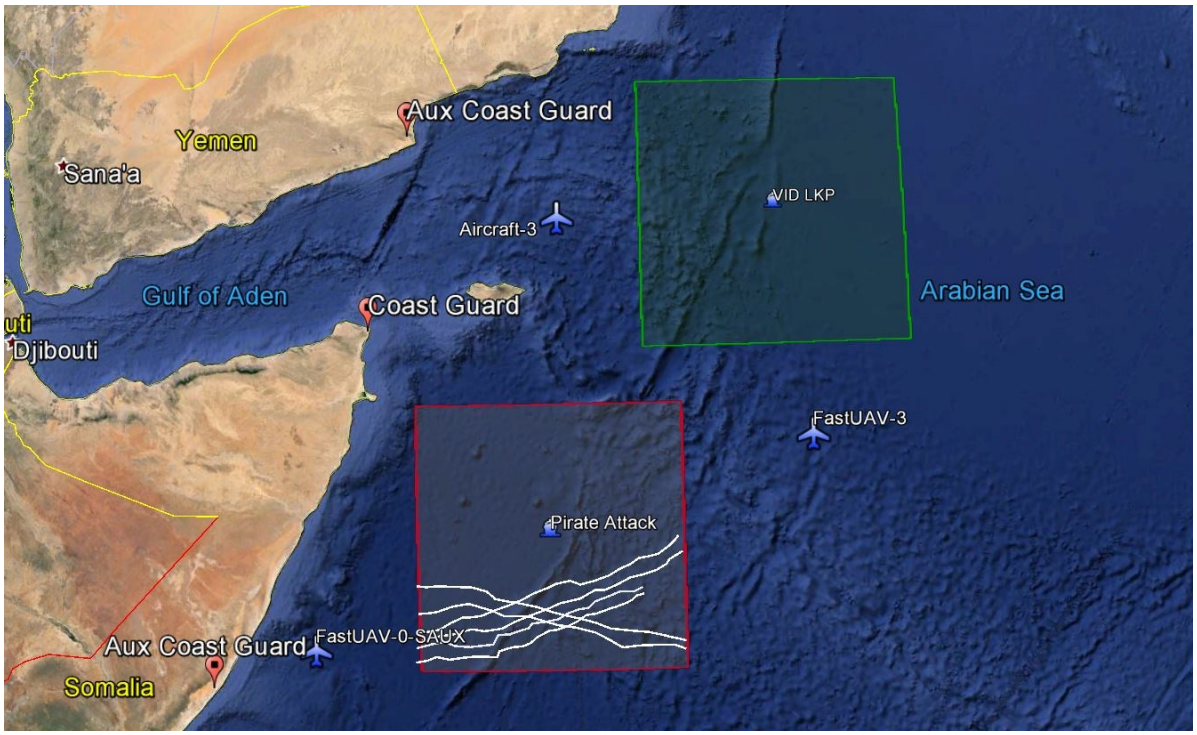


Figure 26: VID and piracy event in the North-East coast of Somalia

The historical incident response reports which were deemed by the system to be the most pertinent to the VID and piracy incident situations are presented in Figure 27 and Figure 28, respectively.

On the 20th of May, coastal radars lost contact with a cargo ship at 13.5375 N, 50.6724 E. Following unsuccessful attempts to contact the crew, a search mission was launched. Aircraft-3 was assigned a square in search pattern. Aircraft-2 was assigned a square in search pattern. SlowUAV-3 was assigned a parallel track line search pattern. Speedboat-3-A was assigned a square out pattern, and Speedboat-3-B was assigned a square out pattern. Tugboat-3-A was assigned a square in pattern. Helicopter-3 was assigned a parallel track line search pattern. Lastly, Tugboat-3-B was assigned a parallel track line search pattern. At the time of the incident, there was moderate rain present with extremely dense clouds.

Figure 27: Most pertinent VID incident and response report in the Somalian region

While underway, an oil tanker was boarded by a group of pirates at 10.7687 N, 54.0480 E on August 17th. The intruders were carrying automatic rifles. The ship's alarm was sounded and the master was able to send a distress call to the coast guard over the VHF radio. The coast guard immediately dispatched response assets. FastUAV-0 was assigned a track crawl search pattern. SlowUAV-0 was assigned a track crawl search pattern. Speedboat-0-A was assigned a track crawl search pattern. Lastly, Speedboat-0-B was assigned a track crawl search pattern. There was no rain and no clouds during the mission.

Figure 28: Most pertinent piracy incident and response report in the Somalian region

The VID report describes an incident taking place during bad weather. As a result, the coastal agency decided to dispatch assets with higher quality sensors (all of the assets had 'very high' quality sensors, with the exception of *Aircraft-2*, which had a 'high' quality onboard sensor). This report tells the system to explore solutions comprising of assets with high quality sensors, via the MR objective function used within NSGA-II. The piracy event report describes an incident taking place during favourable weather conditions, and as a result, the coastal agency deployed assets with lower quality sensors (due to their cheaper operational costs). These assets were also all assigned the track crawl pattern, which is used when the VaL tracks are known ahead of time, as previously discussed in Section 2.5.2.

For this experiment there are four expected trends:

Expected Trend 1. It is anticipated that whenever simulations are run with the soft data enabled (vs. disabled), there will be a higher chance of mission success, as judged by the *PCDRA* value obtained via response simulations.

Expected Trend 2. It is also expected that within the results gathered by running the system with soft data enabled, there will be a correlation between the level to which *MRs* are met and the *PCDRA* value – the higher the MRs are, the higher the

PCDRA values are expected to be. As before, it is expected that the converse might not necessarily be true.

Expected Trend 3. Similarly, there is an expected correlation between the MR objective, and the ME objective, as higher-quality sensors are attached to more expensive platforms, and the VID is taking place during unfavourable weather conditions.

Expected Trend 4. It is furthermore expected that whenever the system is setup to consider multiple, concurrently unfolding situations, the average *AU* value will be lower, as assets will be more optimally assigned to the situations they are most needed for.

5.2.4.1 Scenario Assets

In this scenario, most of the CGAs are docked at coast guard and auxiliary coast guard stations, with a few traversing in the vicinity, as presented in Table 20, Table 21, Table 22, and Table 23.

