



Available online at www.sciencedirect.com





Procedia Computer Science 44 (2015) 476 - 485

2015 Conference on Systems Engineering Research

Metis⁺: an integrated reference architecture for addressing uncertainty in decision-support systems

Bas Huijbrechts^a, Marina Velikova^{ab}, Steffen Michels^b, Roeland Scheepens^c*

^aEmbedded Systems Innovation by TNO, P.O. Box 6235, 5600 HE Eindhoven, The Netherlands ^bRadboud University Nijmegen, The Netherlands ^cEindhoven University of Technology, The Netherlands

Abstract

Deliver "actionable" intelligence instead of just raw information – this is what the *Metis* research project pursues for supporting operational work in domains characterized by constantly evolving situations with a diversity of entities, complex interactions and high-level uncertainty in the information gathered. Operating effectively in such domains requires robust, on-the-fly and context-based information reasoning, which goes beyond human capabilities.

In this paper a real-time reference architecture is presented employing and integrating several state-of-the-art computing technologies for automated and consolidated 'situational understanding'. In particular, outlined are the innovative components (i) for fusing of and reasoning on uncertain information based on probabilistic logic and (ii) for a complementary interactive visualization disclosing the system's line of reasoning inferred from the domain model and provided evidence. The architecture has been realized as a fully demonstrable proof of concept and its applied value is illustrated in a number of real and fictive cases from the domain of maritime safety and security.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Peer-review under responsibility of the Stevens Institute of Technology.

Keywords: automated situational understanding; actionable intelligence; uncertainty; trust; information fusion; probabilisitc reasoning

* Corresponding author. Tel.: +31 88 86 65458.

E-mail address: bas.huijbrechts@tno.nl

⁺ In the Greek mythology Metis is an Oceanid and the Titaness of wisdom and deep thought.

1. Introduction

Variety of domains, such as safety and security, are nowadays characterized by constantly evolving situations with a diversity of entities, complex interactions and high-level uncertainty in the information gathered. Operating effectively in this context requires harvesting the relevant details with the highest information gain and importance providing the right confidence for the current operational goals, which exceeds human perception and comprehension capabilities.

To date situational awareness (SA) systems are mainly developed to provide human operators with situational focus and decision-support by means of clever visualization and configurable operational pictures. However, such support systems are designed to adhere to a purely reactive workflow: it is the human operator who is still providing the operational understanding. Take the human, by far the most scarcest resource in operations, out of the loop and the system will become entirely ineffective.

The *Metis* research [1,2] pursues the opposite: the enablement of the next generation of fully-automated information centric systems for actionable situational awareness. The goal is to deliver operators "actionable" intelligence instead of just raw information. In the project promising state-of-the-art computing technologies are being researched, integrated, and validated, into a fully demonstrable proof of concept boosting future insights in advanced capabilities required for automation of situational awareness systems.

In this paper the results of the *Metis* research are presented from a system engineering and integration perspective. The main focus is on the applied value of integrating the computing technologies into one consolidated architecture possible to be used as a reference architecture for 'situational understanding' system development for similar solution domains. Outlined are, in particular, the innovative architecture components that allow the representation, fusion and reasoning of diverse, uncertain and time-related information, as well as the advanced visualization of the reasoning results in an interactive and intuitive manner.

For the *Metis* research the maritime application domain is chosen as example; an application domain that is characterized by trusted (e.g. insurance data) but sometimes outdated information on one hand, and high-frequent real-time AIS [5] sensor data, possibly subject to (un)intentional tampering, on the other hand.

In section 2 the chosen application domain is sketched out in more depth, and a number of in the project used domain cases are presented to illustrate the added value of the Metis technology to industry. In section 3 an integrated system architecture, sketching how to go from disparate data to an automated and consolidated situational understanding, is presented. Section 4 elaborates on the aspects of the core *Metis* technology, probabilistic reasoning, used for both information fusion and intent hypothesis, and on a concept for visualizing the system's line of reasoning to real humans. Section 5 elaborates on the realized proof of concept system and the first experiment results. Section 6 concludes this paper.

