Experiences and Challenges in Automated Support for Intelligence in Asymmetric Operations

JOACHIM BIERMANN
PONTUS HÖRLING
LAURO SNIDARO

This paper presents some experiences and findings of the NATO RTO Task Group on Information Fusion in Asymmetric Operations. It briefly describes the functional processing steps in military intelligence presenting the underlying aspects of information processing and fusion and revealing main challenges for automatic support of the required functionalities. The extraction and structuring of relevant information from unstructured text documents is shown to be one of the fundamental steps where human operators need assistance. As an example of the state of the art the interactive tools PARANOID and CoALA are presented. They provide the basic information and knowledge structure for all subsequent information processing like Link Analysis and Social Network Analysis. The use and benefit of CoALA will be illustrated by results from a military experiment. Finally, with respect to further research, open questions and new approaches for the support of intelligence production are discussed concerning automatic information structuring and discovery as well as pattern and behaviour based threat assessment. In relation to pattern based threat assessment a third tool, called Impactorium that is developed for threat assessment in military as well as in civilian environments, will be briefly described.

Manuscript received August 26, 2010; released for publication February 8, 2012.

Refereeing of this contribution was handled by Chee-Yee Chong.

Authors' addresses: Joachim Biermann, Dept. for Sensor Data and Information Fusion, Fraunhofer FKIE, 53343 Wachtberg, Germany, E-mail: (joachim.biermann@fkie.fraunhofer.de); Pontus Hörling, Dept. for Information and Aeronautical Systems, Swedish Defence Research Agency (FOI), SE-164 90 Stockholm, Sweden, E-mail: (hoerling@foi.se); Lauro Snidaro, Dept. of Mathematics and Computer Science, University of Udine, Udine, Italy, E-mail: (lauro.snidaro@uniud.it).

1557-6418/13/$17.00

1 INTRODUCTION

The conduct of intelligence is an essential task not only in military command and control but also for homeland security and disaster management. A most accurate awareness of the actual situation, including an assessment of the potential development and threats, is essential prior to all decisions and own activities. An intelligence cell needs the capability to collect, process, and disseminate a wide variety of data and information produced by the full spectrum of technical sensors, human intelligence, i.e. intelligence gathered from humans via observations, interviews etc. (HUMINT), and socio-political sources. There are a number of major challenges for the conduct of intelligence: first, there is a danger that the processing capability will be overtaken by the sheer volume of information that is available in very large quantities and various formats. Second and especially true for asymmetric threats, by its nature the collected information and knowledge mainly are unstructured, typically provided as text documents. Therefore, as an inevitable precondition for being processed automatically, relevant information aspects have to be extracted and structured efficiently so that this type of input can be readily and efficiently exploited for all of its intelligence value without loss of rigor [6] [29]. The urgent requirement for reasoning methods and procedure which give automated support to the further analysis and integration of structured semantic information defines a further challenge. Shortcomings in the ability to make deductions about missing and conflicting information and the current inability to support automatic context based correlation and reasoning about vast amounts of information are drawbacks to providing a coherent overview of the unfolding events.

This paper, extending the concepts presented in [8], describes some results and findings of a series of NATO Research and Technology Organization (RTO) Task Groups on Information Fusion of which the authors are members. By revisiting the intelligence process with particular attention paid to collation and analysis, the requirements for automated support are exposed and examples of existing solutions are presented.

1.1 Structure of the paper

Section 2 will explain the main steps of information processing in intelligence and explains some of the major challenges with respect to support and automation. A short description of heuristic human intelligence processing is given and two main aspects for support are presented. Section 3 then introduces two tool suites for automatic Collation and Link Analysis which focus on the ideas for support presented in Section 2. Finally, some findings from a military trial testing on one of the tools are presented discussing its benefit to the military users. Section 4 and 5 discuss further aspects of intelligence processing which are still unsolved with

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [1] □ (XXX) □ 10-16-13 □ 12:10 PM

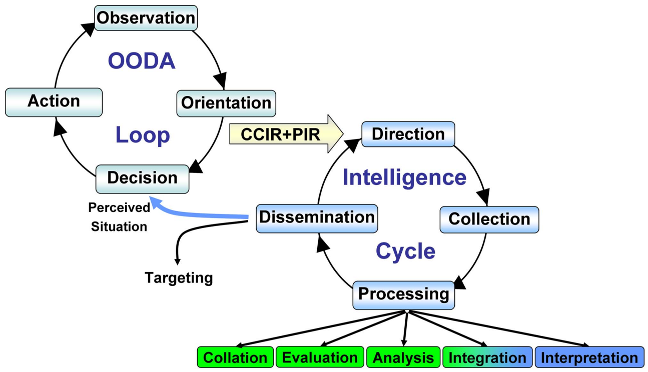

Fig. 1. Intelligence Cycle interfacing with OODA Loop

respect to automation, shortly referring to a third tool. In Section 6 we conclude our findings.

2 THE INTELLIGENCE CONTEXT

Intelligence processing is an important part of Command and Control (C2) because a completely accurate situational awareness of the situation is essential prior to all decisions and successful activities. In order to fulfil the requirements of all the various users in the military area and to provide an appropriate picture of the Area of Operations or Interest in the most timely and reliable fashion, intelligence cells have to process and evaluate all incoming information. A wide variety of information produced by the full spectrum of sensors and human sources has to be collected, filtered, processed and disseminated. The final goal for intelligence is often to provide, roughly speaking, a decision support for assessing our room for manoeuvre in the current situation. So, besides giving timely situational awareness, these working processes are also expected to give well founded assessments on opportunities for own forces, people and the infrastructure we need to protect as well as threats and risks against them. This will be discussed more in detail in Section 5.

2.1 Process flow and functional steps of the Intelligence Cycle

The processing of information for the production of intelligence is performed in a structured and systematic series of operations which is called the Intelligence Cycle. It includes four stages, *Direction—Collection—Processing—Dissemination*, which are defined by the NATO Glossary of Terms and Definitions (AAP-6) [33] as follows:

Direction: "Determination of intelligence requirements, planning the collection effort, issuance of orders and requests to collection agencies and maintenance of a continuous check on the productivity of such agencies"

Collection: "The exploitation of sources by collection agencies and the delivery of the information obtained to the appropriate processing unit for use in the production of intelligence"

Processing: "The production of intelligence through collation, evaluation, analysis, integration and interpretation of information and/or other intelligence."

Dissemination: "The timely conveyance of intelligence, in an appropriate form and by any suitable means, to those who need it."

These four discrete stages are conducted culminating in the distribution of the finished intelligence product. The representation of the military intelligence function in Figure 1 is strongly connected with the so called OODA (Observe, Orient, Decide, Act) Loop as intelligence is an integral part within the military command and control cycle. Boyd introduced the notion "O-O-D-A" and he stated "The process of observation-orientation-decision-action represents what takes place during the command and control process—which means that the O-O-D-A loop can be thought of being the C&C loop." [3]. The intelligence effort is "Directed" by the Commander's Critical Information Requirements (CCIR) from which his Priority Intelligence Requirements (PIR) are derived. Eventually providing the military commander with a most timely and comprehensive situational picture the intelligence cycle supports both the *Orientation* and the *Decision* phase.

The Processing phase is the most essential part with respect to information fusion issues. It is defined as "The production of intelligence through *Collation*, *Evaluation*, *Analysis*, *Integration* and *Interpretation* of information and/or other intelligence." [33]. It is a structural series of actions where the information, which has been collected in response to the directions of the commander (CCIR, PIR), is converted into intelligence products. A more detailed discussion on the principles of heuristic intelligence processing can be found in [6]. It is here, that the intelligence staff needs automation to be more effective in its work. In cooperation with an international group of military experts and based on realistic asymmetric scenarios, the established heuristic procedures of intelligence processing have been analysed to understand the approach of the human experts and their cognitive processes in order to adapt their reasoning principles and methods for automated fusion concepts and procedures.

Figure 2 [31] illustrates the relation of the different processes supporting the Intelligence Cycle (shown in Figure 1) now organised as functional flow and having its focus on the Processing phase. Specifically, Collation is presented in more detail and the steps Analysis, Integration and Interpretation are grouped and renamed to "Link Analysis" which is, in some nations, how Analysts name their job. As mentioned before, the CCIR and other information requirements of the commander and his staff initiate the intelligence processing (see ① in Figure 2). Incoming information first has to be digitised, if necessary, logged and stored into a data management system. This part is covered by ② in Figure 2. The main function of such document management relates to the

Mt2 job no. 2172 jaif Journal of Advances in Information Fusion 2172D02 [2] (XXX) 10-16-13 12:10 PM

