NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

AC/323(IST-040)TP/23

www.rta.nato.int

RTO MEETING PROCEEDINGS **MP-IST-040**

Military Data and Information Fusion

(La fusion des informations et de données militaires)

Papers presented at the Symposium organised by the RTO Information Systems Technology (IST) Panel in Prague, Czech Republic, from 20 to 22 October 2003.

Published March 2004

Distribution and Availability on Back Cover

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

AC/323(IST-040)TP/23

www.rta.nato.int

RTO MEETING PROCEEDINGS **MP-IST-040**

Military Data and Information Fusion

(La fusion des informations et de données militaires)

Papers presented at the Symposium organised by the RTO Information Systems Technology (IST) Panel in Prague, Czech Republic, from 20 to 22 October 2003.

The Research and Technology Organisation (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote co-operative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective co-ordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also co-ordinates RTO's co-operation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of co-operation.

The total spectrum of R&T activities is covered by the following 7 bodies:

- AVT Applied Vehicle Technology Panel
- HFM Human Factors and Medicine Panel
- IST Information Systems Technology Panel
- NMSG NATO Modelling and Simulation Group
- SAS Studies, Analysis and Simulation Panel
- SCI Systems Concepts and Integration Panel
- SET Sensors and Electronics Technology Panel

These bodies are made up of national representatives as well as generally recognised 'world class' scientists. They also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier co-operation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

The content of this publication has been reproduced
directly from material supplied by RTO or the authors.

Published March 2004

ISBN 92-837-0037-6

Military Data and Information Fusion

(RTO-MP-IST-040)

Executive Summary

Military commanders require timely and accurate awareness of the situation in their respective area of responsibility as well as some prediction of the likely intentions of the forces engaged. In addition to sensor based reconnaissance information, new and complex non-military information aspects are influencing the military intelligence cycle now and will do so in the future. Large volumes of information and data will have to be processed. Classic manual evaluation followed by the presentation of results is much too time consuming. It is therefore important to identify key issues affecting data and information fusion problems and to find methods to give automated support to information and data fusion in order to become more efficient in military intelligence.

The symposium provided an interdisciplinary forum for research scientists, military experts, and system engineers to present the state of the art of research and technology in military data and information fusion. Interest was focussed on high-level fusion aspects (level 2, 3 and 4 of the JDL data fusion definition). The broad scope of topics was discussed by two excellent keynote speeches and 23 interesting and well appreciated papers which were presented in 8 technical sessions to an audience of more than 120 participants. The symposium was opened by a keynote speech by LtCol de Chantal (FR). In his presentation entitled "Operational requirements for fusion in the fields of information and intelligence" he emphasised the military point of view and pointed out a real need of information fusion to free man from low level tasks of information and data processing and to provide assistance to military commanders. The operational aspects were also covered by the papers presented in the first session entitled "Military Requirements and Experience", which additionally discussed aspects of interoperability, to guarantee access to all necessary information, transparency of fusion methods, to improve the acceptance, and fusion systems with "Man-in-the-Loop" to keep responsibility with the users.

The second keynote speech entitled "On the Scientific Foundations of Level 2 Fusion" was given by Dr. J. Llinas (US) who addressed two focal topics: (1) suggestions for extensions and modifications of the well-known "JDL" Data Fusion process model, especially related to Distributed Data Fusion (DDF) and Network-Centric Warfare and (2) perspectives on an overarching approach towards developing first theories and then algorithms for Level 2 Data Fusion. The papers presented in the sessions "Fusion System Concepts I and II" also made it obvious that at the moment there are very different views on what 'military information fusion' should be. This underlined the fact that there is presently no sufficient and common understanding of the high level fusion problem domain and its characterising elements. The papers in the sessions "Fusion Methods" and "Fusion Methods for Classification and Identification" presented special algorithms related to level 1, 2, and 3 fusion problems. Among others, the successful incorporation of background information into level 1 data fusion algorithms by evidential reasoning was shown. The interesting studies on the use of Bayesian methods in high level information fusion algorithms presented some encouraging results but also revealed the need for a suitable problem abstraction and domain model for an appropriate mathematical approach to non-numeric problems. The session entitled "Semantic Approach to Information Fusion" presented contributions on text understanding and the semantic principles of domain modelling. The papers showed promising ontological and linguistic methods for treating the problems of symbolic information processing. The paper entitled "Analysis of Free-Form Battlefield Reports with Shallow Parsing Techniques" which presented a method for information extraction from free text documents was given the "Best Paper Award".

The contributions to the sessions on “Applications” and “Applications and Lessons Learned” mostly addressed questions and requirements of intelligent data and information management, retrieval, and visualisation for decision support. It became obvious, however, that no actual substantial high level fusion module is yet operational. The symposium was closed by a panel discussion. Five members of the Programme Committee gave a personnel review of the meeting and additional ideas for future activities in the area of military information fusion. It was followed by an intensive discussion with the audience. A summary of this panel discussion is included in the meeting proceedings.

La fusion des informations et de données militaires

(RTO-MP-IST-040)

Synthèse

Les chefs militaires ont besoin de données précises en temps voulu sur la situation des forces dans leurs domaines de responsabilité respectifs, ainsi que d'un certain nombre de prévisions sur les intentions probables des forces engagées. En plus des informations de reconnaissance fournies par capteurs, le cycle du renseignement militaire est actuellement, et sera à l'avenir, influencé par des aspects nouveaux et complexes, non militaires. Il faudra traiter de grands volumes de données et d'informations. L'évaluation classique et la présentation manuelles des résultats sont beaucoup trop longues. Il est par conséquent important d'identifier les questions clés pouvant avoir une incidence sur les problèmes de fusionnement des données et des informations, et de trouver des méthodes permettant d'automatiser ce processus, afin de le rendre plus efficace pour les besoins du renseignement militaire.

Le symposium a servi de forum interdisciplinaire, permettant aux chercheurs scientifiques, aux spécialistes militaires et aux ingénieurs système de présenter l'état actuel des connaissances dans le domaine du fusionnement des données et des informations militaires. L'accent a été mis sur les aspects fusionnement de haut niveau (les niveaux 2, 3 et 4 de la définition JDL du fusionnement des données). L'ensemble des sujets a été couvert par deux excellents discours d'ouverture, ainsi que par 23 communications intéressantes et bien accueillies, présentées au cours de 8 sessions techniques devant une assistance de plus de 120 participants. Le symposium a débuté par un discours d'ouverture prononcé par le LCL de Chantal (FR). Lors de sa présentation, intitulée "Les besoins opérationnels dans le domaine du fusionnement des informations et du renseignement", il a mis en relief le point de vue militaire, en exposant le besoin réel en matière de fusionnement d'informations pour libérer l'homme des tâche élémentaires de traitement de données et d'informations, ainsi que pour fournir une aide aux chefs militaires. Les aspects opérationnels ont également été abordés dans des communications présentées lors de la première session intitulée "Besoins et expérience militaires", qui portait aussi sur certains aspects de l'interopérabilité destinés à garantir l'accès à toutes les informations nécessaires, sur la transparence des méthodes de fusionnement, dans l'optique de favoriser l'acceptation des ces méthodes, ainsi que sur les systèmes de fusionnement du type "l'homme dans la boucle", pour lesquels la responsabilité doit rester celle de l'utilisateur.

Le deuxième discours d'ouverture, intitulé "Sur les bases scientifiques du fusionnement au niveau 2", a été prononcé par M. J. Linas (US), qui a examiné deux sujets fondamentaux, à savoir : (1) des propositions concernant d'éventuelles extensions et modifications à apporter au modèle réputé de fusionnement de données "JDL", en particulier par rapport au fusionnement de données réparti (DDF), ainsi qu'à la guerre réseaucentrique, et (2) des perspectives d'une approche globale de l'élaboration de théories dans un premier temps et ensuite d'algorithmes de fusionnement de données au niveau 2. De même, il est ressorti très clairement des communications présentées lors des sessions sur "Les concepts de systèmes de fusionnement I et II", qu'il existe des définitions très diverses du terme "fusionnement des informations militaires". Cette constatation sert à souligner le fait qu'à l'heure actuelle, il n'y a pas de perception commune du problème de fusionnement de haut niveau, ni des éléments qui le caractérisent. Les communications présentées lors des sessions "Méthodes de fusionnement" et "méthodes pour la classification et l'identification" décrivaient des algorithmes particuliers applicables aux problèmes de

fusionnement des niveaux 1, 2 et 3. Parmi d'autres exemples, on a démontré l'incorporation réussie de données historiques dans des algorithmes de fusionnement de données de niveau 1 par le raisonnement d'évidence. Les études intéressantes présentées sur l'emploi de méthodes bayesiennes dans les algorithmes de fusionnement d'informations de haut niveau faisaient état de résultats encourageants, même si elles révélaient le manque actuel de modèle convenable de domaine et d'abstraction de problèmes pour une approche mathématique appropriée des problèmes non numériques. La session intitulée "Une approche sémantique du fusionnement des informations" a permis de présenter des contributions sur la compréhension des textes et les principes sémantiques de la modélisation des domaines. Les communications présentées décrivaient des méthodes ontologiques et linguistiques prometteuses pour la résolution de problèmes de traitement d'informations symboliques. La communication intitulée "Analyse des comptes rendus de champ de bataille en texte libre à l'aide de techniques d'analyse syntaxique rapide", qui décrivait une méthode permettant l'extraction d'informations de documents rédigés en texte libre, s'est vu attribuer le titre de "meilleure communication".

Les contributions aux sessions "Applications" et "Applications et enseignements tirés" concernaient pour la plupart la gestion, la recherche et la visualisation de données et d'informations intelligentes pour le soutien de la prise de décisions. Cependant, il est paru évident qu'aucun module fiable de fusionnement de haut niveau n'est encore opérationnel. Le symposium s'est terminé par une table ronde. Cinq membres du Comité du programme ont proposé leurs conclusions personnelles concernant la réunion, ainsi que des suggestions relatives à de futures activités dans le domaine du fusionnement des informations militaires. La table ronde a été suivie d'une séance de discussion animée avec l'assistance. Un résumé de ces discussions figure dans le compte rendu de la conférence.

Table of Contents

SESSION IV: FUSION METHODS
Chairman: Dr. P. SVENSSON (SW)

SESSION V: FUSION METHODS FOR CLASSIFICATION AND IDENTIFICATION
Chairman: Mr. B. O'HERN (US)

SESSION VI: SEMANTIC APPROACH TO INFORMATION FUSION
Chairman: Mr. J. BIERMANN (GE)

SESSION VII: APPLICATIONS AND LESSONS LEARNED
Chairman: Mr. J. BIERMANN (GE)

SESSION VIII: APPLICATIONS
Chairman: Dr. M. SNAJDER (CZ)

Theme

Commanders at all levels and types of military organisations require timely and accurate awareness of the situation in their respective area of responsibility as well as prediction of likely intentions of the participants. These requirements exist throughout the range and variety of potential military missions from conflict-prevention to conflict and post-conflict activities. Additional to sensor based reconnaissance information, new and complex non-military information aspects such as political, ethnical, and criminal information are influencing the military intelligence cycle now and in the future. Large volumes of information and data will have to be processed and classical manual evaluation and the presentation of results is much too time consuming. Especially in respect to Counter-Terrorism intelligent data mining and knowledge based information retrieval and content extraction are gaining importance. It is therefore important to identify key issues affecting data and information fusion problems and to find methods to give automated support to information and data fusion in order to become more efficient in military intelligence.

The symposium will provide an interdisciplinary forum for research scientists, military experts, and system engineers to present the state of the art of research and technology in all different aspects of military data and information fusion in the command and control cycle. Interest is focussed on level 2, 3 and 4 of the JDL data fusion definition. The results of the NATO-wide efforts in this evolving area should be documented in technical presentations, research papers, and application oriented discussions. The conference program will include invited speeches by distinguished experts from research and practice as well as submitted papers. This symposium will be complementary to the planned SET Symposium on ´Target Tracking and Sensor Data Fusion for Military Observation Systems´ which will take place the week before in Budapest, Hungary.

Topics of interest include, but are not limited to:

- Military requirements for data interpretation and information fusion
- Concepts of military data and information fusion for
  - Situation awareness, situation assessment
  - Threat/impact assessment
  - Intelligence preparation of the battlefield
  - Classification and identification of battlespace elements
- General fusion methods
  - Model and template based methods
  - Bayesian and belief approach to data and information fusion
  - Context sensitive information exploitation
  - Content extraction from unstructured data using natural language processing techniques
  - Research challenges toward achieving robust C2,C3 fusion
- Ontological approach to the representation of military knowledge
- Application examples and demonstrators of data and information fusion
  - Architecture of intelligence fusion systems in C4ISR
- Resources: COTS S/W, Data, training materials
  - Scenarios
  - Fusion of geo-spatial intelligence and military information
  - Fusion visualisation
- Lessons learned

Thème

Les chefs militaires à tous les niveaux de toutes les organisations militaires ont besoin de données précises en temps voulu sur la situation des forces dans leurs domaines de responsabilité respectifs, ainsi que de prévisions sur les intentions probables des participants. Ces besoins existent pour toutes les missions militaires possibles qui vont des activités antérieures à un conflit (prévention) aux activités postérieures en passant par le conflit lui même. En plus des informations en matière de reconnaissance fournies par les capteurs, le cycle du renseignement militaire est actuellement, et le sera encore plus à l'avenir, influencé par des aspects non militaires, nouveaux et complexes, comme des informations touchant aux domaines politique, ethnique ou de la criminalité Il faudra traiter de grands volumes de données et d'informations, alors que l'évaluation et la présentation manuelles des résultats prennent actuellement beaucoup trop de temps. L'exploration de données, intelligente et en profondeur, ainsi que la recherche d'informations basée sur la connaissance et l'extraction de contenu gagnent en importance, en particulier dans le domaine du contre-terrorisme. Il est par conséquent important d'identifier les questions clés pouvant avoir une incidence sur les problèmes de fusion des données et des informations et de trouver des méthodes permettant d'automatiser ce processus, afin de le rendre plus efficace aux fins du renseignement militaire.

Le symposium servira de forum aux chercheurs scientifiques, aux spécialistes militaires et aux ingénieurs systèmes pour leur permettre de présenter l'état de l'art en R&T dans tous les domaines de la fusion des données et des informations militaires du cycle de commandement et contrôle. L'accent sera mis sur les niveaux 2, 3 et 4 de la définition JDL de fusion des données. Les résultats des efforts de l'ensemble des pays membres de l'OTAN dans ce domaine évolutif devraient apparaître au cours des présentations techniques, des mémoires de recherche et des discussions orientées applications. En plus des communications officielles, le programme de la conférence inclura des interventions sur invitation de la part d'éminents spécialistes universitaires et de l'industrie. Le symposium se déroule en complément du symposium SET prévu sur « La poursuite des objectifs et le fusionnement des données capteurs pour les systèmes d'observation militaires » organisé au cours de la semaine précédente à Budapest, en Hongrie.

La liste de sujets d'intérêt suivants n'est pas limitative :

- Besoins militaires en matière de fusion des informations et d'interprétation des données
- Concepts de fusion de données et d'informations militaires aux fins de
  - Connaissance et évaluation de la situation des forces
  - Évaluation de la menace/de l'impact
  - Renseignement et préparation du champ de bataille
  - Classification et identification des éléments de l'espace de bataille
- Méthodes générales de fusion
  - Méthodes à base de modèles et de formes de référence
  - Méthodes Bayésiennes et approches basées sur la croyance pour la fusion des données et des informations
  - Exploitation des informations dépendant du contexte
  - Des communications extraction du contexte de données non structurées à l'aide de techniques de traitement du langage naturel
  - Défis de recherche dans l'obtention de la fusion robuste C2, C3
- Approche ontologique de la représentation des connaissances militaires
- Exemples d'applications et démonstrateurs de fusion de données et d'informations
  - Architectures de systèmes de fusion du renseignement pour C4ISR
  - Moyens : COTS S/W, données, supports de formation, scénarios
  - Fusion du renseignement géospatial et des informations militaires
  - Visualisation de la fusion
- Enseignements tirés

Information Systems Technology Panel

CHAIRMAN:

Prof. Ann MILLER
Distinguished Professor of Electrical and Computer Engineering
University of Missouri-Rolla
125, Emerson Electric Co. Hall
Rolla, MO 65409-0040
UNITED STATES

DEPUTY CHAIRMAN:

Prof. Marek AMANOWICZ
Ministry of Defence
Military Communication Institute
05-130 Zegrze
CZECH REPUBLIC

TECHNICAL PROGRAMME COMMITTEE

CHAIRMAN:

Dr. Jürgen GROSCHE
FGAN-FKIE
Neuenahrer Str. 20, 53343 Wachtberg
GERMANY
Tel: +49 (228) 9435 288
Fax: +49 (228) 9435 685
E-mail: grosche@fgan.de

CANADA

Mr. Gaétan THIBAULT
DRDC Quebec
Tel: +1 (418) 844 4000 Ext. 4540
Fax: +1 (418) 844 4538
E-mail: gaetan.thibault@drdc-rddc.gc.ca

CZECH REPUBLIC

Dr. Milan SNAJDER
Military Technical Institute of Electronics
Tel: +420 (2) 20 20 77 31
Fax: +420 (2) 51 62 33 97
E-mail: msnajder@vtue.cz

DENMARK

Dr. Per HUSMANN RASMUSSEN
Danish Defence Research Establishment
Tel: +45 (39) 15 1808
Fax: +45 (39) 29 1533
E-mail: phr@ddre.dk

GERMANY

Mr. Joachim BIERMANN
FGAN-FKIE
Tel: +49 228 9435 276
Fax: +49 228 9435 685
E-mail: j.biermann@fgan.de

ITALY

Dr. Luciano CERIONI
SIRTI Sistemi SPA
Tel: +39 (06) 8288 0419
E-mail: l.cerioni@sirtisistemi.it

NORWAY

Mr. Tore SMESTAD
Norwegian Defence Research Establishment
Tel: +47 (63) 80 73 77
Fax: +47 (63) 80 72 12
E-mail: Tore.Smestad@ffi.no

SWEDEN

Dr. Per SVENSSON
F.O.I.
Tel: +46 (8) 5550 3696
Fax: +46 (8) 5550 3686
E-mail: Per.Svensson@foi.se

UNITED STATES

Mr. Brian O'HERN
USAF/AFRL
Tel: +1 (315) 330 3995
Fax: +1 (315) 330 8063
E-mail: brian.ohern@rl.af.mil

Dr. David L. HALL
Pennsylvania State University
Tel: +1 (814) 865 8711
Fax: +1 (814) 865 7065
E-mail: dhall@ist.psu.edu

Mr. Alan N. STEINBERG
Utah State University
Tel: +1 (301) 229 6631
Fax: +1 (301) 229 6631
E-mail: asteinberg@sdl.usu.edu

Dr. James LLINAS
University of Buffalo
Tel: +1 (716) 645 2357 Ext. 2105
Fax: +1 (716) 645 3302
E-mail: llinas@eng.buffalo.edu

PANEL EXECUTIVE

From Europe:
RTA-OTAN
Lt.Col. A. GOUAY, FAF
IST Executive
BP 25
F-92201 Neuilly-sur-Seine Cedex
FRANCE

From the USA or CANADA:
RTA-NATO
Attention: IST Executive
PSC 116
APO AE 09777

Telephone: +33 (1) 5561 2280 / 82 – Telefax: +33 (1) 5561 2298 / 99

HOST NATION LOCAL COORDINATOR

Dr. Milan SNAJDER
Military Technical Institute of Electronics
Tel: +420 (2) 20 20 77 31
Fax: +420 (2) 51 62 33 97
E-mail: msnajder@vtue.cz

ACKNOWLEDGEMENTS/REMERCIEMENTS

The IST Panel wishes to express its thanks to the Czech Republic members to RTO for the invitation to hold this Symposium in Prague and for the facilities and personnel which made the Symposium possible.

Le Panel IST tient à remercier les membres du RTB de la République Czech auprès de la RTO de leur invitation à tenir cette réunion à Prague, ainsi que pour les installations et le personnel mis à sa disposition.

Besoins opérationnels en fusion en matière d'information et de renseignement

(Operational Requirements for Fusion in the Fields of Information and Intelligence)

Lt.Col. Baron Louis de Chantal
THALES Communications, SAS/SEA
160 Boulevard de Valmy
BP 82
92704 Colombes Cedex
FRANCE

LA FUSION, CLE DE LA MAITRISE DE L'INFORMATION ?

PREALABLES

Le domaine principal d'application de cette présentation est l'Armée de Terre dans son contexte d'opérations. Ce cadre d'étude est cependant facilement transposable aux domaines aérien et maritime, avec des caractéristiques toutefois assez différentes dans le domaine du temps utile, de la représentation et du volume d'information bien moindre.

Cet exposé prend en compte l'information de manière générale en faisant cependant une distinction forte entre la donnée, l'information et le renseignement en raison des diverses implications opérationnelles et techniques. La sémantique est prise en défaut ici en noyant dans le terme flou d' « information » des notions opérationnellement très diverses. Cette distinction se base sur la pyramide du savoir telle qu'elle est comprise en France et sur des études phares de l'OTAN comme celle du groupe RTGonIF/D sur la fusion de l'information. Le terme « information » sera généralement utilisé ici comme générique pour désigné un élément informationnel de n'importe quel niveau de complexité.

Le champs d'action informationnel de l'opérationnel recouvre le spectre total de l'information et de la connaissance dans les champs matériels et immatériels. Il est en effet impossible de parler de « maîtrise de l'information » sans faire une place importante à l'environnement global, en particulier dans la prise en compte de la population et de ses caractéristiques humaines et psychologiques.

En ce qui concerne les relations entre la fusion et le monde opérationnel, il est bien établi que ce dernier ne parle qu'en matière de résultats et non en technique, qui ne représente que des moyens d'aide à l'accomplissement de la mission. L'opérationnel n'exprime pas un besoin technique, mais un besoin de résultat.

Communication présentée lors du symposium RTO/IST sur "La fusion des données et des informations militaires", organisé à Prague, République Tchèque, du 20 au 22 octobre 2003, et publiée dans RTO-MP-IST-040.

Ces résultats sont généralement la somme d'efforts individuels soutenus par des aides techniques adaptées, donc de produits et des clients différents. Il en ressort clairement qu'il n'y a pas un mais plusieurs types de fusion adaptés aux produits nécessaires à l'obtention des résultats recherchés. Le niveau des fonctionnalités développées par le processus de fusion varie donc suivant le type de produit à fournir. Pour l'opérationnel, tout ou partie de ce processus est appelé génériquement « fusion ».

Enfin, La fusion n'est pas une fin en soi. Elle ne crée pas de nouvelles fonctions, elle optimise le travail de l'homme. La finalité de la fusion, comme tous les services et aides aux opérationnels, est la satisfaction d'un besoin suivant les normes opérationnelles et de travail exprimées par l'utilisateur.

Quelle que soit la difficulté technique de la fusion et les prouesses intellectuelles nécessaires pour la mettre au point, un système de fusion doit constituer une réelle amélioration pour l'opérationnel. Les implications liées à l'introduction d'un nouveau système sont naturellement pénalisantes pour l'opérateur et son organisation (réduction d'efficacité temporaire, perte de temps en apprentissage, adaptation, montée en puissance progressive,…). Tout système qui ne serait pas ressenti comme une aide sensible à l'homme, ou même se révèlerait être une gène pour l'utilisateur en raison de ses interfaces Homme – Machine ou de l'utilisation de ses fonctionnalités, perturberait la cohérence de l'organisation ou de l'architecture technique, ou n'améliorerait pas sensiblement une fonctionnalité ne serait pas acceptable.

Pour des raisons de facilité Le mot « homme » dans cet exposé sera utilisé dans sa définition générique de « genre humain » comme le définit le dictionnaire.

BESOIN OPERATIONNEL

Constat

Même si il est actuellement considéré que la sphère informationnelle, menant à la maîtrise de l'information, conditionne le succès des missions opérationnelles, le domaine de l'information n'est pas assez performant. Les principaux problèmes opérationnels relevés concernent les domaines suivants:

- Temps : respect des temps utiles, (processer et délivrer l'information en temps impartis)
- Volume : explosion du nombre d'informations disponibles (problème de ressources humaines)
- Gestion : tri, classement, marquage,…pour pouvoir mettre à disposition 100% de l'information existante (fastidieux, consommateur de ressources humaines, sans plus value intellectuelle)
- Amélioration : augmenter le taux de confiance (cotation) et de qualité (attributs) de l'information
- Elaboration : complexité et diversité des produits à élaborer (puissance de traitement de l'homme limitée pour embrasser tous les critères utiles)
- Complétude : utiliser toutes les informations afférentes à un sujet malgré leur éparpillement (géographique et temporel) et leur source

La supériorité informationnelle dépend de la réponse à tous ces facteurs.

Il semble donc que l'amélioration, à défaut d'optimisation, du traitement de l'information, exige essentiellement des progrès pour :

- concentrer l'information, dès le plus bas niveau, afin d'obtenir un volume moindre à exploiter ainsi qu'un enrichissement de l'information par l'augmentation des attributs attachés et une confiance supérieure par l'intermédiaire d'une cotation réévaluée,
- dégager l'homme des tâches de bas niveau n'exigeant pas de subjectivité ni de plus value humaine. La ressource humaine est un problème majeur dans les armées modernes, et la technique

permet maintenant d'automatiser un certain nombre de tâches. Il est donc intéressant de désengager l'homme de tâches d'exécution pure pour le réinsérer dans les tâches de réflexion. Il convient ici de fixer les limites de compétence et d'autorité entre l'homme et le système, ce que nous aborderons plus loin.

- Optimiser le traitement en matière :
  - de temps de traitement,
  - de qualité du traitement technique et du produit réalisé,
  - de quantité d'information traitée pour pouvoir utiliser la totalité de l'information disponible.
- Offrir des services spécifiques élaborés liés aux diverses fonctionnalités opérationnelles pour supporter le travail de réflexion et de commandement.

La fusion semble pouvoir apporter des réponses à ces défis.

CEPENDANT, même si l'homme n'arrive pas à réaliser seul ces tâches, il reste par défaut le décideur et le responsable. En tant qu'utilisateur, il doit garder au minimum un droit de regard et d'action sur les produits fournis. On ne peut décrire la portée de la fusion si on ne fixe pas ses limites par rapport à l'utilisateur. Cet aspect des relations entre l'homme et le système est abordé plus loin.

Typologie de l'information

L'information n'est pas une fin en soi. Elle répond à un besoin destiné à permettre une action.

Elle peut être brute ou travaillée afin de lui donner une plus value, ce qui implique un traitement différencié des informations pour chaque produit.

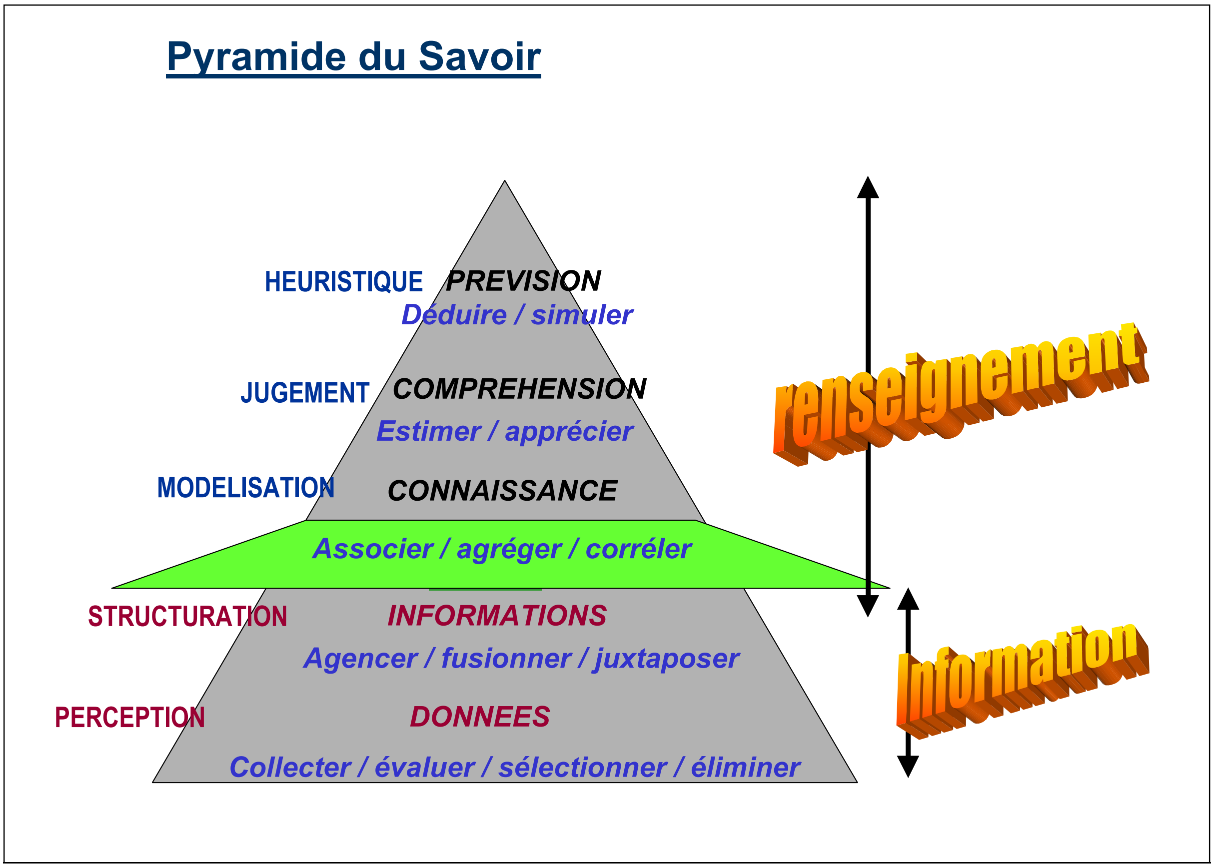
Chaque information est subordonnée à un besoin spécifique qui peut être :

- connu ou éventuel
- individuel ou collectif,
- spontané ou permanent.

De plus, l'impact de la prise en compte de l'environnement dans la doctrine influe fortement sur la sphère informationnelle en touchant l'information elle-même et sa technique de traitement. En effet, la traditionnelle dominance du couple « action – destruction » a été remplacée par celle de la gestion des champs psychologiques. Ce secteur d'activité encore quasiment vierge doit être défriché en finesse, et l'apport d'automatismes et de fusion, incontournable, demande une analyse très pointue des contraintes et limites.

Information passive et active. L'information peut être considérée comme passive ou active en fonction de son utilisation opérationnelle. Elle est passive lorsqu'elle répond à une demande et fournit une information à tendance documentaire, comme les cadres d'ordres et les images de référence. Elle est active lorsqu'elle se positionne comme un acteur quasi autonome et agressif, seule ou dans le cadre d'un système, s'auto alimentant pour obtenir un effet direct sur l'environnement. Elle peut agir dans le cadre d'un système d'arme ou d'un système d'information opérationnel.

Données – informations – renseignements. Pour répondre aux spécificités du besoin opérationnel, l'information peut être répartie en plusieurs niveaux d'intérêt et de traitement : données – information - renseignement (D.I.R.) qui correspondent souvent à des sphères d'activité militaires liées à l'action, la réflexion ou la décision. Ces sphères d'intérêt ne sont pas cloisonnées par niveau.