Table 20: Coast guard assets in Somalia

Asset Type	Sensor Quality	Asset Count	Asset Location (Name or Lat, Lon)	Status
Speed boat	Low	6	Coast guard station	Docked
Speed boat	Very high	2	Coast guard station	Docked
Tug boat	Low	5	Coast guard station	Docked
Tug boat	Medium	1	Coast guard station	Docked
Tug boat	High	2	Coast guard station	Docked
General cargo boat	Low	1	Coast guard station	Docked
Slow UAV	Low	1	Coast guard station	Docked
Slow UAV	High	1	Coast guard station	Docked
Fast UAV	Medium	1	Coast guard station	Docked
Helicopter	Low	1	Coast guard station	Docked
Helicopter	High	1	Coast guard station	Docked
Fast UAV	Very high	1	9.153426, 60.350681	Not docked
Helicopter	Very high	1	13.637996, 55.188994	Not docked

Table 21: Auxiliary coast guard assets in Somalia

Asset Type	Sensor Quality	Asset Count	Asset Location (Name or Lat, Lon)	Status
Speed boat	Very high	2	Auxiliary coast guard station	Docked
Tug boat	Low	1	Auxiliary coast guard station	Docked
Fast UAV	Medium	1	Auxiliary coast guard station	Docked
Helicopter	Low	1	Auxiliary coast guard station	Docked
Helicopter	High	1	Auxiliary coast guard station	Docked
Fast UAV	Low	1	4.880430, 50.019164	Not docked

Table 22: Auxiliary coast guard assets in Yemen

Asset Type	Sensor Quality	Asset Count	Asset Location (Name or Lat, Lon)	Status
Speed boat	Very high	2	Auxiliary coast guard station	Docked

Table 23: Opportunistic response assets

Asset Type	Sensor Quality	Asset Count	Asset Location (Name or Lat, Lon)	Status
Speed boat	Low	1	9.7399, 56.587	Not docked
Speed boat	Low	1	9.0399, 54.587	Not docked
Aircraft	Low	1	14.223, 62.988	Not docked
Aircraft	Low	1	13.933, 50.771	Not docked
Aircraft	Low	1	9.0475, 50.000	Not docked

5.2.4.2 Experimental Results

This section presents the experimental results gathered by running the system with the three aforementioned configurations. The performance metrics gathered for Configuration 1 are presented in Table 24 and Table 25; the metrics gathered for Configuration 2 and

Configuration 3 are laid out in Table 26 and Table 27, respectively. A human operator looking at each of these solution sets could perform an analysis of the same form as the one presented in Section 5.2.3.2, and observe the trade-offs between the different performance metrics for each viable response. For instance, if this operator is presented with the set of viable CoAs outlined in Table 26, and his or her foremost goal is to maximize the probability of mission success, he or she would find him or herself opting for *Response 14*, due it having the highest PCDRA value. If this same operator, however, is really adamant on achieving a shorter mission, yet maintaining a relatively high chance of success, he or she can opt in for *Response 11*, and observe a minute drop in PCDRA (approximately 2%), whilst substantially lowering the mission time (by approximately 20%), at the expense of bringing in 17% more assets, thus, in turn yielding a much costlier mission (approximately doubling the mission expenses). If, instead, the primary goals of the mission were to minimize expenses and time, but maintain a reasonable chance of mission success, the operator could decide on selecting *Response 25* instead of *Response 14*, and experience a 92% cost saving through utilizing 56% less assets, whilst dropping the chance of mission success by about 40%, but completing it approximately 87% quicker.

Table 24: Experiment results with historical incident response data for the VID event

Response Number	ME ($)	MR	MT (min)	USA (%)	AU	PCDRA
1	8,027,566.68	5	42,341.51	4.58	1.00	2,048.19
2	7,885,982.83	5	47,491.15	5.56	1.00	1,126.82
3	8,111,060.31	5	38,833.52	8.17	1.00	2,833.84
4	1,573,602.62	5	10,950.48	11.11	0.32	1,424.56
5	3,915,815.82	5	10,648.65	11.11	0.29	2,347.69
6	3,563,138.08	5	8,824.46	13.40	0.37	667.11
7	3,245,009.30	5	10,469.73	20.59	0.39	646.71
8	3,509,015.96	5	9,229.27	21.24	0.29	3,401.81
9	1,907,264.85	5	9,413.04	21.24	0.32	1,168.59
10	1,207,260.81	4	9,336.63	11.11	0.24	1,480.63
11	3,404,189.50	4	8,094.91	11.11	0.24	3,095.73
12	2,800,506.21	3	6,678.24	11.11	0.24	1,867.62
13	3,334,146.51	2	6,486.07	11.11	0.16	1,084.32
14	1,196,198.52	1	9,100.32	11.11	0.24	33.49
15	1,186,538.17	0	10,386.65	11.11	0.24	636.02
16	2,883,963.26	4	7,865.84	15.69	0.18	3,171.38
17	2,732,420.47	3	8,422.56	14.71	0.29	657.30
18	1,269,493.99	3	8,427.85	16.34	0.18	0.00
19	336,135.06	1	1,965.70	98.04	0.11	1,181.73
20	366,901.93	1	3,009.79	92.16	0.16	2,127.68
21	407,181.68	1	2,847.78	92.16	0.16	0.00
22	410,768.79	2	1,870.10	94.44	0.16	100.70
23	412,792.35	1	2,036.48	93.46	0.16	56.71
24	469,536.18	1	2,565.30	91.50	0.13	2,141.09
25	474,201.65	1	1,830.44	93.79	0.13	0.00
26	474,942.00	1	5,077.53	84.97	0.18	1,411.85
27	478,654.49	3	3,319.32	82.03	0.16	4,551.93
28	865,143.96	1	2,209.75	88.56	0.16	0.00
29	742,958.81	1	2,222.48	90.85	0.16	211.80
30	669,577.18	2	2,349.95	86.93	0.18	654.24
31	495,459.82	1	2,467.71	92.16	0.18	0.00
32	555,995.33	2	2,654.55	93.46	0.16	1,942.44
33	625,604.47	3	2,656.17	97.39	0.16	2,802.88
34	623,066.90	4	2,705.07	93.14	0.18	4,095.74
35	895,379.71	3	2,763.63	87.25	0.13	3,357.93
36	590,809.56	1	2,769.57	89.87	0.13	5,713.06