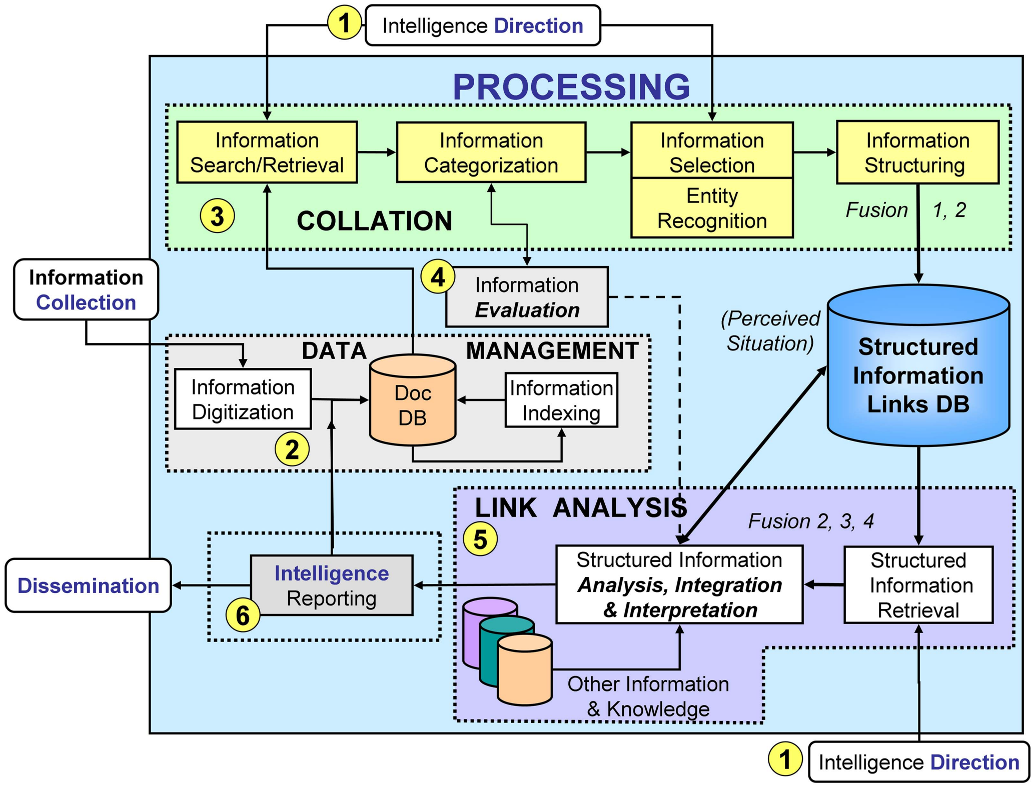

Fig. 2. Intelligence processing from unstructured to structured information

ability of registering and storing structured and unstructured documents in a document database, and of discovering knowledge. The function of knowledge discovery refers to the different ways of searching and retrieving information from large information sources with interactive capabilities of guiding the user through the process. It exploits structures such as semantic networks, ontology, and meta-data to establish links between domain models and information sources, and helps users to find relevant information. These functions directly support the Collation process described below (see also Section 4).

During the *Collation* process (indicated by ③ in Figure 2) information is decomposed into individual items which are grouped according to categories relevant to the context of the mission and cross-referenced with previously processed information items.

From an operational context, it is known that especially in asymmetric operations much of the incoming information is to be found within text documents and is often not in a format suitable for machine manipulation. Therefore, any automated support of the collation step essentially requires the extraction of relevant information from incoming unstructured pieces of semantic information as well as the structured representation of these newly processed information items. One of the main purposes and benefits of the tools described in chapter 3 actually is to support this information structuring (see section 3.1.3).

The *Evaluation* of the reliability of sources and the credibility of collected information is done by intelligence analysts as soon as the relevant information has been extracted and can be annotated directly as a tag to the piece of information or the document (④ in Figure 2). In the context of semantic information from HUMINT and Open Source INTelligence (OSINT) sources, newspapers, journals, the web, blogs, twitter etc., evaluation is most often a very experience-based task with highly subjective results. For these reasons *Evaluation* was not regarded to be done or supported automatically.

Analysis: "…information is subjected to review in order to identify significant facts for subsequent interpretation" [33]. It consists of a number of interacting sub-processes resulting in the analyst answering questions like: "Who/What is it?," "What does it mean?," "Why is it happening?" etc. in order to recognize indicators and warnings.

Integration: "…analysed information or intelligence is selected and combined into a pattern in the course of the production of further intelligence" [33]. It is the process of building pictures of the current and of the predictive situations from all the gathered and analysed information.

Symbol ⑤ indicates where Analysis and Integration of information are conducted. In practise they are very often performed as one combined step and they are not conducted as separated parts of the overall process flow. It is here that intelligence is produced and the fusion of

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [3] □ (XXX) □ 10-16-13 □ 12:10 PM

information takes place. The notion "Fusion 1, 2" in ③ and "Fusion 2, 3, 4" in ⑤ used in Figure 2 refers to the level of data fusion as it is defined by the data fusion model of the US Joint Directors of Laboratories (JDL) [28].

A further important requirement for an intelligence processing system is to be able to support link discovery and analysis. This approach (compare so-called "Storytelling" [44]) depends on the capacity of the system to automatically or semi-automatically allow the identification of a specific object of information and all of its related categories such as the location, the time, the cause, the originator, the subject, etc. Once those links are enabled, identified and validated, analysts will obtain a better and more focused image of the situation. Disparate pieces of information that had little or no value when considered independently could have a whole new meaning when combined and linked to form a pattern. Link creation is carried out during the "Information Structuring" process found in ③ and link discovery and analysis during the "Structured Information Analysis and Interpretation" process found in ⑤ with the information stored into the Structured Information Links Data Base shown in Figure 1. Link analysis is a capability that can support both the collation and analysis processes. We will discuss link discovery and analysis further in Section 3 of this paper.

To summarise: by categorising, classifying, indexing and cross-referencing all information appropriately the intelligence organization avoids losing important information and context. Disciplined and methodical collation enables further analysis to be efficiently performed using link analysis among other techniques. Information systems support this approach depending on their capacity to automatically or semi-automatically allow identification of specific information and all of its related categories such as the location, the time, the cause, the originator, the subject, etc. Once those links are identified and validated, analysts are given better bases for understanding the different key factors influencing the overall situation. Disparate pieces of information that have little or no value when considered independently could have a whole new meaning when combined and linked together thus allowing the emergence of potential key patterns.

2.2 Challenges and main requirements for automated information processing

All the different processes shown in Figure 2 are relevant to the conduct of intelligence but the three processing steps shown in Table 1 [7] were determined as those ones which, on the one hand, are central to the conduct of intelligence, and, on the other hand, were supposed to be capable of being automated.

To be able to build a system for (semi-) automated intelligence processing and decision support incorporating these functionalities, at least the following requirements and challenges have to be met:

TABLE 1
Required functionality for automated information processing in intelligence

Step	Required Functionalities		
Collation	Semantic text extraction	Categorisation	Information structuring
Analysis	Classification Identification	Correlation	Link analysis
Integration	Pattern matching	Aggregation	Fusion

1. Semantic access to all input information

Within the *Collation* step operators have to deal with a continuum of different types of information and all available input information and data should be used for the production of a reliable and most comprehensive operational picture as a base for situation awareness. Therefore it is necessary to be able to get full semantic access to the content of all unstructured text documents. The Battle Management Language (BML) is an unambiguous language which, among others, is used to provide for situational awareness and a shared, common operational picture. It is a promising linguistic approach to structure free text information for decision support and automatic information fusion [37], [36] (see also section 2.4 for BML use in knowledge representation). For a wider use of BML in military Command & Control and simulation see e.g. [35] [10]. In Section 3 two software tools are presented which support the interactive extraction of text from documents for categorising and structuring information relevant for respective information requirements (see esp. Subsection 3.1.3). These tools support the heuristic operating procedures of the human operators.

2. Understanding human reasoning in intelligence processing

Within *Analysis* and *Integration* significant information has to be found within the information set which is provided as the result of the *Collation* step. This significant information has to be put together to a situational picture according to the information requests given by the respective commander. Human analysts develop an appropriate view of the theatre which means they have a mental model of all relevant aspects of own and hostile operations as well as of the activities of the environment. Link Analysis is a technique well known by intelligence analysts and other security organizations that allows for the detection and visualization of interrelated topics to help resolving the "effects-to-cause" puzzles (see Figure 3) which arises when trying to put together all pieces of fractional information to form a coherent and reliable picture of the real situation. Ongoing research in this area is discussed in Sections 4 and 5.

The analysts have to solve many different puzzles at the same time. The underlying problem is the same as in risk assessment and threat detection in civil secu-

Mt2 ☐ job no. 2172 ☐ jaif ☐ Journal of Advances in Information Fusion ☐ 2172D02 [4] ☐ (XXX) ☐ 10-16-13 ☐ 12:10 PM

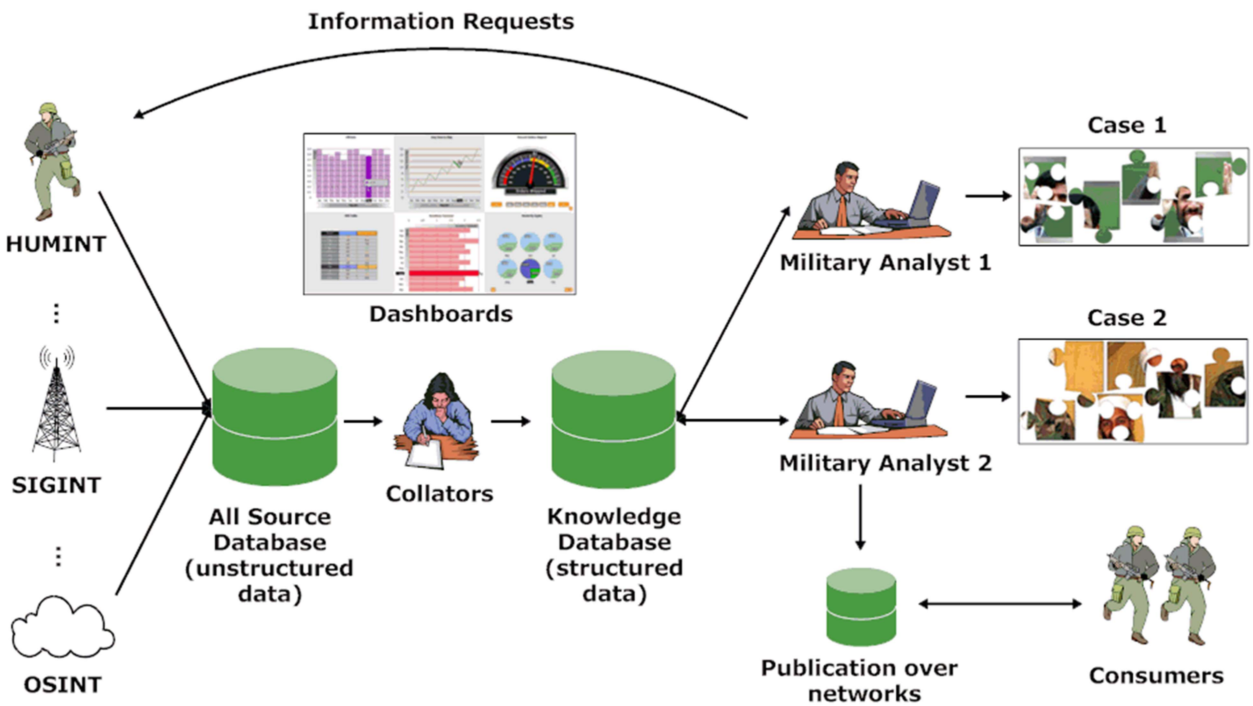

Fig. 3. Solving the many incomplete puzzles in intelligence [48]