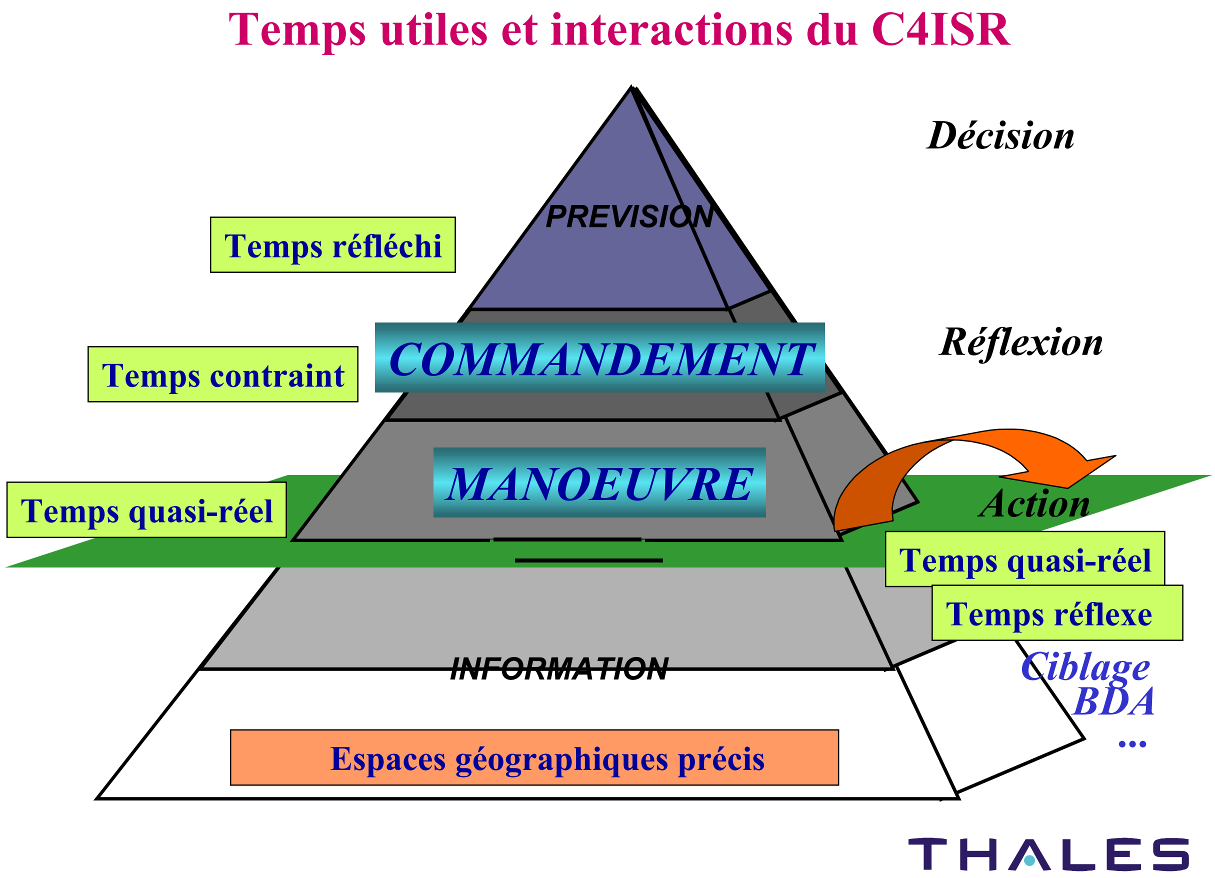
Besoin dans les domaines Action – Réflexion – Décision

Les produits informationnels répondent globalement à 3 catégories de besoin directement liées à l'organisation et l'exécution de l'action militaire: Action –Réflexion – Décision.

En fonction de ces catégories, les clients de tous les niveaux hiérarchiques sont différents (J3 Ops, J2 Rens, Etat-Major,…) et demandent des in formations différentes plus ou moins élaborées en fonction de leurs contraintes propres comme la précision, la granularité, le temps utile.

D'une manière schématique,

- L'information destinée à l 'action est caractérisée par un temps utile réflexe ou quasi réel et une granularité très fine. La zone d'action est généralement assez étroite, et la contrainte majeure est le temps de décision. Le client est généralement issu de la chaîne « opérations /J3» et « appui feu /targeting». La chaîne de surveillance est également abonnée à ce type d'information. Ce niveau a besoin d'informations de type « SAVOIR ». Aux petits échelons, l'information délivrée peut –et doit- être présentée de manière simple et adaptée à l'utilisation réflexe. Le soldat au contact a besoin non pas d'une carte géographique montrant le théâtre complet, mais de la visualisation de son univers en fonction de ses échelles de valeur : temps réflexe, distance de combat, parties vues et cachées, réalité augmentée,…A ce niveau, la fonction « Push » est essentielle et le système doit comprendre le besoin sans requêtes. L'aspect technique du système d'information est totalement transparent pour le client, ce qui sous-entends que le système de fusion doit pouvoir trier l'information et ne livrer que l'information strictement relevante et indispensable. Le traitement de l'information est relativement simple et peut s'arrêter à des coordonnées d'objet.



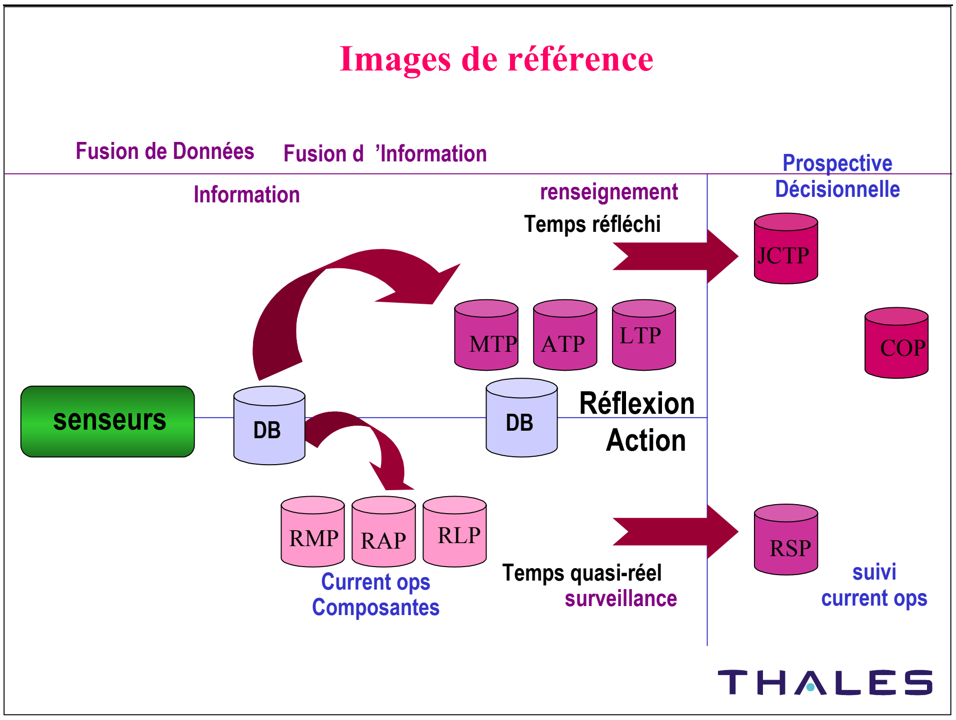
- L'opérationnel a également besoin dans ce contexte d'être aidé par des systèmes autonomes (typiquement l'engagement automatique de cibles clairement identifiées ou l'autodéfense), quasi autonomes (ne demandant à l'opérationnel que des accords de principe) ou de proposition (particulièrement dans les champs psychologiques où la plus value humaine est maximum. En effet, le champ privilégié d'action de l'homme est la gestion de l'homme). Il est clair que la limite de l'automatisation des aides est liée aux critères de responsabilité humaine. La fusion apporte à l'action une aide vitale en réduisant les délais de traitement, en embrassant l'ensemble des informations disponibles trop nombreuses pour être toutes prises par l'homme, et par la proposition de solutions rationnelles « par défaut ».
- La recherche d'informations ancienne de documentation et la comparaison avec la vision actuelle est une fonction particulièrement intéressante pour l'action. A l'extrême, on peut concevoir, dans certaines conditions d'engagement très claires, la constitution de systèmes de systèmes gérés par un centre de fusion et effectuant d'une manière autonome la couverture feu d'un groupe en reconnaissance en ville. Cet aspect sera détaillé un peu plus loin dans le cadre de l'optimisation de l'information.
- <u>L'information destinée à la réflexion</u> s'adresse plus particulièrement aux états-majors de tous niveaux, pour permettre la prise en compte par métier puis d'une manière transverse de grandes masses d'informations. Le temps utile, moins rapide, ainsi que la granularité sont liés au tempo de planification et d'exécution de l'échelon considéré : 3, 6, 12, 24H… Les informations demandées sont relativement objectives et issues du domaine « SAVOIR » pour être transformées en « COMPRENDRE » et « PREVOIR ». L'information de documentation, l'élaboration de référentiels et les systèmes d'analyse et de synthèse dominent. Toutes les étapes du cycle de

planification sont concernées, et les différents métiers et branches (de G/J1 à G/J9, conseillers et chefs de cellules) confrontent leurs informations et produits. Une interaction entre les produits des diverses branches se crée alors. Le besoin en fusion est alors très spécifique, parfois même « à la carte ». Les systèmes doivent donc être adaptés aux fonctionnalités de chaque métier et deviennent duals: création du socle informationnel spécifique métier (gestion du factuel) par une fusion « simple », et création complémentaire d'outils métier plus logiques et spécifiques. Les fonctions « jeu » (et rejeu) prenant en compte des fonctionnalités de métiers différents (par exemple en planification, l'étude comparée de l'impact sur la chaîne logistique d'une attaque amie à 3 contre 1 et à 4 contre 1). Les diverses images de référence, comme nous le verrons plus loin, sont une expression du travail de fusion de ce niveau de réflexion. Elles peuvent être réalisées en travail coopératif entre les niveaux concernés. Les besoins en coordination font appel à ce type de fusion.

- L'information destinée à la décision s'adresse aux responsables de tous les niveaux. Cette information peut être soit liée à une heure limite de fourniture, elle-même liée au cycle du commandement, ou inspirée par l'urgence d'une situation. Contrairement à l'information de réflexion, le temps utile rapide des petits échelons peut impacter celui des échelons supérieurs. Dans le cadre de la prise de décision, l'intégration de dernière minute de critères nouveaux, personnels (caractéristique de la prise de décision impliquant la projection de sa personnalité propre dans le choix des critères) et secrets (feuilles d'instructions personnelles des chefs de détachement) implique la mise en œuvre d'un système de fusion d'information multicritères paramétrable, permettant d'étudier l'impact des variations des critères sur le but affiché, et dirigeant la prise de décision. Le produit délivré par le système n'est qu'une proposition de choix rationnelle, le décideur devant alors juger et s'engager de manière subjective en fonction de ses affinités et de son expérience, mais aussi de son aspect psychologique personnel. Il peut alors aller contre les propositions du système.

- Information de documentation et référentiels: En marge de ces fonctions opérationnelles, le circuit de l'information repose beaucoup sur des bases de données et de connaissance, entretenues de manière permanente avant, pendant et après l'opération. Ces bases de données sont initialement constituées avec les informations de documentation recueillies jour après jour, enrichies et améliorées en permanence pour constituer des bases de référence et des référentiels organisés destinés à suivre les évolutions de situations, d'objets ou d'activités. Les référentiels peuvent être de tous types, texte, image,…mais peuvent concerner aussi bien les champs psychologiques et les modes opératoires. L'exploration de ce dernier domaine par la fusion pourrait se révéler un aide très riche en matière de contre terrorisme et de lutte contre la criminalité.

- Il apparaît clairement que deux grands types de besoin apparaissent, ceux menant à la connaissance pure et ceux menant à l'interprétation et à la prospective : « awareness » contre « assessment », factuel contre subjectif. Ces deux grandes options sont clairement distinguées dans l'élaboration des images de référence, images « recognized » de surveillance avec des temps utile rapides décrivant l'environnement de manière factuelle, et images « tactical » avec des temps utiles plus réfléchis permettant la réflexion. Leur but est de permettre d'interpréter et de projeter dans le futur l'évolution possible de la situation figée à un instant T. Le rythme de rafraîchissement de ces familles d'image est représentatif de leur temps utile habituel.



Optimisation de l'information

Faits et interprétation. Quelles que soient les faiblesses de l'homme en matière de traitement de l'information, le but de toute entité opérationnelle est de transformer l'information générale disponible en informations ayant une signification directement rattachée à son domaine d'emploi. A ce titre, il est tout à fait normal de voir une même information dirigée vers plusieurs utilisateurs pour être interprétée dans le sens de leur action. La nouvelle d'un char ami touché, par exemple, sera perçue par le G3 « Opérations » comme une diminution de son pouvoir d 'attaque, par le G4 « Logistique » comme le besoin d'envoyer une équipe de dépannage sur place pour le réparer, par la santé comme des blessés potentiels à trier et à soigner,…A l'inverse, un mécanicien pour char envoyé sur le terrain fera passer à son chef logistique son compte-rendu de fin de réparation sans mentionner qu'il n'a réparé que la partie mécanique et non les optiques et les transmissions. Le G3 qui recevra l'information pensera alors qu'il peut réintégrer ce char dans son dispositif d'attaque alors que celui-ci est toujours incapable de combattre. Il est donc nécessaire de dissocier le fait strict de son interprétation pour pouvoir valoriser l'information et l'optimiser.

Concentration. L'information délivrée à un utilisateur n'est pas seulement un produit fini, c'est également la concentration dans une seule information de tous les attributs compatibles issus d'autres informations par le biais de la fusion. Parallèlement, cette action entraîne le masquage d'un certain nombre de données afin de diminuer le nombre d'éléments à présenter à l'opérateur. Cette substitution implique cependant la création d'une information plus riche factuellement et plus orientée opérationnellement. La phase d'enrichissement des attributs est une phase très coûteuse en opérateurs et peu valorisante pour l'homme. Elle constitue un terrain d'excellence pour la fusion à plusieurs titres :

- en effectuant le travail à la place de l'homme, elle permet de dégager des ressources humaines plus utiles dans les tâches de réflexion,
- en fusionnant les informations, elle diminue le nombre d'informations qui seront présentées et manipulées par les utilisateurs ultérieurs,
- en spécifiant les attributs, elle permet d'obtenir une précision largement supérieure à celle obtenue par l'homme, et une réutilisation permanente sans limite temporelle.

Réalité augmentée. La réalité augmentée est un des grands besoins du combattant, que ce soit en terrain ouvert ou cloisonné comme la ville. En effet, la compilation de vues diverses d'une même zone par des senseurs différents permet de reconstruire une vision plus globale et complète de la zone. Particulièrement cruciale dans le cas du combat urbain, où le fantassin souhaiterait voir à travers les murs, la fusion multi source permettrait de cumuler en 3D ou 4D les informations multi directionnelles issues d'images, de radars, de textes ou d'autres sources pour les concaténer en une image globale. Cette image fusionnée permettrait, comme le propose le concept français de Bulle Opérationnelle Aéroterrestre (BOA), d'engager une manœuvre préalable des unités ou d'associer un système d'affectation des cibles futures ainsi décelées mais non encore visibles par les unités. La réalité augmentée permettrait de réduire les pertes et d'optimiser les actions.

Fusion multi sources. La fusion multi sources offre comme intérêt particulier, outre l'augmentation des attributs, d'augmenter de manière drastique la crédibilité de l'information face aux manœuvres de déception. De manière doctrinale, il est estimé qu'une information recoupée par 3 sources différentes (image, guerre électronique et humaine) peut être tenue pour vraie. En effet, chacune de ces sources a ses propres faiblesses et peut être leurrée spécifiquement, mais la réalisation cohérente de leurrage pour chacune des sources semble quasi impossible. Sans pouvoir vérifier chaque information, un certain nombre de faits déterminants pour la manœuvre ou le choix des options à prendre sont isolés et vérifiés délibérément de manière multi source. L'analyse technique obtenue par fusion permet de lever les doutes et d'obtenir des réponses.

Champs immatériels. Même si les champs immatériels et l'environnement semblent difficiles à analyser directement, les comptes-rendus textuels ou d'exploitation faits par les diverses sources peuvent faire l'objet de fusion. La prise en compte dans le processus des données de référence ou de documentation permettent alors d'exploiter plus finement les informations, que ce soit par le biais d'une quantification ou de l'étude des évolutions ou différences. L'écueil dans le traitement du domaine psychologique est la complexité du comportement humain et l'impossibilité de maîtriser tous les paramètres qui régissent les hommes minute après minute. Les besoins exprimés dans ce domaine devraient donc être très prudents et mesurés, sous peine d'obtenir soit des banalités, soit des aberrations. Là encore, les systèmes de fusion ne sauront travailler rien d'autre que ce qu'on leur aura donné en entrée.

Référentiels. Le besoin d'interaction entre connaissance et compréhension se traduit par la confrontation des informations issues de diverses branches et structures (PsyOps /CIMIC / Renseignement / Contre renseignement) pour obtenir une vision la plus complète possible de la situation courante. Jour après jour, les informations obtenues et leur confrontation avec le passé permettent de constituer des images de référence ou des référentiels relatifs aux centres d'intérêt de la Force comme les comportements et les modes opératoires dans la lutte anti terroriste. Issus d'interprétations de situations passées, ces référentiels sont examinés systématiquement ou ponctuellement par les J2 dans le cours de leurs réflexions et constituent des aides à la déduction.

Complétude et relevance. Le grand nombre d'informations qui circule ne signifie pas pour autant que toute l'information soit disponible, ni que celle qui existe soit d'un quelconque intérêt. L'intérêt d'un système est de mettre en relation entre elles les informations similaires ou les attributs cohérents afin de dégager des possibilités d'interaction entre les différents objets. L'architecture du système de fusion aura alors un impact sur le résultat opérationnel.

Discrimination. Essentiellement en zone urbaine, le soldat doit faire face à l'imbrication totale des civils, des ennemis et des amis. Cette confusion l'empêche de discerner clairement les zones à risque et peut l'entraîner à prendre une attitude par défaut hostile, source de tirs fratricides et de dommages collatéraux. La discrimination Amis / ennemis/ neutres est donc essentiel pour diminuer le risque et optimiser l'efficacité. L'apport de la numérisation sera essentiel pour permettre de fixer clairement les positions amies et, par déduction, en déduire les ennemis et neutres ensemble. La fusion et le pistage devraient permettre de discriminer encore entre ces 2 catégories, et ne laisser au soldat qu'une faible incertitude.

Restitution 3D+. La bonne connaissance du terrain est essentielle pour le combattant, que ce soit dans la « jungle » urbaine ou en terrain ouvert. La reconstitution en 3D des immeubles, sous-sols et infrastructures à partir de documentation existante mise à jour et construite en 3dimension par l'apport de capteurs spécifiques permet de préparer les missions avec soin et de manœuvrer avec efficacité. Elle permet aussi au combattant de voir son environnement avant de l'atteindre, et de cumuler les visions parcellaires des membres de son groupe pour obtenir leurs positions relatives et les zones dangereuses.

Auto-intoxication. Lié directement au besoin de partager l'information entre divers utilisateurs, le risque de voir revenir plusieurs fois la même information par des chemins différents est grand. Le risque est donc de croire que l'information a été vérifiée par plusieurs origines, donc est vraie, alors qu'il n'en existe qu'une qui a circulé et a été améliorée par différents utilisateurs. L'utilisation indispensable des « boucles courtes » contribue également à ce phénomène. En effet, une information jugée importante et urgente va d'une part être envoyée immédiatement vers son destinataire en court-circuitant toute la chaîne normale de traitement, et d'autre part placée simultanément dans le circuit de fusion pour la valoriser. La version valorisée rejoindra le même destinataire quelque temps après en ayant subi des modifications, mais sera considérée comme complémentaire sir un marquage caractéristique n'est pas prévu. L'auto intoxication peut être particulièrement grave car elle affecte par définition des informations graves et urgentes.

Information active

L'intérêt de la fusion est incontournable dans certains cas. Prenons par exemple la mise en œuvre d'un méta système composé d'un système d'arme dirigé par un système de fusion d'information, agissant en zone urbaine en couverture d'un groupe de combat. S'agissant de combat rapproché avec tir instinctif, on comprend aisément que le combattant n'a pas le temps ni les capacités (réalité augmentée) de réaliser une étude optimisant ses chances de survie.

Schématiquement, le système humain ou automatisé doit passer au travers de fonctionnalités et d'un processus complexes :

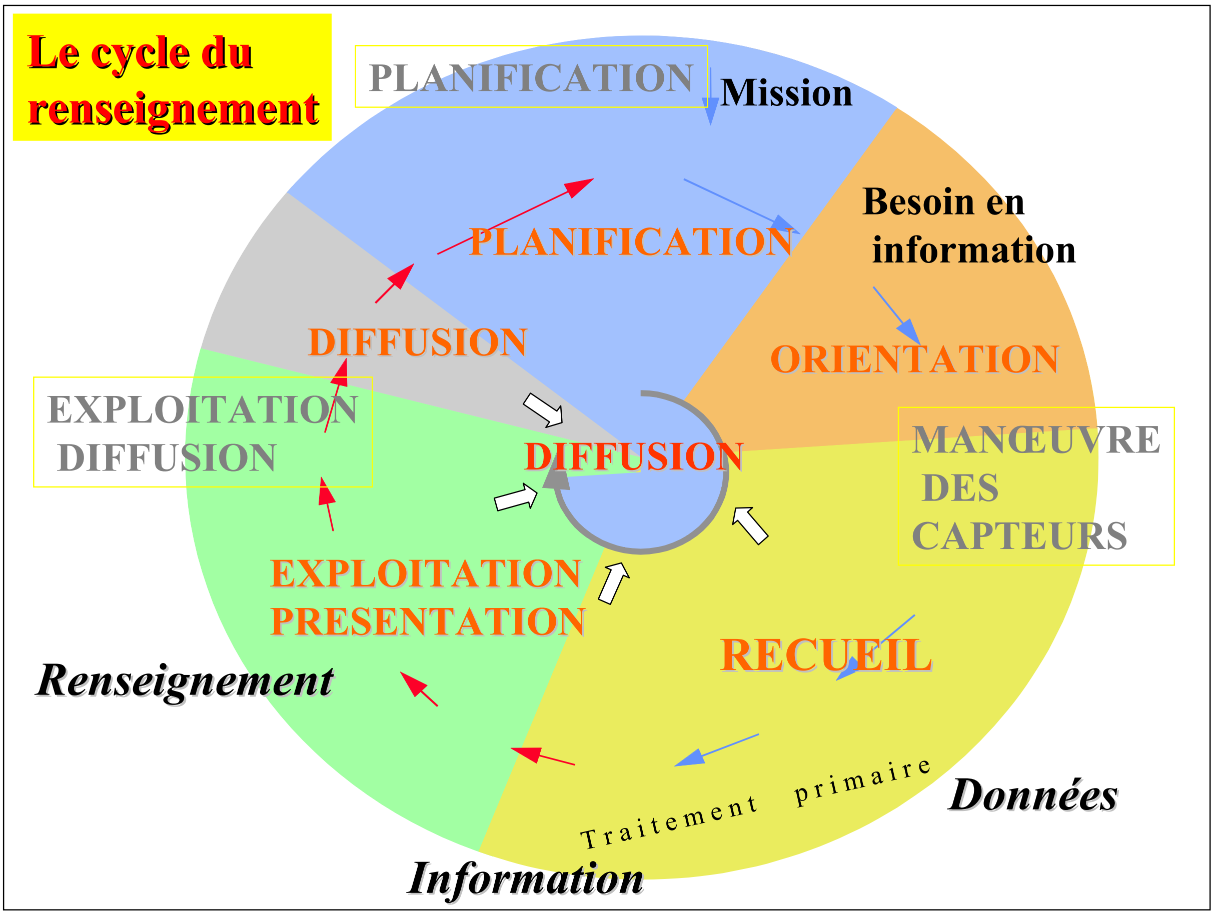
- acquisition du terrain (très dense et 3D),
- analyse du terrain,
- perception de menaces potentielles simultanées et disjointes géographiquement,
- analyse des menaces potentielles,
- confrontation avec des référentiels,
- discrimination et identification,
- choix d'objectif,
- choix de l'arme adaptée à la menace,
- choix du vecteur porteur en fonction de sa situation géographique et de sa capacité à engager la cible,
- décision et ordre de tir incluant les consommations allouées, les consignes diverses,
- recherche du résultat de l'action.

Les principes de fusion sont donc parfaitement adaptées pour répondre à ce besoin sur toute l'étendue géographique, à partir de sources multi capteurs et/ou multi sources, dans le temps requis.

Enfin, la concrétisation de la gestion globale des informations sera assurée dans le cadre futur de la BOA (Bulle Opérationnelle Aéroterrestre) de l'armée française de manière différente suivant les niveaux. Ces informations seront fusionnées à partir d'un travail collaboratif par cercle d'intérêt qui pourra se dégager de toute notion hiérarchique. Ces cœurs de fusion devront alors être obligatoirement couplés avec des systèmes de diffusion proactifs permettant d'extraire les informations relevantes automatiquement ou sur requête au profit de chaque abonné.

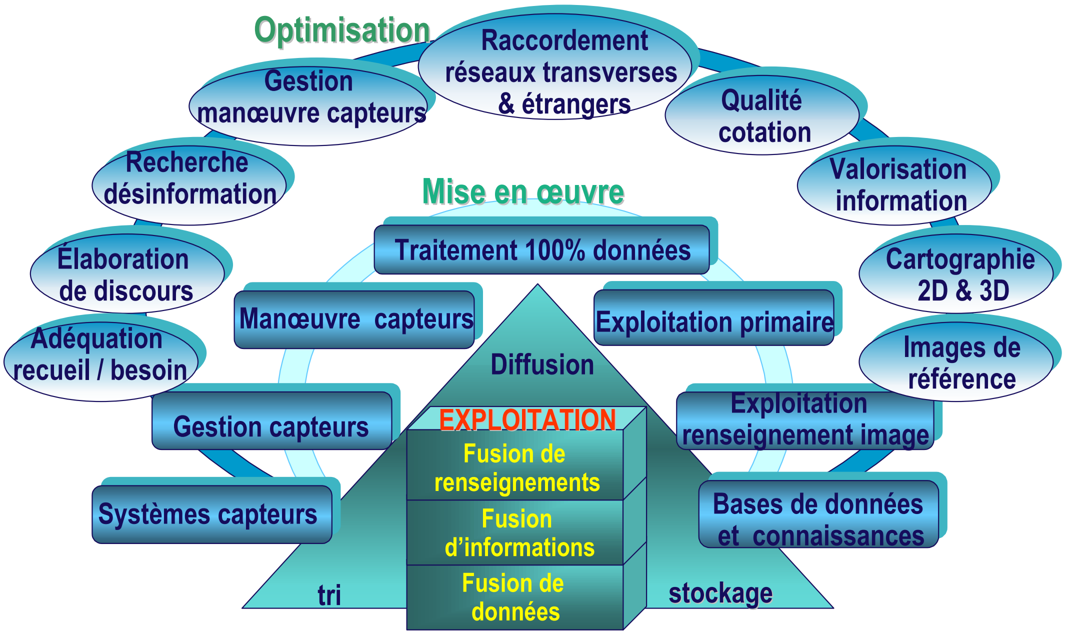
Optimisation des fonctions

Outre l'optimisation de l'information, il convient également de revisiter les fonctions opérationnelles séparément et dans leur ensemble pour les optimiser. C'est une des vocations du système ISTAR dans le domaine du renseignement. En effet, l'optimisation de chacune des fonctionnalités du cycle du renseignement est entreprise pour tirer le meilleur parti de chaque étape et la coordination de l'ensemble est assurée par une cellule spécifique.



Les fonctionnalités basiques de traitement de la donnée sont optimisées par un automatisme plus ou moins total. Un exemple de ces fonctionnalités est présentée ici dans le cadre d'un système ISTAR comme celui dont l'armée française va se doter dans un horizon de 10 ans. Les fonctionnalités évoquées s'assemblent ici comme des pièces indépendantes de Lego, chacune étant autonome d'un cœur principal comportant outre le système de fusion central des organes de gestion de l'information.

Potentiel de l'ISTAR



L'étude OTAN TGonIF (Task Group on Information Fusion) doit prouver par un démonstrateur l'intérêt de la fusion dans certaines fonctionnalités clés du traitement de la donnée et de l'in formation dans le domaine du renseignement, avec une connotation de lutte contre le terrorisme. Cette phase de démonstration suit une étude d'analyse fonctionnelle de la cellule de renseignement de niveau division MIC (Multinational intelligence Cell) sous l'égide du Canada qui a permis de mettre en évidence le fonctionnement d'une cellule de type ISTAR. Ces travaux très concrets, menés par une équipe multinationale intégrée de chercheurs et de conseillers militaires, devraient permettre de concrétiser des avancées dans le domaine des outils d'aide à l'opérateur.

Cette politique d'aide aux opérateurs est applicable à tous les domaines, que ce soit le commandement avec les images de référence par exemple ou la planification avec l'étude des théâtres potentiels, les choix du mode d'action , la confrontation des potentiels et leur analyse multicritère pour décider du dispositif à adopter et des variantes, de la simulation d'engagement et des implications logistiques sur l'action. Chaque métier doit être analysé de la même façon, avec un objectif double d'optimisation de la tâche et de réduction du personnel.

Plus encore que pour l'information, la définition des IHM (interfaces homme machine) Importance du format du display du résultat de la fusion

FUSION

La distinction opérationnelle entre les divers types de besoin rejaillit sur le domaine technique de la fusion.

Les informations destinées à devenir du renseignement sont issues du domaine « SAVOIR » pour être transformées en « COMPRENDRE » et « PREVOIR ». Le traitement de l'information est donc plus élaboré et l'ajout de plus value humaine le complique. La fusion requise devient symbolique, et les fonctionnalités réalisées par la fusion s'ajoutent aux étapes initiales de la fusion numérique. On peut dire que le socle informationnel « SAVOIR », traitant d'éléments très peu subjectifs, est réalisé par l'enchaînement plus ou moins complet des étapes de la fusion numérique, et que le fait de vouloir « COMPRENDRE » impose un ajout de subjectivité humaine qui change le type de traitement possible. Le plus ou moins « 100% automatique » est donc techniquement et opérationnellement concevable quand on parle de traitement d'information simple et factuelle. A l'opposé, la responsabilité de l'homme est engagée lorsqu'on aborde le besoin de subjectivité. Le pourcentage d'automatisme baisse alors pour devenir un système d'aide à la pensée humaine. Opérationnellement, on ne peut pas en tout cas écarter le besoin de posséder une fonction de réglage de l'automaticité de la fusion en raison du rythme de l'opération. Dans le cas d'une planification de division à 24 heures, l'opérateur possède le temps nécessaire pour réfléchir et interagir avec le système pour optimiser les résultats en insérant de nouveaux critères. Dans le cas d'un imprévu, où en phase d'activité intense, l'opérateur doit pouvoir décider de sa direction d'effort en fonction du temps qu'il a et décider de laisser l'automatisme plus libre.

La coexistence de divers temps utiles, et en particulier les temps réflexes et quasi réels, impliquent l'existence d'une boucle courte de traitement - ou de non traitement temporaire -, qui ne doit cependant pas exclure tout cycle de traitement parallèle. Cette double boucle simultanée - le by-pass du circuit de fusion et l'introduction dans ce circuit - génère les problèmes d'auto intoxication évoqués précédemment. Chaque information en entrée doit donc comporter une indication d'urgence entraînant un cursus particulier.

Le niveau de fusion, le type de fonctionnalité traitée dépendent pleinement du but à atteindre. Simultanément, ils influent sur l'organisation de l'architecture et sur les options offertes à l'opérationnel. La fusion de renseignement de documentation est plus complète et comporte des étapes plus avancées que le simple renseignement de combat, qui n'est parfois autre que l'affichage d'une information simple comme la localisation d'un objet inconnu.

Là encore, la fusion ne représente qu'un service pour l'opérationnel, et la partie technique doit se plier à ses exigences. La modularité des résultats et le rejeu avec des paramètres modifiés est un besoin standard des systèmes liés à la réflexion. Cette difficulté est augmentée par le fait que la signification du terme « fusion » n'est pas la même suivant les pays, les entités civiles et militaires et même souvent au sein des divers organismes techniques de chaque pays. L'appréciation du besoin de contrôle humain sur le processus technique subit les mêmes contraintes. Curieusement cependant, le groupe de travail TGonIF/D ne subit pas cette loi malgré la diversité des personnalités qui le constitue, et la cohérence de compréhension est presque parfaite. Ce constat est porteur d'espoirs…

La grande diversité technique de la fusion apparaît dans la confrontation des fusions de différentes sources. La fusion d'image procède par des chemins qui ne sont pas ceux de la fusion textuelle, par exemple. Le mixage de la fusion multi source représente donc la globalisation des problèmes de chaque type de fusion, avec en plus les problèmes de cohérence d'ensemble. Le domaine de la fusion image multi capteur a cependant été bien adressé par certains systèmes comme celui utilisé par l'armée française. Les systèmes de guerre électronique, qui possèdent leur propre logique, sont aussi aidés par des systèmes de fusion assurant la fusion multi capteurs.

RELATIONS ENTRE L'HOMME ET LE SYSTEME AUTOMATISE

Les relations entre l'opérationnel et les systèmes automatisés de fusion sont complexes. Ces deux entités sont complémentaires, car la fusion optimise le rôle de l'homme et l'homme optimise le travail de la fusion. Mais le comportement humain peut remettre en cause l'efficacité des moteurs de fusion, voir même leur utilisation ou leur utilité.

Optimisation de l'homme

La fusion assure directement deux rôles vis-à-vis de l'homme :

- lui redonner son vrai niveau de responsabilité dans le traitement de l'information : En le déchargeant des tâches répétitives et sans plus value humaine, la fusion permet à l'homme de se consacrer à son vrai métier qui est l'exploitation et non le tri et la gestion. L'expérience, le jugement et la décision sont en effet les ressources spécifiques de l'homme qui complètent parfaitement les capacités de calcul et de gestion des systèmes. La ressource humaine est donc mieux employée et ses compétences optimisées.
- lui permettre d'assurer qualitativement et quantitativement sa mission opérationnelle : la mission opérationnelle est confiée à l'homme et non pas à un service. C'est donc lui le responsable de son exécution, par tous les moyens dont il pourra disposer. Par la prise en compte de certaines de ses charges de travail et l'optimisation de tâches, la fusion permet de traiter toute l'information en complémentarité avec l'exploitant, valorise le travail humain et augmente son efficacité globale.