2. The maritime operational picture

2.1. The need for a dependable cooperative system for public safety for the maritime domain

The maritime Dutch Exclusive Economic Zone (EEZ), an area covering 154.011 km² of sea, nowadays resembles more and more a complex and complete 'society' with its own safety necessities to be addressed. As environment it is facilitating an increasing set of critical functions which all require close monitoring and an integrated management with respect to proper utilization. Among these functions are shipping (transport), oil and gas exploration, sand extraction, fishery, recreation, below surface vulnerable infrastructures (cables and pipes), wind turbine parks, land reclamation, endangered environment protection, and military use.

The North Sea is a major contributor to the thoroughfare for the Dutch economic traffic. There are, on a yearly basis, thousands of ship movements, to and from Dutch ports, as well as vessels crossing over towards ports in other countries via shipping lanes that are among the most busiest in the world. Adding complexity, the same shipping lanes are located near extreme sensitive environmental areas such as the 'Wadden Sea' which plays a crucial role in annual worldwide bird migrations. Also the North Sea is playing a growing role in the Dutch energy supply and

distribution, for both offshore industry and as 'breeding ground' for renewable forms of energy, such as wind parks and aquatic biomass. These 'grounds' are increasingly competing with shipping infrastructure for the same scarce sea space. Fig 1 presents a representative sketch of this 'societal playfield'.

Maritime safety demands for a common understanding of all relevant activities occurring at sea where, compared to land, marked delineations, physical barriers and associated fixed sensors often lack or are only virtually present. Monitoring and an integrated management to ensure proper and safe utilization require new and innovative solutions

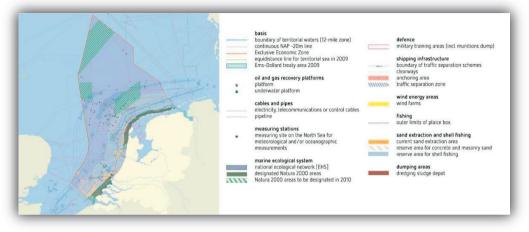


Fig 1 Overview of the current environment utilization in the Dutch EEZ, graphics courtesy Rijkswaterstaat [3]

to enable (maritime) regulatory organizations and authorities to increase the range of their effective operational awareness. However nowadays these organizations and authorities are more and more confronted with the needs for de-manning, so 'do more with less people', whereas the sheer volume of available information (sources) grows exponentially. So an intensified level of automatic, intelligible and scalable information assessment is becoming a prerequisite to operate successful.

Following subsections depict both real life and fictive, but all operational relevant, maritime use cases whereas automatic and scalable system support providing intelligent information assessment can increase the human operator's effectiveness. These in-the-research used cases are sketched out to show the potential (real-life) value of systems, such as Metis, able to discover and track actionable information in presence of uncertainty.

2.2. Case I, tracking alleged smuggling of combat helicopters to Syria

On June 16th 2012, the 'MV Alaed', a Russian-operated cargo vessel under Curaçao (Dutch Antilles) flag, was thought to be sailing through the North Sea after allegedly picking up a consignment of munitions and MI25 helicopters - known as "flying tanks" - from the Russian Baltic port of Kaliningrad with destination Syria [4]. Under the terms of the European Union arms embargo against Syria, imposed in May 2011, there is a ban on the "transfer or export" of arms and any related "brokering" services such as insurance. Now withdrawal of the vessel's insurance cover would make it difficult for it legally to dock elsewhere and could force it to return the cargo to port. Both under pressure of the US administration and UK public opinion the London based insurance company was forced to withdraw the cover after which the vessel returned to the Russian port of Murmansk.

Not long after, a similar vessel appeared on the North Sea sailing towards Syria under a Russian flag and with a different unique MMSI (Maritime Mobile Service Identity) number, which is used to identify the vessel using AIS (Automatic Identification System [5]).

<u>Operational challenge</u>: Due to missing identification information, such as name, for the Russian-flag vessel it is uncertain whether this is actually the vessel 'MV Alaed' earlier observed but with a new identity.

System support: Using additional non-identifying historical information such as vessel size and origin, the system can deduct, with some certainty, that the Russian-flag vessel is the 'MV Alaed' and that its intent of the initial

voyage likely remains unchanged. Hence, the vessel's voyage still remains interesting for further close monitoring by authorities.