rity. Police officers tasked to determine the structure of criminal groups, the territories or business areas they control, and the tactics, logistics and communication they use will follow the same reasoning principles as in military intelligence. Some basic aspects of cognition in heuristic intelligence processing are given in Section 2.3.

3. Knowledge representation and methods for automatic reasoning

To process semantic information from text documents the functionalities listed in Table 1 require an ontological model of the domain of application. It has to be established in a formal and structured way as a sufficiently appropriate and consistent approximation of the real situation and activities. Furthermore, reasoning methods operating on this model have to be developed that are able to cope with incomplete and imperfect information. One successful approach can be seen in using the principles of default or pattern based reasoning deducing from structured semantic information. Some arguments about this topic are given in Subsection 2.4.

There are more challenges in building support systems for intelligence processing, e.g. how to cope with assessing or estimating uncertainty in text information or dealing with deceptive or wrong information. The tools presented in this paper have not been designed to support these highly experience based and subjective information processing steps. Approaches to deal with the before mentioned questions can be found e.g. in [38] [24] [52].

2.3 Heuristic intelligence processing and default reasoning

Intelligence cells will never get all the relevant information they request, but they will be bombarded by partial, false, unreliable, irrelevant, and redundant pieces of information which they will have to filter according to the information and intelligence requirements given to them by the commander. The human brain permanently selects, relates and inserts relevant information guided by its internal, mental abstract model of the world to further develop the perception of the reality which can be understood as a concrete instantiation of the abstract world model. Human analysts use their experience gained from "similar" problems in order to fuse all data and information into a reasonable picture of the situation thus deciphering the meaning of all pieces of input data. This describes roughly the main steps and aspects of cognitive reasoning based on patterns, schemata, learning and experience as it can be found e.g. in [1]. The principles of "default reasoning" are not depending on the specific problem area but are a general human problem solving paradigm.

In heuristic intelligence processing the following constraints have to be stated with respect to the available data, information and knowledge:

- Usually only general and incomplete information is available about the structure, the activities, the situation and the intent of adversaries and other involved factions

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [5] □ (XXX) □ 10-16-13 □ 12:10 PM

- Information gained by reconnaissance is imperfect and incomplete

This means that because of lacking knowledge and information, neither the mental model of the situation ("world"), which analysts have, is perfect, nor are they provided with a perfect actual view on this world by the incoming information stream. In order to deduce a most reliable picture of the situation, in spite of these weak preconditions, analysts practise a method of heuristic reasoning which relies on the assumptions that

- Actors operating in a professional way according to their doctrines, rules, principles, modus operandi, and standards
- Effective operations are planned and organised for the benefit of the mission
- There are conditional dependencies among the phases and steps of an operation
- The real situation picture can be deduced from characteristic patterns of activity or of state (called templates or schemata)
- Templates can be recognised from significant information

It is common military experience and expert knowledge that the production of intelligence can be done by integrating current information based on the assumption of default behaviour. Behaviour modelling, such as doctrinal templating [4], is a descriptive, qualitative method of knowledge representation. The elements of the situation are not only described by their attributes but also by their relations and dynamic behaviour as well as their operational potential and assumed tactical intentions.

For supporting intelligence systems two different kinds of models are relevant [6]:

a) Behaviour models describing the tactics of potential adversary factions and all necessary pre-conditions for their hostile activities;
b) Models describing "normality," the common and unsuspicious behaviour of defined subgroups of the population or other elements and groups relevant to the situation

In case a), analysis is the task of detecting special indicators of activities or status which define by their combination a potentially evolving threat. This approach is used e.g. in low and high intensity conflicts of military or paramilitary type. Threats, like an ambush or an Improvised Explosive Device (IED) attack, are complex sequences and interrelations of different activities. Each of them having their own structure (pattern) and being combined they form the high-level pattern of the final threat.

As a consequence, a pattern to be used in automatic reasoning has to consider and incorporate all significant and characteristic factors in order to build up a template for analysis and integration. Little and Rogova [26] discuss the formal ontological structure of threats as holistic phenomenons possessing three interrelated parts: intentions, capabilities and opportunities (further elaborated on here in section 5, where we will discuss the Indicator concept). They show how these facets of a threat are related to one another, as well as to states of vulnerability. These aspects have to be covered by surveillance and reconnaissance to provide the intelligence staff with corresponding information.

In the above mentioned case b) concerning "normality," the task is to detect deviations from patterns of "normal" behaviour. Snidaro, et al. [45] give an informal definition of an anomalous event as "a deviation from common patterns of activity." This method is used e.g. in combating terrorism to be able to define indicators of suspicious activities [7]. It is an increasingly important topic for decision support, since it can give hints to the intelligence staff towards where more analysis or information is needed.

Intelligence operators often have to deal with information sources that can provide a sequence of unreliable observations or reports due to unfavourable sensing conditions or limited/erroneous projection of real-world observables. In the case of human sources, information could even be deliberately incomplete, erroneous or deceptive. Poor quality information and unreliable sources can have disruptive effects on the fusion process. For this reason an automatic Situation Assessment (SA) system for intelligence purposes should take into account both the reliability of the sources [41], and the quality of their data [40] to regulate the fusion process accordingly. While (automatically) weighting or pruning information is far from being a trivial process, these topics are being actively researched by the fusion community.

2.4 Knowledge representation for automatic information processing

In a decision support system all actual and background information, including the models of behaviour and normality, has to be processed automatically. The representation of information and knowledge can be based on an ontology of the domain. Ontology, as a semantic description of all objects and classes (or categories) and their relations, incorporates taxonomies, attributes of the objects and respective values and constraints, rules and schemata, representing the behaviour defaults [23] [43]. Schemata can be well represented as so called feature-value matrices (FVM). These are sets of features (or attributes) and value pairs. For schemata, on the top level, the features denote the thematic roles of the represented object or class and the values are feature-value matrices themselves that pool the information about the object that fulfils the respective role. From the mathematical view, a feature-value matrix is a finite set of pairs. Each pair consists of a feature and a value. A feature is always a symbol, a value, however,

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [6] □ (XXX) □ 10-16-13 □ 12:10 PM

can be a symbol or a feature-value matrix itself. In addition, the following uniqueness condition holds: every feature has a unique value. In other words, in a matrix there cannot be two pairs which share the feature but not the value. However, different features may have the same value.

These matrices have many beneficial properties. First, objects can be represented which are not specified completely. This is important for fusing partial information. Second, the matrices obey XML schemata which allow further automatic processing. Third, the matrix representation allows 'unification,' a standard computational linguistic algorithm for merging information which we regard beneficial in information fusion [43]. FVM are used beneficially as a technical representation form within the linguistic BML approach to knowledge representation.

3 TOOLS SUPPORTING INFORMATION PROCESSING

Coping with the more or less diffuse adversaries in asymmetric warfare as well as unfolding the structure of criminal networks has resulted in the introduction of several dedicated tools for intelligence analysis in the last decade. They give the ability for analysing social and semantic networks, spatio-temporal pattern analysis, all with different abilities of visualization [9]. One example of the former is Palantir [21] and Detica NetReveal. Other tools focus even more on Visual Analysis such as NetLens [18], IN-SPIRE [39] and Jigsaw [19]. However, such tools still have their limitations [44].

Two examples of existing support tools for actual intelligence processing are presented in greater detail here. The special features which relate to the before mentioned process flow and required functionality are highlighted. In Subsection 3.2 some results of a military trial on intelligence processing using one of the tools are given and the requirements of the military analysts with respect to more elaborated automatic support for analysis and integration are presented.

3.1 Interactive tools for Collation and Link Analysis

CoALA and PARANOID [48] are products of a close and intensive collaboration effort between Defence Research and Development Canada (DRDC), Quebec, Canada, and the Dutch research organization TNO Defence, Safety and Security, Den Haag, The Netherlands. They have been developed in parallel to the activities of the NATO RTO research Task Groups on Information Fusion active since 2000 and are related to the results of these groups. CoALA is based on PARANOID and it was supposed to be in operation in 2009. For different reasons the fielding had been postponed to 2010. These tool suites provide the intelligence personnel with a functionality that supports the collation of free-text documents. It does so by supporting interactive extraction of relevant information from free text source documents and storing that information to a structured database to be further analysed and related to other items of information, thus creating intelligence. In brief the general characteristics of the tools are:

- Rapid collation of unstructured text information into pertinent intelligence products
- Identification of hidden patterns and connections within information to focus analysis on counterterrorism, organised crime, threat assessment and incidents
- Collaborative collection and analysis enabled

A more detailed description of the information processing approach underlying both tools and the implemented functionalities can be found in [30].