Complémentarité lucide

Les opérationnels utilisant des informations ne réalisent pas la somme d'opérations d'analyse et de synthèse successives leur cerveau réalise à leur insu pour transformer chaque bribe d'information anodine en un élément logique d'un puzzle dont ils ne connaissent pas le contour. La décomposition fonctionnelle de ce processus humain incontrôlé, qui intervient en dehors de toute notion de temps, fait souvent ressortir entre 10 et 20 étapes successives qui sont autant d'automatisations à réaliser pour optimiser l'ensemble du processus. Pour un observateur en poste depuis quelque temps dans un endroit, une silhouette de char passant sur un pont sera enregistrée dans un coin de son cerveau avec plus d'attributs qu'il n'en a vu. Son expérience militaire, la connaissance du contexte, les habitudes de son environnement, l'intuition, sa logique, ses anciennes informations feront qu'en 1 seconde et de manière automatique il aura analysé la scène, extrapolé et enregistré non pas les faits mais son interprétation des faits. La faille reste que l'effort fait pour percevoir n'est pas multidirectionnel, que le nombre d'évènements stockés est somme toute limité et que la prioritisation des informations ainsi enregistrées fait disparaître les informations ne semble pas prioritaire dans l'instant, effaçant les traces qui auraient pu servir dans le long terme. Le processus de fusion doit donc décomposer suivant le produit à réaliser les diverses fonctions nécessaires pour arriver au résultat final.

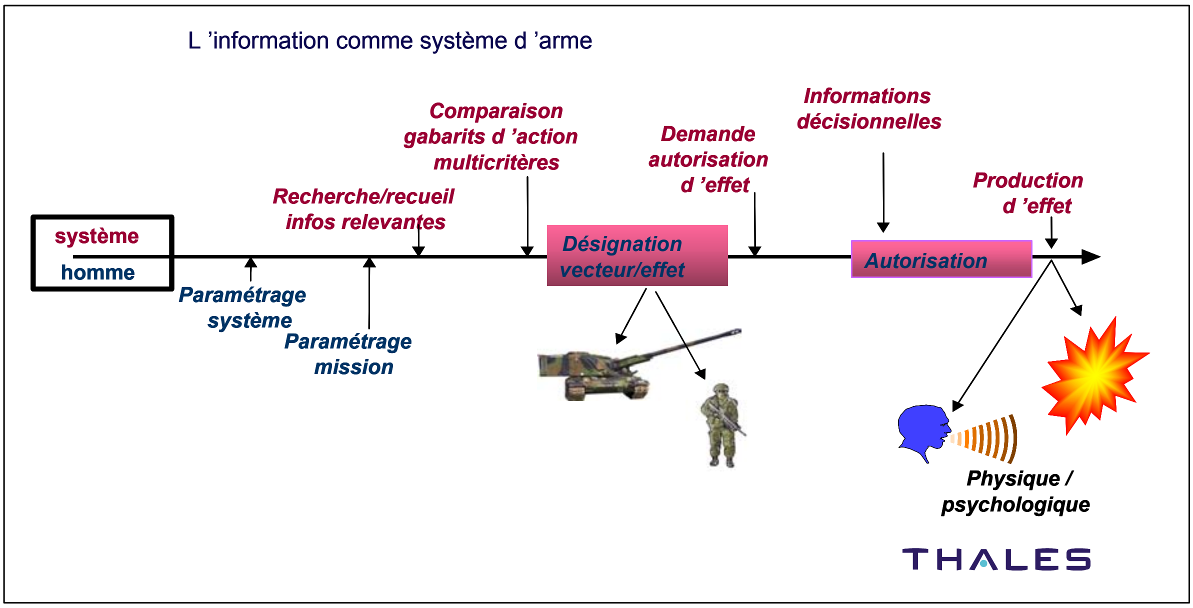
Une des difficultés majeures est donc d'arriver à imiter le plus fidèlement possible le processus intellectuel et psychologique qui anime le cerveau humain. Il est cependant posé que deux cerveaux n'analyseront pas de la même manière, que les conditions psychologiques de chacun interviendront pour modifier la relativité des critères, et que les conditions de travail et de stress agiront de manière différente sur l'avancement du travail. Un opérateur fatigué et souffrant verra non seulement ses capacités réduites mais également son optimisme ébréché. Il aura donc tendance à interpréter de manière pessimiste les faits. Il les jugerait vraisemblablement de manière différente le lendemain si, après une bonne nuit, il venait d'apprendre une victoire ou de recevoir de bonnes nouvelles.

Cet aspect trop méconnu est pourtant un des atouts de la fusion. Au lieu d'avoir une approche négative basée sur la responsabilité humaine et l'habituel discours des prérogatives de l'homme sur la technicité, il convient de réexaminer très sérieusement l'intérêt d'un système contrebalançant la subjectivité naturelle de l'homme, et particulièrement dans le domaine du renseignement, où le fait doit rester strictement un fait pour pouvoir être interprété. En effet, loin de se poser en rival de l'homme, le système de fusion à l'avantage de poser un œil froid et dépassionné sur la réalité perçue. Il est alors l'indispensable complémentarité de l'homme en lui permettant de confronter une vision lucide à sa vision subjective. Cette complémentarité devrait être recherchée systématiquement dans le cadre de la désinformation et de la déception.

Information active

Le domaine de l'information active, comme vu précédemment, est un champ d'action privilégié pour la fusion. La dualité Homme – Système y relève souvent de la dialectique du Maître et de l'esclave : l'homme reste le maître ultime et dirige la machine, mais il est fondamentalement dépendant des systèmes s'il veut optimiser ses chances de réussir sa mission. Dans ce cadre, un système est constitué pour remplir une mission spécifique en soutien de l'homme, avec un temps utile quasi-réel. Initié par les paramètres choisis et introduits par l'homme, le système fusionne de manière autonome ses informations, « décide » de l'action à mener en fonction de ses moyens, de ses référentiels et de ses paramètres et engage l'action. La liberté d'action du système est fixée initialement par l'opérationnel.

Il faut noter cependant que l'enjeu dans ce contexte est non seulement la vie ou la mort du combattant, mais aussi la discrimination entre combattant adverse et population civile imbriquée. Les règles d'engagement préalables et le contexte de l'action constituent donc une limite claire au « tout automatique ».



Interfaces homme-machine

Même si cet aspect, qui ne concerne pas que les systèmes de fusion, semble secondaire, il constitue en fait un enjeu majeur et un vrai défi. Il devrait constituer la vraie finalité des produits

- En amont du système, la qualité de la prise en compte du système par l'opérateur influe directement sur la bonne exécution des tâches. Une bonne compréhension de l'IHM à réaliser consiste à étudier le profil spécifique de l'opérateur standard pour déterminer ses connaissances et ses habitudes. La fracture entre les « anciennes générations » d'opérationnels et « les jeunes » est énorme. A titre d'exemple, l'utilisation de techniques utilisées pour les jeux vidéos aura très peu d'efficacité sur un outil s'adressant à des cadres de plus de 40 ans, mais sera génératrice d'une haute efficacité chez les plus jeunes rompus a ce type de manipulation et de concept.
- En aval du système, la manière de délivrer un produit est déterminante sur l'utilisateur et peut obérer la qualité intrinsèque du produit. Le « display » final, majoritairement délivré sur écran informatique sous forme d'un affichage et de fonctionnalités, est la partie émergée de l'iceberg et conditionne l'efficacité, donc l'appréciation de l'opérateur. Il doit être parfaitement adapté à la logique générique de l'homme mais aussi au besoin spécifique de la fonction et du produit.

En effet, la fusion étant appliquée de manière différente suivant le produit à fournir, le produit final doit refléter ce besoin spécifique.

Besoin d'évolution des mentalités

Il ne faut pas mésestimer la part de résistance au changement de certains opérationnels, et plus spécifiquement de l'Armée de Terre, envers les systèmes automatisés.

L'utilisation des aides, dont les systèmes de fusion, sera soumise à 2 « principes de friction » majeurs :

- les réactions épidermiques liées au changement des habitudes, à la remise en cause de ses savoir-faire, au besoin d'apprendre et d'entrer dans un monde inconnu,
- la remise en cause de la responsabilité personnelle. Ce facteur est particulièrement important dans les cas de décisions importantes comme la mise cause de vies, que ce soit la sienne où celle des autres, liées au métier des armes. L'éthique du comportement réagit dans ce cas à des règles non-dites issues de la fusion personnalisée de nombreux facteurs humains et psychologiques, par essence non quantifiables. Le combattant terrestre est habitué à se heurter individuellement à la complexité des situations d'engagement toujours différentes, dans un milieu résolument polymorphe, composite, flou et peuplé d'innombrables objets. Les opérations en zone urbaine augmentent encore la complexité de la réaction en même temps que le besoin en information. La part de la décision humaine est alors énorme, et fait directement appel à la subjectivité naturelle de l'homme ainsi qu'à son expérience, ses connaissances profondes et jusqu'à son instinct. La conscience personnelle du combattant, ainsi que sa culture, font que ses réactions ne seront jamais totalement cartésiennes et qu'à ce titre, il ne peut donner « carte blanche » à un système automatisé pour lui faire prendre des décisions ou mener des actions dont il devrait supporter les conséquences sur le plan moral ou judiciaire.

CONCLUSION

L'opérationnel recherche déjà toute aide qui pourrait lui permettre d'accomplir sa mission. La fusion en fait partie. Le besoin est réel, mais des difficultés subsistent :

Paradoxalement, la limite majeure du système de fusion est l'homme et non pas la technique. Quelle que soit l'efficacité des moteurs de fusion, le résultat obtenu n'est rien d'autre que le dividende de ce que l'opérateur a entré. Cela sous-entends que le système n 'obtiendra aucun résultat si les bons paramètres n'ont pas été rentrés, et si le besoin opérationnel générant l'élaboration du produit n'a pas été clairement exprimé. En dehors de la réalisation du moteur de fusion, il faut donc se pencher en amont sur les moyens d'obtenir une bonne définition du produit à élaborer et des étapes à respecter, ce qui ne fait pas partie du métier de l'opérationnel. Si l'on se souvient de la comparaison entre le processus instantané du cerveau et les 20 phases consécutives du processus technique, il apparaît clairement qu'il faudrait également élaborer un système permettant à l'opérationnel de décrire son besoin.

Il faut garder également à l'esprit que les processus de fusion ne sont que des moyens, et que la concrétisation de leur travail ne pourra advenir que s'ils sont liés à système performant de diffusion et que la présentation finale des produits est réalisée sous une forme exactement adaptée au besoin et aux procédures de l'utilisateur concerné. Plus que jamais, le fusion est un ensemble de systèmes « à la carte » à disposition des opérationnels.

Il est difficile de conclure sans devoir redire que la fusion reste sous le contrôle de l'homme et de sa mission, que ce soit en amont du processus ou pendant celui ci, et ne représente pour l'opérationnel qu'un moyen d'arriver au bout de sa mission en profitant de cette aide technique.

Dans ce sens, un système de fusion, même s'il est plus coûteux, rend à peu près le même service que la machine à café indispensable à tout bon état-major : on y puise les ressources complémentaires nécessaires pour réaliser le travail demandé…

(English Version)

Operational Requirements for Fusion in the Fields of Information and Intelligence

FUSION, NATURAL COMPLEMENT OF THE OPERATIONAL?

Operational functions related to action, reflection and decision do not receive the necessary support enabling the full completion of the tasks.

The dissemination of raw and processed information at various levels and the support of decision aid systems are crucial requirements for the operational world in the battle for information dominance. This support, which is deeply related to the specific needs of the operational entities, collides with the huge volume of available data, the diversity of the need, the different due times and the specific constraints.

The ever-increasing constraint of the operational tempo and the developing complexity of operations in civilian environment rises the need for processed products and aids, with a recurrent function of compiling objects and parameters.

As human processing cannot achieve its operational tasks in a satisfactory manner, the complementarity of man and system looks more than ever key to better handle the operational situations.

Fusion under its various forms appears to be the natural complement to the operational.

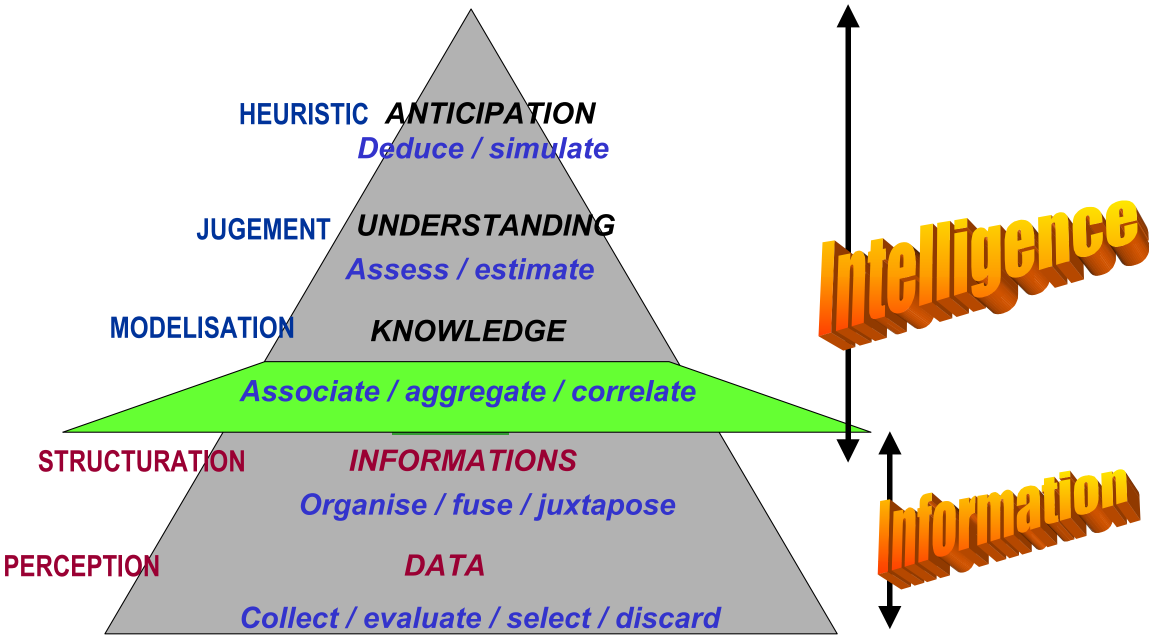
THE OPERATIONAL WHEREAS

The quality of any informational product is linked to some specific constraints, in which time, (split into several different due times), volume of information and complexity are determinant. Due to these factors, a significant enhancement of information would imply the need to concentrate information, optimise the use of human resources and optimise the processing. In addition, the provision of specific services would complete the effectiveness of functionality.

For the military point of view, information is not an end state, but an answer to a requirement for an action, which requires the providing of raw or processed information under different forms. As such, Information is divided in 3 categories: data, information and intelligence, related to the military activity spheres of action / reflection / decision. Its informational content and its due time responds to the specificity of each operational user, thus operational requirement. This specification deeply impacts the processing.

It must be mentioned at this point that Land activity places a new focus on psychology and sees its traditional dominance of « destruction » replaced by the « psychological aspects ».

The constitution of the informational products is the result of a chain of processes, aiming at providing those specific products to clients.



With regard to constitution of products, information can be related to action, to reflection or to decision. The respective attributes of the various kinds of information provided are usually specific:

- Action requires a very short due time (NRT, reflex) and a small granularity. Its main feature is a short decision time and a sharp display. This mainly concerns the Operations Branch (G/J3), including Fire support and targeting.
- Reflection requires variable delays linked to the hierarchic level, a medium to high granularity and a large geographical area. Its main clients are the staffs looking for additional information to perform their specific works.
- Decision is mostly level related but with possible interference between linked levels. Deciders are the commanders of all ranking, dealing with personal contexts and multiple decisional criteria.

THE NECESSARY OPTIMISATION

Optimising both information and functions is the condition to obtain a better informational material. This optimisation must revisit the traditional fields of information, but also those related to immaterial fields, and particularly all matters pertaining to environment, to the completion and relevance of information, to discrimination, display and self intoxication, which appears to be a crucial problem.

Global management of forward information and constitution of referential as well must also be checked.

To accompany information enhancement, it is indispensable to revisit any and each function in order to reduce the volume of human resources wasted in the completion of "small jobs", speed the process and proceed the full scope of potential information related domains.

With regard to implementation and efficiency, an effort should be made to group the common tasks, add the missing parameters, secure the basics of each task and ease the operational use. This has to be done in conjunction with the review of the operational reality and its new operational contexts and technologies.

VARIOUS FUSION FOR VARIOUS NEEDS

It is becoming evident that the various requirements of the multiple clients constitute that much divers kinds of fusion products, thus of fusion process. At this stage, it is obvious to consider that fusion is not one but plural, and any step of its full process can become finality for a user due to its personal requirement.

Concentration will however remain one permanent feature of fusion, allowing the reduction of the number of available data and the enhancement of its attributes. In such way does it also create the conditions for a higher value and confidence in the information, then in rating.

Fusion will allow to realise some particular products as enhanced reality, which is a "must" for urban operations, and multi sources fusion through cross cueing.

The realisation of autonomous information systems, self-generating actions to sub systems, could be envisaged through he concept of Active information. Such system could prove very useful in specific contexts, under the reserve of keeping man setting very precise parameters.

Human-system relations appear to be a lucid complementarity as it totally mixes different stages of work and allows to mix experience, culture, context and personality to technical process.

As mostly automated fusion should be kept to process low level information, and contribute to support the operators for humanly valued ones.

HUMAN MACHINE INTERFACE

The major stake for military personnel is the real achievement of tasks. The way products are constituted is of no interest for him as long as he gets what he required. To this end, products must imperatively respect the natural skills and knowledge of each operator in its own domain, and propose the adapted personalised services. The user's final display must be logical and specific.

On the operational side of the house, there is a need for mentality evolution.

Two major "friction principles" currently exist:

- The epidermic reactions to any change to skills and habits,
- The crucial issue of personal responsibility, which is a major issue for Land Forces as it impacts the ethics of comportment, the freedom of reaction and the human responsibility for fire.

CONCLUSION

Even if fusion has still a long way to go, the major limit of fusion is human, not technical. Operational input is determinant and the choice and relativity of the parameters implies the effectiveness of the processes and the completeness of tasks.

If there is a real need of fusion to free man from low-level tasks, complete the jobs and provide assistance, Fusion process is only a means devoted to the operational success, a technical asset converging towards operational success.

Defensive Planning for Combined Forces

David A. Griffith
Air Force Research Laboratory (AFRL)
Information Directorate (IF)
Rome, NY 13441-4505
USA

david.griffith@rl.af.mil

ABSTRACT

This paper is based on the experience gained by personnel of the Information Directorate of the US Air Force Research Laboratory's Rome, NY Research Site, during the development, validation, testing and fielding OF the Joint Defensive Planner (JDP) in the USAF Theater Battle Management Core Systems (TBMCS). JDP was used, for the past three years, in its successive developmental spirals, by the US/NE/GE Extended Air Defense Task Force during the Joint Project Optic Windmill (JPOW) exercises in NATO. JDP is currently being evaluated by personnel of the United Kingdom (UK) Joint Force Air Command HQ and the UK Air Warfare Centre for potential use in UK Systems. How this all came about is the essence of this paper that addresses many of the topics associated with this Symposium.

INTRODUCTION

This paper "Defensive Planning for Combined Forces" addresses the theme of this Symposium: "Commanders at all levels and types of military organizations require timely and accurate awareness of the situation in their respective areas of responsibility as well as prediction of likely intentions of participants".

Of equal, or greater, importance is that Commanders at all levels have a **consistent** situation awareness, although the level of detail may vary, depending upon position in the Chain of Command or organizational functions to be performed. The overall goal is to provide a single, joint software planning tool to assist Theater Air and Missile Defense (TAMD) staffs of the Combatant Commander (COCOM), Joint Force Commander (JFC), Area Air Defense Commander (AADC), Regional Air Defense Commander(s) (RADC), and Service Component Commanders to collaboratively develop the operational level joint TAMD plan to counter air and missile threats. This was an important JDP design goal that was met during development and fielding. Figure 1 illustrates the operational context of one TAMD planning tool being used at the strategic level (Joint Force Campaign Planning) and down to the operational level (Situation Monitoring and Plan Repair).

The resultant Area Air Defense Plan (AADP) documents the AADC's plan for integrating and coordinating joint air and missile defense. It details how TAMD operations will support the Joint Force Air Component Commander's (JFACC's) Joint Air Operations Plan and the JFC's Campaign Plan. This paper reflects on JDP Lessons Learned and incorporates them as general guidelines to describe a process to develop defensive planning tools for combined forces.

Joint TAMD planning includes campaign level deliberate and crisis action planning. Combatant commanders translate national and theater strategy into strategic and operational concepts through the development of

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

theater campaign plans. Deliberate planning prepares for a potential contingency based upon the best available information and using forces and resources apportioned to the COCOM by the Services. Deliberate planning is conducted principally in peacetime to develop joint operation plans for contingencies identified in strategic planning documents. Crisis action planning is based on current events and conducted in time-sensitive situations and emergencies using assigned, attached, and allocated forces and resources. Crisis action planning follows procedures that parallel deliberate planning, but are more flexible and responsive to changing events. The planning tools should be adaptable to various configurations (i.e., in garrison, aboard ship, in exercises, and in deployments).

This paper addresses several of the topics of interest of this Symposium. The seven following sections relate these topics to defensive planning.

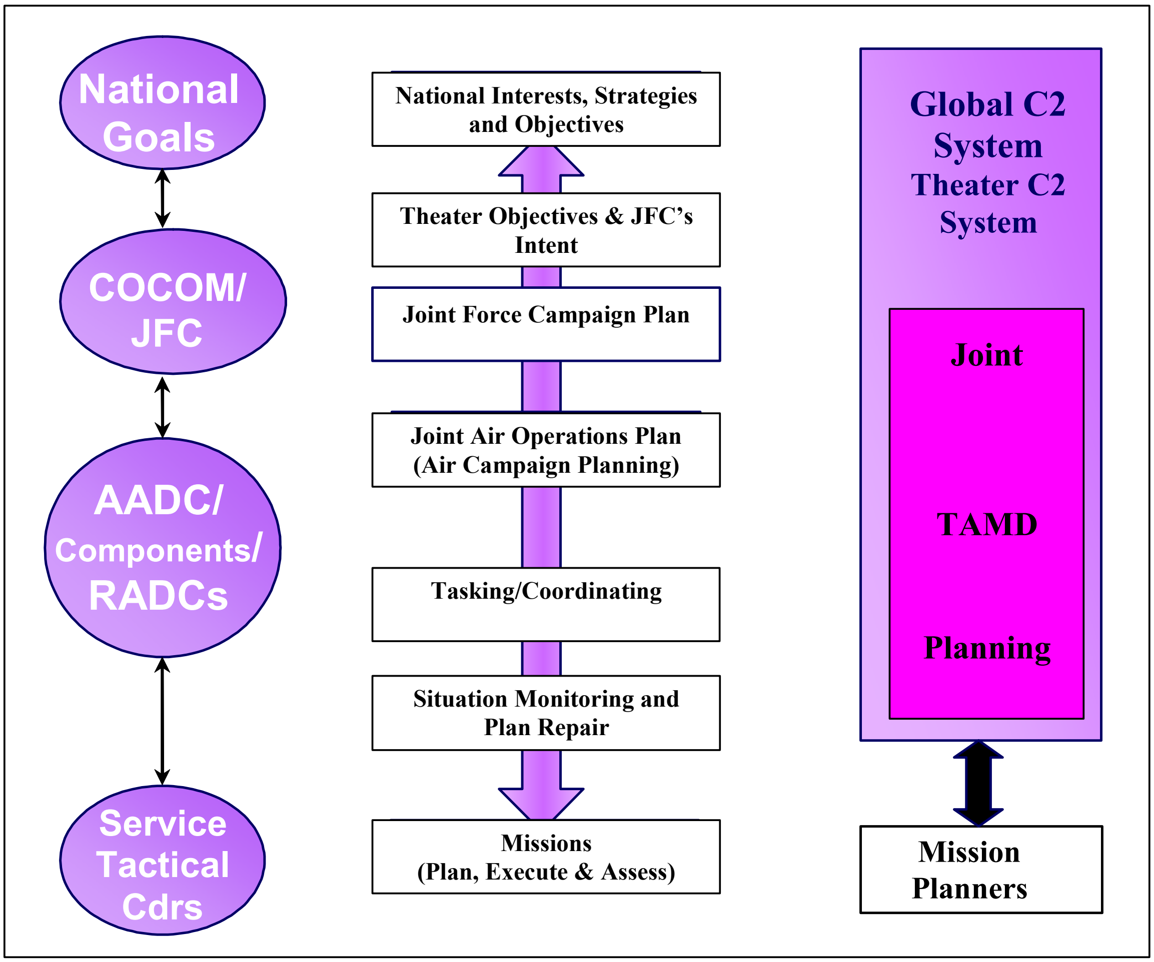


Figure 1: TAMD Operational Context.

1.0 MILITARY REQUIREMENTS FOR DATA INTERPRETATION AND INFORMATION FUSION

General military requirements are to create an Area Air Defense Plan for the entire theater of operations for the entire campaign which may consist of many phases. Defensive planning and operations are critical in the early phases of a campaign as offensive capability builds up. New and complex non-military information aspects such as national, alliance and coalition political and economic objectives and strategies must be translated into military objectives (Effects Based Operations) and strategies that finally result in military tasks. These military tasks must be traced back to the original stated objectives and strategies and must be consistent with the Rules Of Engagement (ROE) and legal aspects of the operation. The objectives and strategy statements must be electronically captured at the highest Command level and must be made available at all Command levels as more detailed and responsive planning ensues. The defensive planning tool should capture these objectives and use them as criteria during plan development while also allowing the commanders at lower echelons to refine the guidance from higher echelons.

In terms of planning tool development, the general requirements described above must be decomposed into more detailed, or derived, functional design requirements to the level that trained and skilled planners and other defensive operators need to perform their jobs. These requirements are generally referred to as "Workstation Requirements". This is accomplished by forming a Joint User Group (JUG) of defensive operational expert users from as many services, agencies, and organizations as possible. The JUG guides the development process, evaluates sequential prototypes from both technical and operational viewpoints, and approves the final product. This process assures the final product meets operational objectives. The Group also makes informed judgements involving cost and time constraints against technical and functional performance. Figure 2 represents this process. The process of rapid evolutionary prototype (i.e. spiral development) demonstrations is conducted with a large group of user representatives. This is the preferred way of verifying very early, and continuously through the development cycle, that the planning tool is performing the required functions, that the correct data is available, and is fused into information the user needs to see. At this point the visualization aspects should be derived.

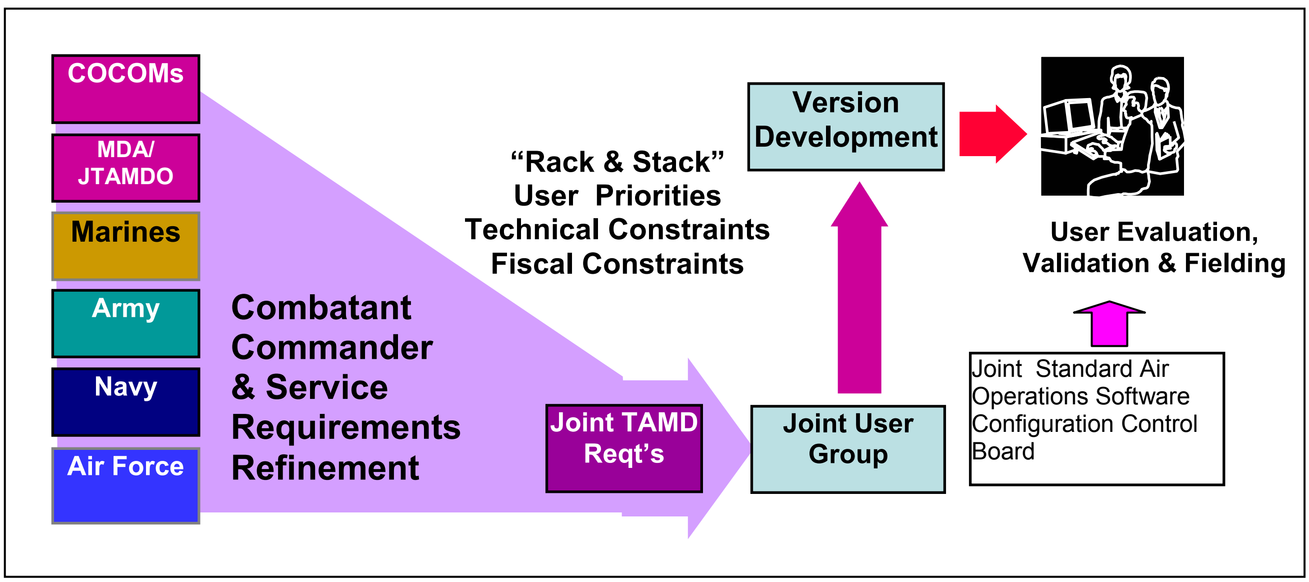
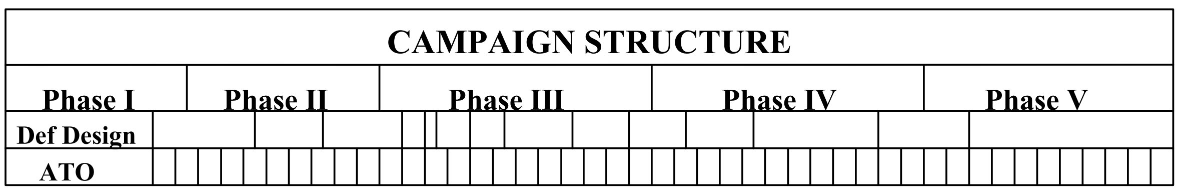


Figure 2: Joint Users Group (JUG).

Figure 3 illustrates another important aspect of defensive planning. A campaign may consist of many phases from pre-deployment until forces return to garrison. Also, the phases may begin as Operations Other Than War (OOTW) and could conceivably escalate to a combat situation, possibly asymmetric, perhaps with more defensive than offensive operations, and then revert to peace keeping. The AADP must serve the purpose through all phases of the campaign. Figure 3 shows a five phase campaign. Defensive designs must be prepared for each phase consistent with the situation during that phase. More than one defense design may be needed during each phase, dependent upon the changing threat, situation assessment, political and military objectives and other factors. Note that a defense design may overlap many 24 hour Air Tasking Order (ATO) days. There may also be more than one defense design in effect during a single ATO day. Aircraft that fly Combat Air Patrol (CAP), or are associated with other defensive missions, must be tasked as part of the ATO and must adhere to the Airspace Control Plan (ACP).



- **AADP Covers the Campaign**
- **The OPS tempo of the Defense is often different from the Offense**
- **The AADP correlates to the Master Air Attack Plan (MAAP)**
- **A "Defense Design" correlates to an "Air Battle Plan" Covers one or more ATO cycles**
  - **Changes as prioritization of the Defended Asset List changes**
- **Defended Asset List (DAL)**
  - **Developed by AADC staff**
  - **Combines inputs from the supporting & supported CCs**
  - **Approved by the JFC**
  - **Becomes the master list of 'things' to defend**
- **Prioritized Defended Asset List (PDAL)**
  - **May be a subset of the DAL**
  - **Varies by time, situation, and JFC objectives**
  - **Forms basis of Defense Design**

Figure 3: Defensive Planning within the Campaign Structure.

2.0 CONCEPTS OF MILITARY DATA AND INFORMATION FUSION

Military data on both friendly and adversary forces is required. Data input should be automated as much as possible. Not all data is available in easily accessible data bases. Therefore, this requires that simple user friendly Graphical User Interface (GUI) capabilities be incorporated that allow operators to set up or edit cases of interest. Representations of friendly resources, Command Control (C2), surveillance and shooters, are critical to conducting "what if" scenarios. Threat locations and potential avenues of approach, or courses of action, need to be quickly entered. Lists of defended assets need to be nominated and ranked in priority order with respect to criticality, vulnerability and recoverability with respect to a specific threat weapon system (see Figure 4). Finally, the map based graphical representations need to be easily understood by defensive planning commanders and staffs. The system must be adaptable to a variety of situations including Operations Other Than War, Peace Keeping, combat or in instances of Counter Terrorism.