Note: not breaking EU embargo conditions anymore, the vessel reached Syria and offloaded its cargo in the end.

2.3. Case II (fictive), automatic monitoring of an environmental protected area

In the area north of the Dutch 'Wadden Sea', considered being a Particularly Sensitive Sea Area so requiring severe environmental protection, tankers above 10.000 Gross Tonnage (GT) are not allowed to sail. Now a vessel, identifying itself only by its name the 'Kang Long', is detected in this area.

<u>Operational challenge</u>: Due to missing identification information, such as vessel type and gross tonnage, it is uncertain whether the vessel poses an environmental threat.

System support: Using additional information sources, the system is able to identify that the vessel can be either an oil tanker with hull id IMO: 9277890, 5.926 GT, or a bulk dry carrier with hull id IMO: 9240835, 29.935 GT. Using this information and the domain knowledge incorporated in the system's reasoning model, the system deducts that the vessel does not pose a threat with a very high probability, and thus it achieves false-positive alarm reduction.

2.4. Case III, harbor authorities dealing with ageing evidence

The English Channel is one of the busiest traffic routes of the world, both English and French authorities are dealing with over 2.000 vessel movement updates per second, but is also used for other purposes such as fishing.

<u>Operational challenge</u>: Fishing vessels 'loiter' when trawling nets, but do that mainly in fishing zones. Smugglers also 'loiter', but do so to pick up recently dropped contrabands, and do that mainly out of 'vision' in quite zones where vessels are not expected to sail under normal conditions. As contrabands are expected to be passed on further due time, smugglers are best to be caught in the act, or preferably not long after.

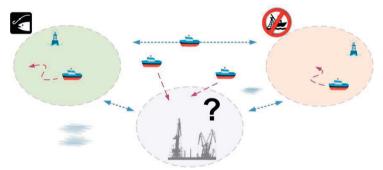


Fig 2 Outline for the harbor authority dilemma (.. now imagine monitoring hundreds of vessels continuously over time)

Harbor authorities try to monitor on smuggling by having dedicated high sensitive sensors deployed in aforementioned zones to register odd behavior, but can only take action by doing inspections on vessels returning from international waters into the harbor safety perimeter. What if multiple fishing vessels, as sketched out in Fig 2, for which loitering has been automatically detected more or less recently, enter the harbor at the same moment. Which one will be the first candidate for inspection, and why?

<u>System support</u>: Using prior knowledge on the continuous decreasing relevancy of evidence over time (for example typical contrabands offload characteristics), the system can automatically reason on the probability of each vessel being a threat. Using the probability the system is still able to weigh the evidence against other relevant factors instead of just ignoring it above a certain threshold time.

2.5. Case IV, dealing with information on ownerships changes

It is known that a vessel has been involved in illegal activities 5 years ago.

<u>Operational challenge</u>: How relevant is this piece of information when it is taken into account that vessels themselves (being just dull assemblies of steel) are never the real offenders. It is known that the crew is responsible for real behavior, that ownership is typically subject to a high frequency change (as depicted in Fig 3), and new owners hire typically fresh crew.

System support: Knowing the vessel name's age, the system is able to determine whether recorded historical behavior is still relevant: the name change / ownership change is not always registered but information on name changes is mostly available, and ownership changes are typically reflected in name changes. So a prior distribution as depicted in Fig 3 could be used to drive a system's reasoning to determine the relevancy of the information.

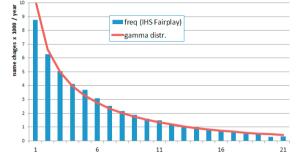


Fig 3 Frequency of vessel name changes, within years after the last name change (measured on the full population of 40.077 of 100+ GT vessels as registered in IHS Fairplay [6])

3. An integrated information system reference architecture

In Fig 4 the result of the *Metis* pursuit for shifting from 'mental reasoning' to 'automated reasoning' is depicted. In the diagram, of what is considered being representative for typical context-aware situational understanding systems, multiple artificial intelligence (AI) technologies are given a functional purpose.