3.1.1 PARANOID

PARANOID (Program for Analysis Retrieval And Navigation On Intelligence Data), developed by TNO. In this tool techniques for searching, storing and analysing information are being implemented and tested. This tool suite supports the process of specifying the total functionality for an operational processing system for intelligence, such that it reflects the workflow of intelligence staff. PARANOID processes information in support of Peace Support Operations (PSO), but is equally applicable to other areas such as counter-terrorism operations, the fight against fraud, and the acquisition of business intelligence.

The functions of PARANOID reflect the workflow in the intelligence process, starting with the definition of information need through to the storage of the intelligence products. Three main functional areas have been defined:

Profiles: In this function the user is able to define certain factors, such as time and space definitions, certain types of events, and particular individuals that have to be taken into account while processing the incoming information.

Documents: This function carries out a range of different operations on all incoming information. One example is the storage and transformation of structured and unstructured data from documents into a structured database, carried out by applying different information extraction techniques.

Analysis: There is a need for different types of analyses to be able to support the different sub-processes of Processing: link analysis, pattern recognition, trend analysis and threat/risk analysis. There is also a need to be able to visualise the data and results. This should be possible not only by using a geographical information system, but also through a number of innovative ways of navigating through a network of different types of related data and information.

3.1.2 CoALA

The Collation and Link Analysis (CoALA) tool is an evolutionary specialized collation tool suite for intelli-

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [7] □ (XXX) □ 10-16-13 □ 12:10 PM

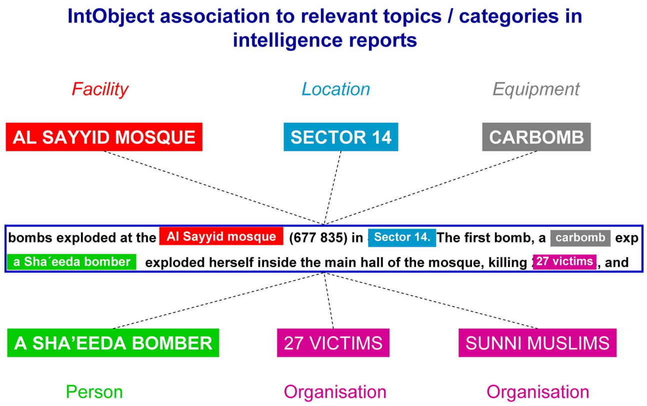

Fig. 4. Interactive information extraction and structuring

gence analysts based on PARANOID and developed by DRDC. It provides expert applications to exploit unstructured information and populate a structured intelligence database that allows detailed analysis and production of intelligence. CoALA version (1.0) provides a suite of functions that support the collation and analysis process. The key functions are as follows:

Document management: Basic document management functions such as importing, registering, storing and disposing of documents.

Information Management: CoALA includes a structured knowledge database that provides a means to record common pieces of information and intelligence in an organized fashion that support the retrieval of that information and intelligence. With respect to the amount of cases to be managed during a military mission there is a function to get detailed information about the record management. This management is closely tied to the management of Priority Information Requirements (PIR) and other Information Requirements (IR) related to the management of intelligence collection.

Data Collation: Capabilities that allow pieces of information to be related to each other, grouped in related categories, and stored into the knowledge database.

In detail CoALA provides an information categorization tool. This function is used to identify the various objects found in the text and associated them with a class of objects predefined by the user. Based on the intelligence analysts' experience, the following classes had been chosen: person; equipment, facility, organization, event and geo-object (location, map reference, coordinates).To establish a structured information set there is a function to create, manage and visualize relationships between objects in a central knowledge database.

Data Analysis: To conduct link, pattern, geospatial and temporal analysis of information and intelligence. The results are stored into the knowledge database. There is a function to analyze and create working assumptions supported by visualization of the content of the master knowledge database. This visualization is done using various tools, including:

- Charts of information objects and their interrelations (Link Chart);
- Timeline charts for events and their interrelationships;
- Matrix of links and relationships existing between various types of objects. The most common example is the matrix of what is known in an organization;
- Basic geospatial visualization of geo-referenced objects (GIS);

Intelligence production management: Simple means to capture and manage the IR/PIR list and to link the intelligence production back to it. The tool allows for any intelligence products (assessments, analytical charts, briefings and reports) to be stored in the knowledge database with references to all of its supporting material. Furthermore CoALA (v1.0) provided

- a limited printing tool.
- a function for exporting Link Charts to the commercial i2 Analysts' Notebook product.
- collaborative work in real time via a common database (MS SQL 2005).

3.1.3 Information extraction and structuring

One of the core concepts for good analysis in both tools is the collation concept: the extraction of relevant information from unstructured information into structured knowledge. The extraction of information is predominantly done by interactively tagging relevant parts of sentences from documents ("Statements") and linking them to the so called "Intelligence Objects" or "IntObjects."

Int Objects are elements of categories of domain items as Persons, Organisations, Location, Equipment and Facilities. Figure 4 shows an example of a statement (in the rectangle) that is linked to other IntObjects. The Statement contains different IntObjects that are linked in a standard way ("related to"). Figure 4 gives an example how the relationships between IntObjects, like between the Person "A Sha'eeda Bomber" and the location "Sector 14" is established by extracting and tagging the single information items.

The newly established set of structured information is to be integrated into the knowledge base (KB) which represents the so far perceived situation. The KB is searched for already existing IntObjects which are the same or may be the same as one of the elements of the newly structured information set. Figure 5 shows that two IntObjects "A Sha'eeda Bomber" and "Carbomb" are already known within the KB. They are offered to the operator to verify and confirm that the already known IntObjects in the knowledge network are identical to those ones which are part of the newly structured information set. If this is true the new IntObject structure is merged into the KB unifying the identical IntObjects which results in more comprehensive and/or

Mt2 ☐ job no. 2172 ☐ jaif ☐ Journal of Advances in Information Fusion ☐ 2172D02 [8] ☐ (XXX) ☐ 10-16-13 ☐ 12:10 PM

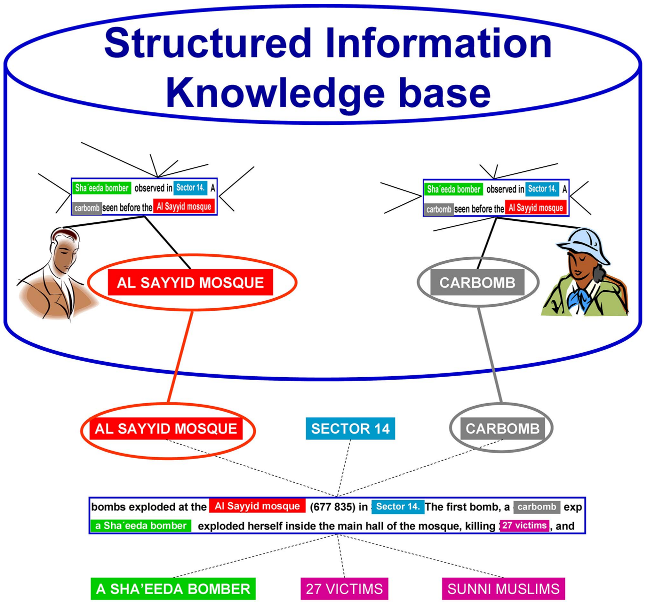

Fig. 5. Integrating the new structure into the knowledge base

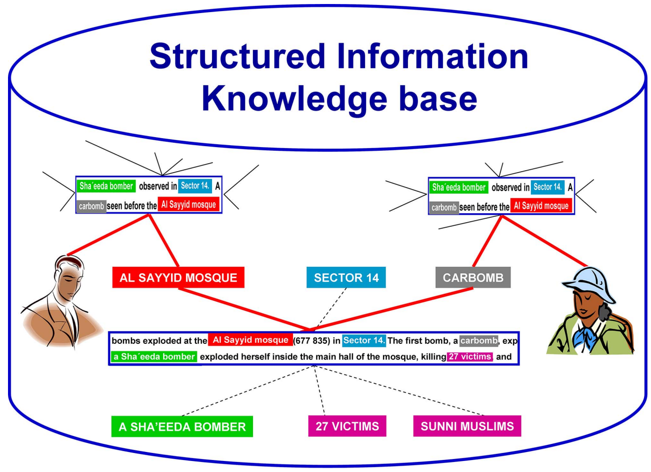

Fig. 6. The merged information reveals new relations in the knowledge network

more precise and/better confirmed knowledge about the respective IntObjects. The graphical representation of the result is shown in Figure 5. The special benefit of this information integration lies in the now established connection between the two persons, shown by the red line in Figure 6. These two actors now are related to one another by this merged-in information structure.

3.2 Expert trial on intelligence processing

The investigation of the RTO Task Group had been carried out with the support of an international group of military advisors. They focused on the structure and process flow of the conduct of intelligence, the human cognitive methods and practical procedures on how to process the collected information and available knowledge. This analysis was based on several scenarios, starting with a more conventional low intensive operation dealing with the Kosovo conflict and finally using an Iraq-type asymmetric operation. The insight gained into the main character of the conduct of intelligence did not change over the varying conflicts and the necessary steps which have to be done in the course of the production of intelligence seemed to remain the same. This is at least true for the more abstract point of view of a paper work analysis. But there was no certified and detailed information on how the processing of intelligence is carried out under real conditions by analyst experts of the intelligence branch. In particular there was only little information about the detailed breakdown and organisation of the work, the sharing of information and partitioning of responsibility.