The term "Information Fusion" in this paper is not used in the familiar intelligence context of bringing together reports from disparate sensors, or other information sources, and "fusing" into a single coherent interpretation. Rather, fusion means bringing together Friendly Order Of Battle and Enemy Order Of Battle data elements, capabilities of friendly and enemy air and missile systems, both surveillance and weapons, and Geographic Information Systems (GIS), in such a way that several candidate defense designs can be planned and evaluated against a variety of threat Courses Of Action (COA). The planning tool operator is supported in a software architecture by a collection of Wizards, Text and Data Editors, graphical display products including Power Point, Hyper Text Markup Language (HTML) and others. This collection of capabilities is designed and configured to support the operational user who always has the capability to interact with, and override, the computer operation, as part of the planning capability.

3.0 APPLICATION EXAMPLES AND DEMONSTRATORS OF DATA AND INFORMATION FUSION

One of the early defensive planning tasks is to prepare the Defended Asset List (DAL) which includes all the entities that need to be defended in the Joint Operations Area during the campaign. In the terminology of this paper, "assets" are the entities to be defended and "resources" are the entities that are used to defend the assets. The next step is to rank order or prioritize the assets, for a specific period of time, in terms of criticality, vulnerability, and recoverability. A simple example is given in Figure 4. Values in the range 0.0 to 1.0 are applied to each factor based on the capabilities of the threat weapon system. The values are then weighted. For this example the weights applied are in the ratio 3:2:1. The result of this activity is the prioritized DAL (PDAL).

Asset	Criticality Weight - 3	Vulnerability Weight - 2	Recoverability Weight - 1	Ranking	Priority
Airbase	0.9	0.9	0.6	0.85	1
C2 HQ	0.9	0.7	0.9	0.83	2
City	0.6	0.9	0.9	0.75	3
Port	0.6	0.9	0.8	0.73	4
Cultural Center	0.9	0.1	0.4	0.55	5
Logistics Site	0.5	0.5	0.1	0.43	6
Civilian Airport	0.4	0.3	0.1	0.32	7

Figure 4: PDAL Results from Analyzing Criticality, Vulnerability, and Recoverability.

Note that the "cultural center" is given a very high criticality rating of 0.9, but is only number 5 on the PDAL. It may be of great value to the local population for historical, religious, or political reasons, but may have little military value. Or it may just not be a viable target for enemy strike capability, at least in the current operational phase, and therefore has a very low vulnerability rating.

4.0 THREAT IMPACT/ASSESSMENT

The air and missile threat can be defined through Intelligence Preparation of the Battlespace (IPB) or the more advanced Predictive Battlespace Awareness (PBA). In actual system use, qualified air and missile defense planners will have the final say in refining the threat activity and possible Enemy Courses of Action. The planning tool software should be configured to retrieve (import) and display detailed enemy order of battle data from intelligence databases. The Import Wizard function allows the operator to view and import data without having to enter the data manually.

5.0 SITUATION AWARENESS, SITUATION ASSESSMENT

Situation awareness will be determined by a combination of data availability, information availability, and geographic mapping systems overlayed with both friendly and enemy status. Critical to defensive planning is enemy launch points, for both air and missiles threats, possible avenues of approach, trajectories, assets to be defended, and resources available to defend these assets. The planning tool software should be configured to retrieve (import) and display detailed friendly order of battle data from air operations databases, using the Import Wizard, as above.

Situation assessment depends upon the fusion of critical performance of all available sensors in the area, based on their location, including land, sea and airborne sensors. This information should be used to display overlapping surveillance coverage patterns as affected by terrain elevation obstructions. For defense laydown planning the position of mobile or relocatable sensors should be evaluated in a series of “what if” scenarios. Even more important is the fusion of critical performance of all available land, sea, and airborne shooters in the area, based on their location and capabilities to engage the incoming threat. The objective is a defense design that best utilizes limited defensive resources tasked to counter projected enemy courses of action against rank ordered strategic and operational friendly assets.

6.0 REPRESENTATION OF MILITARY KNOWLEDGE

Military knowledge in this context is represented in the final products. The AADP is the most important. It includes the DAL and PDALs. PDALs is plural since there may be many PDAL instances as the campaign progresses and the situation evolves. There may be many Defense Designs for the same reasons. The AADP is part of the overall Theater Operations Plan (OPLAN), or could be tailored to specific contingency operations. It is extensible to different phases of the campaign and could be modified for various time periods within each phase, as the situation changes. The AADP is a force level plan which is distributed to the tactical level operators for detailed planning and implementation. Actual implementation, as in precise location of fixed and mobile surveillance and shooter resources, is the responsibility of the respective Service tactical commanders. However these locations need to be fed back to the Force Level for plan updates. Likewise aircraft orbits and ship patterns need to be fed back to keep the AADP as current as possible.

Military knowledge and situation awareness is also represented by the graphics in Figures 5, 6, and 7.

Many of the topics described above are illustrated in Figure 5, which represents an entirely made up scenario. The capabilities assigned to the sensors and shooters are not based on actual capabilities but are meant to show some of the graphics capabilities in a scenario that is not very realistic. It shows aspects of what a defense design might look like. Enemy Air Avenues of Approach (AAoA) are shown in triple red lines. A potential Missile Threat Origin (MTO) is located and the large red circle indicates the range of missiles launched from this site. Note that several of the assets on the PDAL (Figure 4) are currently in threat of enemy activity while some are not. An airborne warning and control system aircraft is positioned in an orbit for surveillance of both enemy AAoAs. A combat Air Patrol (CAP) fighter is positioned to protect both the AWACS aircraft, C2 HQ and Airbase 3, through Fighter Engagement Zones (FEZ) 2 and 3. Likewise CAP 1 is protecting Air Base 1. The pie shaped "wedges" from the CAP aircraft indicate typical acquisition radar range and kinetic range of their weapons. The shields near the assets indicate they are being protected. Similarly a Navy ship is protecting the port. Friendly Surface to Air Missile (SAM) sites 2 and 3 are located to protect the City and the C2 site. The C2 site is partially obscured by the words SAM 3 in Figure 5.

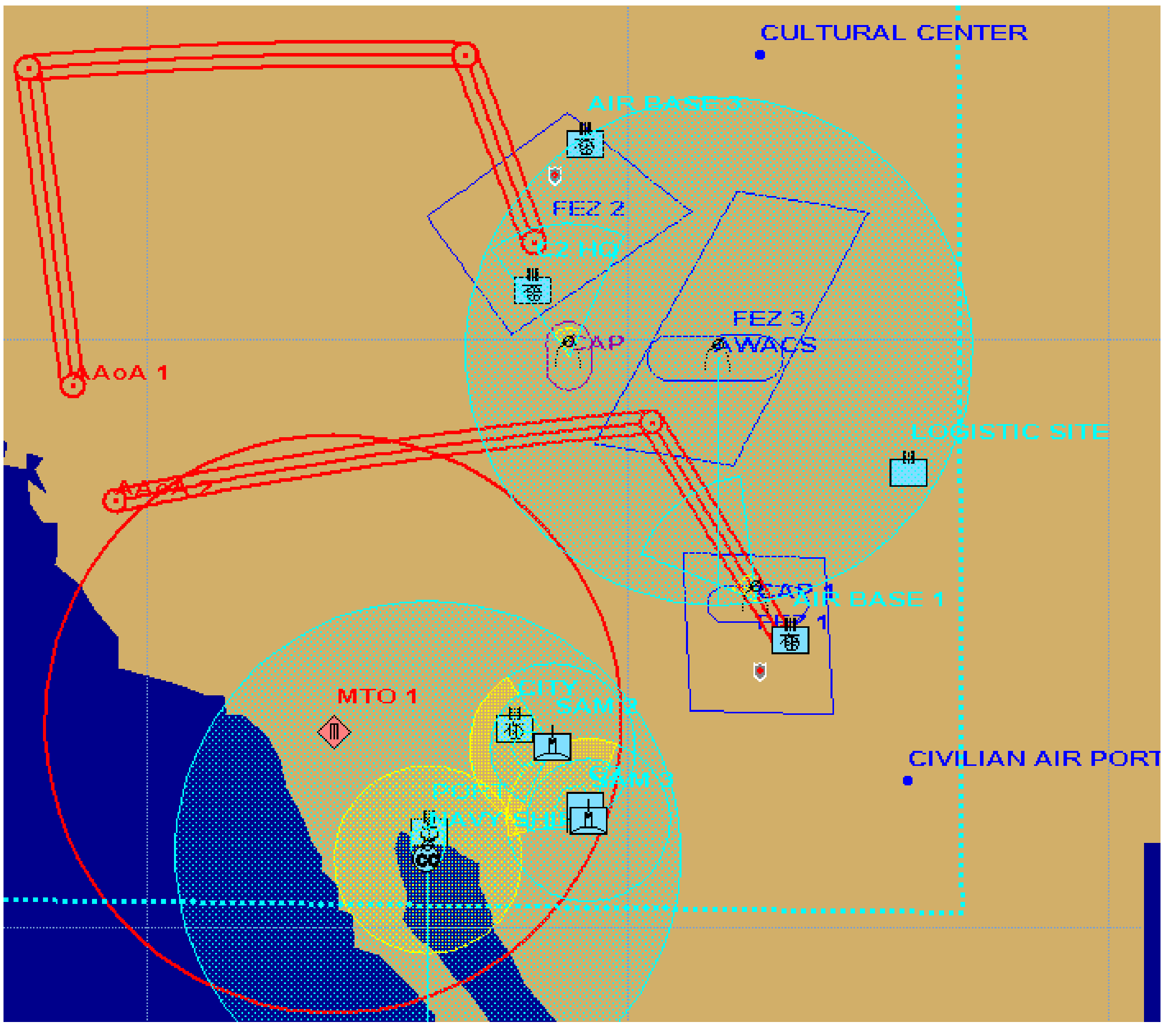


Figure 5: Map Based Planning.

Figures 6 and 7 show sensor coverage to detect enemy aircraft that might use the AAoAs indicated by the red lines. Both ingress and egress routes are shown. Here altitude, radar cross section, and terrain effects are parameters included in the calculation. These calculations can be included on the map display of Figure 5 to complete the Defense Design. This feature is not shown. Several Defense Designs may be developed and evaluated through the application of appropriate Measures of Effectiveness and Measures of Performance.

The Defense Design selected is included in the Area Air Defense Plan for each phase of the campaign. Non-selected Defense Designs can be archived and reused as alternatives in the future.

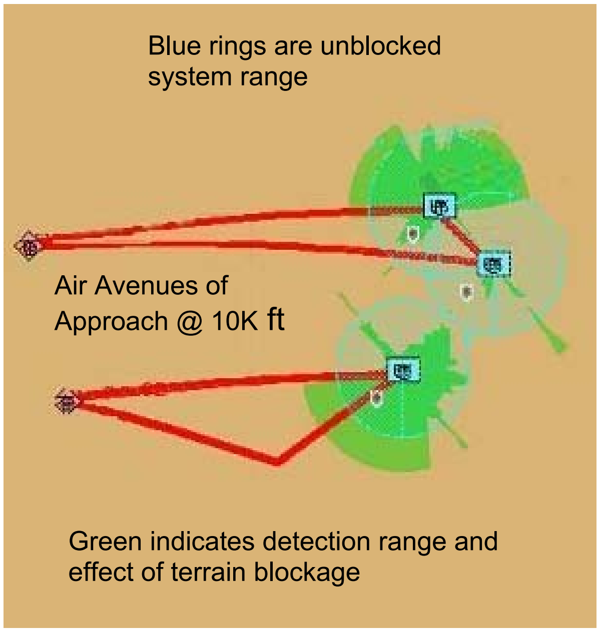


Figure 6: Terrain Considerations at Low Altitude.

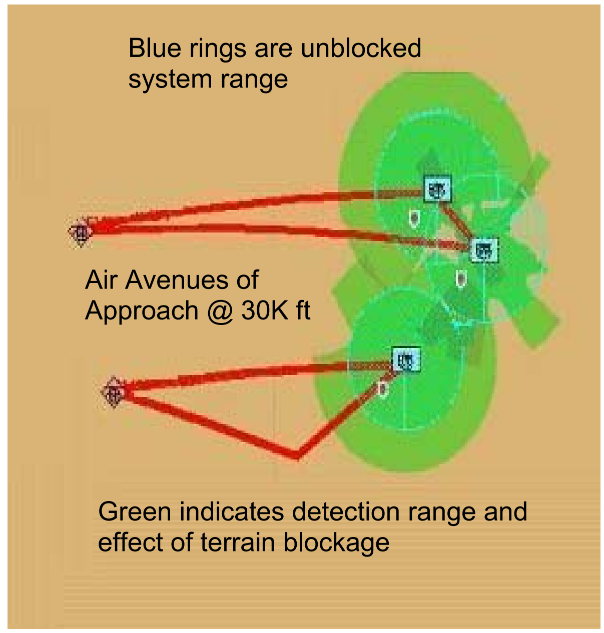


Figure 7: Terrain Considerations at High Altitude.

7.0 ARCHITECTURE OF FUSION SYSTEMS

Having considered all the above, the operational, system, and technical architecture of the planning tool should be robust, and fully implemented. The architecture is represented in Figure 10. The selection, integration and maturation of the technologies as the planning tool evolves provides the requisite functionality. Consideration is given to data, data bases, relational and/or object oriented data base management systems, information/knowledge bases and management, displays, programming languages, distributed collaboration, interactive collaboration, local or wide area collaboration, effects of narrow band communications connectivity between collaborating sites, security, geospatial information systems, and other technologies. One other architecture consideration is the use of Commercial Off-The-Shelf (COTS) products. This consideration must be tempered by the cost of licenses and the length of time the COTS product will be supported by the vendor. The COTS license for the JDP client server architecture allows up to five clients to create five different defense designs at the same time, or all five could be collaboratively working on the same design. The collaborative nature of the tool allows one operator to make a change from his/her client. The change is propagated through the server and stored in the data base, and immediately distributed to all clients on the network. Commanders at all levels can not only see the same consistent view, they can help create it.

A critical part of the architecture's flexibility is the use of smart static table (SST) representations of C2, surveillance and shooter capabilities. The use of SSTs versus detailed models and simulations provides faster analysis times and greater latitude in system representation. The SSTs are accessed and the results are available in a timely manner. Also using SSTs ensures that each operator using same data will always get the same consistent result which is not always the case with a Monte Carlo simulation approach.

Other features include compatibility with Microsoft Office products. Graphical and text information is readily imported into Power Point to create briefings. The AADP is produced as an HTML document for distribution or web portrayal. Text can easily be formatted or adapted for transmission during staff coordination or for a presentation. Text can be cut and pasted from the planning tool to several Microsoft Office products. Figure 8 is the content of the AADP and Figure 9 shows an example of a somewhat expanded AADP pasted in a word document.

Area Air Defense Plan

Copy No.
Issuing Headquarters
Place of Issue
Date/Time Group of Signature

JOINT AIR DEFENSE PLAN:
REFERENCES:
COMMAND RELATIONSHIPS:
1. Situation:
- a. Guidance.
- b. Adversary Forces.
- c. Friendly Forces.
- d. Non-Allied Forces.
- e. Rules of Engagement

2. Mission
3. Air Defense Operations
- a. Operational Concept
- b. Coordinating Instructions

4. Logistics
5. Command, Control, and Communications.

Figure 8: ADP Content Figure.

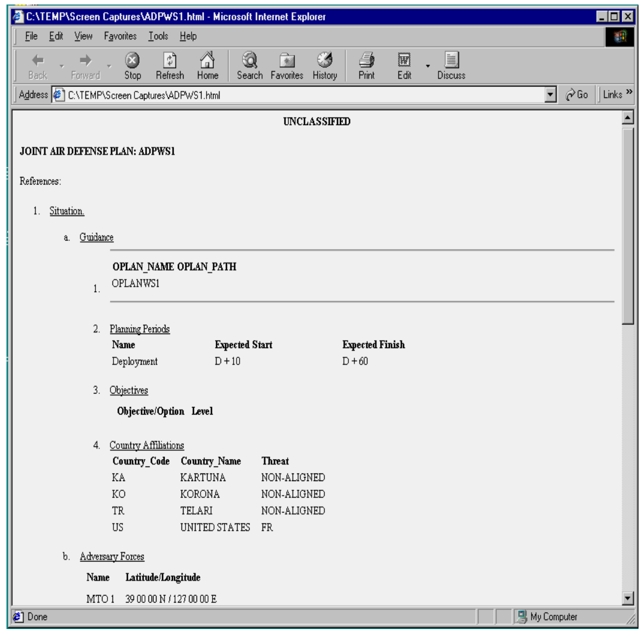


Figure 9: AADP Pasted in Word Document.

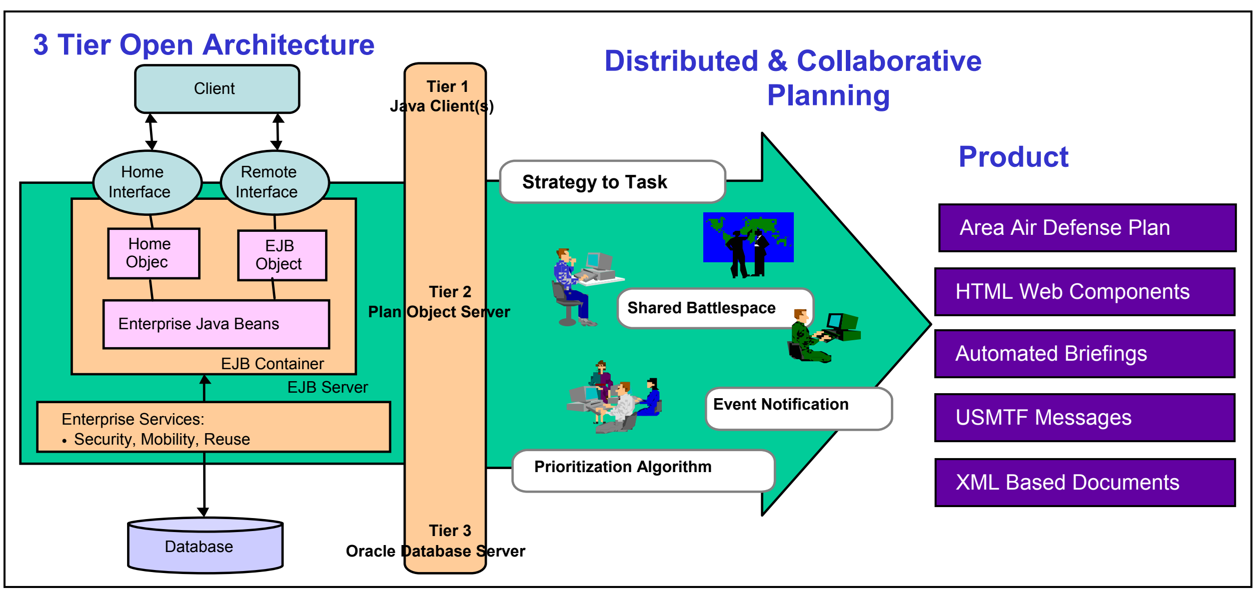


Figure 10: Technical Architecture of Fusion System.

8.0 SUMMARY

The stated intent of this paper entitled "Defensive Planning for Combined Forces" was to addresses the theme of this Symposium: "Commanders at all levels and types of military organizations require timely accurate awareness of the situation in their respective areas of responsibility as well as prediction of likely intentions of the participants". To accomplish this goal the paper addressed the topics of interest to this symposium with lessons learned from an exhaustive, comprehensive and successful development and fielding of JDP. Figure 11 combines in pictorial form the essentials of the defensive planning process beginning with Global/Theater objectives and ending with task specific outputs.

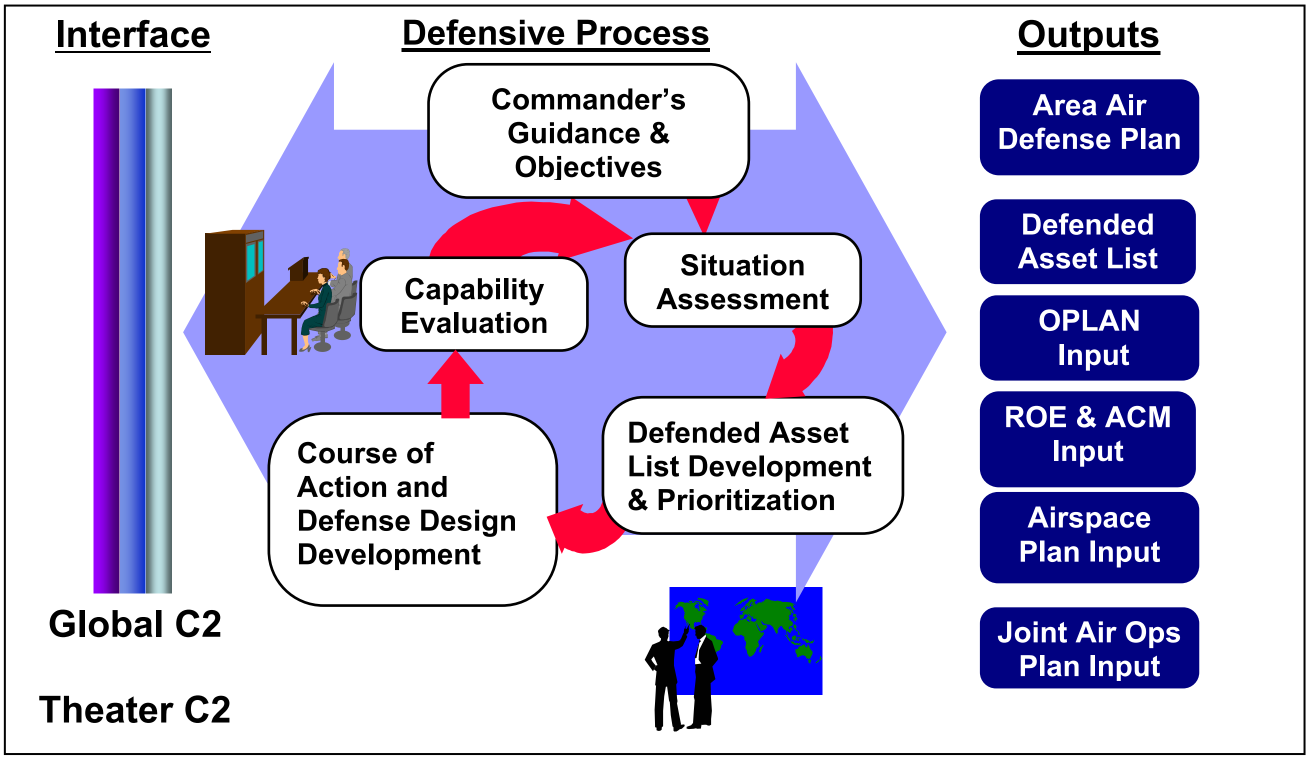


Figure 11: TAMD Operations Planning.

9.0 CONCLUSION

The Joint Defensive Planner planning tool was developed to provide the capability to satisfy stated US Joint Service defensive planning requirements using the process described above. This goal was met with JDP being fielded as an integral part of the Theater Battle Management Core System. A stand alone laptop version has also been used in exercises and real world operations.

The open architecture and the technologies incorporated represent capabilities that we believe could be somewhat easily extended to Combined Force and even coalition operations. Of particular importance is the capability of the smart static table representations to allow instances of C2, surveillance and shooter resources to be input by system and/or data base administrators. A worldwide terrain data base is included. The client server architecture is amenable to collaboration with allied or coalition partners. The AADP created in the planning tool can be cut and pasted into a variety of formats for other users.

A Wide-Area Surveillance Prototype System from Identification Fusion Perspective

LtCdr. A. Metin Balci, PhD
Turkish Navy (TN) Software Development Center
TN Research Center Command
Pendik-ISTANBUL
TURKEY

Email: tnsdc@superonline.com

Along with our efforts to prepare an operational prototype for a wide area surveillance system, identity fusion was considered as one of the major system components. Many problems have been detected about how the sensors produce the identity information, how they sent this information, registration of the information, detecting the conflicts, combining the appropriate information in an effective way, the presentation of the final fused information to the operators, and consequently interpreting the results for the coordination of sensors for further identification efforts. On the other hand, the appropriate way of integrating identity information to the rest of the fusion process constitutes the other important context. In this paper, we report our experience for a wide area surveillance system from identification fusion perspective based on the prototype system developed in Turkish Navy Research Center Command.

1.0 INTRODUCTION

Different type of threats, changing hostile activities, dense operational areas, littoral waters, and asymmetric threat expectations increase the significance of the classification and identification process in a maritime surveillance system. Although many different types of evolving sensors are utilized in modern surveillance systems, it is still a challenging task to establish and maintain the "recognized tactical picture".

As the threats and threat evaluations have been changing, associated identification process has become more complex in military surveillance systems. Littoral Warfare and operations other than war concepts has created different and more challenging requirements in both the content and timing requirements of the identification information of the tactical picture.

As the requirements are getting more challenging, the surveillance systems are getting better especially by adapting their sensors to a "network centric", "network oriented" or "network enabled" systems. But unfortunately, despite very nice definitions and explanations, we observe that there are still some very important technical challenges to achieve in network centric applications.

Different multi sensor data fusion techniques are utilized in establishing the recognized maritime tactical picture. In most general sense, the first level of multi-sensor data fusion, known as parametric fusion is considered in two main stages: Positional fusion and identity fusion. On the other hand, the interactions of the second and third level fusion, namely situation assessment and threat evaluation is not always a clear cut especially when the identification fusion is considered.

In this paper, we cover some of the problems that we faced in a prototype maritime surveillance system from identification fusion perspective. To cover the material in an organized way, we group the problems

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

as technical and operational problems; where as technical problems can further be grouped as architectural and fusion process aspects.

We introduce the problem in more detail with a brief explanation of our prototype system in section 2. In section 3 we cover the technical problems covering both architecture and fusion process aspects. Section 4 deals with operational problems. Section 5 concludes our report.

2.0 PROBLEM DEFINITION

A wide-area surveillance system is a good example of multi-sensor, multi-site fusion system with potentially many different type and number of systems, located in different geographical locations and covering many different types of "targets".

Such a surveillance system has been prototyped in Turkish Navy Research Center Command to try out various techniques in actual system development. Main components of the prototype system are:

- Distributed sensor simulation system
- Fusion node prototype
  - Sensor Integration Units
  - Fusion and Command Tactical System (FACTS)
- Communication media simulation system

In the distributed sensor simulation environment you can create off-line scenarios. These scenarios can be modified during the run-time as they provide the ground-truth information for the sensor simulators. When these scenarios are run, sensor simulators receives the ground-truth information, evaluate their coverage, and creates sensor tracks with a medium fidelity, which we consider as satisfactory for our prototype purposes.

After the appropriate track management, the sensors report their track information to the fusion center, via very realistic communication media simulation system, which is in our case a land-based communication system.

In the prototype fusion node, where all of the information is collected and processed, FACTS, has the main processing functionalities. It has mainly two different modules:

- Multi-sensor data fusion module (MSDF), and
- Tactical level fusion module (TM)

where MSDF module performs actual fusion process and TM provides services for operator input and presentations for higher levels evaluations and interactions.

At this point we need to emphasize the differences between the organic and non-organic sensors in such a fusion system. Sensors which are controlled, operated and monitored directly within the fusion system are considered as organic sensors, where as sensors or systems which provide information to the fusion system indirectly, mostly after an evaluation process are considered as non-organic sensors.

Within this context, MSDF has been designed and implemented in order to fuse data and information from different number and types of organic sources. Modern fusion algorithms and techniques can be studied within this module. On the other side, TM is designed to integrate non-organic data and operator evaluations into the system.

In a wide-area maritime surveillance system that can be simulated effectively with the above described prototype system, we faced many problems starting from very basic concepts regarding the simulation process up to the distribution of the recognized maritime tactical picture established. We choose evolutionary prototyping method. It was useful since it has provided a common environment to exchange ideas and evaluate the results before the system gets unmanageable. Another important system wide problem was to decide about the alternative algorithms to be implemented in the prototype system. Different risk evaluations, concerning performance, infrastructure, integration, and available resources has led sometimes the selection of not the best algorithm evaluated.

Two main important modules of parametric fusion, positional and identity fusion are designed and implemented in an integrated but modular manner. Although there are some recent efforts in the literature to integrate the identity information in the early stages of the positional fusion, even in the association phase, to get better association results, we have utilized the methods which integrate the identity information in later stages of the fusion process, especially after the positional fusion process.

In order to explain the problems from identification process, which we concentrate in this paper, we divide these problems in two different groups as technical and operational. Technical problems are further detailed as architectural and identification fusion process related.

3.0 TECHNICAL PROBLEMS

There are some technical challenges to be taken into account in any large-scale fusion system. These problems, which we named as architectural problems are listed below for the sake of completeness and detailed in [2]. These architectural problems are:

- Different types of sensors,
- Different types of systems, subsystems,
- Different types of communication media and communication requirements,
- Different types of information processing techniques,
- Effective presentation to the operators and to the command team, (Human Machine Interface),
- Effective sensor and resource management.

On the other hand, there are some problems that we faced related to identification process specifically. These problems are defined in [3] to some extend. We enhance these explanations based on our experience in the rest of this section.

First of all, due to its significance we emphasize that sensor related issues constitutes a great problem by itself, due to:

- Different types of sensors reporting different types of identification information,
- Unknown characteristics of the sensors,
- Uncertainties associated with their reports,
- Difficulties to enforce the sensor manufacturers to think appropriately for fusion purposes.

Most of the sensors are generally designed in more of a stand-alone system. As the integrated systems, systems of systems and finally network-centric warfare concepts evolves, it becomes inevitable to require some changes in the specification of the sensors. Acknowledging the increasing significant role of the “system engineers” role in large-scale projects, system engineers have to consider many different topics, which will definitely include the very detailed capabilities of each sensor providing some kind of data to the fusion process. In many cases, knowing only the capabilities of the sensors will not be enough, but the algorithms and even some implementation details of the sensors will be necessary for the success of the overall system performance.

As the basic step of the parametric fusion process, association (or in some literature correlation) serves both for the positional and identification fusion purposes. A successful association mechanism is the key element for a successful identification process. Assuming that association has been done successfully, we can enhance identification fusion problems as follows:

- Processing combat id (hostile, friend, suspect, etc) and identification information (type A, Class B, platform X) separately is a major design decision in the system. It is partially due to the fact that the significance of these two different levels of classification has different operational requirements. In any case, the first goal is to provide a combat id to the operator. Combat id is more fundamental and it has some well-known regulations and rules defined in operational documents. On the other hand, identification information provides more information about a specific target. Although there is a close relationship between combat id and identification information, these two can be processed independently up to a certain decision level. Identification information after having some conflict checks with the combat id, will be used for situation assessment and threat evaluation purposes.
- Concentrating on the identification information and its fusion process, we need to consider different identification information produced by the sensors. For most of the sensors, it is not very well defined how they produce the identification information associated with a specific track. Hence very careful study of the sensor data is required. Most of the time, understanding of the internal sensor processing is vital for a successful fusion process.
- When these data or information is sent to Fusion center or fusion node, appropriate registration mechanisms are to be done effectively. This covers converting possibly different types of information, such as possibility, probabilistic, linguistic and likelihood values to a common information domain.
- Detecting the possible conflicting information reported by the same sensor and resolving these conflicts, either automatically or via interactions with the operator constitutes the next step in the fusion process. Conflicting reports from a specific sensor may affect the result of the identification fusion process significantly, hence we think that before making such reports effective on the fusion process, potential conflicts have to be checked and resolved. Actually it is a way of having a safety-point in the identification fusion process, since we are not sure about the quality of the identification data/information sent by the sensors automatically. In order to be able to keep the system reliable for the end user, we think that it is worthwhile to put some more effort (computer and operator if needed) to check for the inconsistent reports.
- Similar conflict checks are also to be performed within the reports of different sensors for a specific target. Detecting the possible conflicting information reported by different sensors and resolving these conflicts is necessary for maintaining reliable results.
- After these conflict checks identification information combining process takes place. Bayesian and Dempster-Shafer models form two main approaches for combining the identification information. We have also got some experience with the Transferable Belief Model (TBM) [4], which is actually an interpretation of Dempster-Shafer model. Although we have not a clear-cut idea on which algorithm to utilize for the final system, TBM gives promising results for scenarios with specific problems.
- The presentation of the combined information to the operator in an understandable format is still an important issue. The operator needs to have access to the information sent by the sensors, the identification information calculated by the system (if necessary after a pruning process) and

some resultant identification information that is either generated by the system or operator. Operator needs to be able to modify the resultant identification information even if it is generated by the system automatically.