In detail, based on (1) prerequisites such as a 'mission' or 'aim' the system will (2) collect data from its context, typically a wide variety of heterogeneous information sources with diverse characteristics. This data is correlated and harmonized into a semantic homogeneous and unambiguous information model (not explicitly presented in the diagram) including meta aspects such as time, provenance, trustworthiness of the source, etc. For unstructured, typically human readable, information sources such as (maritime) news publications, natural language processing (NLP) and relevancy extraction and clustering techniques are used to come to meta-level data such as machine representations for timelines of events [7].

These information elements require (3) fusion with respect to their disparate sources, and operational reasoning on high-level domain hypotheses. Both aspects are described in more depth in section 4. Within chosen application domain (maritime safety) so called 'intent hypotheses', such as smuggling and piracy, are pursued. As the system is explicitly reasoning on such high-level hypotheses, the situational awareness is promoted to a sense of 'situational understanding', which is (4) visualized in a (mission) operational picture emphasizing the domain specific object (vessels) that require most attention. For *Metis* the latter is done rather straight forward by representing the vessel using a glyph visualizing the for-the-operator-most-relevant probability distribution, and augmenting glyph in size proportionally with the level of attention it requires (see Fig 7).

To ensure a correct level of 'situational understanding' these systems require, under the hood, some sense of self-reflection, for which the system needs to be aware of the system's quality of information and the system health with respect to its information flows. These aspects are typically represented by concepts such as trust in information, belief and confidence in understanding the reasoning on information and could be implicitly inferred in the reasoning models. Within *Metis* self-reflection is expressed by endorsing the mission operational picture (visualization) with the uncertainty on the understanding in the vessel glyph tooltip (see Fig 7). Furthermore using (5) human machine interfaces (HMI) operator(s) can improve the system's understanding and self-reflection. Within

the realized proof of concept system this is addressed by providing the system's reasoning rationale, described in more depth in section 4.2.

Next to the systems self-reflection capabilities, which could facilitate (6) an automatic feedback loop into the system's information reasoning, the system requires, (7) based on its perceived system health, some level of reconfiguration of its context perception. This is done by (8) dynamically improving and managing the data collection feed towards the system's information reasoning. This dynamic re-configuration addresses the quality of service properties of the system, so it is addressing the correct and optimal usage of information resources. Based on the system's configuration (incl. mission) the data collection is re-configured so that a satisfactory level of understanding can be achieved using a strategy of selectively collecting information only from a subset of the sources instead from all sources [8]. Selection of a proper subset is based on redundancy in information between the information sources. Furthermore the scalability aspects of the system are addressed here as (i) prioritization is used to focus the system's information collection on the information for which the highest information gain is to be expected and/or (ii) information overload of the system reasoning is avoided by filtering on expected information relevancy.

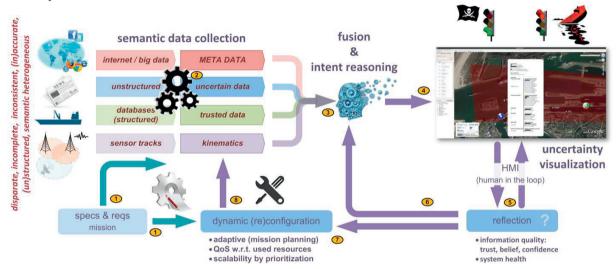


Fig 4 The Metis 'situational understanding' reference information architecture

4. Automatic system reasoning addressing uncertainty

If one would define the typical aspect of human reasoning upon uncertain information the following characteristics would emerge:

- ability to deal with **trustworthiness**, even when knowing that the information source is not always fully trustworthy, the relevancy of the evidence can be weighed;
- using approximation due to **laziness**, when setting up a complete theory in the domain of interest is just too exhaustive;
- applying **ignorance**, when a complete theory in the domain of interest simply lacks, or if present but one might be uncertain about particular objects or events because not all necessary evidence is available.

4.1. Representing and reasoning under uncertainty

In section 3 the need for alignment and fusion with respect to disparate heterogeneous sources, and operational reasoning on high-level domain hypotheses, in a domain characterized with a dynamic number of observations as

evidence, was sketched out. As solution a human like approach in reasoning is pursued by opting the use of **probabilistic logic reasoning**, in machine learning known as *statistical relational learning*.