Up to the mid of the last decade intelligence cells in operation have been using mostly standard office tools to manage information and data without any specific functionality and support for exploiting its intelligence value. It had been recognised that collators do not perform their tasks efficiently under these circumstances. They tend to transfer the burden to intelligence analysts who have to complete the collation process. In order to carry out a knowledge elicitation a Subject Matter Experts (SME) trial with domain experts coming from Afghanistan was arranged by the Canadian Forces. Six intelligence experts, using the CoALA tool suite, constituted an All Source Intelligence Cell. Their task was to work on a set of CCIRs and PIRs based on the context of a so far unknown asymmetric scenario. The intelligence trial was performed:

- to analyse whether the military understanding of the conventional processing steps in intelligence had been carried over to asymmetric operations, to observe the real workflow and processing steps of the experts,
- to observe the processing of unstructured text information carried out by experts experienced in asymmetric operations using the support of an interactive collation tool,
- to analyse the way of human deduction and main reasoning principles of the experts,
- to observe in how far and to which extent the supporting functionality provided by CoALA are accepted by military experts,
- to validate the usability, capability and potential of the interactive supporting tool CoALA, the acceptance by the intelligence experts from the Canadian Forces and to get recommendations for further enhancement and development of collation, analysis and integration functionalities.

The behaviour of the experts which was observed first was caused by the fact that during the last years, the All-Source Information Cells (ASICs) have been overwhelmed by unstructured text information. Often no IT support has been available or IT support was too unqualified or unusable for the task of structuring this input according to established and proven intelligence requirements. It therefore was natural for the *collators* in

Mt2 ☐ job no. 2172 ☐ jaif ☐ Journal of Advances in Information Fusion ☐ 2172D02 [9] ☐ (XXX) ☐ 10-16-13 ☐ 12:10 PM

our trial to take the task just to read the messages, identify important information mentioned within the messages, inform the analysts about the interesting observations and organize the messages in a way that they could be accessed easily when required. The linking of intelligence objects was only carried out by the *analysts*. They were re-reading the messages directed to them by the collators and, which from the point of view of the individual analyst, were of interest to the very Priority Intelligence Requirement (PIR) or Information Requirement (IR) he/she was working on. Therefore, at the beginning of the trial, the collators were told by the leading officer just to tag the statements and other intelligence objects, but not to link the statement to objects. This easily could and should have been done using the Collation tool CoALA to keep the connection between newly created intelligence objects and the constituting and justifying statement or message. Later on, the analysts started complaining about the fact that they only could retrieve "standalone" intelligence objects, as there had been no links established to be analysed. The analysts were almost doing the Collation process again. Therefore, after some time, the collators were asked to establish the links between processed statements and other intelligence objects. Establishing this "new" work schedule, the ASIC personnel only returned to the well defined and commonly performed procedure and work share as they used to be before the overwhelming flood of information degraded the role of the collators to just tagging the information. It was observed that PIRs were the leading factor in directing the information processing for intelligence. Processing of any information which could not be related to the list of CCIRs and PIRs was not observed. Nevertheless it would be of interest how analysts cope with developments outside the scope of interest.

Despite several difficulties the SMEs encountered during the trial, they were able to reach their "operational" goals. They answered the CCIRs and they were able to give detailed recommendations and assessments on the CoALA functionality, although they might not have fully experienced and tested the full potential of the CoALA concept of information extraction and structuring.

The semi-automated tool functions which support the extraction and structuring of information are going far beyond the low level requirements of the NATO AAP-6 [33] Collation step definition, which only claims for "grouping of information." The support given by information structuring tools like those presented above will enable users to establish and persistently keep relations between pieces of information and to give the rational for these associations. By this, a complex information structure is being developed in a cooperative way to be used commonly by all users in the ASIC. This persistent knowledge gives insight in the dynamically developing situation and serves as bases to all further intelligence processing, as it has to be done during the analysis and integration step.

The assessment of the military intelligence experts concerning the usability of the CoALA tool was as follows: Despite the fact that it was still under development, the members of the intelligence branch considered it usable enough. The tool was perceived as being easy to learn and intuitive because of the new paradigm introduced to directly support the process of transforming unstructured information to structured information. It followed user interface architecture best practices but did not yet implement all the Microsoft Windows GUI practices. One function that collators really appreciated was that the text document processing tool retains objects once initially categorized. This speeded up collation work. The tool also supported collaborative work via a central server and knowledge database. This let the user know immediately if a certain object was already retained and recorded in the system, thus preventing pointless duplication.

4 AUTOMATIC KNOWLEDGE DISCOVERY

This section discusses other available technologies and current research directions for automatic knowledge extraction from data. The early stages of intelligence processing are largely deductive detection processes, performed by intelligence operators who look for relevant information in the intelligence information repository assisted by software applications that support information search and retrieval. Indexing and cross-referencing are processes that can be performed automatically, even by off-the-shelf software, as the documents are filed in the database. These simple steps already add value to the database as they provide means for retrieval of digitised documents and navigation capabilities within the information repository. COTS (commercial off-the-shelf) software is therefore already available to support the "Information Indexing" phase of the "Data Management" processing block of Figure 2. The specialized tools described in Section 3.1 offer dedicated functionalities to ease this early stage of the intelligence operator's work.

Knowledge extraction from unstructured data is of course a topic of paramount importance not only for intelligence processing. For example, the topic is very relevant to contemporary search engines that aim at indexing all sorts of documents and media. To be mentioned is the Unstructured Information Management Architecture (UIMA) standard for the development, discovery, composition, and deployment of multi-modal analytics for unstructured information search technologies [5]. These efforts are in line with the processing steps involved in the "Collation" phase.

However, it would be extremely valuable for intelligence purposes to have a document management system which is able to perform batch knowledge discovery. That is, to automatically mine the data with the purpose

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [10] □ (XXX) □ 10-16-13 □ 12:10 PM

of aggregating, linking and relating information without a specific directive from the operator. This early form of knowledge discovery is called *structure discovery* and could provide precious "new" information to the operator as it could hint on hidden or unknown patterns of relations in the data. This abductive discovery process aims at finding the best explanation of relationships that describe the data. This batch processing would automate the "Information Structuring" phase of the Collation step.

Once structured information has been extracted from data, data mining techniques can be used to discover knowledge from data. An example for this kind of discovery is detecting patterns, associations, and correlations that occur frequently. Frequent patterns can include item sets, subsequences, and substructures. Subsequences can refer to sequential patterns (e.g. a temporal sequence of events) while the presence of substructures (e.g. graphs, trees, lattices) in the data can suggest interesting relational patterns among entities. In particular, graph mining techniques are mostly based either on the Apriori or pattern-growth algorithms [17]. The above mentioned data mining techniques work with structured data which are typically organized in databases and discover relations typically on a statistical basis according to the number of occurrences of certain patterns. Recent research efforts are focused on methods and techniques for handling inherently uncertain and compositionally structured data. In the intelligence context, all sources of information are likely to provide data affected by some degree of uncertainty. It could be measurable uncertainty as in the case of sensory data measuring some physical quantity or much less measurable as in the case of a human observer.

Statistical Relational Learning (SRL) is an emerging research area that, building on probability and statistics, aims to represent reason and learn in domains with complex relational and rich probabilistic structure [15]. SRL encompasses a number of formalisms that have been proposed over the years, from early attempts within the inductive logic programming community [25], to Machine Learning techniques such as Bayesian Networks, Markov Networks and Conditional Random Fields, to mixed (and more recent) approaches such as Markov Logic Networks [15]. The latter is an example of the emerging trend in SRL combining first order logic and Markov Networks into a new formalism.

Pre-processing of intelligence information could be an interesting application of SRL techniques given their ability to model (possibly uncertain) dependencies between related instances. Distinguishing "significant" information from "noise" in the continuous flow of input data to any reasoning system is a key step to be carried out in order to seed the interpretation process. This initial partition can be performed by matching a priori defined models. This implies that a certain number of possible "explanations" of observed data should already be available to the system as (human) expert provided knowledge. An unsupervised probabilistic approach, on the contrary, can succeed in identifying a structural model (which can explain regularities in the data) using just a very basic form of prior knowledge if none at all. Recent studies in cognitive sciences show how achieving significant degree of success in "comprehension" needs discerning the underlying regularities in the world. This process seems to require some (inductive-abductive) constraints in order to cope with sparsity and noise in data and information [20]. According to these cognitive theories, the best the human mind can do in inferring from available data is to make the "best possible guess" guided by prior probabilities about which world structures are most likely.

A Bayesian approach seems to mimic human reasoning over structures, relations and links, and it is possible to provide a detailed computational account of how a number of basic structural forms can be inferred from various types of data (feature sets, similarity matrices, and relations). This can be applied to different areas of interest, covering higher-level problems like inferring causal structures, learning about hidden properties or objects, and interpreting the meaning of words [20]. As already mentioned, the process of deductive detection of patterns or "significant" information implies already having a model according to which data can be judged as such, that is having some strong a priori assumption over the situation under investigation. This is what is needed by logic-based approaches or graph matching algorithms used in data mining.

The algorithm proposed by Kemp and Tenenbaum's exploits Bayesian inference to identify a hierarchical model that best accounts for the observed data and generates candidate models from graph grammars. The model with maximum posterior probability given the data [20] is taken as the most likely explanation of observed patterns. This framework allows alternative forms to compete with one another to explain any given set of data rather than requiring an a priori assumption about the form appropriate for a specific dataset. For example, the technique allows inferring structure from relational data as in the case of frequency of communications between a group of persons leading to the discovery of social cliques or hierarchical tree structures (eventually discovering lead roles within an organization). Discovered structures can dynamically be adjusted as new information is collected and filed in the database. A similar approach could be applied as a batch pre-processing to intelligence data greatly augmenting the value of the information contained in the repository as it can direct the attention of the collation operator and provide precious clues for later higher-level processing by intelligence analysts.