Table 25: Experiment results with historical incident response data for the piracy event

Response Number	ME ($)	MR	MT (min)	USA (%)	AU	PCDRA
1	1,205,877.58	4	5,194.59	0.00	0.26	47.85
2	782,033.43	4	5,595.24	0.00	0.37	567.65
3	1,891,831.75	4	3,034.84	0.00	0.18	0.00
4	2,489,054.34	4	2,808.51	0.00	0.42	0.00
5	525,687.49	4	7,885.31	2.13	0.29	2,562.11
6	1,361,847.08	4	4,133.14	4.26	0.32	82.31
7	724,486.06	4	5,185.42	4.26	0.37	0.00
8	686,499.20	4	6,947.19	8.51	0.18	9,114.95
9	1,261,975.06	4	4,936.31	8.51	0.39	391.56
10	756,360.87	3	9,190.38	0.00	0.29	7,229.74
11	1,583,621.63	3	4,223.74	0.00	0.37	85.64
12	727,407.69	3	9,340.28	0.00	0.34	3,689.20
13	520,107.23	4	7,801.61	8.51	0.37	5,244.28
14	658,711.67	3	5,812.16	8.51	0.26	158.95
15	786,615.66	4	5,099.45	10.64	0.26	40.11
16	552,121.98	3	7,428.26	10.64	0.42	367.75
17	482,646.10	4	7,239.69	12.77	0.29	3,738.67
18	1,098,136.80	3	4,874.27	12.77	0.29	0.00
19	30,528.07	3	457.92	91.49	0.18	1,424.49
20	33,519.04	4	502.79	97.87	0.29	816.39
21	41,056.21	4	615.84	95.74	0.24	1,540.06
22	47,687.49	4	715.31	93.62	0.26	422.33
23	56,135.47	4	495.31	93.62	0.34	28.32
24	56,528.07	4	847.92	78.72	0.32	3,271.76
25	58,658.32	4	279.33	97.87	0.26	0.00
26	79,886.41	4	319.55	95.74	0.21	0.00
27	83,085.41	3	395.64	93.62	0.26	0.00
28	91,854.85	3	102.06	95.74	0.32	14.06
29	124,221.54	4	138.02	91.49	0.42	47.82
30	124,170.82	4	177.39	97.87	0.26	0.00
31	215,906.23	4	308.44	87.23	0.26	0.00
32	101,741.33	4	406.97	93.62	0.26	0.00
33	123,102.32	2	434.48	89.36	0.24	0.00
34	111,914.05	4	447.66	91.49	0.37	13.02
35	334,721.19	4	478.17	85.11	0.32	0.00
36	103,005.28	4	490.50	87.23	0.26	0.00

Table 26: Experiment results with historical incident response data for the concurrent VID and piracy events

Response Number	ME ($)	MR	MT (min)	USA (%)	AU	PCDRA
1	4,834,063.04	9	15,932.18	5.56	0.50	2,157.89
2	3,476,417.78	9	19,426.49	5.56	0.50	2,324.64
3	2,760,890.87	9	24,591.68	5.56	0.53	3,931.71
4	3,326,576.50	9	21,696.21	6.62	0.47	3,801.53
5	4,380,268.82	9	19,222.61	8.75	0.50	3,756.52
6	3,266,435.12	9	20,358.44	9.40	0.47	3,050.36
7	6,819,214.81	9	15,512.54	9.81	0.50	1,898.66
8	3,958,445.49	9	17,459.30	9.81	0.45	2,073.21
9	3,069,187.84	9	23,252.14	10.87	0.50	5,139.91
10	3,303,858.68	8	18,625.57	5.56	0.50	1,761.11
11	4,363,752.32	8	15,248.65	5.56	0.55	5,983.43
12	6,170,019.36	8	12,636.36	5.56	0.50	5,087.69
13	2,162,415.50	8	19,402.55	5.56	0.42	2,945.65
14	2,162,683.92	7	19,059.54	5.56	0.47	6,111.16
15	3,562,036.83	7	14,875.21	5.56	0.42	1,733.54
16	6,078,016.77	7	12,558.36	5.56	0.37	1,207.70
17	4,490,155.94	7	13,361.84	5.56	0.32	1,960.40
18	1,536,382.22	6	15,233.23	5.56	0.29	5,939.57
19	47,441.37	4	418.60	99.51	0.16	0.00
20	82,272.60	2	1,234.09	96.81	0.18	4,699.35
21	110,011.96	5	833.52	99.35	0.24	0.00
22	158,258.43	4	692.48	95.74	0.13	627.89
23	168,796.34	3	2,095.24	91.17	0.29	1,727.25
24	176,579.65	2	2,368.03	90.10	0.18	777.33
25	178,111.40	3	2,498.49	90.18	0.21	3,644.22
26	197,966.64	5	1,187.80	98.37	0.16	0.00
27	208,540.81	4	1,734.72	94.03	0.37	322.85
28	613,272.41	3	1,114.06	90.02	0.13	140.08
29	482,314.79	4	1,279.73	94.20	0.24	0.00
30	794,417.29	4	1,370.88	88.13	0.29	0.00
31	610,605.52	6	1,373.27	92.88	0.34	266.78
32	239,448.10	6	1,387.07	97.55	0.24	261.69
33	467,097.27	5	1,393.23	96.40	0.34	0.00
34	499,409.47	5	1,393.71	95.17	0.24	0.00
35	617,746.59	4	1,571.96	91.18	0.32	0.00
36	277,642.18	6	1,584.64	96.16	0.26	0.00

Table 27: Experiment results without historical incident response data for the concurrent VID and piracy events