For the *Metis* system a reasoning engine was constructed using expert knowledge representing the relationships between objects (vessels), their (identity) attributes and observed / intended behavior. Based on the heterogeneous nature of and uncertainty in the reported information, as well as the dynamic number of reporting sources, a *first-order probabilistic logic* approach was chosen as effective modelling technique. Its advantage is the combination of powerful knowledge representation provided by first-order logic rules with the explicit expression of uncertainty in terms of probabilities [14]. It was extended with discrete random variables to handle probabilistic uncertainty known as 'distributional clauses' [11]. Support for continuous random variables have been added later on to explicitly address the continuous property of some of the modelled information, for example ageing of relevancy.

The core of the expert knowledge captured in the reasoning models is twofold (see Fig 5). First, the **prior information** has to be modeled, so the (statistical) view representing the normal real-world behavior. For example

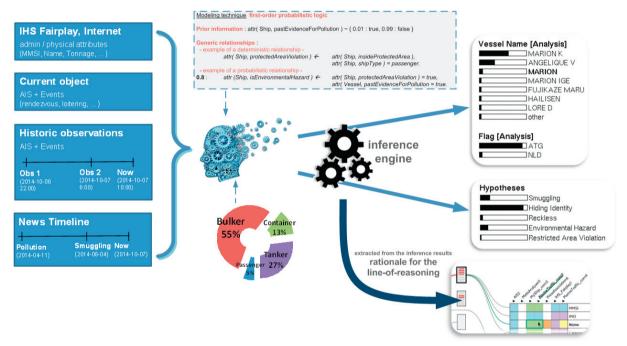


Fig 5 Metis reasoning engine based on probabilistic logic models

how ship types are distributed (bulker 55%, container 13%, ..) under normal conditions for the monitored domain. Typical knowledge as depicted in Fig 3 will be modeled as prior information. Second, **generic relationships**, either expressed deterministic or probabilistic, are used to capture knowledge about objects, their attributes, high level hypotheses (such as possible intents), and information reported from various sources. A strong feature of this modeling approach is the capability to perform information fusion by modeling the relationships between the reported information from various sources and the true object's attributes. In addition, the expressive nature of the probabilistic logic language developed and applied in the *Metis* system allows the use of variety of data types (discrete, continuous, binary) as well as the representation of temporal information at different granularities.

Once a reasoning engine is built it can be employed to perform inference, i.e., answer questions of interest, called queries, for example 'is the vessel involved in environmental pollution?'. This is done as follows: for a certain object (vessel) of interest the system collects relevant information from the available information sources, which is fed to the probabilistic reasoning engine as evidence to answer queries. Missing values in the evidence are handled using marginalization and the prior information. The success probabilities of the queries are computed by summing

of the probabilities of all possible choices for the random variable values under which the query can be derived. Exact **inference** is done by translating the distributional clauses to *Problog2* [12] as was prescribed in depth in [10].

4.2. Exposing the system's line of automatic reasoning

It is expected that any human operator will perceive the automatic reasoning mechanism described in section 4.1 as a black box mechanism that cannot be trusted. Exposing the system's line of automatic reasoning is therefore considered to be a critical step in successfully introducing any automatic reasoning technology in the operational field. To achieve this an interactive reasoning rational visualization was designed [13,15] in Metis as shown in Fig 6. The visualization provides an intuitive and self-explanatory insight into the system's line of reasoning so that a human operator can confirm or reject the system's assessment at once.

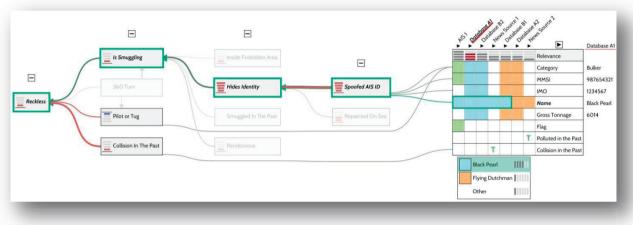


Fig 6 Visualization of the system's reasoning rationale

The visualization consists of two parts: On the left is the explanation graph, which depicts how the relevant operational hypothesis are connected in the domain model, and how each hypothesis influences others (by the thickness of the arrows) for a specific case. Hypotheses that are not relevant for the given conclusion are faded into the background. On the right is the evidence matrix. Each column represents an observation and each row an attribute. The colors are used to show where the observations agree and where not, providing an instant grasp of where information sources conflict.