The support of later stages of information processing could benefit from the use of graphical models to express the probabilistic consequences of causal relationships. The scientific research community is currently

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [11] □ (XXX) □ 10-16-13 □ 12:10 PM

discussing whether these models could serve as the basis for learning causal relationships from data. The prospect would be to have a Bayesian learner working backwards from observed patterns of correlation (or statistical dependency) to make probabilistic inferences about the underlying causal structures likely to have generated those observed data. This process would be very similar to what is intended as creative abduction [34]. It is possible to use the basic principles of Bayesian inference over data which is represented by samples from an unknown causal graphical model and the hypotheses to be evaluated are different candidate graphical models. A brief survey of some cognitive approaches which we believe might be considered to support the automation of information fusion tasks for intelligence analysis is given in [12] [13] where also links to historical philosophical foundations are given.

5 THREAT AND RISK ASSESSMENT

There are lots of definitions of the concepts of *threat* and *risk* on, for instance, the web. They have in common that they both represent something that might happen in the future that will influence us in a negative way. One common definition, that for statistical decision theory, is found on the Wikipedia: *risk* = (Probability of an event occurring) X (Impact of the event if it occurs). For several (infinitely) possible and observable events it is, simply put, a sum (integral), over the impacts (loss function) X the probabilities (probability density function) for the observable events. Something that influences us in a positive way could on the other hand give an *opportunity* for own action. In practice, there are normally a limited set of events that can reasonably be expected and where the impact can be estimated. This is the case we discuss below.

In the military case, there is often a more or less well defined "adversary" which imposes a *threat*. Here, the threat can be formalized as a combination of the adversary's *capability* to attack us, their *intensions* to attack, and if they can find an *opportunity* to attack us [26]. As earlier described in the paper, in a tool like CoALA a network of IntObjects and their relations are continuously built up to reflect the semantic content of a set of intelligence reports in many different "qualitative" dimensions besides the more "quantitative" time and space. Now, how could this network and the patterns emerging in it be used for forensic (history), situation (now) and threat (future) assessment? We will do this by introducing the concept of *Indicators*.

5.1 Definition and usage of an Indicator

An indicator in its most general form can be defined as something that signals (indicates) the presence of something else. Here its definition is limited to a direct observation of a maybe seemingly less relevant event or a state that can indicate something more serious (primary)

Event *has happened*, is *going on*, or is *about to happen*

State *has been realized*, is *becoming realized* or *will be realized*

Hence, the indicator is a secondary effect of the past, present or future primary event or state, simply called the *primary* below, which has not been observed directly so far. The indicator concept can be used both for detecting present or forecasting future primaries that might be threats or opportunities, as well as used in abductive reasoning [16] [49]. In the last case, indicators are regarded as consequences or effects of primaries that have already happened, as in forensic investigations. Experienced persons can often assess, or hypothesize about, what has been or is going on, or if the risk of something happening is increased, by taking notion of such indicators.

A single observation like "There are no people in the square when it is normally crowded." can be an indicator as well as the fused result of several different observations leading to some conclusion or hypothesis like "There seem to have been repeated correlated money transfers from X to a known IED expert Y via Z's account in bank W." The primaries in these two cases can be a forthcoming shooting on the square, being known to the local population, and a forthcoming bomb attack, respectively.

How can we, using indicators, formalize this building of hypotheses about primaries? Imagine that a decision maker has a *"monitor list"* of primaries in the form of events or state changes that, in our asymmetric scenario, are regarded as more important than others to prevent, or exploit. They might impose plausible and extra serious threats, or positive opportunities. This list is assumed to be compiled by SMEs that in some way are familiar with the situation at large. The list, perhaps sorted according to priority or probability, might contain many primaries, more or less related to each other. Intelligence reports on the presence, or explicit absence, of indicators must now be exploited in order to somehow assess the probabilities of the different primaries in this list.

5.2 Coupling Indicators to Primaries

The couplings between indicators and a primary can for example, but not necessarily, be achieved via a Bayesian Network (BN) built by an experienced person who knows which indicators tend to influence a primary, and which indicators are more important than others and should have higher weights. A combination of indicators with different weights, and maybe also observed absence of expected indicators (Negative Information) feed into a BN, and if the output is higher than some threshold, an alert corresponding to the primary modelled by that BN as a root node is issued. A BN could be built, and be extended or modified during a mission when situation-specific knowledge grows, or several BN fragments managed separately by domain

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [12] □ (XXX) □ 10-16-13 □ 12:10 PM

experts could be put together to a BN tailored to match the specific mission or case [49]. There are other ways than BNs to couple the influence of indicators to primaries, but to obtain trust in the system, it must be easy to understand why a certain primary might suddenly be alerted in the system by, for instance, "clicking" on it in a Graphical User Interface. If inferences cannot be followed easily it would render the tool useless; no reasonable decision maker could take decisive decisions based on threat alerts generated by a tool whose way of functioning is regarded anything similar to a black box. Then the inference path used must be displayed in an easy-to-understand way. Furthermore, there will often be an interest to study the history of a monitored feature (primary or indicator) and see how its probability has changed in time during the inflow of intelligence reports. It is often difficult to judge about an absolute probability number, should it be, say, 0.4 or 0.6, so the rate of change can be more interesting.

Indicator weights can besides a preset importance level also be related to the frequency of similar observations as well as a preset value in the leaves of the BNs on how much a certain category of observations affects an indicator. As well, an observed indicator of a future primary must have a decay time constant associated with it depending on what it is assumed to indicate, or if the indicator itself is more of a state than an event. An indicator (explosives found) typical of a discrete upcoming event (bombing) of course decays more quickly than one (bad harvests) typical for a more permanent state (famine) and must soon enough decrease its influence on BN's representing discrete events. What decay times to set for different types of indicators of course varies, and has to be judged by SMEs.

5.3 How to display the risk—Impactorium

So, it would be of great benefit to have a mechanism that continuously shows if the estimated risk has increased that some primary is realized. At FOI, a tool called Impactorium [46] [14] [2] [11] has been developed.[1] Impactorium has a display idea based on the so called Impact Matrix (IM). In the enterprise world, the IM has been used for risk visualisation for a long time. An IM is a 2D plot area with a "coordinate system" for the primaries where the (horizontal) X dimension represents the severity (impact or consequence) *if* it happens, and the (vertical) Y dimension the a priori probability for it to happen. We do not elaborate on the X dimension more in this paper, and the impacts have to be judged by SMEs and are normally not a subject for change in time. The primaries in the assumed monitor list mentioned above, on the other hand, are assessed concerning their probabilities using incoming intelligence reports as sources to BNs. They should now be moved along the Y axis of the IM according to their updated probabilities. In Impactorium we have tested a slightly different visualization technique; monitored features are positioned along the Y dimension according to their a priori probabilities. When the monitored features change probability by the influence of the indicators in the BN due to incoming intelligence, their symbols in the IM are initially not shifted in position along the Y axis. Rather they automatically change colour continuously between green (improbable) and red (strong alert). When corroborating information might later be received, they can be shifted accordingly in the IM. If not received, their colour fades back to their earlier states with the earlier mentioned decay time constant. Primaries, indicators and intermediate nodes in the BN can of course also be displayed in a "monitor list," sorted according to present probability.

The tool allows for Impactorium clients to access intelligence reports in a common database, as well as to design or use pre-designed BNs to connect indicators to primaries. Instead of BNs, simple mathematical operators like *mean*, *min* or *max* can be applied in the network nodes. Different analysts can focus on subsets of primaries by keeping them and their associated indicators in personal analysis object baskets, much like the IntObject basket in CoALA. As Impactorium is a semi-automatic tool, the operator is now able to browse the inference chains in the BNs to see details on why the alert emerged. In situations of time pressure the visualization could be done on the fly where the BN issuing the alert is visualized automatically and the most important nodes for the alert in it is highlighted.

User studies have been performed with Impactorium as well [32] [47] and the tool has since then been further developed concerning web service API and user interfaces. Impactorium is still something of a research test bench, and not yet an as well developed product as CoALA or PARANOID), but a plan for how it could be implemented in the Swedish Armed Forces has been produced.

5.4 Relating Impactorium to CoALA

As mentioned, BNs are one way to link observations or intelligence reports via indicators to potential primaries, which is the way it is done today in Impactorium. Another way would be to continuously monitor the structure of a semantic network while it is built up as in CoALA. Instead of letting one or several, maybe fused, intelligence reports trig one or several indicators (as is the case today in Impactorium), one could try to identify patterns in the CoALA network that are known beforehand to indicate threatening situations. This could be done for instance by graph matching techniques [42] [27]. Impactorium currently has a somewhat more causal event-chain analysis approach, but extending it with pattern-recognition techniques would be very interesting; this is elaborated a bit in the next

[1] This tool was not described in Section 3.1 above since it was not a part of the SME trial with PARANOID/CoALA and is not really a dedicated collation and link analysis tool.

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [13] □ (XXX) □ 10-16-13 □ 12:10 PM

section. A maybe semi-automatic Mixed-Initiative reasoning, pattern-recognition functionality for identifying such network structures should then be the equivalent to the BNs causing certain types of observations to trig indicators in Impactorium today. How could this be achieved? Experienced people have models to which they compare a new situation they are confronted with, and to link cause and effect. Earlier experienced cases, maybe in different mixings, serve as models used to assess the type and characteristics of the new situation. This can to a large extent be compared to case-based reasoning. An implementation of this mental model-building and matching process into some algorithm, following the ideas in Section 4 of this paper, would make it possible to obtain a coupling between the outputs of a fusion level 2 tool like CoALA to the input of a level 3 risk analysis tool like Impactorium.