- On the other hand the content of the identification information presented to the operator is also important. Especially in DS and TBM methods, basic probability assignments (mass values), Belief and Plausibility values calculated after the combining process, are difficult to be visualized by the operators, especially by the end users. Hence either a very detailed training shall be given to the operators, or automatic conversions have to be implemented for more custom values such as probability values. A conversion method is proposed in [4] by Delmotte and Smets.
- An additional consistency check is very useful between the resultant identification information and the other data gathered from different sources such as intelligent reports, operational messages and even information via a Geographical Information system.
- In almost every step there is a conflicting check requirement. This is mostly due to the requirement to maintain the high level of reliability through out the system. In most cases, having no result is considered as better than producing a wrong result. Hence we prefer to process the data automatically as much as we can, but if there is a doubt the task is to alert the operator about the problematic case.
- The effect of the identification information fusion on the positional fusion results, especially on the association process has also to be taken care of by the system. Especially a conflict report is approved by the operator for a track reported by more than one sensor then operator shall be provided with the tools/services to maintain the association relation appropriately.
- As it can be followed all the issues that we mentioned are strongly related to the organic sensor information process. There are also some additional problems in the Tactical Level Fusion process from the technical challenges of the identification fusion perspective:
  - The management of identification information retrieved from more than one link shall be handled appropriately.
  - For the conflict resolution between different links and organic sensors, we utilize operator support. But automation of identification information assignment can be performed to a certain extend without any operator interaction.

4.0 OPERATIONAL PROBLEMS

While there are some technical challenges in the identification fusion process, operational requirements constitute the other important aspect of the identification fusion process. Our explanation will start from the sensors, continue with the communication media and conclude with the fusion node itself.

The independent sensors and their operators must have a clear idea about the general overall fusion concept. They may even be warned against the results of their ignorance in their scope. On the other hand, this may also affect the track exchange policies of the fleet, which may cause some extra difficulties to maintain the existing link communication. Nevertheless, the important thing is that a track that is not important for one sensor may be vital for another unit. Hence we need to exchange the track information to the extend possible without any filtering if possible. Since this will most likely cause a bottleneck in the communication media, alternative communication possibilities or operational procedures shall be utilized.

The significance of the communication capabilities can be seen very easily in such an integrated surveillance system. Most of the time, alternative communication mechanisms are to be established both to collect data and to distribute the resultant tactical picture effectively.

Secure voice capabilities between appropriate commanders and secure coordination networks for specific purposes, such as EW or AAW is of special interest. Again due to the limitation of the available communication channels, some of the coordination may be supported via text messaging between the operators or even between the commanders.

As we mentioned briefly, assigning an appropriate combat id for a track is more important than producing an identification information for that specific track. But in any case, they need to be complementary. Any conflicting case is a potential issue to investigate for the tactical picture compilers. This process can easily be automated to save time and man power.

Commander/ Decision makers shall not be expected to know all the details of the surveillance and identification fusion process, but instead there shall be some other people who will be in charge of coordination of the surveillance process and the evaluation of the intermediate results. In most cases, it is a good idea to have one or more identification officer to support the commander for these tasks.

Identification officer should have detailed information about the capabilities, limitations, and even the underlying algorithms of the integrated sensors and communication systems of the overall surveillance system.

Operational organization of the fusion node or in more general sense fusion center is very important to achieve the targeted goal. Although the organization is to be flexible with respect to the given scenario, the workflow of the fusion center is very important for the successful result. Within this context, the presentation of the tactical picture internally, the filtering mechanisms, switching between different operator console screens, the video-wall applications, internal communication capabilities are all to be considered as parts of workflow of the fusion center.

It is our experience that to enable some operators to access the detailed information gives a good confidence over the system. Hence the related operators, including the identification operator, shall have access to the sensor reports, their details and possibly the history of the sensor reports; the calculated identification information produced by the sensor, the pruned version of the calculated identification information if necessary and the resultant identification information.

In order to establish and maintain the recognized tactical picture appropriately, the coordination of the sensors and available resources shall also be considered as an important operational aspect. Especially the allocation of extra resources to identify a specific track requires a comparison of potential threat and necessary resource allocation. We think that there will be some automated tools to support the decision makers for this specific resource allocation problem.

The presentation of the compiled picture to the operator and command team is still another problematic issue. There are no generally accepted standards to present the tactical picture so that operators and command team can evaluate the current situation effectively. On the other side, in terms of human machine interface (HMI) presentation is not the only issue. Our concern is that the reliability and confidence of the overall system as observed and evaluated by its operators and users is much more important and vital. We have observed many systems not being used effectively just because of the missing feelings of the confidence. Hence, great care shall be given, not to disturb the confidence of the operator to the system, as a continuous process starting from analysis of the system up to and including maintenance phase of the system.

5.0 RESULTS

Our experience in the prototype system showed that sensors, their characteristics, and (sometimes) unreleased parts have significant effect in selecting, designing and implementing the appropriate fusion system. As there is no well-defined procedure for identification fusion process, we benefit our modular architecture so that we can plug and test any kind of identification fusion algorithms in our system or we can enhance our existing implementation. One of our conclusions is that belief theory deserves more attention in the identification fusion implementations with its superior characteristics fitting our purposes.

REFERENCES

[1] David L. Hall, J. Llinas "Handbook of Multisensor Data Fusion" , CRC Publishing, 2001.

[2] Balci, M. "Problems in Integration of ES Sensors for EW Situation Awareness", 6th Armed Forces Communication and Electronics Association (AFCEA) /Turkish Chapter International Seminar on Effective ….. , Ankara, Turkey, May 2003.

[3] Bosse, E., Roy, J., Grenier D., "Data Fusion Concepts Applied to a Suite of Dissimilar Sensors", Proceedings of the IEEE Canadian Conference on Electical and Computer Engineering, CCECE'96, Calgary, May 1996.

[4] Delmotte F., Smets P., "Target Identification Based on the Transferable Belief Model Interpretation of Dempster-Shafer Model", http://iridia.ulb.ac.be/~psmets.

[5] Muhammet Altun, Metin Balci "Target Tracking with Integrated ESM Sensors", NATO-RTA SCI-116 Symposium on Multi-Platform Integration of Sensors and Weapon Systems for Maritime Applications, Norfolk, Virginia, United States, October 2002.

The Integrated Data Environment: A New Tool for Interoperability and Effective Data Integration for Command and Control

Jon Wilkes
NATO C3 Agency
Postbox 174
2501 CD The Hague
THE NETHERLANDS

jon.wilkes@nc3a.nato.int

ABSTRACT

This paper describes an ongoing effort at NC3A to provide one integrated database which contains data from a number of different sources. Initially, these sources are legacy NATO systems. Later, other systems, including messaging interfaces of a wide variety, and national systems, will be added. A common data model is used as the lingua franca between systems. A COTS product has been identified that creates translator boxes to provide interfaces to and from the legacy systems.

1.0 INTRODUCTION

The NATO C3 Agency has responded to customer requirements with the Integrated Data Environment project, which has been evolving over the past three years. The intention of the effort is to provide one integrated database which contains data from a number of different sources; in the first place these will be legacy internal NATO systems. Later, other systems, including messaging interfaces of a wide variety, and national systems, will be added as requirements and political concurrence allow. It is foreseen that IDE will play a significant role in the Core Capability package for the Bi-SC AIS. This paper is based heavily on the work started by the late Martin Krick.

2.0 THE PROBLEM

Many of the data exchange problems that have confronted and bedevilled NATO for the past few decades have arisen from the fact that early systems were conceived, developed and implemented as stand-alone, or "stovepipe", systems by groups of users and technicians whose requirements horizon extended no further than the immediate needs of the system on which they were engaged. In the early days, interoperability of data models was not even considered relevant.

As time progressed, and the initial desirability of being able to pass information from one system to another became a more firm requirement, many mechanisms were devised to address these issues, but always with the caveat that the software within the in-service systems, seen to be of such acquisition cost as to be untouchable for interoperability needs, could not be modified to assist in the process of bringing systems together to provide for any meaningful direct exchange of data. In addition, because early systems were so expensive, and therefore made available only to the smallest possible community of users, and because many of the more senior users had no ADP facilities at all, or at most a simple teletype, these early mechanisms were specified to be able to be used in manual environments, leading to the definition of a range of messages. Once again, these message definitions were aimed at encapsulating the specific needs of the group of users responsible for the definition of each message; correlation between messages was not a driving force in the definitions.

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

3.0 PREVIOUS STUDIES

Many studies were carried out when the nature of the problem became so large that it could no longer be ignored; these studies stressed the need for common standards for data definition, but could not provide low-cost solutions and their conclusions were therefore ignored. In essence, they proposed a "data fusion" approach, which is nowadays seen to be both impractical and unnecessary.

4.0 THE DATA FUSION APPROACH

The principle behind a data fusion approach is to define a single data model, and implement a single database, which will encompass the entire set of data currently held in all existing systems. This approach has some advantages, but also has many more major drawbacks which make it an impractical proposition. If we take two or three existing systems, and create a new database which holds all the data previously held in the three individual databases in accordance with a new all-encompassing data model, then the new database will not be the same as any of the old ones. Each application suite in the original systems must therefore be re-written.

It might be possible to create a database interface package for each system to make the new database appear as the old database, but that too would be substantial effort (and there would be no *ab initio* guarantee of feasibility) and would represent an additional load for the original system which it might well be ill-suited to handle.

A further major, and potentially even more serious, disadvantage is that if another legacy system were to be added to the fusion set, it may impose changes on the data model which would have a knock-on effect on all current systems within the fusion set. This would lead to potentially exponentially soaring costs, and to management problems of equally soaring complexity. Little wonder that the NATO committees of the time were not persuaded to follow down this route!

The perceived advantages and disadvantages of the data fusion approach can be summarized as:

- single view of all data
- single physical database from which all applications can draw data

whereas the disadvantages are:

- need to agree the (large) data model between 19 nations and all NATO HQs and Agencies
- immediate impact on all legacy systems which are required to conform to the new global data model
- applicable only for a small number of systems (three or maybe four)
- ongoing management overhead for the fusion schema
- complexity increases dramatically with the number of systems
- process becomes unmanageable with large numbers of systems

It should be noted that the advantages are not matched by any known requirement for all data to be perceived in a single view, nor that there should be a capability of providing a single database implementation which would hold all data; these advantages represent theoretical technical possibilities only. By contrast, the listed disadvantages are very real, not least the political problems associated with the first of those listed. Corresponding agreements in related areas are not famous for the speed with which such agreements have been reached nor for the technical clarity of the final agreements.

5.0 OTHER MORE RECENT STUDIES

In the last ten years, other initiatives have been taking place on a lower profile basis, and the fruits of their endeavours are now beginning to become visible in a number of places; national implementations based on these initiatives have been put in place and have become sufficiently mature for reasonable projections to be made. Principal of these initiatives is the multi-nation ATCCIS[1] study, sponsored and led by NATO, with active participation at varying levels by eleven nations. The major outputs of the ATCCIS study to date have been:

- a wealth of well-documented analysis
- a fully specified data model for information exchange
- an ATCCIS Replication Mechanism (ARM) for selective transfers of data between two or more ATCCIS-conformant databases

The primary achievement of the data modellers is that they recognised that they were endeavouring to specify a data model to facilitate the exchange of information rather than for the design or development of systems; thus the level of detail of the model is appropriate to information exchange, and much low-level data, which would typically be found only in specialist systems, was not included. This separation of "local" data and "global" data has been one of the foundation links of the NC3A work on the Integrated Data Environment.

6.0 THE INTEGRATION APPROACH

The separation of local and global data leads immediately to the concept of an IDE which addresses only some of the totality of data held in all existing (and future) systems. It also leads directly to the recognition that the IDE can be established (either as a virtual database or as a real one) for new purposes, and that the existing systems can be left with their current databases and database management systems – be they rudimentary or advanced – with the immediate benefit that no changes to those systems are required. Indeed, it became one of the design objectives of the IDE work that the IDE concept should be seen to be non-intrusive from the perspective of any legacy system.

In the integration approach, data are translated from the native (legacy) environment to the common data model of the IDE, so that the translated data subsets reside in a single database or transmission mechanism with one common data model describing all data. We may think of this common data model as a "lingua franca". The integration approach offers as advantages:

- single view of all global data
- no impact on legacy systems
- no requirement to have a single database
- all future applications can draw global data from existing databases
- process remains manageable with large numbers of systems
- ongoing management overhead for the integrated database is much smaller than for the fusion approach
- technology is mature and in use in large commercial organizations

[1] The common ATCCIS Generic Hub 4 data model was forwarded to NATO in 1999. NATO initiated a standardisation process for this data model, now called the Land C2 Information Exchange Data Model (LC2IEDM). The respective STANAG 5532 (ADatP-32) has been submitted as draft and is expected to be agreed in 2001.

and as disadvantage:

- as of end 2001, the technology has not been proven within a NATO operational system (but a demonstrator has been produced, and is clearly scalable to full operational use).

It may be seen that almost all of the disadvantages of the fusion approach have been stood on their heads for the IDE approach. The single view of all data, which was never supported as an operational requirement, has been scaled down to become a single view of all global data, for which operational requirements most certainly exist. The previous high impact, in terms of both cost and operational implications, of the fusion approach, has become a zero impact on those systems. And, the management problems remain tractable.

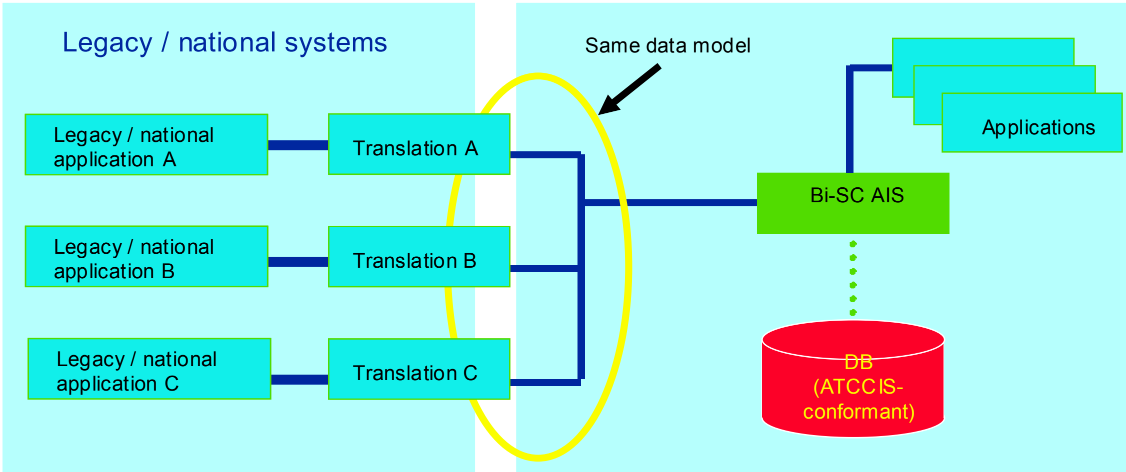
On the disadvantage side, the technology has not yet been tested in a full NATO operational environment, but a four-system demonstrator has been produced, and the technology is scalable to encompass a very large number of systems. In particular, the technology ensures that the management problems remain at the one-system level, and therefore do not grow as the number of systems being integrated expands.

7.0 ALTERNATIVE TECHNIQUES

There are two techniques available to implement the IDE function, Data Mediation and Data Translation. Data Mediation works by first making associations of the meta-data of the data sources and the data sink, and then automatically converting source data to the sink on the basis of these pre-determined associations. In principle, this is a very powerful technique; however, at the present time the technology is still in the research stage, with academic institutions producing small-scale demonstrations. No proposals for a full-scale demonstration have come to our notice at this time. The technology is thus considered to be far too immature to be considered for introduction to NATO at the present.

By contrast, Data Translation is a very much more mature technology which has been in use in commerce for some time. Most of those applications have been for data warehousing applications, but some applications have been for genuine data integration applications. Where the translation process is carried out on a one-translator-per-system basis, there are very few problems about scaling to multiple systems. The scaling problems are mainly associated with the suitability of the sink data model for the spread of data types to be found in the source systems; in this respect, the highly generic nature of the ATCCIS data model is of immense benefit in minimising such risks. Finally, it must be emphasised that both techniques act on the conversion of data on a one-for-one basis. Data aggregation, data fusion and other application-level functions are outside the scope of both technologies.

8.0 THE IDE ARCHITECTURE



This diagram gives a very simplified overview of the IDE architecture resulting in the use of translation techniques on a Translator-per-System basis. Data from each legacy or national system is processed by its own local translation process to the target (sink) data model and added to the data model of the target system by normal database update techniques. The translation mechanism is a process, implemented as a software package; although for simplicity it is shown in this slide as though it were a separate system, it could equally well be hosted on the legacy system if that were to prove to be the preferred option. However, to emphasise the “No Impact” concept, we always show it as a separate system.

Because the translator process will only translate data about which it has been provided with appropriate translation data (which is another form of meta-data), it acts as a simple form of guard against the accidental translation of data which is not to be released. However, the translator process makes no claims to be an approved guard, and additional security devices would normally be expected to be fitted by national authorities to protect national systems which may contain nationally-sensitive data. These would typically be positioned between the national system and the translator.

Both the initial configuration of the translator, and any subsequent upgrades or changes to a national system will require detailed analysis of the source system in order to specify the translation meta-data. For this reason, the configuration of the translators is expected always to be done by the nation concerned. The diagram thus shows the translators residing in the national management domain, with the exception of the specification of the output format (ATCCIS conformant) which is essentially public domain.

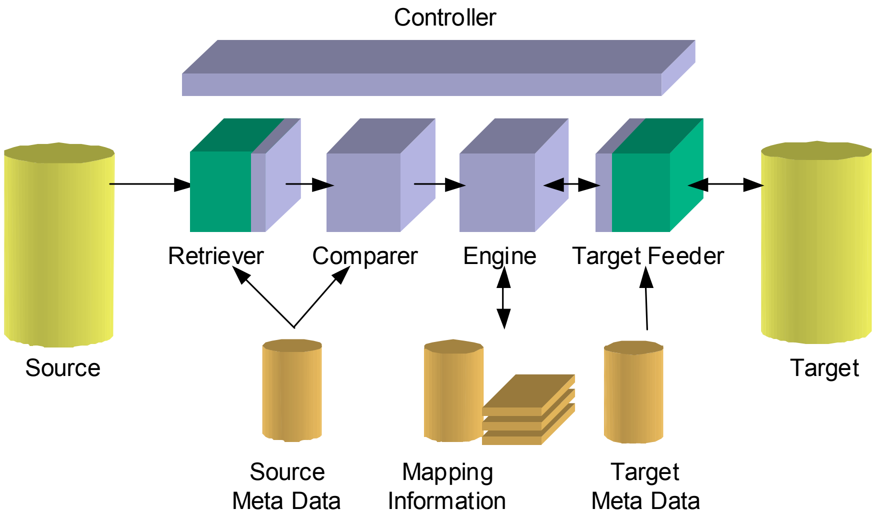
9.0 WORK DONE BY NC3A

The preliminary study on data mediation carried out in 1998 showed that the technique held potential for complex translation situations, and for the tracking of changes to databases. A simple demonstration system was created, using the most rudimentary meta-data, which was shown at JWID-99. Much interest was demonstrated by visitors at the ability to show data from three different systems out of a common database in response to a single query, with the consequential ability to provide for integrated data solutions.

Evaluation of a contractor report made clear that, although the concepts behind the Data Mediation technology were both powerful and useful, the technology was very immature with no commercially available implementations of a data mediator product, and little prospect of any such products appearing in the market for some considerable time. Data mediation may have benefits for special situations in the future, yet to be assessed and proven.

At the same time, an investigation was made of other products, all of which proved to be Data Translator systems, and it was determined that this offered a better approach for the near term. A contract was let for the development of a demonstrator using translation technology for display at JWID 2000. Problems with the suitability of the translation proposed by the contractor meant that only a very limited demonstration could be mounted at that time, but a very good tool has since been developed by the contractor as a COTS product, which has proven to be very successful and very flexible. A demonstration held at NC3A in late November 2000 showed the capabilities of this tool, and the design gives confidence for its use in many other situations, including message-oriented environments. A major demonstration was held at JWID 2001 and JWID 2002 at SHAPE.

The diagram below shows the architecture of the translator box produced by the tool.



10.0 THE SELECTED DATA MODEL FOR IDE

The selection of the ATCCIS data model, in the form known as the SHAPE Land Command and Control Information Exchange Data model (LC2IE DM) proved to be a sound choice. The complex nature of this data model means that the specifications of the translations are themselves more complex, but no instances were found in the work on the four NATO legacy systems where translations could not be specified with alacrity and accuracy.

The NATO Data Administration Group Reference Model is also based on the ATCCIS model, and is under strict Configuration Management; the LC2IE DM should similarly be placed under CM while it is being used as an interim measure before the full availability of the NATO Reference Data Model. At the same time, some of the work of the NDAG could usefully be retro-fitted to the LC2IE DM to make it into a Joint product, a JC2IE DM; the experience of NC3A and their contractor suggests that the minimal changes for the interim product would be small and easy to define and implement. For the November 2000 demonstration mentioned above it was necessary to add only four low-level entities (Naval unit, Air unit, Naval facility and Air facility) and to extend the range of a set of domain values to cover maritime and air factors. The total work took less than a couple of days; to repeat this work under full CM control would take less than one week. The future ATCCIS Generic Hub 5 may address the problem.

11.0 THE TOOLS USED FOR IDE DEVELOPMENT

Mention has already been made of the shortcomings of the original analysis tools proposed by the contractor. These tools were designed for data warehousing applications where the primary focus of the tools was to analyse data – often dirty data – for which a data model did not exist. In the IDE situation, data models existed and were well documented (although there were some instances where the semantics of the data were not fully defined). Additionally, in data warehousing applications, the emphasis on fitting all source data into a single data model in the destination system does not apply. It is thus not surprising, with the benefit of hindsight, that the tools were found to be unsatisfactory for the IDE situation.

The analytical process involved in determining the translations required is both a skilled process and one which requires time. An analyst familiar with both the source system and the destination data model can complete several source tables each day if the source data model is "clean" and the semantics are fully defined and supported by exemplar data samples. Loose source data models, or a lack of semantic definition, or a lack of sample data, will slow the process to a considerable extent. The tool developed by the contractor provides considerable assistance in converting the results of the analysis into translation rules; future versions are expected to provide some additional assistance to the analysis itself, but cannot fully replace the need for analysis or the analyst.

12.0 SUMMARY

NATO and the nations still have a plethora of incompatible data systems which are likely to remain in service for many years. A fusion approach is not appropriate, and is likely to prove unmanageable and unaffordable.

The Integrated Data Environment provides a response to this information management challenge that is both manageable and affordable, and is eminently suitable for an incremental growth approach.

Commercial off-the-shelf tools are available which support IDE and thus support Coalition interoperability, NATO to NATO interoperability, NATO to nations interoperability, and Coalition HQ to nations interoperability.

Jon Wilkes is a Senior Analyst Programmer in the Communications and Information Systems Division at the NATO C3 Agency in The Hague. He has been specialising in the interoperability of C2 systems, and the associated problems of the definition of data, for several years at the NC3A.

More recently he has been assisting investigations into ways of implementing mechanisms for providing for interoperability between non-compatible systems which have led to the development of the IDE concept and the creation of tools to support the concept.

On the Scientific Foundations of Level 2 Fusion

Dr. James Llinas
Research Professor, Director
Center for Multisource Information Fusion
State University of New York at Buffalo
315, Bell Hall
Buffalo, NY 14260
USA

llinas@eng.buffalo.edu

This Keynote speech will actually address two focal topics: (1) certain suggestions for extensions and modifications of the well-known "JDL"Data Fusion process model, and (2) perspectives on an overarching approach toward developing first Theories and then Algorithms for Level 2 Data Fusion.

1.0 EXTENSIONS AND MODIFICATIONS TO THE JDL DATA FUSION MODEL

The JDL Data Fusion process model has certainly served the Data Fusion (DF) community well as a means for describing and understanding the nature of the DF information process. It has not however been recently reviewed as regards application to the framework of a distributed information and fusion environment. Distributed Data Fusion (DDF) will no doubt form one important underpinning for the realization of Network-Centric Warfare, and it is in this motivational context that the preliminary ideas herein are offered. Certain other ideas involving some suggested modifications of the model are also offered. The (incomplete) suggestions are shown in Slide 4 in the Keynote presentation. There are two categories of suggestions as just remarked, one for the DDF condition and one regarding new thoughts for an improved specification and characterization of the DF process.

1.1 Extensions Regarding DDF

Four major extensions are suggested: one related to the "Pedigree" function, one related to an "Adjudication" function, one related to the need for both local fusion and network-based fusion algorithms, and one related to the sharing of information across the nodes of the network. The Pedigree function is defined as comprising those sub-functions and items of information necessary to maintain formal mathematical (or, generally, algorithmic) integrity for fusion of the received information at the receiving node. The Adjudication function comprises those sub-functions necessary to address internodal conflict in fused state estimates (eg as required to arbitrate and evolve the "Common Operating Picture"). These functions are shown notionally in Slide 4 as they would be involved in internodal processing in a DDF architecture. Note that we see the Pedigree function as necessary for both the internodal exchange of either measurements or fused state estimates, whereas the Adjudication function is presumed only applicable to the exchange of fused estimates. We suggest making obvious the fact that in a netted environment there will typically be a need for two classes of fusion processing, one for local fusion (eg of node-organic sensor data), and one to fuse the local information or estimates with that received from the network. Finally, we characterize the overall processes involved with internodal sharing of data or information as "Information Sharing Strategies or ISS"; these

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

functions contain whatever logic necessary to govern the protocols by which nodes decide to send particular information to other nodes or to broadcast it, etc. That is, the ISS is the logic which determines “who gets what information, how rapidly, in what form, etc”.

1.2 Modifications of Basic DF Processing

We also see a deficiency in the current JDL model characterization regarding the role of quality control, and inter-level processing. A frequent criticism of the model is that it appears to hint at sequential processing; this of course was never intended but admittedly does come across in the diagrammatic construct of the model. We suggest to repair this by making obvious the need for internodal bilevel processing, as shown in Slide 4. We also envision a two-step qulaity control process before a node issues an estimate: one that checks whether a just-formed estimate is of adequate quality (if not a Level 4 directive for additional information might be issued, for example), and what we label as a “Belief Revision”process that checks for both internodal consistency and consistency with recent global estimates.

Lastly, we suggest that the basic DF model be modified to make obvious the co-processing of both inductively-based and deductively-based inferencing processes. This suggestion is driven by the new need for addressing problems in which the degree of a priori knowledge about adversarial behavior is very limited such as terrorism problems and the variety of asymmetric warfare problems. We are not the first to suggest this; as noted in Slide 6, Waltz has previously suggested the need for co-processing of Data Mining and Data Fusion operations.

2.0 PERSPECTIVES ON AN OVERARCHING APPROACH TOWARD DEVELOPING FIRST THEORIES AND THEN ALGORITHMS FOR LEVEL 2 DATA FUSION

The views that are depicted in Slides 9 onward offer some ideas for moving forward in framing a formalized approach to dealing with Level 2 fusion problems. The suggestions revolve initially about the assertion that part of the fusion community’s problem in making headway in both scientific understanding of Level 2 concepts, in development of Theories related to Level 2, in algorithmic specification, and then application-oriented accomplishments is that we have not been adequately specific in the characterization of the elements of “situational constructs”. Slide 11 argues this point, in noting that the terms shown there are (more or less) as specific as we have been in describing what Level 2 is about. The need for specificity is thought to follow directly from the fact that, if we are to estimate Level 2 states of interest, we must define what those states are. In the end, we see this as a need for an Ontological specification of Level 2 state elements. An ontological specification is in itself an Existential Theory, as it is descriptive of how we see the real (truth) world for any Level 2 application. We then remark on the qualites of a Theory (Slides 17,18) and, drawn from the Ontological literature, suggest how Ontology may help in the framing of necessary theories (Slide 19). It is suggested in Slide 20 that progress has been made at Level 1 without formalized ontological analysis because the real-world concepts are adequately specific and adequately well-understood such that engineering theories and algorithms can be nominated to form desired state estimates. In Slide 21 we argue that that is not the case for Level 2. The next Slides through Slide 30 discuss the important interaction (yet to be better or well-understood) between formalized processes by which application and user-specific information needs are defined (such as by Cognitive Work Analysis, CWA), and the formalisms of Ontological specification. These two related but not equivalent approaches to detailed specification of the information elements of a Level 2 conceptualization must be harmonized in some way; ie we must be able to relate “user language”to “ontological language”. Slides 31 and 32 simply depict the concern that if we must develop mission-specific ontologies (as the ontological community suggests that practical ontologies are of course application-specific), then there is a scope problem of constructing a wide base of ontologies.

Slides 33 onward to Slide 52 review a variety of works in the literature that specifically address various of the basic notions we have so far (withoput a proper ontology) ascribed to Level 2; these slides point out that there is a body of theoretical literature that exists that in fact attempts, albeit not in the defense applications of interest here, to offer various formalisms to deal with situational states, behaviors of individuals and groups, with human intentionality, etc. It is suggested and remarked that the fusion community has apparently not drawn possibly-relevant ideas for its own purposes from these possibly-helpful theoretical foundations. Summarizing in Slide 53, we suggest that, following the formation of a relevant ontology, there would be suggested various relevant theories for describing the relationships among situational components and their attributes that can then be used as a basis for eventual algorithm development. Earlier, in Slides 20 and 21, we have asserted the distinction between Theories and Algorithms in that theories are descriptive of the real world in a context that is unconstrained by limits of observability, whereas algorithms are in essence approximations to the theories governed by or constrained by particular observational limitations.

Technical Survey and Forecast for Information Fusion

Per Svensson
Swedish Defence Research Agency
FOI
SE-172 90 Stockholm
SWEDEN

"One of the greatest opportunities of the current Revolution in Military Affairs is to design and build computer networks that can enhance - and replace - humans in many aspects of the data 'fusion' process."

Admiral William Owens,

Lifting the Fog of War, Farrar Straus & Giroux, 2000.

ABSTRACT

This paper summarizes the results of a study forecasting the development of information fusion technology over the next 20 years [1]. The study was carried out on behalf of the Swedish Armed Forces, as part of an effort to estimate future developments in some technology areas of high relevance for the development of a Network-Based Defence system.

1.0 INTRODUCTION

Information fusion is still in the research stage and it is not yet clear how its basic methodology should be structured. The study describes the current situation based on a thorough and critical analysis of the open research literature. Based partly on published assessments by other authors, partly on the author's own experience, an estimate is then made of the likely progress in this area over a 10 and a 20 year period.

The combined effect of rapid technical development in sensor technology, the establishing of new communication technologies globally, and the literally exploding improvement of the price-performance ratio for computer systems, forms the technological background of the efforts by a growing number of nations to create a Network-Based Defence (NBD).

This move towards NBD implies that military decision makers will receive an ever increasing information flux to try to cope with. Information fusion is about providing support in the task of weighing together and interpreting this information, i e, to instantiate, perhaps even build, good models of a complex, uncertain reality more or less in real time.

New and rapidly improving capabilities to automatically collect large sets of data, and to move large sets of data across communication networks, have raised new questions regarding opportunities to automatically sort, classify, and interpret data to form an operational picture. In this process, data from various sources are used: sensor data, as well as data on probable behavioral patterns of the opponent, cultural and geographic characteristics of the area of operations, etc. Such data have previously been *fused* to an operational picture through time-consuming manual analyses and discussions. As the availability of sensor data explodes as a result of technological advances, this manual fusion process becomes a bottleneck when establishing the operational picture. If sensor data can not be processed and automatically

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

fused with information already known from other sources, the increased availability of sensor data in the widest sense will be of little use. Also, it is only through the use of computerized information processing that end users can be made to benefit from the quality, robustness, and conceptual precision of the methods being developed by information fusion research.

Like in other large-scale and ambitious applications of information technology there may exist a risk of user alienation if systems become too autonomous and do not communicate adequately with its users during the analysis and forecasting processes. This risk is closely related to an increasing need for a mathematical and scientific education of those intelligence officers who will be the users of future information fusion systems.

Table 1: General Forecast for the Development of Automated Information Fusion

Situation in 2002	Estimated situation in 2012	Estimated situation ca 2022
Immature technology. Methodological basis of information fusion not yet established.	Methodology for situation analysis developed for many military applications.	Effective systems for threat and situation analysis integrated in NBD.

2.0 TECHNOLOGY AND SYSTEMS FOR INFORMATION FUSION

Information technology – primarily software – supporting the task of creating in real time a correct situation description, or more precisely, model, adequate for the decision support needs of an operative or tactical military commander, will be called here *information fusion systems.* The concept covers those parts of the information fusion process which are intended to support analysts and decision makers in their planning and execution of intelligence collection, as well as their analysis, of intelligence from all sources, i e processes which are more closely related to intelligence processing than to sensor data processing or target position estimation.