In the example in Fig 6 there is obviously some inconsistency between the unique MMSI and IMO identifiers, which is directly reflected by a high probability for "Hiding Identity", which again only results in a minor contribution for the "Is Smuggling" operational hypothesis.

5. Proof of concept

The *Metis* system architecture is realized in a proof of concept (see Fig 7) which was successfully deployed as a plugin for Thales' command-and-control industrial platform *Tacticos*. Using a realistic virtual traffic generator developed in the *SeaBILLA* project [9] the system is driven hard on available (synthetic) ground truth data and established operational models, to benchmark and validate the investigated Metis technologies.

For experiments outside the restricted industrial setting the system was virtualized as a software-only stack (see Fig 8) where commercial and industrial confidential component functionalities are adequately replaced by open source components. Within this realization the focus was purely on the reasoning on non-kinematic vessel attributes. Driven by a real-time AIS data (approximate 1.5 Gb of real-time AIS messages / day) distributed by the AIS Hub (www.aishub.net) an experiment was set up to continuously monitor all vessel activities within the Dutch EEZ. Note

that no ground truth is available for such real-time AIS data. During the summer of 2014 the experiment was run and improved in cycles of approximate one week duration, assessing from 7k up to 12k vessels per week.

Main objective of the experiment was to get better confidence for the real operational value for the utilized technologies, and to validate and continuously improve by learning the domain reasoning models. Even though the experiments resulted in learning with respect to gaps in the used domain knowledge, only a few really interesting cases were found up to now which after manual analysis mostly proved to be false positives. A root cause analysis concluded that the used domain reasoning models still require to be matured with realistic operational knowledge before becoming effective.

6. Conclusions

To date, uncertainty has not methodically been addressed by decision-support systems. Uncertain information sources are either ignored being too uncertain, or included as being true when exceeding some predefined threshold. This paper described a consolidated reference architecture for systematic and automatic handling of uncertain information. It presented an approach for integrating technologies addressing most of the research challenges, that have been stated in the CSER 2013 *Metis* paper [2], to be applied in application areas such as surveillance, marketing, healthcare, machine diagnostics, forensics, semantic web (social media), etc.

Future directions in the Metis research will be focusing on the scalability of the approach on one hand, and on translating the results to the industry, for example using Domain Specific Languages (DSLs) to enable easier capturing of domain expert knowledge, on the other hand.

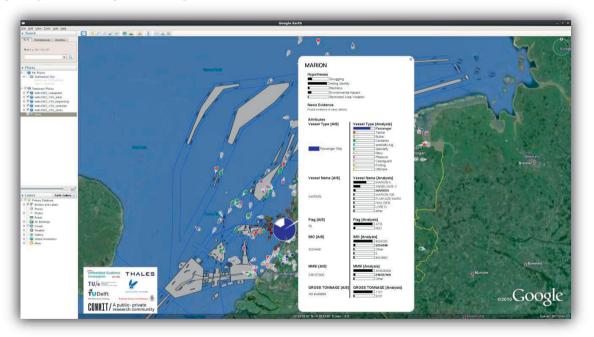


Fig 7 Screenshot of the operating Metis proof of concept (incl. glyph tooltip with uncertainty distributions)

Acknowledgements

The research is carried out as part of the *Metis* project under the responsibility of the Embedded Systems Innovation by TNO (TNO-ESI) with Thales Nederland B.V. as the carrying industrial partner. Participating in the research are the Radboud University Nijmegen (RU), Eindhoven University of Technology (TU/e), Free University Amsterdam (VU), and Delft University of Technology (TUD). The *Metis* research is supported by the Dutch national program COMMIT/.