5.5 Other applications of Impactorium

Impactorium have so far been developed as an impact assessment tool for two cases: assessing potential *ongoing*, but still not directly observed, primaries, and potential *future* ones. An example of the latter is [47]. An example on the former is the work going on in the EU FP7 projects "Support" and "Contain" for assessing and controlling threats in sea port areas and against containers in the container logistics chain, respectively. These works are still unpublished, but use a method very similar to the idea described here: Port surveillance sensor- or container status sensor reports are tagged semantically; in this case using RDF [51] triples, describing the observed events on a semantic level appropriate for the observations done by the specific sensors. An ontology, for example described as an RDF schema, defines the event types that could be "instantiated" by the sensors. Triples expressed in XML from different sensors in the port stream into a RDF stream complex event processor [50] [22] which, from these triples and further entailments done using implicit knowledge in the ontology, builds semantic networks that describe the situations to which the sensors have contributed with observations. Predefined queries on these networks, defined using SPARQL (an RDF network query language) acts as the patterns to be searched for in the growing network. This idea is very similar to a monitoring function for networks in CoALA described in the previous subsection. When a pattern matches, perhaps within a spatial and temporal window, an alert is issued, and analysts using Impactorium can via web services subscribe on such events depending on what type of alert is relevant for their respective role (representing customs, a freight company, port security etc.).

We would also like to include "forensic" reasoning: the primary might already have happened in the *past*, and we want to reason abductively from the indicators which have followed as consequences of the hypothesis of a primary similar to [49]. This is to say that Impactorium should allow the operator to enter anywhere in the temporal chain of intelligence reports and associated alerts of indicators and follow the inferences done by the tool.

As mentioned, primaries can be events or states. States can change discretely or more continuously. Most military actions are executed as activities together aiming at some higher goals, or *effects*. This is a central concept in the Effect-Based Approach to Operations (EBAO) paradigm, but has of course always been important in military thinking. An effect is in principle a change of state as a result of non-planned events or planned actions or activities. The way Impactorium works can help personnel that are responsible for monitoring these changes of states to do this monitoring, and when reporting on it, be more clear and concise about what causal chains are assumed to be the reason for the change. In this case the assessed outcome of executed actions is fed into the indicators of a BN or similar network that defines the influence of the success or failure of actions on the primary, here being the effect one wants to obtain. This is a mode of usage more operative than tactical and it suits threat assessments that are done by intelligence staffs rather than by the staff responsible for direct execution of military activities.

6 CONCLUSIONS

Tools like CoALA or PARANOID are accepted and appreciated by the military community. They give support for the processing and exploitation of unstructured semantic information as well as for some additional functionality analysing the established structured information set. However, up to now these interactive tools mainly just assist the human operators in their semantic exploitation of the information and their reasoning about the meaning and the consequences of the determined situation. To support situation awareness and threat and impact assessment more research on the discovery and update of behaviour pattern and system structures as well as on the principles of pattern and behaviour based reasoning, especially for imperfect data and information has to be performed. How to then alert and focus users on emerging threats and risks found accordingly, like in the Impactorium tool, is another important issue.

REFERENCES

[1] J. R. Anderson
Cognitive Psychology and its Implications.
Palgrave Macmillan, 6th Edition, 2005.

[2] M. Andersson et al.
The FOI C4ISR Demonstrator 2008.
In *Proceedings of the 12th International Conference on Information Fusion*, Seattle, WA, USA, July 2009, pp. 1604–1612.

[3] W. S. Angerman
Coming Full Circle With Boyd's Ooda Loop Ideas: An Analysis Of Innovation Diffusion And Evolution,
Thesis, AFIT/GIR/ENV/04M-01, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, USA, March 2004, http://www.dict.mil/cgi-bin/GetTRDoc?AD=ADA425228.

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [14] □ (XXX) □ 10-16-13 □ 12:10 PM

[4] R. T. Antony
Principles of Data Fusion Automation,
Artech House, Inc., Norwood, MA, 1995.

[5] Apache UIMA project
Homepage: http://uima.apache.org/index.html.

[6] J. Biermann
Understanding Military Information Processing—An Approach to Supporting the Production of Intelligence in Defence and Security
in: Shahbazian, Elisa; Rogova, Galina; De Weert, Michael J. (Eds.) "Harbour Protection Through Data Fusion Technologies," *Proceedings of the NATO Advanced Research Workshop on Data Fusion Technologies for Harbour Protection* Tallinn, Estonia 27 June–1 July 2005, Springer, 2009.

[7] J. Biermann et al.
From Unstructured to Structured Information in Military Intelligence: Some Steps to Improve Information Fusion.
NATO RTO SCI-158 Panel Symposium on "Systems, Concepts and Integration (SCI) Methods and Technologies for Defence Against Terrorism," 25 to 27 October 2004, London, UK.

[8] J. Biermann, P. Hörling, and L. Snidaro
Automated Support for Intelligence in Asymmetric Operations: Requirements and Experimental Results.
In *Proceedings of the Twelfth International Conference on Information Fusion*, Seattle, U.S., July 6–9, 2009, pp. 1592–1599.

[9] J. Bohannon
Counterterrorism's New Tool: 'Metanetwork' Analysis.
Science, Vol. 325, pp. 409–411, 24 July 2009.

[10] A. Brook
UK Experiences of Using Coalition Battle Management Language.
In *15th International Symposium on Distributed Simulation and Real Time Applications* (DS-RT), Salford, UK, 2011.

[11] J. Brynielsson, A. Horndahl, F. Johansson, L. Kaati, C. Mårtenson, and P. Svenson
Analysis of weak signals for detecting lone wolf terrorists.
In *Proceedings of the European Intelligence and Security Informatics Conference* (EISIC 2012), 2012, pp. 197–204.

[12] G. Ferrin, L. Snidaro, and G. L. Foresti
Structuring Relations for Fusion in Intelligence.
In *Proceedings of the 12th International Conference on Information Fusion*, Seattle, U.S., July 6–9, 2009, pp. 1621–1626.

[13] G. Ferrin, L. Snidaro, and G. L. Foresti
Revisiting the role of abductive inference in fusion domain.
In *Proceedings of the 13th International Conference on Information Fusion*, Edinburgh, U.K, July 26–29, 2010.

[14] R. Forsgren, L. Kaati, C. Mårtenson, P. Svenson, and E. Tjörnhammar
An overview of the Impactorium tools 2008.
SWIFT 2008—Skövde Workshop on Information Fusion Topics, Skövde, Sweden (2008), http://www.his.se/PageFiles/6801/Forsgren_SWIFT2008.pdf.

[15] L. Getoor and B. Taskar (Eds.)
Statistical Relational Learning, Adaptive Computation and Machine Learning.
MIT Press, Cambridge, Massachussets, 2007.

[16] T. Goan, N. Kartha, and R. Kaneshiro
Effective Behavioral Modeling and Prediction Even When Few Exemplars are Available.
Proc. SPIE Vol. 6235, Signal Processing, Sensor Fusion and target recognition XV, 623511 (2006).

[17] J. Han and M. Kamber
Data Mining: Concepts and Techniques.
Morgan Kaufmann, 2006.

[18] H. Kang, C. Plaisant, B. Lee, and B. B. Bederson
NetLens: Iterative Exploration of Content-actor Network Data.
Information Visualization. vol. 6, n. 1, 2007, pp. 18–31.

[19] Y. Kang, C. Gorg, and J. Stasko
How can visual analytics assist investigative analysis? Design implications from an evaluation.
IEEE Transactions on Visualization and Computer Graphics, Vol. 17, no. 5, 2011, pp. 570–583.

[20] C. Kemp and J. B. Tenenbaum
The discovery of structural form.
Proceedings of the National Academy of Sciences, Vol. 105, n. 31, pp. 10687–10692, 2008.

[21] H. Khurana, J. Basney, M. Bakht, M. Freemon, V. Welch, and R. Butler
Palantir: A Framework for Collaborative Incident Response and Investigation.
In *Proceedings of the 8th Symposium on Identity and Trust on the Internet*, 2009, pp. 38–51.

[22] S. C. Komazec
Sparkwave: Continuous Schema-Enhanced Pattern Matching over RDF Data Streams.
In *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems* (2012), pp. 58–68.

[23] K. Kruger, J. Biermann, M. Frey, and U. Schade
Automatic Information Fusion from Diverse Sources.
In *Proceedings of the Military Communications and Systems Conference* (MCC) 2007, Bonn, 2007.

[24] K. B. Laskey, D. A. Schum, P. C. G. Costa, and Terry Janssen
Ontology of Evidence.
In *Ontology For the Intelligence Community*, 2008/12/3, p. 20–62, http://ceur-ws.org/Vol-440/Proceedings.pdf#page=20; 2013-05-05.

[25] N. Lavracm and S. Dzeroski
Inductive Logic Programming: Techniques and Applications.
Ellis Horwood, New York, 1994.

[26] E. G. Little and G. L. Rogova
An Ontological Analysis of Threat and Vulnerability.
In *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, July 2006.

[27] E. Little, K. Sabmbhoos, and J. Llinas
Enhancing Graph matching Techniques with Ontologies.
In *Proceedings of the 11th International Conference on Information Fusion*, July 2008, Cologne, Germany, pp. 1890–1897.