Response Number	ME ($)	MR	MT (min)	USA (%)	AU	PCDRA
1	11,513,181.05	N/A	58,510.49	5.31	0.68	962.22
2	8,901,833.17	N/A	45,922.80	5.52	0.47	2,438.34
3	13,672,103.81	N/A	70,527.69	12.99	0.61	4,357.51
4	5,588,108.09	N/A	29,937.43	16.83	0.42	2,549.64
5	5,662,395.05	N/A	36,686.29	17.74	0.45	2,449.84
6	10,093,867.21	N/A	58,952.20	17.81	0.58	2,729.82
7	14,819,678.87	N/A	64,503.97	20.83	0.55	2,301.55
8	10,440,040.30	N/A	46,018.40	23.49	0.71	1,676.61
9	6,839,576.67	N/A	44,472.14	24.68	0.58	1,744.59
10	3,204,248.81	N/A	35,054.10	24.59	0.45	2,430.50
11	7,785,889.30	N/A	27,565.89	24.56	0.63	888.18
12	6,290,527.37	N/A	29,561.66	24.43	0.55	399.07
13	320,982.25	N/A	3,776.19	93.59	0.29	559.89
14	346,909.19	N/A	1,940.12	97.42	0.24	2,062.12
15	376,949.73	N/A	3,809.79	92.43	0.13	0.00
16	427,679.63	N/A	3,670.54	91.33	0.26	2,669.21
17	430,426.37	N/A	2,922.62	93.82	0.21	3,349.09
18	458,576.79	N/A	2,949.04	91.17	0.29	0.00
19	482,209.57	N/A	2,016.22	95.99	0.29	2,066.21
20	498,922.49	N/A	3,479.90	91.37	0.21	1,267.41
21	548,303.57	N/A	6,108.44	89.87	0.24	0.00
22	554,922.14	N/A	5,353.02	86.44	0.24	710.89
23	570,916.96	N/A	4,269.25	87.21	0.32	1,371.80
24	646,477.34	N/A	2,733.89	93.62	0.24	286.64
25	817,129.30	N/A	2,063.48	91.04	0.18	0.00
26	793,929.33	N/A	2,479.98	95.42	0.24	312.37
27	687,945.33	N/A	2,603.53	94.64	0.29	903.91
28	963,492.26	N/A	2,760.56	90.80	0.18	149.68
29	929,249.82	N/A	3,443.25	90.84	0.32	1,321.18
30	954,728.10	N/A	3,681.70	88.59	0.26	0.00
31	1,389,045.33	N/A	3,891.62	87.53	0.29	157.00
32	1,709,989.41	N/A	3,966.70	85.12	0.34	499.94
33	740,869.56	N/A	4,099.92	83.44	0.26	0.00
34	2,163,312.56	N/A	4,228.47	80.44	0.21	2,747.29
35	680,549.66	N/A	4,722.85	88.07	0.39	11.44
36	1,491,999.69	N/A	4,908.18	81.72	0.29	0.00

Figure 29 presents the normalized values for *PCDRA*, *ME*, and *MR* from Table 26. it can be observed that meeting *MRs* with a degree of 7 or more presents a 3.335 times increase in *PCDRA* values as compared to the rest of the responses (i.e. the ones with *MR* degree of less than 7); this increase comes at the expense of costlier responses – approximately 10 times more expensive on average. Both of these observations fall in line with *Expected Trend 2* and *Expected Trend 3*. This increase in *PCDRA* and *ME* for high values of *MR* is attributed to the fact that the requirements being derived from the soft data for the VID situation are to select sensors with higher qualities (due to the prevailing weather conditions), which happen to be mounted on costlier platforms.

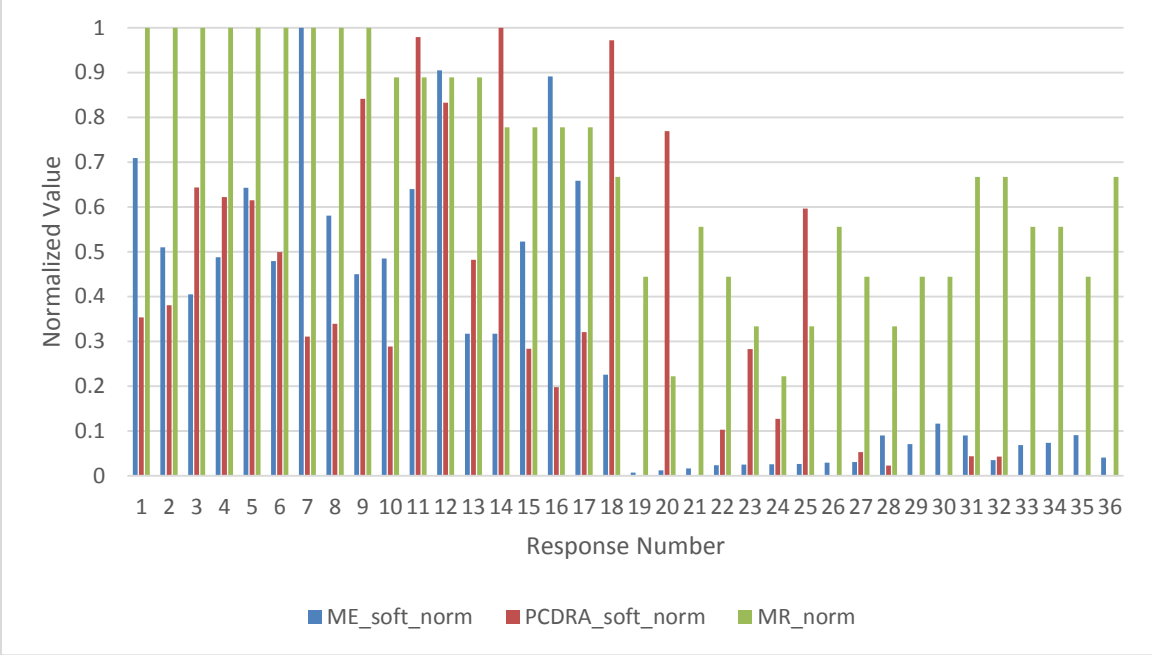


Figure 29: Normalized soft data PCDRA, ME, and MR for the concurrent VID and piracy events

Figure 30 presents the *PCDRA* values obtained by running the system with and without soft incident response data. Running the system with soft data enabled yielded a

60% increase in *PCDRA*, which translates to a substantial increase in the probability of the missions benefitting from soft data being successful. These observed results match the anticipated *PCDRA* relations outlined in *Expected Trend 1*.