The purpose of developing and deploying systems for information fusion in tactical C^2 and intelligence systems is thus to provide support to commanders and intelligence officers in their continuous work of collecting, processing, and conceptualizing information about various potential threats during an ongoing military operation. Sensors and multi-sensor systems have important roles as "subcontractors" in this process, however, in focus here is the issue of how to shape computer supported methodology for *real-time modelling*, i e, representation, estimation, interpretation, and forecasting of threat extent, structure, capability, and options, given what is known about the opponent's situation and behaviour, or doctrine, as well as the geographical environment. Note that the intent of the opponent will usually be unknown, even "unknowable", to the fusion system, whose task is instead to establish and evaluate all his options to act. The underlying idea is that from a sophisticated analysis and evaluation of these options using mathematical modelling, it should usually be possible to radically but rationally reduce the number of likely enemy actions to a small set, manageable by human decision makers, and whenever this is found not to be the case, to redirect in real time one's own intelligence collection resources so as to eventually reach that goal.

Reconnaissance and analysis resources, as well as fusion results such as situation and forecasting information, must be accessible from all levels in the operative organisation. In dialogue with its users, the system needs to be able to make good proposals to resolve resource conflicts as they occur. Also, it needs to function in situations when communications have low capacity, even when they are intermittent, although necessarily with reduced precision of its result. Thus, the feasibility of information fusion depends strongly on the availability of a seamless and flexible information network. Without such a network, neither input to nor output from fusion processes will reach their destination in time. Quantitatively, the message delivery time and bandwidth of each link in the network at each point in time will be important factors influencing the total capability of the information fusion system.

3.0 INFORMATION FUSION TERMINOLOGY AND PROCESSES

Language usage in the fusion field has been and still is varying, so that there is reason to define the terminology used in this paper (and in the Swedish R&D community for NBD).

Sensor data fusion, or using a practical abbreviation, *sensor fusion*, deals with creating a target situation description (or "picture") based on data received from the sensors. Static data, such as sensor and target properties or threat libraries, are also sometimes used. Sensor fusion primarily involves target tracking, but also target identification based on sensor data. *Multi-sensor fusion*, in particular, denotes such sensor fusion where more than one sensor is employed to create target tracks or situation picture. The purpose of multi-sensor fusion is to combine information from several sensors in order to obtain greater robustness, precision, and range from the sensor systems. *Multi-sensor multi-target fusion* is sometimes used when one wants to emphasize the use of a set of several sensors to create and maintain a target situation picture containing many different perceived targets which may perhaps also appear and disappear randomly.

Information fusion employs, in addition to all available target information, other kinds of intelligence data, as well as other relevant kinds of data. This is all fused during *situation assessment* into a higher-level *situation model* (here too sometimes called "situation picture", a usage we have frequently found misleading). Interpretation of the situation – threat, vulnerability, tactics etc. – is performed in the *impact assessment* process. The *resource management* process provides feedback to the sensors as well as to the assessment processes. In our terminology, Levels 0 and 1 in the so-called JDL model do not belong to information fusion and should be viewed as sensor fusion processes.

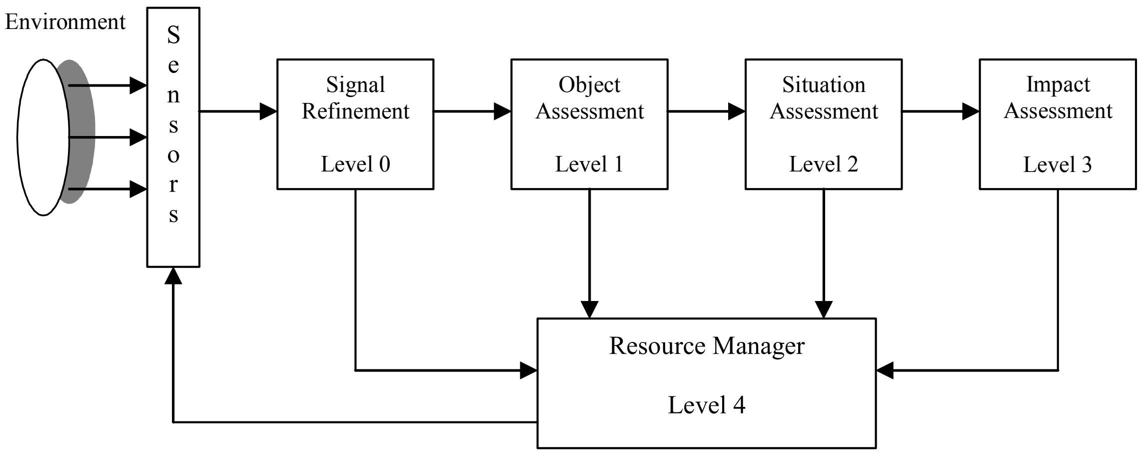


Figure 1: JDL Model.

3.1 Level 2: Situation Assessment

Results from sensor fusion processes (level 0 and 1) are passed to a decision support system for situation and impact assessment. During situation assessment a model is reconstructed of the situation which caused the observed data and events. A set of alternative hypotheses about the perceived situation is also generated. The hypotheses are evaluated with respect to the observations and are ascribed probabilities. The goal of this analysis is to find, characterize, and rank the most likely hypotheses. Situation assessment is a chronologically ordered process where data arrive over time and the analysis is successively refined towards a gradually more detailed and probable best hypothesis.

3.1.1 The Aggregation Problem

If, like during situation assessment, one strives to understand an ongoing operational development, the relevant level of analysis is usually not separate physical objects but groups of organizationally related objects. Frequently, the goals of a military operation are such that no single unit is indispensable for completion of the task. Further, a group of platforms may present a different – and much greater – threat than the sum of all the threats presented by each individual platform. Thus, in addition to potentially meeting the need of the decision maker to have the situation presented in a way which avoids too much detail, aggregation of vehicles or other platforms to military units promises to be an important step towards the goal of unmasking the opponent's intent and find effective countermeasures.

The aggregation problem for sets of objects [2] has certain similarities to the classification and identification issues for single objects which are treated during object assessment. Based on observations of single objects (or, more accurately, of certain attributes of an object), one wants to deduce which kind of unit is observed. The difficulties are in some respects similar. How do you determine if two or more observations belong to the same unit (object)? What kind of a priori knowledge do you need to determine which unit (object) the various objects (attributes) belong to? How should this a priori knowledge be represented, how should the evaluation be carried out, and how do you best represent the uncertainty of the conclusions?

In the aggregation problem there is the added complication that unit types belong to some (usually hierarchical) organizational structure. Different objects may belong to different units, different units may belong to different companies, etc. Certain identification on one level may be combined with uncertain identification on another level. Also, the structures at each level are dependent of each other. Higher units may contain certain specific smaller subunits and at the same time some subunits may in themselves be organization-determining. Methods which can deal with such dependencies are therefore likely to be particularly interesting.

3.1.2 Level 3: Impact Assessment

Impact or *threat assessment* is a process which should be viewed from several perspectives.

Here risks and opportunities are analyzed for own forces to meet the opponent in an effective manner. Results from the previous situation assessment is combined with received indications about the opponent's intent as well as prestored information from technical and doctrinal databases to deliver a complete impact assessment.

In impact assessment possible threats are identified and ranked according to the character and time horizon of the threat. Until now, research has largely been focused on air and sea applications where the threat analysis takes its departure from own aircraft or ship survival. In simpler cases the threats from adverse forces are evaluated based on distance and speed of approach of the threat. Here one also needs to take the speed of own aircraft/ship into account. Doctrinal data play an important role. Knowledge about weapon systems and their ranges may help refine threat evaluation, and more complex doctrinal data on, e g, tactical behaviour might also be used in the modelling.

In a longer time perspective the threat analysis may be extended from dealing only with the security of the own unit to that of other targets of interest to the opponent. Here too doctrinal information about the opposing units will play an important role, however, needs to be extended with geographical data about avenues of approach and trafficability.

To estimate the probability that a potential threat is real is a critical part of impact assessment. Potentially, each enemy unit present within a certain area a possible threat to any valuable targets in the area, but obviously certain combinations of units and threats are less likely. The generation of threat models therefore needs to be sensitive to, and provide a representation of, those factors which are likely to influence the probability assessment of various threats, such as distance to target, direction of movement, military strength and type of unit.

3.2 Level 4: Adaption and Resource Allocation

Multisensor systems become ever more important as a component in a number of applications, military as well as civilian. Since a single sensor can usually perceive only limited and partial information about its environment, one needs to make use of several coordinated sensors of the same or different types, to obtain sufficiently many comprehensive local views with different focus and from different perspectives, in order to create a complete, integrated view of a complex phenomenon [3]. This presumes that an underlying model of the observed phenomenon exists and may be exploited. The value of multi-sensor systems therefore consists in their ability to create greater opportunities for machine perception and to improve the perception of the state of the observed environment, compared to what could be achieved by a single-sensor system.

On a higher level, sensor management deals less with direct control in real time, although this functionality may well be desirable in the future, than how best to plan the use of (combinations of) mobile sensor platforms such as, e g, *Unmanned Aerial Vehicles* (UAV). The basic issues, however, remain the same: sensor resources are to some degree given, however, the information needs have to be defined. Here, impact assessment plays an important role. In the simplest cases the issue is how to cover a set of areas in the most efficient way. Since maybe not all areas can be covered certain areas will have to be prioritized with respect to current threat patterns.

In dedicated sensor systems resource allocation essentially consists of direct sensor control: how are sensors to be controlled in order to track a target or cover an area in the most efficient way. Basically, this is an optimization problem where sensor resources are weighed against information needs. Sensor resources are given, but the information need may be defined in different ways.

4.0 RESEARCH STRATEGY

The key research strategy issue in this context is to identify problem areas and specific research issues which can be tackled by a coherent methodology, and at the same time have, or can be expected to acquire, tactical, operative, or strategic relevance.

Although fusion methodology is of great interest in a number of application areas, we focus here on military applications. This characterization is however far from sufficient when addressing the question whether information fusion is a possible, respectively an important, technology in network-based defence. Those problems which may need fusion support during an international peace-keeping operation are different from those which need to be solved when defending against an armed attack on one's own territory. In another dimension, issues of strategic intelligence during peacetime are largely different from those which need to be answered at the tactical level in wartime. Applying information fusion to the defence of a country's infosphere is again different from information fusion applied to the forecasting of an enemy army unit's movements and intents.

This does not imply that there is no ground for considering all these problems as related, and by doing so, assigning the information fusion concept a broad, cross-disciplinary meaning. It does imply, however, that the technical forecast we are trying to produce has to be made with specific applications in mind.

Although it is obviously impossible in general to foresee with certainty what an enemy unit is going to do, on the other hand no commander of a ground-based unit has much flexibility to rapidly change the situation of his unit. The emergence of increasingly swift observation, command, and impact processes will therefore lead to an increased need for tools that can provide timely and accurate short-time predictions.

However, to be able to make longer-term forecasts than those purely kinematical and dynamical predictions which multisensor fusion provides, one has to exploit two complex scientific areas:

- quantitative, or in some cases qualitative theories for reasoning under uncertainty
- knowledge about those limitations which the environment and the organization and doctrine of the enemy impose.

5.0 HOW CAN INFORMATION FUSION BE AUTOMATED?

In *Handbook of Multisensor Data Fusion* [4], this issue is discussed by Hall och Steinberg.

5.1 Situation Assessment

Improved methodology for situation assessment, a central information fusion subject area, considered by these authors to be immature today, will probably require a better understanding of how to select and use existing methods for knowledge representation (like rules, frames, scripts, and fuzzy logic), in conjunction with a better appreciation of the strengths and weaknesses of human reasoning about such problems. One example is the need to employ reasoning based on negative information, i e information that has not been observed but should be expected in a hypothetical situation. Another promising area is considered to be development of tools for compensating known weaknesses in the way humans manage information (e g, humans often look for information which can confirm a stated hypothesis, rather than information that may refute it).

5.2 Impact (Threat) Assessment

Regarding methods and systems for impact assessment the situation is said to be similar but even more immature. The problem of predicting an opponent's intent is a basic issue in impact assessment, and it can be systematically modelled in many cases when the opponent's behavior follows a known doctrine. In modern wars and conflicts, however, doctrine is frequently poorly developed, unknown, or even non-existent.

Impact assessment is dependent on our ability to recognize behaviors, and furthermore requires methods for evaluating both enemy intent given a certain behavior, as well as for modelling one's own vulnerability against various kinds of threats. Since interpretation of threats and possibilities can seldom be made unequivocally, it is important to endow a system for impact assessment with the ability to represent and manage many alternative interpretations, all with some measure of likelihood attached. Here, one's dependence on doctrinal knowledge and behavioral rule systems is even greater than during situation assessment. This we consider to be one of the greatest obstacles against the development of methods that can be applied across a wide spectrum of unconventional threat scenarios. It is likely, e g, that new methods must be created for mapping, and later perhaps modelling, of organizational structure and threat behavior of terror and guerilla organizations.

5.3 Process Control

Process control involves, e g, methodology which has long been applied to controlling single sensors and homogeneous sensor systems. In information fusion the concept has a wider meaning, however: how can the entire intelligence collection process be optimally controlled in real time? In this area, there is much work left to do, although significant progress can be expected also in the shorter term.

6.0 FORECAST

Our own view of the state of development in the areas of situation and impact assessment differs in a few but presumably significant respects from the one presented in the cited paper by Hall and Steinberg. An important recent trend has been the emergence of new theoretical approaches, creating significantly better opportunities for managing uncertainty and ambiguity in information systems. In general, new representation and analysis methods from the discipline called management of uncertainty are appearing more or less continuously, some of which, e g Bayesian Networks and Abstract Hidden Markov Models, promise to be effectively applicable to information fusion problems, such as multi-agent plan recognition [5].

In particular, we believe that new developments in finite set statistics, new particle filtering methods, and clustering methods for force aggregation, in combination with new methods for multisensor management, advanced simulation systems and terrain models, will lead to important advances regarding detection, classification, tracking, and prediction of complex set objects, such as ground forces on various organizational levels. In light of these and other scientific advances, our assessment can be considered somewhat more optimistic than that of Hall and Steinberg.

We believe that the problems that primarily have to be solved to enable significant automation of information fusion processes are in large measure scientific methodology issues about how to formulate, model, and solve certain generic problems, whose structure is often currently insufficiently known. If one recognizes this, it becomes easier to understand why research in this area will require considerable resources, including probably calendar time, before practical results can be obtained across a broad application area.

In conclusion, we estimate that within ten years there will exist methodology for information fusion which will be useful in practice for the Swedish defence, in particular for situation assessment in ground warfare scenarios, and that an additional 10 years of focused research may lead to the establishing of effective, productivity-boosting such systems within the context of a network-based defence. This prediction presumes that the quality and quantity of this research does not diminish but instead successively improves during the next 5 to 10 years.

ACKNOWLEDGEMENTS

The author is indebted to Egils Sviestins, SAAB Tech Systems, John Cantwell, KTH, and Johan Schubert, Swedish Defence Research Agency (FOI) for oral and written exchanges which have strongly influenced the views presented in this paper.

REFERENCES

[1] Svensson, P. (2003) *Technical survey and forecast for information fusion 2002-2020*. FOI-R–0798–SE (in Swedish).

[2] Cantwell, J., Schubert, J., Svensson, P. (2000) *Information Fusion. Proposal for a "strategic research kernel*" 2001-2005. FOA-R–00-01632-505–SE (in Swedish).

[3] Xiong, N., Svensson, P. (2002) Multi-sensor management for information fusion: issues and approaches. *Information Fusion* 3, 163-186.

[4] Hall, D.L., Llinas, J. (eds.) (2001) *Handbook of Multisensor Data Fusion*. CRC Press, Boca Raton, FL, USA.

[5] Suzic, R. (2003) Representation and Recognition of Uncertain Enemy's Policies Using Statistical Models. *This conference*.

Introducing the Canadian Information Centric Workspace Concept

Gaetan Thibault
Defence R&D Canada (DRDC Valcartier)
System of Systems Section
2459 Pie-XI Blvd North
Val-Belair (Quebec) G3J 1X5
CANADA

Gaetan.Thibault@drdc-rddc.gc.ca

François Le May
DMR Consulting
A Division of FUJITSU Consulting
2960 Blvd Laurier, Office 400
Sainte-Foy (Quebec) G1V 4SI
CANADA

Francois_Le_May@dmr.ca

ABSTRACT

Intelligence, Surveillance, Target Acquisition and Reconnaissance (ISTAR) is an evolving information operations (IO) concept in the Canadian Land Force. ISTAR provides the commander with a system to collect and process required information for producing intelligence on the threat and knowledge on the environment during operations, as well as knowledge needed to identify, acquire and engage targets. The various processes used to collect and analyze the information are the result of numerous individual systems some of which have only been recently introduced in the field while many others are still in development as a result of advances in the information age. This compendium of systems makes ISTAR a "System of systems", as opposed to a single system. This paper presents the new Canadian information centric workspace concept that provides a more coherent information management approach to better support the Commander in both its tactical intelligence and operations activities at brigade level. The info-centric workspace concept aims at offering a seamless collaborative environment enabling the ISTAR staff to perform their tasks using different applications / services through a standardized Human Computer Interface (HCI).

INTRODUCTION

The explosion of information technologies has set in motion a virtual tidal wave of change that is in the process of profoundly affecting both organizations and individuals in different aspects. This means that military organizations also face a tidal wave of transformation of an irresistible force [1] that, at the same time, offers unprecedented challenges [2]. The military does not have much choice. Resisting transformation is futile. However, accepting transformation in only the technological aspect is also not a valid option. Today, improvements in processing power and communications means make information technologies even more attractive and cost-effective for organizations to implement. Willingly or not, we have entered the information age. As Owens puts it [3], for a long time, information has been inseparable from commanders, command structures, and command systems. Information is no longer the prerogative of commanders and command structures but has become necessary to all participants in a mission [4].

Many armies have by now learned that when introducing Command and Control (C2) information technologies (IT) to their organization, a series of changes occur in a number of areas and if these changes are not properly taken into consideration in the planning stages of the transformation process, then these changes will become hindrance in the accomplishment of the missions thus planting the seeds for the overall rejection

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

of the system. The areas that will be affected and need to be considered in the transition have been regrouped by Champoux in her thesis [5] into three main perspectives as illustrated in Figure 1 and are: a) Systems, b) Users, and c) Processes. What is meant by "systems" are the hardware and software components related to Information Technologies (IT) that, when put together according to a set of requirements and specifications, make up IT systems. The term "users" refers to the people and their skills, education, training, experience and organization. The term "processes" refers to the Doctrine, Standard Operating Procedures (SOP), and Techniques, Tactics and Procedures (TTP). The successful business solution will be the one achieving best harmony between the three perspectives: Users - Processes - Systems.

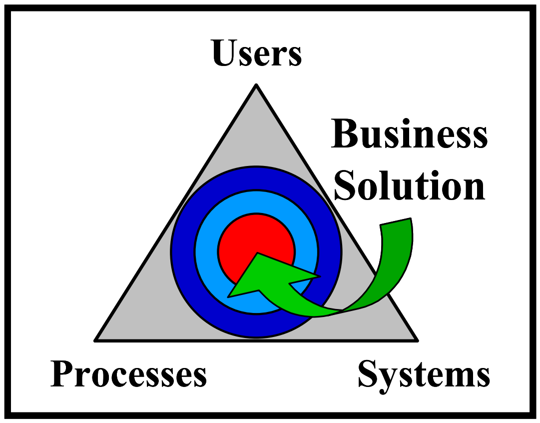


Figure 1: System of Systems Harmony Triangle: Users – Processes – Systems.

THE CANADIAN ISTAR CONTEXT

In the Canadian Land Force Information Operations doctrine [4], the definition of ISTAR (Intelligence, Surveillance, Target Acquisition, and Reconnaissance) is: "a system where information being collected through systematic observation and sensing is integrated with that collected from specific missions, and is processed in order to meet the commander's information requirements." ISTAR integrates sensor capabilities and the intelligence process that provides the direction and processing of sensor data. Therefore, the ISTAR constitutes a "System of systems" that is managing and fusing data to serve the command function through integration of a wide range of sensing capabilities and information functions and processes. Considering this "System of systems" approach, the complexity of introducing automated data fusion tools is strongly related to the nature of the available information, its pattern of dissemination, and the organizational adaptation capacity.

If the Canadian Land Force is to be successful in fielding an efficient ISTAR "System of systems", many ingredients have to work together. We thus need to address many factors. As with any other system, the ISTAR System encompasses the three main perspectives explained earlier: the systems, the users and the processes by which these users use the systems. Based on the results obtained by Champoux [5] in a different context, the authors recognize the fact that all three perspectives have to be addressed concurrently by the Canadian Land Force for a successful implementation. Experience shows that the introduction of new information technologies and their capabilities into organizations is potentially risky unless accompanied by a planned change transition in a number of key areas. This paper discusses two of the aspects illustrated in Figure 1: a) the new ISTAR business concept addressing the "processes" perspective, and b) the necessary transformation in the "systems" perspective for a successful business solution implementation. This paper will concentrate on the ingredients needed to create the conditions where the data fusion tools can be successfully introduced while exploiting information age technologies.

THE INGREDIENTS FOR "PROCESSES" TRANSFORMATION

The ISTAR system must be able to effectively collect, process and act upon information to ensure the commander has enough of the knowledge he needs to conduct a faster and higher quality decision-action cycle. The information should be collected and processed from data into useable knowledge so that the commander can develop an understanding of the current situation in order to prepare future operations. Since, data fusion processes are to be found imbedded at different levels within many of these systems ranging from collectors up to analysis centers, this movement of information must be achieved through a rigorous process stream that allows for the most effective transformation into an intelligence product. One challenge is to bring these data fusion processes to work collaboratively as distributed processing centers.

Within the operational context, it is assumed that ISTAR produces intelligence in response to commander's requirements as defined within the Operational Planning Process (OPP) [6], the Intelligence Preparation of the Battlefield (IPB) [7] and the Targeting Process [6]. These processes provide direction on the employment of the ISTAR system and provide the framework within which ISTAR information and intelligence products are exploited. As mentioned by Cain and Walker in their study of ISTAR processes [8], few countries in the American, British, Canadian and Australian (ABCA) coalition have developed a complete intelligence system comparable to the emerging Canadian ISTAR system. Lessons learned from allied ISTAR systems are useful in pointing the way toward an improved Canadian ISTAR system. However, in these systems, the volume of information to be collected and acted upon is increasing with the inclusion of modern sensors on the battlefield. Further use of existing systems is likely to overwhelm the analysts to the point that modern sensing technology may become a hindrance to effective decision-making.

The principal goal of an ISTAR capability is to collect and transform information into knowledge for the decision-makers. The idea of the information sphere comes to hand when one tries to illustrate the ISTAR business process. The ISTAR doctrine [5] speaks of Global Information Environment (GIE) while the knowledge management literature refers to the information sphere. According to Knight [9], « the info-sphere is a theoretical representation of all the information that exists. From this vast selection of information, the decision-maker is seeking "operational information", that is, the relevant information that is needed to make decisions in a given operational context. In reality, the operational information is scattered all over the info-sphere, and in many cases one will not be able to identify a piece of information as "operational info" until it is needed. Recognizing the term "operational information" is the starting point in building a systematic approach to finding, processing and using all of the information that must be made available to decision-makers.» From an information system designer's perspective, the objective is to build a system that maximizes the efficiency and value of the actions "to access, collect and assess information" with the minimum of friction so that decision-makers get the intelligence they need (Figure 2). If organizations really want to exploit all of the richness available from the info-sphere, the authors believe that we must get away from the mechanization of processes and adopt a global approach taking into consideration the three perspectives of Figure 1.

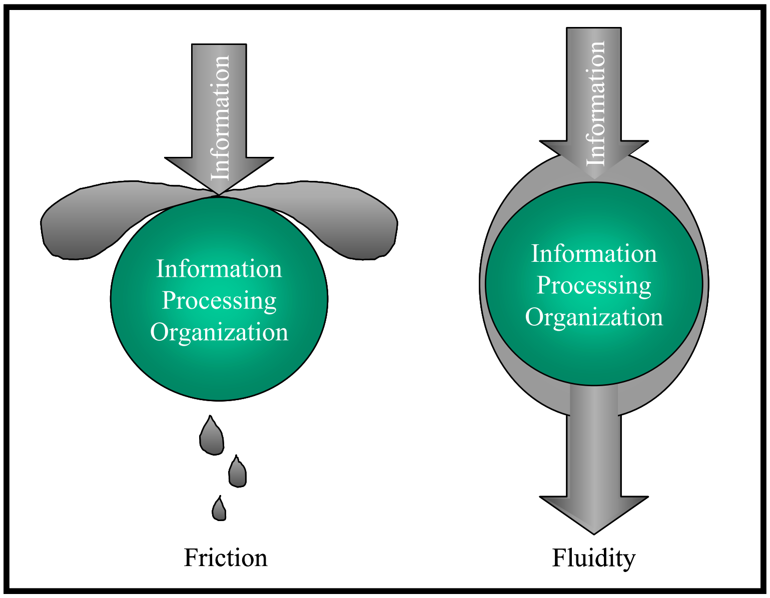


Figure 2: The Process Transformation enables organizations to reengineer their information processes by removing friction while adding more fluidity.

By reviewing the current "Processes" involved in ISTAR, it became clear to the authors that because the overall ISTAR process is not an independent process from Command and Control (note that "Command and Control" is a military term for leadership and management), an improved ISTAR business process must be introduced to facilitate the processing of information and to manage the tasking and re-tasking of ISTAR system components. Operational forces and decision-makers are constantly coming across gaps in their knowledge that need to be filled, and they need a mechanism to raise these requirements in a systematic manner. A similar mechanism is also required for ISTAR analysts that are constantly sifting and comparing information as they often identify information gaps that need to be filled in order to make sound assessments. In addition to these *ad hoc* information needs, a standing set of information requirements governing routine activities and planned operations must be in place and managed at all times. Rather than having separate techniques and procedures to manage all of these requirements, a common feedback or requirements loop is needed to properly task the collectors and ensure that responses are provided in a timely and prioritized manner. In order to render the ISTAR "System of systems" more efficient as a whole, the authors have found that a new ISTAR function needs to be created: an ISTAR Coordination function. This new ISTAR function will provide more than information brokerage services in terms of providing appropriate information to requesters but also providing a resource management capability for information collection. Not surprisingly, this coordination function is in line with the definition of the Joint Directors Laboratories for data fusion level 4: Process Refinement [10]. This process refinement is an element of resource management where adaptive data acquisition and processing are used to support mission objectives. This provides the rationale for having a specific ISTAR Coordination Function composed of two closely tied teams: an Information Collection Management Team and an Intelligence-Surveillance-Reconnaissance Assets Management Team.

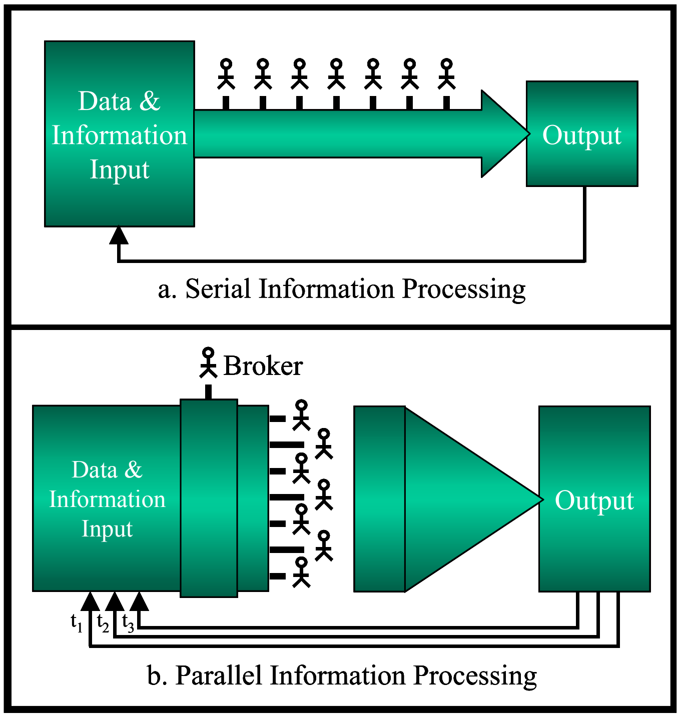


Figure 3: Organizations employing information workers in parallel processing (collaborative sharing) topology proven to produce output of better quality and quantity than the serial processing topology.

To introduce an efficient ISTAR Coordination function, we took the approach to implement a seamless collaborative environment based on a shared network architecture. This will allow all functions comprising ISTAR to work together and enable the different analysts to perform their tasks and extract information using different applications through a standardized Human Computer Interface (HCI). This defines the "information centric workspace" concept. This concept also assumes that core application components will plug into the workspace environment in a similar fashion irrespective of the military functions being integrated. The transformation brought about by the "information centric workspace" concept is now possible because of new modern capabilities associated with information technologies and most notably those associated with information sharing, collaboration, and visualization. We expect this concept will improve our ability to bring all of our information and all of our knowledge and experience to bear. Amongst other things, these modern capabilities enhance the possibility to do collaborative fusion and analysis e.g. parallel information processing (Figure 3) that has proven to increase the situation awareness of all connected information workers from a recent study by Blain [11]. It is now understood by the Canadian ISTAR team that concurrent information processing such as data and information fusion requires collaborative work under asynchronous information management. Thus, the conditions necessary for success of ISTAR in the Information Age revolve around an organization characterized by asynchronous information flows that are not unduly constrained, where the key parts of individual components can be self-synchronized allowing collaborative fusion and analysis (Figure 3). These are the characteristics associated with true integrated processes. One objective is to put the Information Centric Workspace concept accessible at all levels in the chain of information transformation so that fusion and analysis can be performed whenever and by whoever requires it and that the results become input to the next transformation iterations in the "System of systems". The authors believe that to successfully implement automated data fusion tools into C2 systems, it is necessary to adopt such a "System of systems" approach. In this way, the whole System design could take into consideration the needs of all users in the information transformation chain.

The conditions for transformation in the "processes" perspective imply the reengineering of some information processes if organizations are to really exploit the richness of the info-sphere. What is needed in the Canadian ISTAR context is to bring the notion of information brokers to manage the balance between information requirements and information collection. This balance is related to Process Refinement of the JDL data fusion level 4. Hence, the ISTAR Coordination function provides information brokerage services coupled with resource management for information collection. Important evolutions in information technologies had to occur to enable information sharing, collaboration, and visualization. These evolutions made possible the emergence of the new concepts of data centric and network centric warfare which in turn enabled the emergence of the info-centric workspace. These evolutions allow the possibility to do collaborative fusion and analysis in distributed processing centers. The info-centric workspace approach when coupled with the data-centric and the network-centric organizations should have a multiplier effect on the availability, quality, timeliness and management efficiency of the information. However, before being able to reap the fruits from the information centric workspace concept, it is also necessary to investigate the other ingredients needed to implement such a concept from the "Systems" perspective.

THE INGREDIENTS FOR "SYSTEMS" TRANSFORMATION

Over past decades, the military have developed several different vertically specialized functions that supported different Command and Control business processes. When developing information systems to support these military functions without first possessing an overall architectural vision or a road map of all involved functions and processes, it is often discovered too late that these systems while correctly supporting their vertical military functions do not necessarily provide an integrated "System of systems" solution. In this situation, the Canadian Land Forces inherit "stove-pipe" systems that become cumbersome to use and a nightmare to support both from a training and maintenance perspectives. From lessons learned, the requirement for interoperability between systems became the priority. One of the early attempts at integration was through the processing of formatted structured messages that allowed a first level of interoperability. Although promising this approach was not sufficient at the end because of the lack of a reference model such as an agreed common information exchange data model enabling true data sharing capability between systems. While the processing of formatted structured messages was being developed and implemented, communications capability continued to improve in terms of speed, bandwidth and connectivity. Now that systems can be networked together, the development of the Network-Centric Warfare (NCW) concept [12] in the US became possible. This NCW concept is based upon a new model for automated support to command and control (part of "process" transformation), one that features sharing of information (synchronization in the information domain) and the collaborative processes necessary to achieve a high degree of shared situational awareness. Alberts [13] in another publication independently reinforces Blain's [11] findings about collaborative work sharing that «An additional benefit is the increase in richness of the awareness created. This increase in richness occurs as a result of the efforts to reconcile differences in fact and/or perspective that result from (1) more sources of information, (2) increased sharing of information, and (3) collaboration. »

Chaum [14] goes even further by arguing that to fully exploit such NCW concept, there must exist a common reference data model: «The net-centric exchange of operational context, requires a shared top-down information reference model to ensure consistency of content and meaning, without which there is no shared understanding. Properly chosen, a reference model can conceptually link all elements of the battlespace and provide a framework for sophisticated automated reasoning and effective Joint and Coalition communications.» This means that if an Army wishes to exploit the new capabilities offered by information age technologies, it must do at least four things: a) develop an information business vision, b) develop a top-down "System of systems" architecture supporting the vision, c) adopt both a data-centric and d) a network-centric system approach.