[28] J. Llinas, C. Bowman, G. Rogova G., A. Steinberg, E. Waltz, and F. White
Revisiting the JDL Data Fusion Model II.
In: *Proceedings of the 7th International Conference on Information Fusion*, June 2004, Stockholm, Sweden, pp. 1218–1230.

[29] M. D. Moskal, M. Sudit, and K. Sambhoos
The role of information fusion in providing analytical rigor for intelligence analysis.
In *Proceedings of the 14th International Conference on Information Fusion*, 5–8 July 2011 pp. 1279–1284.

[30] NATO AC/323(IST-065)TP/423 RTO-TR-IST-065 Technical Report on "Information Fusion in Asymemetric Operations," November 2012.

[31] NATO RTGonIFD IST-038/RTG-016
Information Fusion: Supporting Military Intelligence by Structuring Information
Proceedings of the NATO RTO IST Final Demonstration on Information Fusion For Command Support (IST-055/CDT), 10 November 2005, The Hague, The Netherlands.

[32] M. Nilsson, J. van Laere, T. Ziemke, P. Berggren, and B. Kylesten
A user study of the Impact matrix, a fusion based decision support for enhanced situation awareness.
In *Proceedings of the 11th Int. Conference on Information Fusion*, Cologne, 2008, pp. 440–447.

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [15] □ (XXX) □ 10-16-13 □ 12:10 PM

[33] North Atlantic Treaty Organization NATO Standardization Agency (NSA)
AAP-06 Edition 2013 "NATO Glossary of Terms and Definitions" (English and French).
http://nsa.nato.int/nsa/nsdd/APdetails.html?APNo=1689&LA=EN (15.5.2013).

[34] C. S. Peirce
Collected Papers of Charles Sanders Peirce.
Harvard University Press, 1931.

[35] J. M. Pullen
Battle Management Language Enables Rapid Integration of Command & Control with Simulation.
In *14th International Symposium on Distributed Simulation and Real Time Applications* (DS-RT), Fairfax, VA, USA, 2010.

[36] K. Rein and U. Schade
Battle Management Language as a "Lingua Franca" for situation awareness.
Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), New Orleans, LA, USA, 2012.

[37] K. Rein, U. Schade, and M. R. Hieb
Battle Management Language (BML) As an Enabler.
International Conference on Communications (ICC '09), Dresden, 2009.

[38] K. Rein, U. Schade, and S. Kawaletz
Uncertainty Estimation in the Fusion of Text-Based Information for Situation Awareness.
In E. Hüllermeier, R. Kruse, F. Hoffmann (Eds.) *Information Processing and Management of Uncertainty, In Knowledge-Based Systems. Applications*, Vol. 81 *Communications in Computer and Information Science*, Springer Berlin Heidelberg, 2010.

[39] PNNL. Pacific Northwest National Laboratory
INSPIRE Visual Document Analysis.
Last accessed: May 20, 2013, http://in-spire.pnl.gov.

[40] G. Rogova and E. Bossá
Information quality in information fusion.
In *Proceedings of the 13th International Conference on Information Fusion*, Edinburgh, U.K., July 2010.

[41] G. Rogova and V. Nimier
Reliability in information fusion: literature survey.
In *Proceedings of the 7th International Conference on Information Fusion*, ISIF, June 2004, Stockholm, Sweden, pp. 1158–1165.

[42] K. Sambhoos, J. Llinas, and E. Little
Graphical Methods for Real-Time Fusion and Estimation with Soft Message Data.
In *Proceedings of the 11th International Conference on Information Fusion*, Cologne, 2008, pp. 1621–1628.

[43] U. Schade, J. Biermann, M. Frey, and K. Kruger
From Battle Management Language (BML) to automatic information fusion.
In: V. Popovich, M. Schrenk, and K. Korolenko (Eds.) *Information Fusion and Geographical Information Systems* (Lecture Notes in Geoinformation and Cartography), pp. 84–95, Berlin: Springer.

[44] M. Shahriar Hossain, P. Butler, A. P. Boedihardjo, and N. Ramakrishnan
Storytelling in entity networks to support intelligence analysts.
In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012, pp. 1375–1383.

[45] L. Snidaro, I. Visentini, K. Bryan, and G. L. Foresti
Markov Logic Networks for context integration and situation assessment in maritime domain.
In *Proceedings of the 15th International Conference on Information Fusion*, Singapore, July 09–12, 2012, pp. 1534–1539.

[46] P. Svenson, T. Berg, P. Hörling, M. Malm, and C. Mårtenson
Using the impact matrix for predictive situational awareness.
In *Proceedings of the 10th Int. Conference on Information Fusion*, Québec, Canada, July 2007.

[47] P. Svenson, R. Forsgren, B. Kylesten, P. Berggren, W. Rong Fah, M. S. Choo, and J. K. Yew Hann
Swedish-Singapore studies of Bayesian modeling techniques for tactical intelligence analysis.
In *Proceedings of the 13th International Conference on Information Fusion*, Edinburgh, UK, July 2010.

[48] G. Thibault, M. Gareau, and F. Le May
Intelligence Collation in Asymmetric Conflict: A Canadian Armed Forces Perspective.
Proc. 10th International Conference on Information Fusion, Québec, Canada, July 2007.

[49] S. P. van Gosliga and I. van de Voorde
Hypothesis Management Framework: a flexible design pattern for belief networks in decision support systems.
UAI 2008, Helsinki, Finland, 2008.

[50] Wikipedia
Complex Event Prosessing
http://en.wikipedia.org/wiki/Complex_event_processing.

[51] Wikipedia
Resource Description Framework:
http://en.wikipedia.org/wiki/Resource_Description_Framework.

[52] E. J. Wright and K. B. Laskey
Credibility Models for Multi-Source Fusion.
In *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, July 2006, pp. 274–280.

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [16] □ (XXX) □ 10-16-13 □ 12:10 PM

Joachim Biermann received his degree in numerical mathematics (Dipl. Math.) from the Department of Mathematics of the University of Cologne in 1982. From 1982 to 1987 he was assistant professor with the same department working in the area of numerical mathematics, especially on methods for parallel and vector computing, finite-element multigrid methods for fluid dynamic problems, and computer science.

Since 1988 he is with the Research Institute for Communication, Information Processing, and Ergonomics (FKIE) an institute of the Fraunhofer Gesellschaft and has been mainly working on knowledge-based exploitation and information fusion in the context of intelligence processing and decision support. Since 1992 he has been actively participating in several NATO activities on information fusion in intelligence, particularly in all NATO RTO IST Task Groups on Information Fusion (1999–2009). In 2003 he became deputy department head of the department of Sensor data and Information Fusion at the FKIE and leader of the research group 'Threat Recognition for Defence and Security.'

His actual research interest is related to the analysis and modelling of heuristic information processing procedures, especially for unconventional or asymmetric threat, the integration of context and background information into multi sensor data fusion and the development and application of according fusion methods in defence and security applications.

He is a member of the Armed Forces Communications and Electronics Association (AFCEA), member of the Advisory Board of the "Information Fusion Research Programme" of the University of Skövde, Sweden, and a member of the Board of Directors of the International Society of Information Fusion (ISIF).

Pontus Hörling received the M.Sc. degree in physics and chemistry from Uppsala University, Uppsala, Sweden in 1988, and the Ph.D. degree in physics from the Royal Institute of Technology, Stockholm, Sweden, in 1996. Since then he has been with Swedish Defense Research Agency (FOI), the last 8 years as Senior Scientist. Here, he has mainly been involved with computer science, software development and simulation. His research interests are information processing for command and control systems, sensor- and information fusion, effects-based planning and simulation of such processes. Besides being involved in the NATO RTO task group that delivered the results reported on in this article, he has been the Swedish PoC in the NATO STO (former RTO) IST Panel since spring 2007.

Mt2 □ job no. 2172 □ jaif □ Journal of Advances in Information Fusion □ 2172D02 [17] □ (XXX) □ 10-16-13 □ 12:10 PM

Lauro Snidaro received his M.Sc. and Ph.D. in Computer Science from University of Udine in 2002 and 2006 respectively. Since 2008 he is an Assistant Professor at the Dept. of Mathematics and Computer Science, Univ. of Udine. His main interests include Data/Information Fusion, Computer Vision, Pattern Recognition, Machine Learning, Multimedia, Video understanding and annotation. He actively publishes in international journals and conferences and has co-authored more than 70 papers. He was the appointed Italian member of the NATO Task Group on Information Fusion Demonstrator (TGonIFD) IST-038/RTG-016 for 2003–05, appointed Italian member of the NATO Information Technology Panel Task Group IST-065/RTG-028 on "Information Fusion in Asymmetric Operations" (RTGonIFAO) for 2007–2009, and is the appointed Italian member of the NATO Exploratory Group on "Information Filtering and Multi-Source Information Fusion" for 2011–2013. He cooperates with international industries and research centres and is involved in several international projects on multi-sensor data fusion. He has been invited speaker at international symposiums and meetings in Armenia, Bulgaria, Spain, United States. In 2011 he was co-chair of the 16th International Conference on Image Analysis and Processing (ICIAP). He has co-organized several special sessions at the International Conference on Information Fusion and the IEEE International Conference on Advanced Video and Signal-Based Surveillance. He is special issue organizer and editorial board member of the Information Fusion journal. He serves as a computer science project reviewer for the Italian Ministry of University and Research. He regularly serves as a reviewer for 19 international journals and 18 conferences.

Mt2 ☐ job no. 2172 ☐ jaif ☐ Journal of Advances in Information Fusion ☐ 2172D02 [18] ☐ (XXX) ☐ 10-16-13 ☐ 12:10 PM