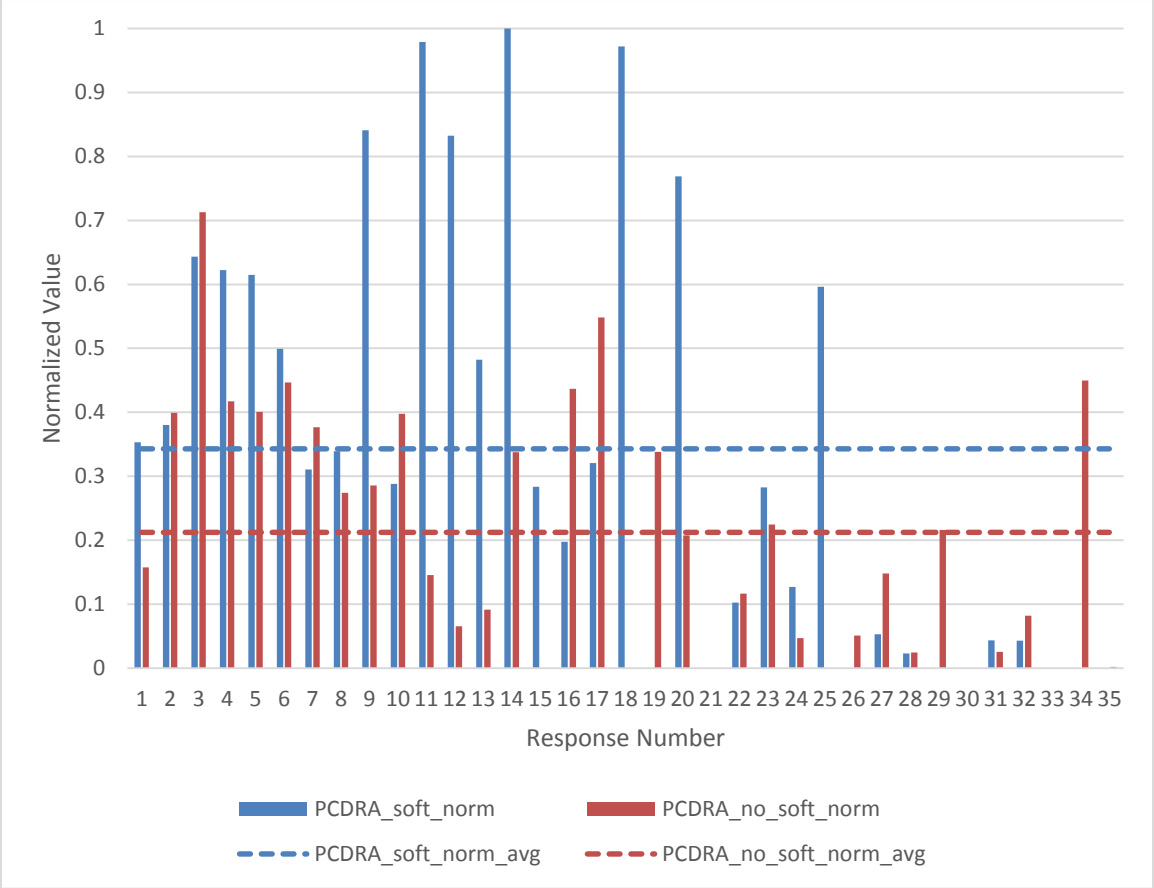


Figure 30: Normalized soft vs no soft PCDRA comparison for the VID and piracy events

Figure 31 presents the average *AU* values obtained by running the system in the three different configurations. The findings presented are in line with *Expected Trend 4*. The average *AU* in the multi-situation configuration was approximately 35%, whereas, in the single-situation VID in and single-situation piracy incident, it was approximately 27% and 30%, respectively. At best, the set of 27% and set of 30% of assets used in the two

sequential situations is disjoint (i.e. the intersection of the 27% and the 30% set of assets is the null set); this would amount to a total of 57% of average *AU* – significantly higher than the 35% average utilization in the multi-situation simulations. This desired decline in average *AU* occurring when the system is considering concurrently unfolding situations also translates into lowered average response costs – about 20% lower, as compared to the sum of the averages of the costs in the sequential, single situations. In practice however, these two sets will be rarely disjoint, meaning that if the system were only considering single situation scenarios, there would be a conflict in the selected assets between the different situations (i.e. one asset being chosen to concurrently participate in more than one situation), inevitably complicating the response selection process left to the human operator. These results thus present a tangible benefit from an *AU*, a cost-effectiveness perspective, and ultimately, a response selection perspective.

Figure 31: Average AU values for single VID, single piracy, and concurrent VID and piracy events

5.3 Chapter Summary

The experiments conducted in this chapter validate the proposed soft-data-augmented SA and CoA techniques through a series of maritime domain experiments. Furthermore, performance metrics are proposed and utilized for each of the techniques. As it pertains to SA, the proposed approach is used to proactively identify potential VID situations through the use of two real-world soft incident data sources. The proposed IFE metric is then used to demonstrate and quantify the exact benefit provided to the fusion process by each of the soft sources within these two real-world scenarios. The proposed L3 fusion (CoA) methodology is validated through two synthetically generated experiments, containing VID and pirate incidents. This methodology was demonstrated to successfully generate viable responses with a number of conflicting objectives, whilst providing a higher chance of mission success, whenever soft incident response data was utilized in the system. Additionally, it was demonstrated how the CoA methodology is able to more optimally assign assets to ongoing incidents, through its multi-situation handling capabilities.

Chapter 6. Concluding Remarks

This study focused on extending an existing RMF's L2 and L3 fusion capabilities through the inclusion of soft data for the purpose of increased SA and improved CoA generation. At the L2 level, the RMF's (SA) risk model was enhanced via the augmentation of two risk features, the *Degree of Distress* and the *Regional Hostility Metric*, with the injection of information derived from soft maritime incident reports. At the L3 level, the RMF's capabilities were enhanced by proposing a new soft data-augmented CoA subsystem architecture.