Canada has adopted a new ISTAR vision and it is developing a top-down "System of systems" architecture supporting that vision. A first revolution occurred when the Canadian Army decided to impose a data-centric approach to all Land Force systems through the adoption of the NATO Land Command and Control Information Exchange Data Model (LC2IEDM) to promote contextual information (operational context) exchange and interoperability between all systems in 2001. A second similar revolution occurred when the ISTAR project decided to also adopt a collaborative network-centric approach at Battle Group and Brigade command levels in 2002. Yet, this is not enough. We need to go a step further. What is now proposed is a further evolution of these concepts to the Human Computer Interface (HCI) and applications level with the adoption of a collaborative info-centric workspace environment.

This further evolution is nevertheless facing one major constraint in that the workspace environment must work in a distributed environment while being flexible enough to support different decision making processes in different combat situations. The Commander of the Army clearly stated this objective for the approval of the ISTAR project: "ISTAR is to enhance the capability to exercise effective Command & Control guarantying success in all operations." The information-centric workspace will contain a set of tools that will be used as necessary and appropriate depending upon the role and the level of command of the user in the information production chain. A difficult objective to achieve will be to deliver the info-centric workspace at all levels in the information chain so that the same set of tools is available to perform fusion and analysis in support of the intelligence and SA. In order to improve the commander's ability to understand and conduct operations, he must be better informed just not more informed. This is the difference between "too much" and "just enough" information to enable the creation of the right knowledge about and sufficient understanding of the situation [15]. A shift to an intelligent pull approach, where the users get to shape their information space, clearly reduces the probability that users will be overwhelmed with information of little or no relevance. On the other hand, producers of information cannot possibly know all of the uses of the information they collect, nor the importance of the various details or lack of details, so posting before processing is not a solution. Perhaps, giving the possibility to information workers to obtain on-demand underlying data details may alleviate the danger of information bottlenecks.

One of the anticipated benefits of the workspace approach is to provide more flexible work sharing to the organization but at the expense of more user training. For example, in the case of a serial information processing organization (as illustrated in Figure 3), the users learn one specific function or occupy one specific position therefore requiring less training but providing less flexibility for the organization. In this scenario, reduced flexibility also means reduced adaptability to different operational contexts (from the classical low-medium-high level of conflicts to the asymmetrical type of warfare) with different levels of resources to perform the work. In the case of a parallel information processing organization, when using an info-centric workspace, it is expected that one person receiving proper training with the different tools will be able to perform a multitude of tasks but may have to take a longer period of time to accomplish them. This type of organization will increase its flexibility to handle the tasks but may not be able to afford the extra user training required because of lack of resources (trainers and trainees). In the ISTAR TD project, we anticipate that this situation will be alleviated by the implementation of the Electronic Task Support Services (ETSS) tools as an integral component of the workspace. More on the ETSS can be found in another paper presented by Champoux [16]. Furthermore, to reduce the training time required by the users and to maximize usability of the system, some tools of the Workspace must have a consistent Human Computer

Interface (HCI). This implementation results in an aggregation of fully integrated tools available in the workspace as opposed to a set of individually developed stove-pipe systems brought together on a desktop. The information workers will be provided with an enhanced capability to adapt the tools to fit their operational needs and their specific tasks.

Understanding that this new paradigm was a necessary deviation from the old method of acquiring a functional capability, the first step in the project was to develop a "System of systems" Architecture (SoSA). A preliminary study was conducted to model current and future systems in order to extract the common services required by the majority of them. The results of the study produced a clear vision for the environment and a clear path to achieve it. It was recommended that this path be evolutionary instead of being revolutionary so the design of the solution had to take into consideration as many of the current operational and future systems as possible including their limitations. Finally, a "System of systems" architecture was proposed and endorsed by our army sponsors. Figure 4 presents the high level view of the Canadian ISTAR Info-Centric Workspace concept supported by nine groups of services and a data service layer regulating the access to five types of databases for non-structured, structured and special data formats.

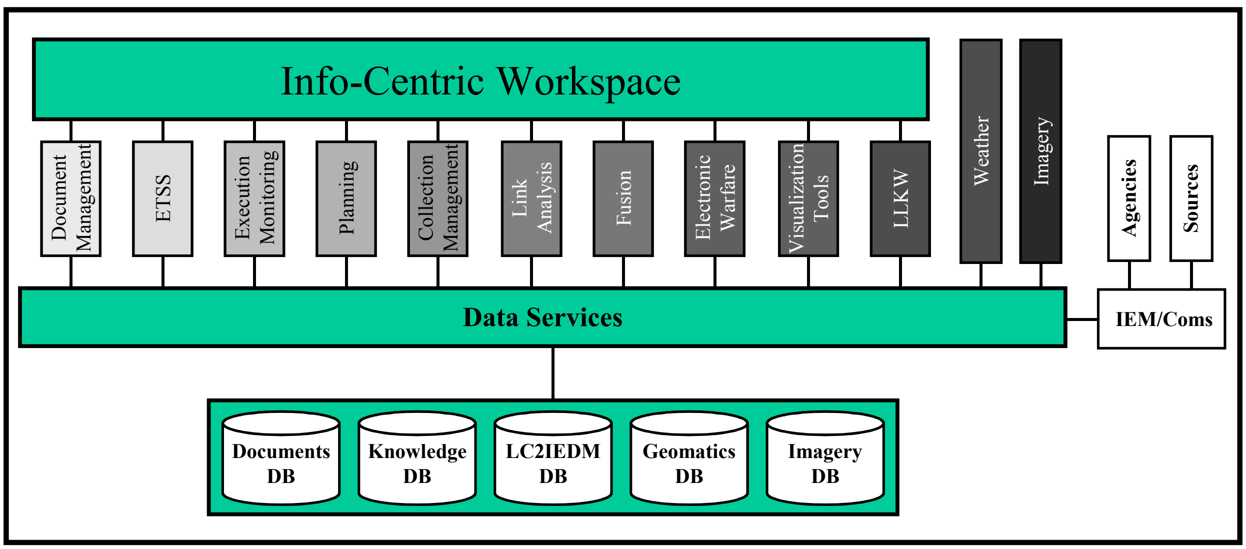


Figure 4: The proposed ISTAR Information Centric Workspace System of Systems Vision.

The SoSA definition is relatively simple to accomplish if the mission and objectives of the System are clearly defined *a priori*. This means that the organization must have a clear vision of where the transformation process will lead it. If one is not rigorous during this phase, System Principles and Orientations for system design can translate into a lot of wasted effort [17]. As an example, if a given system has specific requirements and approach for managing security access to data and processes then it is more than likely that they will impact all of the software development. So building a "System of systems" represents a challenge in determining the necessary and appropriate Principles and Orientation to the problem. In the domain of science and engineering the notion of "System of systems" has always existed. In biology, for example, a human cell is a system in the human body that itself is a system in the animal kingdom. Then that animal kingdom can be considered a system of systems. That is to say that in order to properly address a problem, one has to adopt the right perspective [18]. Once identified, the next step consists in a de-composition into components and services that make up the different parts of the "System of systems". Even though this represents a classical and well-known approach to system design, its success is still an art rather than a science and is the purview of a few dedicated professionals who have refined their art through years of experience.

The ISTAR project obviously brought elements of science in the equation by adopting methodologies and standards into its project management procedures. Understanding that standards are not sufficient by themselves, a suitable methodology is mandatory to provide for proper synchronization and harmonization amongst team members through the use of a common language and agreed checklists that are especially useful to manage large projects distributed over different physical locations. Nevertheless, success rests upon the core team of selected people that must be knowledgeable in that kind of military business. Methods and methodology ensure the quality of the products, hence their transition ability into the field but the validity of the solutions found is based on the team's expertise with its capacity to solve complex problems. However, adaptation to a particular context requires people knowledgeable about the specific operational requirements in order to document the right kind of information. Composing with all these factors is the science of project management.

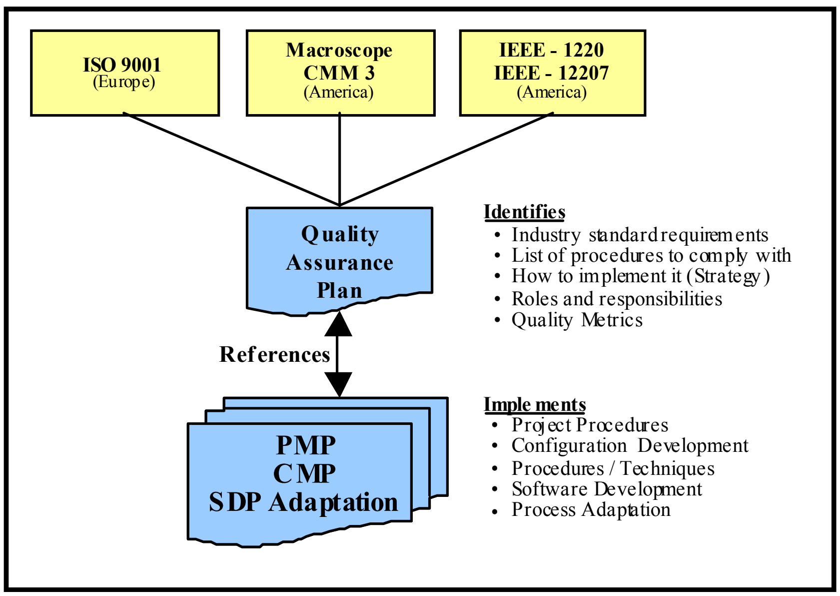


Figure 5: Canadian ISTAR Project Adopted Methodology.

In view of the particular complexity of the SoSA and of the nature of the work to be performed, a methodology was thus chosen and adapted to the Canadian Department of National Defence (DND) context. Figure 5 illustrates the different references that were used in order to build the ISTAR final methodology. The selected methodology was based on "Macroscope™" [19] from Fujitsu Consulting, which is one recommended by Gartner's Group [20], and the Computer Maturity Model (CMM) level 3. Some adaptations were done to take into consideration different standards such as IEEE-12207 (software development life-cycle) [21] and IEEE-1220 ("System of Systems") [22]. From this adapted methodology, the software development and implementation techniques were developed including a Quality Assurance Plan (QAP), a Project Management Plan (PMP), a Configuration Management Plan (CMP) and an adapted Software Development Plan (SDP).

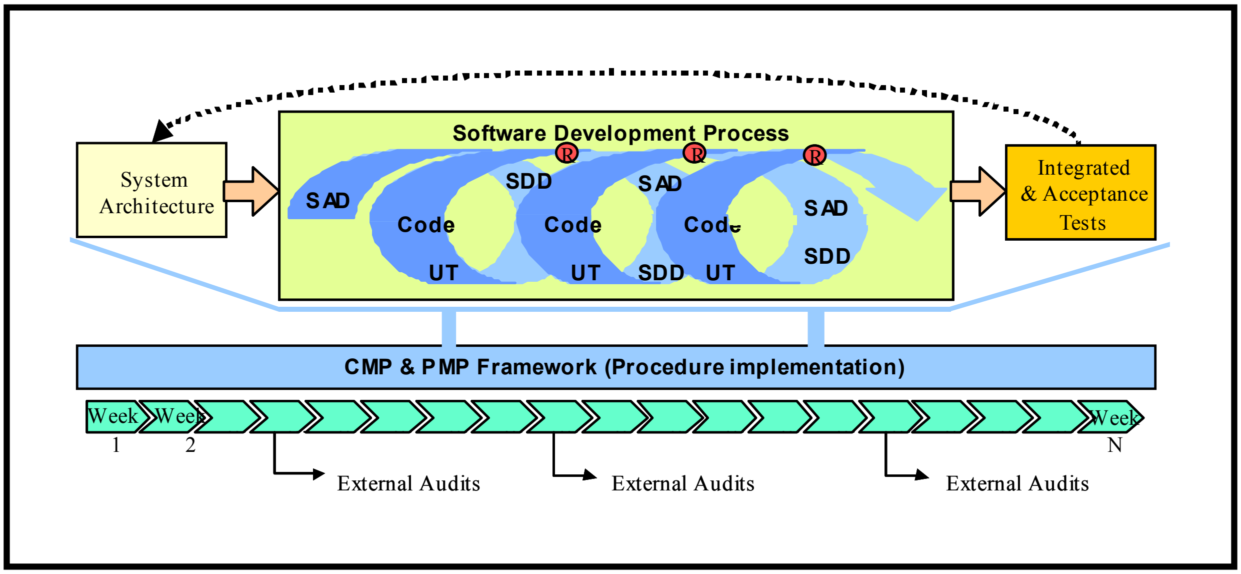


Figure 6: Software Proto-Cycling Adopted Methodology.

Figure 6 illustrates the proto-cycling approach that has been retained. This proto-cycling technique combined with a rigorous Configuration Management Plan (CMP) (including validation tests throughout the development process using test beds in appropriate context) provides a formal incremental system release approach that is better than the traditional waterfall model. This technique allowed all the different perspectives to evolve at the same time and to provide a balanced "System of systems" phased delivery that had periods of 12 to 18 months instead of multi years. This aspect of system delivery becomes a very important issue when fielding complex command and control systems.

One of the most interesting features of "Macroscope™" in this context is the use of Joint Application Development (JAD) sessions with subject matter experts [23]. The JAD technique also had to be adapted to a context of limited user availability. During the JAD sessions, the three main perspectives illustrated in Figure 1 had to be considered and weighed: the users and their capability to absorb the new technology, the procedures and processes that need to be adapted and the diverse required functionality of the systems. This JAD approach had the benefit of providing a continuous training environment for the users, of facilitating user acceptance, and of tailoring the system to user needs. This technique enabled all the different perspectives to evolve at the same time providing a balanced "System of systems". By performing JAD sessions in this fashion, it provided two additional benefits: a) the means to do effective and efficient requirements capture, and b) a value rating for the different requirements was possible thus reinforcing the capacity to perform true value management [24, 25] during the project. What our experience demonstrated is that methodologies and standards to perform this kind of work exist and are available providing they are properly adapted to each particular development effort.

We have seen so far that the project took a top-down approach, it identified a vision, it adopted a methodology, methods and standards, it designed a solution that took into consideration the current and selected future system components, articulated a SoSA, and finally, through its proto-cycling methodology had taken a phased delivery approach. This approach combined with the parallel information processing context as discussed earlier yielded to a development in two phases. Phase I of the project covers the Collection Management (CM) Service, phase II will address three other different aspects: a) development of an Analysis and Fusion Tool Kit that will be distributed in the infrastructure, b) the integration of link analysis service, and c) the enhancement of the Battlespace visualization service currently available.

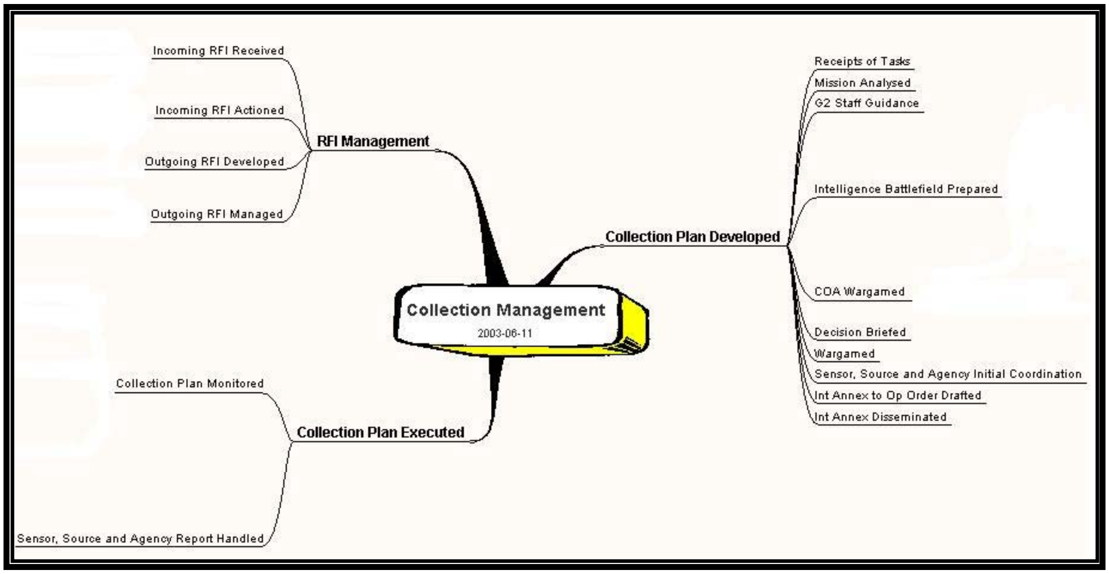


Figure 7: The Collection Management Processes (Levels 0, 1 and 2).

Since the vision developed for ISTAR required the development of the information broker function first, the emphasis was put into the development of the Dynamic Collection Management Service of the ISTAR Coordination Function. Through JAD workshops with user representatives, the Collection Management (CM) process was captured and documented using knowledge management techniques and tools. The captured CM processes are illustrated in Figure 7. The three main processes for CM are: Collection Plan Developed, Collection Plan Executed, and Request for Information Management. These processes have been refined further many levels down and provide a fundamental knowledge element for the ETSS. Again, it was discovered that information technologies when properly introduced in a balanced way (recalling Figure 1) do not change what the users are trying to accomplish but rather change the way to accomplish it. While procedures tend to structure how you do things in an orderly fashion, we found that, when properly implemented, the system could better support the event driven humane way of doing things which is not always so orderly.

Once the different methods in which the CM tasks are executed under different situations have been captured, the information is fed into the ETSS [16]. This service enables on-line contextual help so that when software modules (for each iteration) are tested, the users can readily visualize the procedures, techniques and products related to each specific task. With this overall methodology, the project team can make sure that all the ingredients making a “System of systems” do evolve in synchronization and harmony. Ultimately the ETSS will hold enough information as to provide an on-line contextual help structured in a knowledge network linked to the application. In addition, the JAD technique was used to enrich the ETSS to support Computer Based Training for users on-line. In one controlled experiment, this approach demonstrated that the average time for training could be reduced by a factor of up to five times the traditional tutorial training [11].

Finally, there is another important facet we have not covered so far in this paper. We said that when introducing IT to an organization, processes and procedures will have to be adapted or to be changed.

In military organizations, this refers primarily to Doctrine and secondarily to Standard Operating Procedures and Techniques, Tactics and Procedures known as SOP and TTP. Alberts [13] hits the mark when he wrote: «Military organizations and individuals feel bound to honor doctrine and thus create a mindset and environment not conducive to disruptive innovation. This is because when the nature and distribution of information changes, radical new ways of doing business and complications in the old ways of doing business emerge. …This is why it is so vital that the doctrine community be involved at the beginning of the new concept development process and to stay involved throughout this process.» Also involving the doctrine community early will also facilitate the key process of embedding doctrine into the new systems that in turn will contain and help support the evolution of doctrine. As already experienced in the Canadian Army with the establishment of the new Lessons Learned Knowledge Warehouse (LLKW) [16] on the army infrastructure network, the doctrine is changing from one of publishing "the way" it should be done to a dynamic process of collaboratively learning and sharing best practice.

In summary, based on the current results of our work, we are tempted to believe that the conditions for transformation in the "systems" perspective imply that to fully exploit the new capabilities offered by information age technologies, a military organization like an army should consider at least these five things: a) develop an information business vision, b) develop a top-down "System of systems" architecture supporting the vision, c) adopt a data-centric vision with a common reference model, d) field a distributed network-centric capability and e) embark on an information-centric workspace system approach. The necessity to choose a suitable methodology supported by recognized standards coupled with a project team composed of knowledgeable people are the cornerstones for success. We have selected a methodology for proto-cycling software development based on a phase delivery approach. This approach had the benefit to enable on going user training, user acceptance, and system's tailoring all at the same time during the validation testing sessions. This technique allowed all the different perspectives to evolve at the same time and to provide a balanced "System of systems" phased delivery that had periods of 12 to 18 months instead of multi years.

CONCLUSION

The definition of a perfect "information centric workspace" is when all of the services existing in a "System of systems" are all working on the same distributed network in a manner as to offer a seamless collaborative environment. This enables the different analysts to perform their tasks and to extract information using different services through a standardized Human Computer Interface (HCI). The info-centric workspace approach also provides major improvements in facilitating system's component integration from a user training and a maintenance points of view. The info-centric workspace aims at managing information to avoid information bottlenecks and to improve the commander's capability to understand and conduct operations.

We have seen that ISTAR supports the command function through the integration of a wide range of sensing capabilities and information functions and processes. ISTAR constitutes a "System of systems" that encompasses three main perspectives: the systems, the users and the processes by which these users use the systems. The authors believe that to successfully implement automated data fusion tools into C2 systems, it is necessary to adopt a "System of systems" approach. Another condition of success is to move away from developing systems the old fashion way towards a new information driven "System of systems" approach. In this way, the whole System design could take into consideration the needs of all users in the information transformation chain. These are the ingredients needed to create the conditions where automated data fusion tools can be successfully introduced. The next objective of our research group will be to design, develop and integrate specific fusion tools supporting the intelligence analysts involved in ISTAR processes.

REFERENCES

[1] Alvin and Heidi Toffler, “War and anti-war: The Third Wave”; Bantam Books, New York, 1980, Unclassified

[2] Alvin and Heidi Toffler, “War and anti-war: Survival at the Dawn of 21st Century”; Little, Brown and Company, Toronto, ON, 1993, Unclassified

[3] Owens, William, A., “Introduction to Dominant Battle-Space Knowledge: the Winning Edge”; edited by Stuart E. Johnson and Martin C. Libicki, National Defence University Press, Washington, 1995, Unclassified

[4] Information Operations, B-GL-300-005/FP-001, Director Army Doctrine, Kingston, Ontario, 18 Jan 1999, Unclassified

[5] Champoux, Pierrette, “Étude des tâches et de l'utilité d'un outil à la phase « d'analyse » dans un projet de géomatisation”, (Study of tasks and usability of software tools during the analysis phase in a geomatics project); Ph. D. Thesis, Département des sciences géomatiques, Faculté de foresterie et géomatique, Université Laval, Québec, 289 p., 1997, Unclassified

[6] Land Force Information Operations, Intelligence, Surveillance, Target Acquisition and Reconnaissance (ISTAR), B-GL-352-001/FP-001, Director Army Doctrine, Kingston, Ontario, 01 August 2001, Unclassified

[7] Land Force Information Operations, Intelligence, B-GL-357-001/FP-001, Director Army Doctrine, Kingston, Ontario, 30 Jan 2001, Unclassified

[8] Cain, D.E., and Walker, M. K., “The ISTAR Process Model”, Royal Military College of Canada, Land Force Technical Staff Programme VI, Kingston Ontario, 22 May 2001, Unclassified

[9] Captain (N) Knight, D. W., “Picturing the Puzzle, A Knowledge Management Model for Military Operations”; non published paper, April 2002, Unclassified

[10] Steinberg, Alan N.; Bowman, Christopher L.; White, Franklin E.; “Revisions to the JDL Data Fusion Model”, Presented at the Joint NATO / IRIS Conference, Quebec, Canada, October 1998, Unclassified

[11] Blain, Dominique; “Mesurer et expliquer l'impact du partage en direct des situations militaires du prototype ASIP sur la qualité du renseignement” (To measure and explain the impact of collaborative sharing of military situations on the quality of intelligence using the ASIP prototype), DRDC Valcartier Technical Memorandum 2003-085; September 2003, Canada, Unclassified

[12] Alberts, David S.; Garstka, John J.; and Stein, Frederick P.; “Netwrok Centric Warfare: Developing and Leveraging Information Superiority”. Washington, DC, CCRP. August 1999, Unclassified

[13] Alberts, David S.; DoD Command and Control Research Program, 2002, Publication 96-14441, Unclassified

[14] Chaum, Erik; “Achieving the Joint Vision: The Role of Operational Context in Future Systems”, US Naval Undersea Warfare Center Division, Newport, 2002, Unclassified

[15] Thorp, John, "The information paradox: Realizing the business benefits of information technology"; McGraw-Hill, New-York, 1999-2003, Unclassified

[16] Champoux, Pierrette; Thibault, Gaetan; Trudel, Martin; "A Lessons Learned Knowledge Warehouse to Support the Army Knowledge Management Command-Centric", NATO RTO Military Data and Information Fusion Symposium, Prague, CZ, 20-22 October 2003, Unclassified

[17] Defining System Principles and Orientations, Macroscope Productivity Center, deliverables P140S – Owner Requirements, P230S – User Alternatives, and P240S - User Principles; Fujitsu Consulting (http://us.fujitsu.com/services/Consulting/Methodology/Macroscope/) Unclassified

[18] de Rosnay, Joël; "Le Macroscope", Ed. Seuil, 1979, ISBN : 2020045672 Unclassified

[19] Macroscope from Fujitsu Consulting: (http://us.fujitsu.com/services/Consulting/Methodology/Macroscope/) Unclassified

[20] GartnerReasearch, 2002 "Methodware Magic Quadrant for Software Processes", 27 August 2002, Matt Light, Note Number M17-6239 (http://www3.gartner.com/Init) Unclassified

[21] Institute of Electrical and Electronics Engineers, Standard 12207 – IEEE standard for Industry Implementation of International Standard ISO/IEC 12207;ISBN 1-5593-7999-5; Product No.: SH94603-TBR; IEEE Standard No.: 12207-1997 SET, (WWW.ieee.org), 236 pages; 1997, Unclassified

[22] Institute of Electrical and Electronics Engineers, standard 1220 – 1220-IEEE Standard for Application and Management of the Systems Engineering Process; ISBN 0-7381-1543-6; IEEE Standard No.: 1220-1998 (WWW.ieee.org), 84 pages, Unclassified

[23] Fujitsu Consulting Macroscope, Joint Development Workshop Technique (JWT): Macroscope Productivity Center Reference: (http://us.fujitsu.com/services/Consulting/Methodology/Macroscope/) Unclassified

[24] "Guide to the Project Management Body of Knowledge"; PMBOK® Guide – 2000 Edition, ISBN 1880410230, Unclassified

[25] "Value Management", PMI on-line reference: http://www.cpm-pmi.org Unclassified

A Knowledge-Based Approach to Information Fusion for the Support of Military Intelligence

Joachim Biermann
FGAN - FKIE
Neuenahrerstraße 20
53343 Wachtberg
GERMANY

Email: j.biermann@fgan.de

ABSTRACT

Intelligence cells have to process and evaluate current information to deduce timely and most reliable an appropriate picture of the battlespace. The presented approach of knowledge based information fusion is focussing on the heuristic human evaluation process. Military information processing is modelled as a context dependent and template-based heuristic reasoning process and the real task is modelled by a closed world representation. The developed method of information fusion is basically independent from a specific military scenario. It is based on the assumption that forces are organised in a structured manner and that they operate in a military reasonable and typical way. Respective rules and doctrines can be used as matching templates for knowledge based reasoning. By this approach, the analysis and fusion of incomplete and imperfect information of military reports and background knowledge can be supported substantially in an automated system.

Keywords: *Template based reasoning, heuristic reasoning, situation awareness, military intelligence, decision support*

1.0 INTRODUCTION

1.1 The Objective of Military Intelligence

In military command and control a most accurate situational awareness of the battlespace is essential prior to all decisions and activities. This is a basic military requirement which is independent from the ever changing and variety of potential conflicts. Within the global and military information environment, the challenge to future military operations will be to manage large volumes of information rapidly to portray the results in a prompt and meaningful manner. Intelligence cells have to process and evaluate incoming reports to deduce timely and most reliable an appropriate picture of the battlespace. This task includes to determine the most actual location, strength, and activities of all engaged faction forces and to deduce their likely intensions. A wide variety of information produced by the full spectrum of sensors and human sources has to be collected, filtered, processed and disseminated.

1.2 The Military Intelligence Cycle

This is done in a structured and systematic series of operations which is called the Intelligence Cycle (IC). It is the framework within which four discrete operations are conducted culminating in the distribution of the finished intelligence product. The sequence is cyclic in nature since intelligence requires constant reappraisal and updating if it is to remain current and relevant to a commander's needs. The operations are

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

discrete because, as information begins to flow, is processed and disseminated as intelligence, the operations will overlap and coincide so that they are being conducted concurrently and continuously rather than sequentially. The operations or 'phases' are shown in Figure 1 and are defined by the NATO Glossary of Terms and Definitions (AAP-6) [11] as follows:

1) *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."
2) *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."
3) *Processing*: "The production of intelligence through collation, evaluation, analysis, integration and interpretation of information and/or other intelligence."
4) *Dissemination*: "The timely conveyance of intelligence, in an appropriate form and by any suitable means, to those who need it".

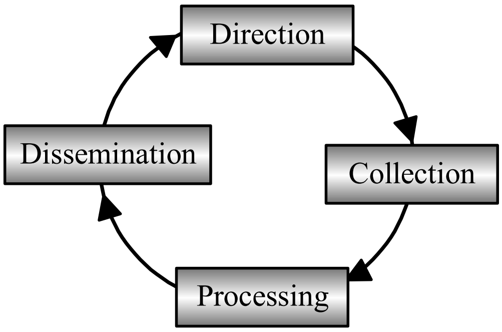


Figure 1: The Intelligence Cycle.

This paper will present an approach to give automated support to the analysis and integration steps of the "Processing" phase which in military terms is the phase where information is converted to intelligence.

1.3 The Objectives of Information Fusion in Intelligence Processing

Current information fed into intelligence cells by reports originating from all different sources of reconnaissance (HUMINT, SIGINT, IMINT, OSINT) are related mostly to particular observable elements of the mission's area of interest. Each of these reports gives a narrow view on a specific aspect of the respective local situation, for example a small group of moving vehicles, the detection of an abandoned training area, or the theft of explosives. The military commander however is interested in a comprehensive and complete picture of the overall situation, where all available information is taken into account and all relevant facts are fused to an integrated description of the situation. Such a deduced picture is supposed to reveal objects which are more complex and represent e.g. higher military hierarchy's units and/or activities on an appropriate operational level, e.g. march columns, indicating the mobilisation of higher formations and (massive) force deployment. Such high level domain objects themselves are not observable by the means of reconnaissance. They can only be deduced from the current information of the developing situation and a sufficient knowledge about the information context of the domain. Thus the sequence of processing steps of the "Processing" phase of the IC produces an increasingly aggregated and fused picture of the situation with a growing level of abstraction of the constituent objects. Starting with physical objects like tanks or a stolen vehicle they end with abstract objects like the situation of the battlespace or a threat assessment of a bomb attack. These processing steps are related to level 2 (situation assessment) and level 3 (impact assessment) fusion problems, according to the definitions of the revised Data Fusion Model, which is maintained by the Data Fusion Group of the Joint Directors of Laboratories (JDL)

Technical Panel for C^3I [16]. In level 2 an observed object is associated to a particular unit of the adversary forces (see section 3.2 Classification), hypotheses of the role of this unit in the concert of all ongoing activities are established and a most accurate description of the overall situation is developed (see section 3.3 Template Based Aggregation). This is done predominantly during the analysis and integration steps of the "Processing" phase of the IC which are the main focus in our own project [3] and of this paper. In level 3 any impact of the determined situation on own interests, especially the potential threat on own and friendly forces is assessed. This is done within the interpretation step of the "Processing" phase and not covered by this paper. Some aspects of threat assessment are considered in [10].

A comprehensive introduction to automation of data fusion, especially to tactical data fusion, can be found in [1]. In [5] a general presentation of different aspects of information fusion in different domains is given.

2.0 FUNDAMENTALS OF THE HEURISTIC INTELLIGENCE PROCESSING

A main intellectual and cognitive part in the production of military intelligence is done by comparing the partial information aspects given by the reported low level objects with the so far deduced perceived picture of the current situation, relating them to the most likely aspects, aggregate them to high level situation objects, and finally integrate these into the picture of the situation. As a metaphor this information fusion problem is compared by Antony [1] to the jigsaw puzzle problem. The very difference to the game situation however is, that, as nobody knows the real battlespace situation, the intelligence cell has to deduce the situation picture from the pieces without knowing a picture of the real solution as it is available in a game. To overcome this lack of a guiding true solution the processing steps are based on knowledge of the background of the situation which is called basic intelligence. The AAP-6 [11] definition of basic intelligence is: “Intelligence, on any subject, which may be used as reference material for planning and as a basis for processing subsequent information or intelligence”. All information about engaged factions describing their structure and strength, their equipment and their initial deployment is of interest. It is continually updated in peace and in the course of operations. The principal use of basic intelligence is to set the scene at the outset of operations and to meet intelligence requirements dealing with unchanging facts such as battlespace terrain and climate which may be raised in answer to new requirements in the course of an operation [12].

2.1 Basic Assumptions

Up to now the production of military intelligence is based on the experience that all hostile forces, groups or elements are organised and have an organisational structure, operate in mission co-ordinated and, with respect to their objectives, in a determined and reasonable way. Military leaders understand war as a highly complex system of separable activities, as it is shown in Figure 2 [13].