This publication is the result of the work of the *Metis* project team; thanks for numerous valuable discussions and contributions go to Jesper Hoeksema (VU), Peter Novák (TUD), Jan Laarhuis and André Bonhof (Thales Naval Systems), and Dave Watts, Teun Hendriks and Piërre van de Laar (TNO-ESI). Special thanks go out to Tom Regeling (Intelligent Systems / Altran) for most work in exploring and realizing the full integration of all ideas with respect to the realized proof of concept.

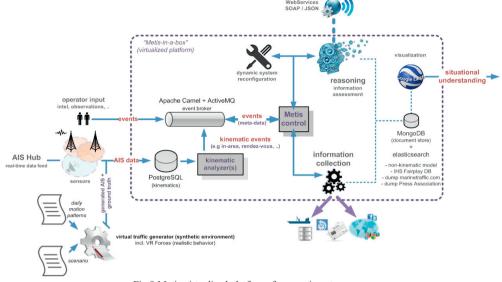


Fig 8 Metis virtualized platform, for experiment purposes

References

- 1. M. Velikova, P. Novak, B. Huijbrechts, J. Laarhuis, J. Hoeksma and S. Michels, *An Integrated Reconfigurable System for Maritime Situational Awareness*. In proceedings of the conference on Prestigious Applications of Intelligent Systems (PAIS), 2014.
- T. Hendriks, P. van de Laar, Metis: Dependable cooperative systems for public safety. In CSER, eds., C.J.J. Paredis, C. Bishop, and D. A. Bodner, volume 16 of Procedia Computer Science, pp. 542–551. Elsevier, (2013).
- Dutch ministry of infrastructure and the environment (Rijkswaterstaat), Policy Document on the North Sea 2009 2015. www.rijksoverheid.nl/bestanden/documenten-en-publicaties/brochures/2010/08/12/beleidsnota-noordzee-2009-2015engels/12pd2010g202.pdf
- Sunday Telegraph, US enlists Britain's help to stop ship 'carrying Russian attack helicopters' to Syria. www.telegraph.co.uk/news/worldnews/middleeast/syria/9336170/US-enlists-Britains-help-to-stop-ship-carrying-Russian-attack-helicoptersto-Syria.html
- 5. AIS (Automatic Identification System), en.wikipedia.org/wiki/Automatic Identification System
- 6. IHS Fairplay Maritime Intelligence, www.ihs.com/products/maritime-information/ships/world-register.aspx
- J. Hoeksema, W.R. van Hage, A case study on automated risk assessment of ships using newspaper-based event extraction, Detection, Representation, and Exploitation of Events in the Semantic Web. 1123, 42–52, (2013).
- 8. P. Novák, C. Witteveen, On the cost-complexity of multi-context systems. CoRR, arXiv:1405.7295, (2014). http://arxiv.org/abs/1405.7295.
- 9. SeaBILLA "safe waterways" project, www.seabilla.eu
- S. Michels, M. Velikova, A. Hommersom, P.J.F. Lucas, A decision support model for uncertainty reasoning in safety and security tasks (2013). In IEEE International Conference on Systems, Man, and Cybernetics, pages 663–668.
- B. Gutmann, I. Thon, A. Kimmig, M. Bruynooghe, and L. de Raedt, *The magic of logical inference in probabilistic programming*. Theory and Practice of Logic Programming, 11, 663–680, (2011).
- J. Renkens, D. Shterionov, G. Van den Broeck, J. Vlasselaer, D. Fierens, W. Meert, G. Janssens, and L. de Raedt, Problog2: From probabilistic programming to statistical relational learning. In NIPS Probabilistic Programming Workshop, (2012).
- 13. R. Scheepens, S. Michels, H. van de Wetering and J. J. van Wijk. Poster: Rationale Visualization for Decision Support. At IEEE VIS 2014.
- 14. S. Michels, A. Hommersom, P.J.F. Lucas, M. Velikova and P. Koopman, P. (2013, August). *Inference for a new probabilistic constraint logic.* In Proceedings of the Twenty-Third international joint conference on Artificial Intelligence/ (pp. 2540-2546). AAAI Press.
- R. Scheepens, S. Michels, H. van de Wetering and J. J. van Wijk. *Rationale Visualization for Safety and Security*. In review for the Eurographics Conference on Visualization (EuroVis) 2015.