Performance metrics were introduced on both the L2 and L3 levels, and used to quantify the value brought to the system by the inclusion of this type of data. The experimental analysis conducted as part of this research demonstrated how the injection of the soft data into the two fusion levels yielded higher mission-specific measures of performance (as defined at each respective level). To the best of my knowledge, this work is the first one to apply soft data to automated CoA generation and performance evaluation in the maritime domain.

Parts of this work has been published in the form of Association for Computing Machinery (ACM) and Institute of Electrical and Electronics Engineers (IEEE) conference proceedings in [72] and [73], respectively, as well as orally presented at a Canadian Tracking and Fusion Group (CTFG) workshop[18].

[18] http://www.ctfg.ca/

6.1 Looking Forward

Although the proposed soft data-augmented L2 and L3 system has been run with real-world (both hard and soft) data, and has been observed to generate feasible solutions, it has yet to be tested out against the actual risks and threats present in a real environment. Due to the wide breadth of techniques and technologies that this system relies upon, there exists a vast landscape of potential future developments. Data analytics algorithms could be employed to analyze the soft maritime incidents in order to identify incident trends for the purpose of anticipating future attacks and thus generating proactive CoAs, so as to prevent or deter these potential attacks (e.g., in the case of anticipated smuggling or pirate attacks). The search patterns used in this research are part of the chromosome encoding, but are currently only mutated between assets; future work could entail decomposing these search patterns into collections of sub-pattern elements, and evolving (mutating and crossing over) these finer granularity elements for the purpose of generating new (and potentially more effective) patterns. In a real-world deployment of the system, it becomes imperative that the asset track generator's primary concern is asset collision avoidance; such a constraint has not yet been imposed on generated asset tracks. Another valuable improvement is to extend the MCDA framework (under which the NSGA-II algorithm resides) to include constraint handling, so as to avoid generating infeasible solutions (e.g., solutions with extended overlap in search subgrids assigned to assets). Currently, when soft data is available, the MR objective is only extracted from the closest-to-the-current-situation incident report (via the NN algorithm); however, an improved approach could be to have an asset requirements generalizer, which extracts mission requirements for the current situation from multiple incident reports, via, for instance, a kNN approach or an

Artificial Neural Network (ANN). Further automated analysis of the soft data could be performed so as to judge historical mission success levels, and attempt to identify which response elements cause missions to fail. Such response elements could then become chromosome constraints in the NSGA-II when generating suitable CoAs. For the purposes of generating suitable responses adhering to more than four mission objectives, a new category of optimization algorithms, Many-Objective Evolutionary Algorithms (MaOEAs) [75], can also be explored.

References

[1] R. Abielmona, "Tackling big data in maritime domain awareness," *Vanguard,* pp. 42-43, August-September 2013.

[2] *U.S. National Concept of Operations for Maritime Domain Awareness,* 2007.

[3] E. Blasch and S. Plano, "DFIG level 5 (user refinement) issues supporting situational assessment reasoning," in *Information Fusion, 2005 8th International Conference on*, 2005.

[4] E. Blasch, I. Kadar, J. Salerno, M. M. Kokar, G. M. Powell, D. D. Corkill and E. H. Ruspini, "Issues and Challenges in Situation Assessment (Level 2 Fusion)," *Journal of Advances in Information Fusion,* vol. 1, pp. 122-139, December 2006.

[5] R. Falcon, R. Abielmona and E. Blasch, "Behavioral learning of vessel types with fuzzy-rough decision trees," in *Information Fusion (FUSION), 2014 17th International Conference on*, 2014.

[6] J. C. Rimland and J. Llinas, "Network and infrastructure considerations for hard and soft information fusion processes," in *Information Fusion (FUSION), 2012 15th International Conference on*, 2012.

[7] A.-L. Jousselme, A.-C. Boury-Brisset, B. Debaque and D. Prevost, "Characterization of hard and soft sources of information: A practical illustration," in *Information Fusion (FUSION), 2014 17th International Conference on*, 2014.

[8] G. A. Gross, D. R. Schlegel, J. J. Corso, J. Llinas, R. Nagi, S. C. Shapiro and others, "Systemic test and evaluation of a hard+ soft information fusion framework: Challenges and current approaches," in *Information Fusion (FUSION), 2014 17th International Conference on*, 2014.

[9] D. L. Hall, M. McNeese, J. Llinas and T. Mullen, "A framework for dynamic hard/soft fusion," in *Information Fusion, 2008 11th International Conference on*, 2008.

[10] A. H. Razavi, D. Inkpen, R. Falcon and R. Abielmona, "Textual risk mining for maritime situational awareness," in *2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2014.

[11] R. Falcon, R. Abielmona and A. Nayak, "An Evolving Risk Management Framework for Wireless Sensor Networks," in *Proceedings of the 2011 IEEE Int'l Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA)*, 2011.

[12] R. Falcon and R. Abielmona, "A Response-Aware Risk Management Framework for Search-and-Rescue Operations," in *2012 IEEE Congress on Evolutionary Computation (CEC)*, 2012.

[13] C. K. Ioannis Chapsos, "Strengthening maritime security through cooperation," *IOS Press,* vol. 122, 2015.

[14] ISO, "Risk management: Principles and Guidelines," *International Organization for Standardization,* no. 31000, 2009.

[15] P. H. Foo and G. W. Ng, "High-level Information Fusion: An Overview.," *Journal of Advances in Information Fusion,* vol. 8, no. 1, pp. 33-72, 2013.

[16] J. Montewka, S. Ehlers, F. Goerlandt, T. Hinz, K. Tabri and P. Kujala, "A framework for risk assessment for maritime transportation systems--A case study for open sea collisions involving RoPax vessels," *Reliability Engineering & System Safety,* vol. 124, pp. 142-157, 2014.

[17] J. R. W. Merrick, J. R. Van Dorp and V. Dinesh, "Assessing Uncertainty in Simulation-Based Maritime Risk Assessment," *Risk Analysis,* vol. 25, no. 3, pp. 731-743, 2005.

[18] X. Tan, Y. Zhang, X. Cui and H. Xi, "Using hidden markov models to evaluate the real-time risks of network," in *Knowledge Acquisition and Modeling Workshop, 2008. KAM Workshop 2008. IEEE International Symposium on*, 2008.

[19] K. Haslum and A. Arnes, "Real-time Risk Assessment using Continuous-time Hidden Markov Models," in *Proceedings of Int'l Conference on Computational Intelligence and Security*, 2006.

[20] A. Mazaheri, J. Montewka, J. Nisula and P. Kujala, "Usability of accident and incident reports for evidence-based risk modeling--A case study on ship grounding reports," *Safety science,* vol. 76, pp. 202-214, 2015.

[21] R. Falcon, R. Abielmona and S. Billings, "Risk-driven intent assessment and response generation in maritime surveillance operations," in *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2015 IEEE International Inter-Disciplinary Conference on*, 2015.

[22] N. A. Bomberger, B. J. Rhodes, M. Seibert and A. M. Waxman, "Associative learning of vessel motion patterns for maritime situation awareness," in *Information Fusion, 2006 9th International Conference on*, 2006.

[23] B. J. Rhodes, N. A. Bomberger, M. Seibert and A. M. Waxman, "Maritime situation monitoring and awareness using learning mechanisms," in *Military Communications Conference, 2005. MILCOM 2005. IEEE*, 2005.

[24] R. Laxhammar, "Artificial intelligence for situation assessment," 2007.

[25] M. Guerriero, P. Willett, S. Coraluppi and C. Carthel, "Radar/AIS data fusion and SAR tasking for maritime surveillance," in *Information Fusion, 2008 11th International Conference on*, 2008.

[26] L. Snidaro, I. Visentini and K. Bryan, "Fusing uncertain knowledge and evidence for maritime situational awareness via Markov Logic Networks," *Information Fusion,* vol. 21, pp. 159-172, 2015.

[27] F. Johansson and G. Falkman, "Detection of vessel anomalies-a Bayesian network approach," in *Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on*, 2007.

[28] R. Laxhammar, G. Falkman and E. Sviestins, "Anomaly detection in sea traffic-a comparison of the Gaussian mixture model and the kernel density estimator," in *Information Fusion, 2009. FUSION'09. 12th International Conference on*, 2009.

[29] R. N. Carvalho, R. Haberlin, P. C. G. Costa, K. B. Laskey and K.-C. Chang, "Modeling a probabilistic ontology for maritime domain awareness," in

Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on, 2011.

[30] A. Bouejla, X. Chaze, F. Guarnieri and A. Napoli, "A Bayesian network to manage risks of maritime piracy against offshore oil fields," *Safety Science,* vol. 68, pp. 222-230, 2014.

[31] H. Shao, N. Japkowicz, R. Abielmona and R. Falcon, "Vessel track correlation and association using fuzzy logic and echo state networks," in *Evolutionary Computation (CEC), 2014 IEEE Congress on*, 2014.

[32] N. Le Guillarme and X. Lerouvreur, "Unsupervised extraction of knowledge from S-AIS data for maritime situational awareness," in *Information Fusion (FUSION), 2013 16th International Conference on*, 2013.

[33] G. Pallotta, M. Vespe and K. Bryan, "Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction," *Entropy,* vol. 15, no. 6, pp. 2218-2245, 2013.

[34] C.-H. Chen, L. P. Khoo, Y. T. Chong and X. F. Yin, "Knowledge discovery using genetic algorithm for maritime situational awareness," *Expert Systems with Applications,* vol. 41, no. 6, pp. 2742-2753, 2014.

[35] L. Vanneschi, M. Castelli, E. Costa, A. Re, H. Vaz, V. Lobo and P. Urbano, "Improving Maritime Awareness with Semantic Genetic Programming and Linear Scaling: Prediction of Vessels Position Based on AIS Data," in *Applications of Evolutionary Computation*, Springer, 2015, pp. 732-744.

[36] M. Riveiro and G. Falkman, "The role of visualization and interaction in maritime anomaly detection," in *IS\&T/SPIE Electronic Imaging*, 2011.

[37] M. Riveiro, G. Falkman and T. Ziemke, "Improving maritime anomaly detection and situation awareness through interactive visualization," in *Information Fusion, 2008 11th International Conference on*, 2008.

[38] E. Blasch, E. Bosse and D. A. Lambert, High-Level Information Fusion Management and Systems Design, Artech House, 2012.

[39] G. L. Rogova and E. Bosse, "Information quality in information fusion," in *Information Fusion (FUSION), 2010 13th Conference on*, 2010.

[40] J. Llinas, "Information Fusion Process Design Issues for Hard and Soft Information: Developing an Initial Prototype," in *Intelligent Methods for Cyber Warfare*, Springer, 2015, pp. 129-149.

[41] A. Preece, D. Pizzocaro, D. Braines, D. Mott, G. de Mel and T. Pham, "Integrating hard and soft information sources for D2D using controlled natural language," in *Information Fusion (FUSION), 2012 15th International Conference on*, 2012.

[42] M. P. Jenkins, G. A. Gross, A. M. Bisantz and R. Nagi, "Towards context aware data fusion: Modeling and integration of situationally qualified human observations to manage uncertainty in a hard+ soft fusion process," *Information Fusion ,* vol. 21, no. 0, pp. 130-144, 2015.

[43] D. McMaster, R. Nagi and K. Sambhoos, "Temporal alignment in soft information processing," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, 2011.

[44] G. A. Gross, R. Nagi and others, "Test and evaluation of data association algorithms in hard+ soft data fusion," in *Information Fusion (FUSION), 2014 17th International Conference on*, 2014.

[45] J. Gómez-Romero, M. A. Serrano, J. García, J. M. Molina and G. Rogova, "Context-based multi-level information fusion for harbor surveillance," *Information Fusion ,* vol. 21, no. 0, pp. 173-186, 2015.

[46] J. G. a. J. L. L. Snidaro, "Context-based Information Fusion: A Survey and Discussion," *Information Fusion,* vol. 25, pp. 16-31, 2015.

[47] A.-C. Boury-Brisset, "Ontology-based approach for information fusion," in *Information Fusion, 2003. Proceedings of the Sixth International Conference of*, 2003.

[48] M. M. Kokar, C. J. Matheus and K. Baclawski, "Ontology-based situation awareness," *Information Fusion ,* vol. 10, no. 1, pp. 83-98, 2009.

[49] B. Khaleghi, A. Khamis, F. O. Karray and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion,* vol. 14, no. 1, pp. 28-44, 2013.

[50] P. Smets, "Imperfect information: Imprecision and uncertainty," in *Uncertainty Management in Information Systems*, Springer, 1997, pp. 225-254.

[51] K. J. Laskey and K. B. Laskey, "Uncertainty Reasoning for the World Wide Web: Report on the URW3-XG Incubator Group.," in *URSW*, 2008.

[52] K. Premaratne, M. N. Murthi, J. Zhang, M. Scheutz and P. H. Bauer, "A Dempster-Shafer theoretic conditional approach to evidence updating for fusion of hard and soft data," in *Information Fusion, 2009. FUSION'09. 12th International Conference on*, 2009.

[53] S. Fossier, C. Laudy and F. Pichon, "Managing uncertainty in conceptual graph-based soft information fusion," in *Information Fusion (FUSION), 2013 16th International Conference on*, 2013.

[54] D. Stampouli, M. Brown and G. Powell, "Fusion of soft information using TBM," in *Information Fusion (FUSION), 2010 13th Conference on*, 2010.

[55] G. Gross, R. Nagi and K. Sambhoos, "A fuzzy graph matching approach in intelligence analysis and maintenance of continuous situational awareness," *Information Fusion,* vol. 18, pp. 43-61, 2014.

[56] B. Khaleghi, A. Khamis and F. Karray, "Random finite set theoretic based soft/hard data fusion with application for target tracking," in *Multisensor Fusion and Integration for Intelligent Systems (MFI), 2010 IEEE Conference on*, 2010.

[57] M. Liggins II, D. Hall and J. Llinas, Handbook of multisensor data fusion: theory and practice, CRC press, 2008.

[58] J. L. Graham, D. L. Hall and J. Rimland, "A COIN-inspired synthetic dataset for qualitative evaluation of hard and soft fusion systems," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, 2011.

[59] D. L. Hall, J. Graham, L. D. More and J. C. Rimland, "Test and evaluation of soft/hard information fusion systems: A test environment, methodology and initial data sets," in *Information Fusion (FUSION), 2010 13th Conference on*, 2010.

[60] I.-G. Todoran, L. Lecornu, A. Khenchaf and J.-M. Le Caillec, "Information quality evaluation in fusion systems," in *Information Fusion (FUSION), 2013 16th International Conference on*, 2013.

[61] E. Blasch, K. B. Laskey, A.-L. Jousselme, V. Dragos, P. C. G. Costa and J. Dezert, "URREF reliability versus credibility in information fusion (STANAG 2511)," in *Information Fusion (FUSION), 2013 16th International Conference on*, 2013.

[62] E. P. Blasch, R. Breton and P. Valin, "Information fusion measures of effectiveness (MOE) for decision support," in *SPIE Defense, Security, and Sensing*, 2011.

[63] D. L. Hall, M. D. McNeese, D. B. Hellar, B. J. Panulla and W. Shumaker, "A cyber infrastructure for evaluating the performance of human centered fusion," in *Information Fusion, 2009. FUSION'09. 12th International Conference on*, 2009.

[64] P. C. G. Costa, K. B. Laskey, E. Blasch and A.-L. Jousselme, "Towards unbiased evaluation of uncertainty reasoning: the URREF ontology," in *Information Fusion (FUSION), 2012 15th International Conference on*, 2012.

[65] E. Blasch, A. Josang, J. Dezert, P. C. G. Costa and A.-L. Jousselme, "URREF self-confidence in information fusion trust," in *Information Fusion (FUSION), 2014 17th International Conference on*, 2014.

[66] A. M. S. Authority, "National Search and Rescue Manual," 6 2014. [Online]. Available: https://natsar.amsa.gov.au/documents/NATSAR-Manual/Australian%20National%20SAR%20Manual%20June%202014%20FINAL.pdf. [Accessed 12 07 2016].

[67] D. Timmins, "Canadian Coast Guard National Search and Rescue Manual," 6 2000. [Online]. Available: http://loki.cgc.gc.ca/cansarp/sarmanuals/nsm.pdf. [Accessed 12 07 2016].

[68] E. Blasch, P. Valin and E. Bosse, "Measures of effectiveness for high-level fusion," in *13th Conference on Information Fusion ({FUSION})*, 2010.

[69] F. Naumann and J. C. Freytag, "Completeness of information sources," in *Proceedings of the International Workshop on Data Quality in Cooperative Information Systems (DQCIS '03)*, 2003.

[70] P. Paatero, "Least squares formulation of robust non-negative factor analysis," *Chemometrics and intelligent laboratory systems,* vol. 37, no. 1, pp. 23-35, 1997.

[71] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE transactions on evolutionary computation,* vol. 6, no. 2, pp. 182-197, 2002.

[72] A. Plachkov, R. Abielmona, M. Harb, R. Falcon, D. Inkpen, V. Groza and E. Petriu, "Automatic Course of Action Generation using Soft Data for Maritime Domain Awareness," in *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*, Denver, 2016.

[73] R. Falcon, R. Abielmona, S. Billings, A. Plachkov and H. Abbass, "Risk management with hard-soft data fusion in maritime domain awareness," in *Computational Intelligence for Security and Defense Applications (CISDA), 2014 Seventh IEEE Symposium on*, 2014.

[74] J. Li, B. Li, K. Tang and X. Yao, "Many-objective evolutionary algorithms: A survey," *ACM Computing Surveys (CSUR),* vol. 48, no. 1, p. 13, 2015.

[75] J. Llinas, S. G. Dastidar, C. Bowman and K. Sambhoos, "Achieving "Fairness" in Data Fusion Performance Evaluation Development," 2006.