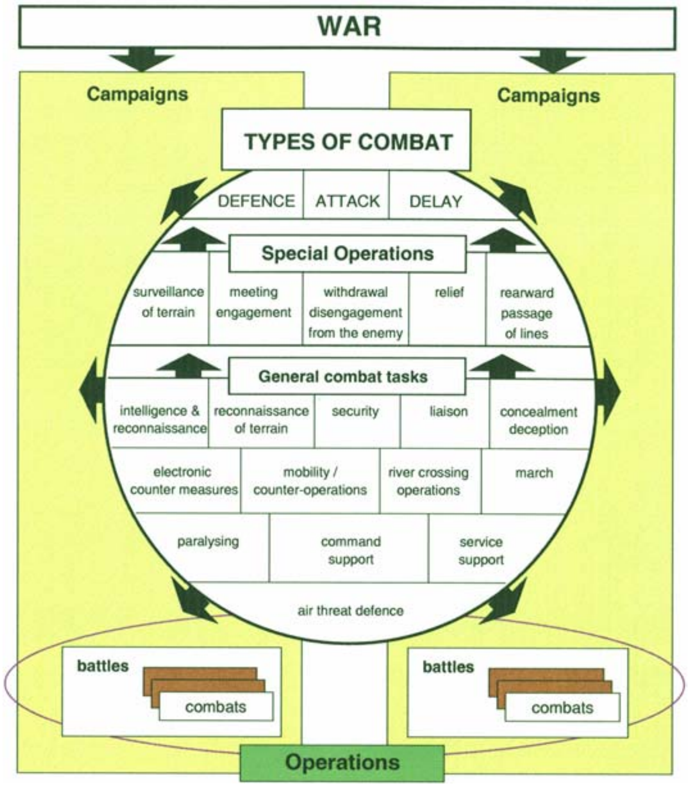


Figure 2: War as a Structured System of Separable Activities.

Individual missions and single activities will be planned based on some kind of doctrines, modus operandi and rules reflecting their structure and their organisation, so they can be distinguished from each others and recognised by significant information. This assumption has been confirmed by experience not only for conflicts with regular military forces but also is applicable for para-military factions, certain criminal activities [1], [17]. Even in Counter Terrorist Operations the search for anomalies in patterns of standard behaviour as significant facts for an identification of terrorists can be interpreted as a template-based approach [14].

Doctrines and rules proved to be a good basis for military success but they also give a good basis for an approach for information processing and situation awareness. Intelligence officers practise by default a method of heuristic reasoning which relies on their knowledge about standards and rules of the adversary and the assumption that the opposing forces will act according to these regulations and constraints. It is common military experience and expert knowledge, that the production of intelligence can be based on putting current information into such an information context. Information dominance is a battle winning factor, thus it is a main concern of every intelligence activity to build up sufficient knowledge about the adversary. The lack of basic intelligence or background knowledge does not prove the absence of e.g. an organisational structure or principles of behaviour, it only proves the urgent need of reconnaissance. A huge variety of information of every kind is needed and a typical list of basic intelligence categories will include:

- organisational structure and equipment,
- operational doctrines and activity patterns,
- political parties and political structures (coalitions),
- sociologic, religious, and ethnic information,

- local places and habits,
- geography, infrastructure, climate, ...

These information aspects span the information context space in which information processing is done in intelligence cells. The relevance of the individual aspects may be dependent from the specific scenario and situation and the importance of a certain category of information is difficult to assess as long as no sufficient understanding of the overall situation has been established. Information concerning the organisation and strength of adversary forces always will be of fundamental interest. The information about the historical, religious and social situation of the population gained importance during the last years in the course of 'Operation Other Than War' (OOTW) missions, e.g. during the Balkan conflict [17]. Depending on the evolving situation and the accordingly changing own information and intelligence requirements the list of necessary categories may be expanded and the importance of its elements may change. The real information fusion problem in principal is an open context problem, but in a real mission the full information context never will be known completely.

The cognitive processes of the intelligence officers about the supposed default behaviour of the enemy are typically based on a limited number of intelligence aspects thus defining a closed world model of the problem space and they constitute an experience based default reasoning process. Default behaviour of an adversary allows to develop generic doctrinal templates of his activities which have to be adapted to real mission environment and to define significant information indicating a certain military activity which is related to this behaviour. A major part of the information processing in intelligence cells is done based on this heuristic method which is the fundamental idea of template based information fusion. (For a basic discussion about doctrinal templating see also Antony [1].) The template-based default reasoning approach is the basis to develop automated evaluation algorithms to support the human intelligence processes.

The matrix in Table 1 shows the functional template of a march column from a classical scenario [3]. It describes in general the components of a march column by their military hierarchy and function. It is used in our experimental system, beside templates on the spatial structure and dynamic features, as a basis for the aggregation of single reports about moving troops into march columns.

Table 1: Functional Template of a March Column

	Head of column	Inner column components	End of column
Type of troops	reconnaissance security units	combat units combat support units air defence and security	logistics air defence and security
Branch	armoured reconnaissance antitank defence	Armour, mechanised infantry , non armoured infantry, multi rocket launcher, artillery, engineer AA, SAM	transportation & maintenance AA, SAM
Command level	≤ platoon or company	≤ battalion	≤ platoon, company or battery
Type of vehicle	APC, ARV	MBT, APC, AAA	VEH, AAA

(AA Anti Air; AAA Armoured Anti Air; APC Armoured Personnel Carrier; ARV Armoured Reconnaissance Vehicle; MBT Main Battle Tank; SAM Surface Air Missile; VEH Vehicle)

More complex template structures in the context of low intensity conflicts have been used in our recent analysis on information fusion for threat assessment [9] [10]. The underlying information environment was expanded significantly and structured into ability and threat related aspects.

2.2 Quality of the Current Information

Topical information about the situation given by reports may be imperfect in various respects and it is typically incomplete, imprecise, uncertain, and vague and the sequence of incoming reports will not necessarily be in a chronological order. (For a short definition of the different aspects of defectiveness see [5]) It is a main benefit of the template-based approach that some deficiencies from incompleteness and imprecision of the current information can be compensated partly. This is caused by the general and more qualitative nature of the template-based information fusion focussing on the deduction of more abstract high level objects of the problem domain. A lot of calculations concerning the correlation of two low level objects cannot be carried out properly because of the imperfectness of the information or would only produce less usable ambiguous results. In the context of the template-based aggregation approach they could be avoided. To give an example, it is not necessary to show that two low level elements observed at different times at different places are identical objects. The military interest is to know whether or not they belong to the same higher level element or activity, because this will answer the question about location, strength and behaviour of the adversary. Target correlation is not a the goal of level 2 fusion problems e.g. at a brigade level, but object aggregation and role identification. Template-based aggregation is a suitable method to cope with this problem.

3.0 KNOWLEDGE-BASED PROCESSING METHODS

The principle feasibility of automated information fusion for the support of intelligence processing has been proven in a former project which was based on a classical cold war scenario [4]. Our actual approach of a knowledge based information fusion is focussing on the heuristic basics of the human cognitive processing of reports in an army intelligence cell and broadening the view to actual military missions. Military information fusion is modelled as a context-dependent and template-based heuristic reasoning process. The aspects of spatial and temporal reasoning which are included in our ongoing research are covered in [10].

The hereafter presented knowledge-based methods can be interpreted as parts of the IC processing phase as it has been analysed and modelled by NATO RTO IST-015/ TASK GROUP 004 ON INFORMATION FUSION (2000 - 2002) [15] and may particularly be related to the Collation and Analysis & Integration steps, as they are shown in Figure 3.

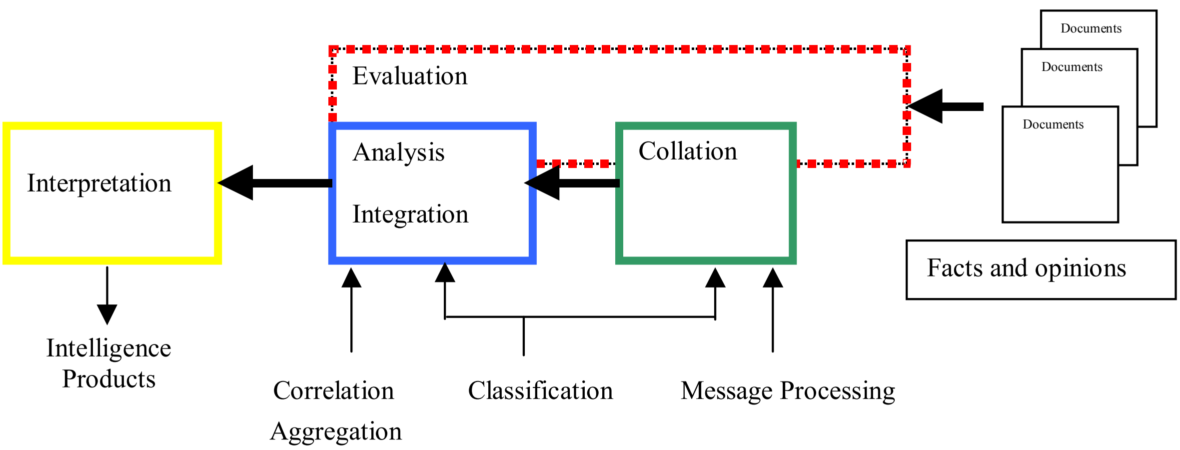


Figure 3: Intelligence Cycle Processing Phase – Representation of Processes [15].

3.1 Context Dependent Collation

During the collation step of the processing phase of the IC, a first semantic analysis is performed associating any new report to an appropriate category of intelligence. The number and definition of the categories depends on the type of operation which is conducted. These categories are deduced from the commanders intelligence requirements and define the main situation aspects and the information context which has to be covered to be able to give a description of the situation that sufficiently reflects the abilities and intentions of the opposite parts and to make an assessment. The association of a single report with a category of information is just a loose correlation. No further semantic interpretation is done at this step. Formatted reports do name a main category of information by the use of key words. Structured text reports also support this categorisation. For free text reports there has to be a pre-processing to exploit the semantic content. It is a premise of our approach that a semantic understanding and categorisation of the current information is given prior to our processing, as it is e.g. possible with formatted reports. If this premise is not true, a substantial automated support to an object oriented processing and information fusion is doubtful.

3.2 Classification

Classification is one of the fundamental tasks of an intelligence cell. There is a strong interest to get to know which unit of the adversary forces was engaged when an activity or event was observed, answering the initial question: "Who or what is it?" Thus a basic functionality in information processing is the correlation of observed situation/domain elements to a domain model from basic intelligence. For a more detailed definition of classification see [15]. A generic solution method for this matching problem was developed, which is independent from the special scenario [7]. The algorithm takes e.g. the known structure of the opposing forces and their list of equipment as initial parameters. Thus the precondition of the method is the same as for the real processing. Without any knowledge about the opposite organisation no classification is possible. It is proven that the computation based on this method, which is a logic two-rule system, is

- complete, so every solution to a matching problem is calculated,
- correct, so every calculated solution solves the matching problem,
- finite, that is the computation stops in finite time.

The combinatorial complexity problems of the pure theoretic algorithm may be reduced by the use of constraints, which can be dependent or independent from the domain. They are expressed as constraining rules and enhance the performance of an algorithm based on the abstract method and the quality of the results. A scenario independent rule is, to start by matching first the most significant elements of an observation. This will speed up the calculation and has no influence on the result as the method always finds every solution. But it is a scenario dependent and perhaps subjective decision, to define a threshold for significance. Heuristic or doctrinal knowledge about deployment rules, e.g. that certain units typically operate jointly but they will not intermix their respective areas of activity, can be used as domain specific constraint. Such rules will reduce the number of hypotheses about the classification of one observation and by this the complexity of following processing steps, and the quality of the result will be enhanced. Figure 4 graphically shows the schema of the classification by an example.

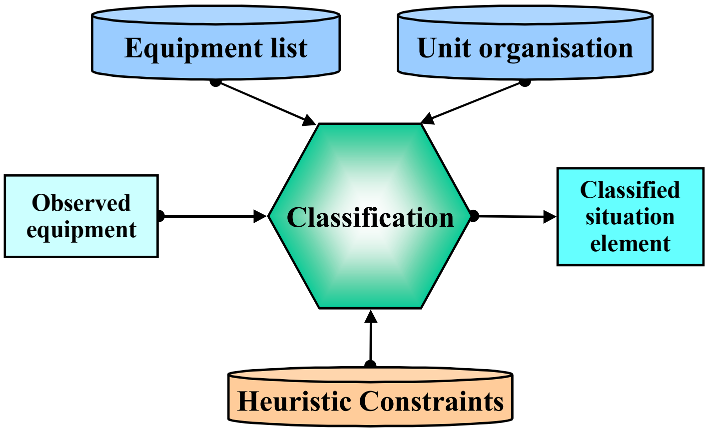


Figure 4: Schema of Classification.

3.3 Template-Based Aggregation

The further analysis of the information of the classified situation elements and their fusion to higher level situation objects is done in three steps. Starting with the generation of so-called aggregation-objects, the generation of aggregation-hypotheses follows and finally fused solutions are determined.

The premise for human and automated aggregation is that there is sufficient knowledge about high level situation objects or activities, so that suitable templates can be developed, describing the structure and standard behaviour of the engaged factions in the context of the current scenario.

The main ideas of this approach will be presented in the following part of the paper. A detailed description of these processing steps is given in [6].

3.3.1 Generation of Aggregation-Objects

Depending on the category of information, to which a lower level situation element is associated, it is compared to the templates of higher level objects to test whether it fits into any part of these general descriptions. The structure and the attributes of the low level object have to match at least parts of those of the higher level object. This information processing is not related to other low level objects of the perceived situation. It is an analysis of the individual low level element in the context of knowledge given by the basic intelligence. If this test is passed the hypothetical relationship is kept as a new attribute of the specific low level object which is then called an ‘aggregation-object’. Attributes of the low level object may be related to different templates of the high level object and/or, being less significant, they may fit to different parts of a single template. So the result of the comparison may not be unique. All resulting hypotheses will be kept and taken as input for the following aggregation step.

For example, information about moving vehicles will be compared with the functional template of a march column shown in Table 1. If the type of the vehicles is given then they can be classified in respect to the type of troops they may belong to (which might be ambiguous). This will then allow to compare to the functional march column template to deduce to which part of a march columns the vehicle group may belong to. Depending on the type of troops this may be ambiguous so that different hypotheses have to be kept. Additionally a known spatial extension of the observed moving vehicle group or the minimal extension of the induced type of troops may constrain the matching test with the functional template.

3.3.2 Aggregating Hypotheses

Aggregation-hypotheses which are generated from aggregation-objects represent higher level abstraction objects of the perceived situation which may be aggregated on their part in a multiple level aggregation procedure. A new aggregation-object either can be integrated into an existing aggregation-hypothesis, thus updating it and confirming its evidence from a military point of view, or if it does not fit to any of the existing hypotheses, a new one is generated.

Figure 5 shows as an example the graph of all aggregation-hypotheses build out of 10 aggregation-objects ao_i , i = 1(1)10. The vertices represent the single aggregation-objects ao_i, an edge between two vertices indicates that the two aggregation-objects together match the underlying template. In Figure 5 [ao_1 - ao_7] and [ao_7 - ao_8] both are correct level-1 aggregation-hypotheses enclosing two aggregation-objects.

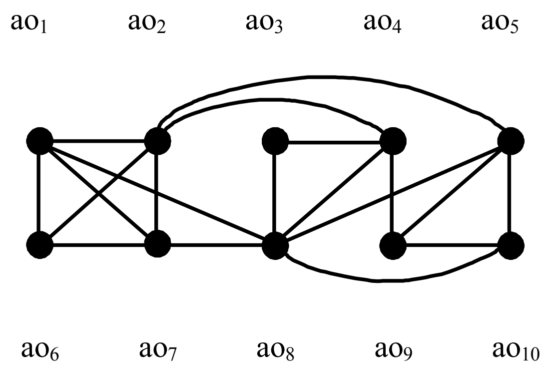


Figure 5: Graph of aggregation-hypotheses.

Several aggregation-objects can be aggregated into one hypotheses only if they all simultaneously fit the underlying template which has to be tested by a specific template dependent function. As in Figure 5 [ao_1 - ao_8] also is a correct level-1 aggregation-hypothesis the triple [ao_1 - ao_7 - ao_8] is a correct level-2 hypothesis. The matching relation of aggregation-objects and their respective template is not necessarily transitive, e. g. in the case of march columns on a road net. Therefore the integration of a new aggregation-object into an existing aggregation-hypothesis may require to test the power set of all combinations of the component objects of the old aggregation-hypothesis and the new aggregation-object. As in the worst case the number of hypotheses doubles for the integration of each new aggregation-object special effort is taken on the internal representation of the set of hypotheses and the updating algorithm. An efficient recursive algorithm to create new hypotheses and to update the set of all valid hypotheses is given in [6].

3.3.3 The Fusion Step

The set of all formal correct aggregation-hypotheses will contain contradictory elements if some aggregation-objects have been aggregated concurrently into different hypotheses which cannot be true all at the same time. A military situation determination has to give an unambiguous picture of the battlespace, which has to be based on a consistent set of hypotheses concerning the different situation aspects respectively. The main objective of the fusion step is to find all consistent sets of aggregation-hypotheses.

Definition: Fused Solution, Minimal Fused Solution

For a given finite set AO = { ao_i | $i \in I$} of aggregation-objects and the corresponding set AH(AO) of all formal correct aggregation-hypotheses a Fused Solution FS is a subset of AH which is a partition of AO. FS is called Minimal, if there exists no different subset $FS^* \subset AH$ which is a Fused Solution and $((FS^* \subset FS) \wedge (FS^* \neq FS))$ is true.

According to this definition a Fused Solution FS has the following properties:

1) Each aggregation-object is contained in one of the aggregation-hypotheses in FS
2) All aggregation-hypotheses in FS mutually do not share any element (aggregation-object)

Which aggregation-hypotheses finally will belong to the Fused Solution may depend on an assessment which is based on additional constraints such as operational effectiveness. In the case of a march column this does mean, that it is reasonable to move along the shortest path and not to split up into many different march columns. Heuristics of this kind are used to constrain the calculation of a fused solution. In many cases a minimum constraint corresponds to military heuristics about effectiveness and efficiency of the deployment of forces which hold under the general assumption of reasonable operating adversary factions. Even if no obvious inherent minimum constraint is given, the calculation of a minimum number of hypotheses which explains the information situation is a valuable hint to the minimum number of high level situation elements which are at least necessary to give rise to the perceived situation picture. By this, a lower limit of the number of high level situation elements is given.

The determination of a Minimal Fusion Solution can be formulated as an abstract graph problem which is not dependent from the application domain and the originating intelligence process [8]. Starting from the graph representation of all formal correct aggregation-hypotheses, as it is given e.g. in Figure 5, the algorithm has to find all minimal clique partitions of the graph. The requirement for cliques corresponds to the requirement that in a correct aggregation-hypothesis all aggregation-objects mutually match the underlying template, which is represented by complete sub-graphs. For the example graph in Figure 5 the unique minimal solution is given by Figure 6.

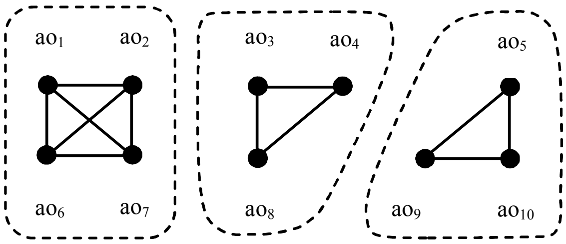


Figure 6: Graph of a Minimal Fusion Solution.

The three cliques $C_1 = \{ ao_1, ao_2, ao_6, ao_7 \}$, $C_2 = \{ ao_3, ao_4, ao_8 \}$, and $C_3 = \{ ao_5, ao_9, ao_{10} \}$ define a partition of the set AH of all formal correct aggregation-hypotheses. In general the minimal clique partition problem has no unique solution. As a result of the fusion step we will get the set of all Minimal Fused Solutions. In [8] it is proven that the developed algorithm, which is based on a Branch-and-Bound method, is total and correct, which means that every solution of the minimal clique partition problem is determined and every determined solution is a Minimal Fused Solution.

The whole three-step process of template-based aggregation and fusion is intended to run continuously and to develop iteratively for each new aggregation-object a new set of aggregation-hypotheses and a new set of Minimal Fused Solutions.

3.4 Hypotheses Management

The set of all aggregation-hypotheses which is calculated in every iteration step contains every formal correct solution of the matching problem. In an operational fusion system the question of maintaining the set of all hypotheses may become important, because system performance may decrease rapidly if the number of hypotheses increases by combinatorial effects. Pruning of hypotheses could be done either interactively under the complete control of the user, or automated, or as a combination of both of them. Some strategies for the pruning of hypotheses are discussed in [6]. The central idea of the pruning policy is to prefer those hypotheses which are part of the set of Minimal Fused Solutions for several iteration steps. In a military interpretation these hypotheses are confirmed several times and thus have gained stronger evidence. So, basically this is again a domain specific heuristic which constrains the solution space.

4.0 CONCLUSION

Just as in the case of human intelligence processing also automated knowledge-based information fusion depends on the availability of a minimum of basic intelligence in terms of knowledge about structure, equipment and default behaviour of the adversary factions. Comparing classical high intensity conflicts, civil war conflicts, crisis reaction operations, terrorist activities or organised criminal activities there are great differences in respect to the type and amount of information and intelligence which is available. Especially at the beginning of a new conflict even the information aspects and according the kind of templates which are of relevance for a situation determination and assessment may hardly be known. Furthermore the production of intelligence used to be and will be based on a varying but limited number of categories of information of basic intelligence which are supposed to be sufficient to understand the situation. As a consequence a future supporting automated fusion system has to be an interactive system with the human operator in the loop. An interactive interface to a fusion system should allow to introduce new or to modify existing aggregation templates to tune the system to the requirements of the specific scenario. Inappropriate templates or constraints which are caused by weak initial basic knowledge have to be adapted to the real situation, which was proven to be feasible in our ongoing research project [3]. The intelligence officer as the system user will be responsible for this tuning and he is responsible for the selection of a specific hypothesis from a set of formal correct Minimal Fusion Solutions which are offered to him by the system. The presented knowledge-based method for an automated fusion system is intended to give support for human decisions not to substitute the intelligence officer.

Intelligence processing of information which is related to standard and default behaviour of the observed factions can be supported by constraint template-based matching methods. This approach is basically not dependent from the special scenario. Procedures based on this approach will determine all formal correct results without any subjective prejudice and probably in a time which is much shorter as for a human analyst. This will free time for the intelligence officer to do his proper task which are the more sophisticated processing steps of situation interpretation and impact assessment. The interactive adaptation of procedures based on the template-based method is one of the challenging problems to develop flexible information fusion systems.

REFERENCES

[1] D. S. Alberts et al., *Understanding Information Age Warfare*, DoD Command and Control Research Program, August 2001.

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

[3] J. Biermann, *Das Experimentalsystem HADES - Methodische Ansätze zur wissensbasierten Fusion strukturierter Informationen*, FKIE-Bericht Nr. 21, Wachtberg, 2000.

[4] J. Biermann, *HADES - A Knowledge-Based System for Message Interpretation and Situation Determination*, in: A. Pasqual del Pobil, J. Mira, M. Ali (Eds.) Tasks and Methods in Applied Artificial Intelligence, 11th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA-98-AIE, Lecture Notes in Artificial Intelligence 1416, Springer Verlag Berlin, 1998.

[5] I. Bloch, A. Hunter, *Fusion: General Concepts and Characteristics,* International Journal of Intelligent Systems, Vol. 16, No. 3, pp. 1107-1134, 2001.

[6] T. Hoffmann, *Musterbasierte Informationsfusion in HADES*, FKIE-Report Nr. 7, Wachtberg, 1999.

[7] T. Hoffmann, *Ein Reduktionssystem zur Klassifikation von Feindbeobachtungsmeldungen*, FFM Bericht Nr. 481, Wachtberg, 1997.

[8] C. Karg, *Ein Verfahren zur Synthese von Konvoibewegungen*, FKIE-Report Nr. 10, Wachtberg, 1999.

[9] F. P. Lorenz, J. Biermann, *Knowledge-Based Fusion of Formets: Discussion of an Example*, Proceedings of the Fifth Int. Conf. on Information Fusion, Fusion'2002, Annapolis, MD, pp. 374-379, 2002.

[10] F. P. Lorenz, J. Biermann, *A Man-in-the-Loop Support Concept for Military Ambush Threat Analysis Based on Reconnaissance Reports*, submitted to: NATO RTO Symposium on Military Data and Information Fusion IST-040 / RSY-012, Prague, Czech Republic, 20.-22.10.2003.

[11] NATO Standardization Agency (NSA), *AAP-6(2002) NATO Glossary of Terms and Definitions*, http://www.nato.int/docu/stanag/aap006/aap6.htm, 2002.

[12] NATO, *AJP 2.0 Allied Joint Intelligence Counter Intelligence and Security Doctrine*, NATO/PfP Unclassified, Ratification Draft 2, 2002.

[13] R. Oestmann, *Dazu befehle ich ...!*, Handbuch für militärische Führer, 3. Auflage. Walhalla Fachverlag, Regensburg Berlin, 2000.

[14] P. Lefever, *Beyond Interoperability*: Part 2, in: D. Potts (ed.), Occasional Paper No 45 - *The Big Issue: Command and Combat in the Information Age*, The Strategic and Combat Studies Institute, Shrivenham March 2002.

[15] D. Thibault et al. (eds.), RTO Task Group Report, NATO IST-015/ Task Group 004 On Information Fusion, Final Activity Report 2000 - 2002.

[16] A.N. Steinberg, C.L. Bowman, F.E. White*, Revision to JDL Data Fusion Model*, Joint NATO/IRIS Conference, Quebec, October 1998.

[17] L Wentz (ed.), Lessons from Bosnia: The IFOR Experiment, DoD Command and Control Research Program CCRP, 1997.

Information Fusion and Extraction Priorities for Australia's Information Capability

Martin Oxenham[1], John Percival[1], Richard Price[2] and Dale Lambert[2]
[1]Intelligence, Surveillance & Reconnaissance Division
[2]Command & Control Division
Defence Science and Technology Organisation
West Avenue, Edinburgh
South Australia 5111
AUSTRALIA

{martin.oxenham, john.percival, richard.price, dale.lambert}@dsto.defence.gov.au

ABSTRACT

The Commonwealth of Australia's strategic policy on defence recognises the need for Australia to further develop and enhance its information capability. While this is mainly a reflection of the uptake of information technology by the military (the Revolution in Military Affairs), in more recent times it has also been in response to the global threats of terrorism and the utilisation of weapons of mass destruction or long-range ballistic missiles by rogue states. The areas which have been identified as being of primary importance for fostering Australia's emerging information capability are intelligence and surveillance capabilities, communications, information warfare, command and headquarters systems, and logistics and business applications. To meet the challenges that the development and enhancement of this information capability raise, advanced information processing techniques for fusing and extracting data or information are required. In this paper, a holistic model for integrating the relevant technologies of data fusion and data mining is proposed and several of the current information fusion and extraction initiatives at Australia's Defence Science and Technology Organisation supporting intelligence, surveillance, and command and control are outlined.

1.0 INTRODUCTION

Australia's strategic policy, as laid down in its Defence White Paper [1], identifies the need for Australia to further develop and enhance its information capability. It states that "Effective use of information is at the heart of Australia's defence capability. In part this is a reflection of a worldwide trend, as information technology is transforming the ways in which armed forces operate at every level. All forms of capability are being transformed by the innovative use of information technology. But this trend is more significant to Australia than to many other countries. Our strategic circumstances mean that innovative applications of different aspects of information technology offer Australia unique advantages." [1, p. 94, para. 8.78]

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

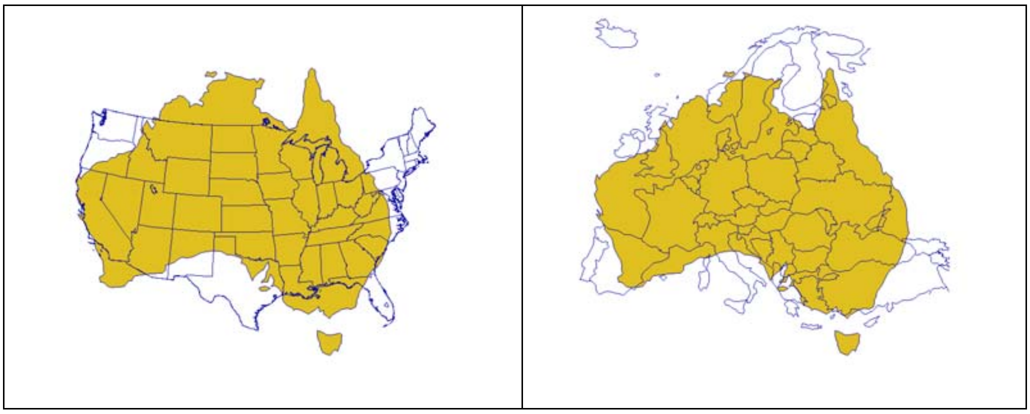


Figure 1: Geographic Comparison of Australia to the USA, the UK and Europe[1].

So what are Australia's strategic circumstances? In terms of geography, Australia is an island continent that ranks sixth in size on a worldwide scale. It covers an area comparable to both that of the USA and of Europe, as depicted in Fig. 1. However, in terms of population, Australia may be regarded as a small nation. Australia has a current population of only 19.9 million, which is 14 times less than that of the USA, 30 times less than that of the whole of Europe, and between 2 and 4 times less than that of the UK, and some of Europe's larger nations including France, Germany, Italy and Spain. The percentage of Australia's population in the permanent defence force is comparable to the percentages for the other countries listed above, but when the defence reserves are factored in, Australia is way below par (refer to Table 1). In terms of actual numbers, Australia's combined defence force is substantially smaller than the average crowd at some premier Australian sporting events such as the Australian Football League's grand final which typically attracts close to 100,000 spectators.

Table 1: Australia compared to other nations based on the sizes of their populations and defence forces in July 2002 [3].

Nation	Population	Size of the Permanent Defence Force	Size of the Defence Force Reserves	% of the Population in the Permanent Defence Force	% of the Population in the Combined Defence Force
Australia	19,546,792	55,200	27,730	0.3	0.4
USA	280,562,489	1,371,500	1,303,300	0.5	1.0
UK	59,778,002	212,400	191,000	0.4	0.7
France	59,765,983	317,300	419,000	0.5	1.2
Germany	83,251,851	332,800	344,000	0.4	0.8
Italy	57,715,625	265,000	72,000	0.5	0.6
Spain	40,077,100	186,000	447,900	0.5	1.6

The majority of the Australian population resides on the eastern seaboard, which is also problematic from a defence perspective, since the most vulnerable stretch of Australia's vast coastline and territorial waters is that which lies to the north. This presents major challenges to the Australian Defence Force and the various Australian intelligence agencies which jointly serve to protect Australia's sovereign territory and national interests. Current defence imperatives for the nation include border protection against smuggling (of people, drugs and other contraband) and illegal fishing, peacekeeping efforts such as in East Timor and increasingly homeland defence in the wake of terrorist attacks by extremist groups, and the threat of the

[1] The images have been produced using Albers' equal-area conic projection [2, p.49].

utilisation of long-range ballistic missiles and weapons of mass destruction by rogue states [4]. Accordingly, "maintaining first-rate intelligence capabilities, developing a comprehensive surveillance system providing continuous coverage of our extended air and sea approaches, developing an integrated command system covering operations at all levels and in all environments, and maximising the efficiency of our logistics systems and management processes by cost-effective investment in information technology applications" [1, p. 95, para. 8.83] are specific goals of Australia's emerging information capability. To achieve these goals, innovative methods for automatically extracting and fusing information from voluminous, disparate and distributed data sources are required to assist defence force personnel and intelligence analysts to perform their duties more efficiently and effectively. The purpose of this paper is to outline some of the initiatives at Australia's Defence Science and Technology Organisation (DSTO)[2] which are striving to meet these goals.

The remainder of the paper is structured as follows. In Section 2, the concepts and terminology used in the paper are presented, and a holistic model for integrating data fusion and data mining is proposed. In Sections 3-5, several of the pertinent information fusion and extraction initiatives at the DSTO in the areas of wide area surveillance, intelligence analysis, and command and control are described. Finally, in Section 6, some concluding remarks are made.

2.0 DATA FUSION AND DATA MINING

The purpose of an information capability for the defence and intelligence communities is to support the establishment and enhancement of the operators' and analysts' situation awareness, and ultimately to facilitate their or others' timely and effective decision making. To assist in the processing of the large volumes of data and information required for meeting these goals, the key enabling technologies of data fusion and data mining have emerged over the last decade or so.

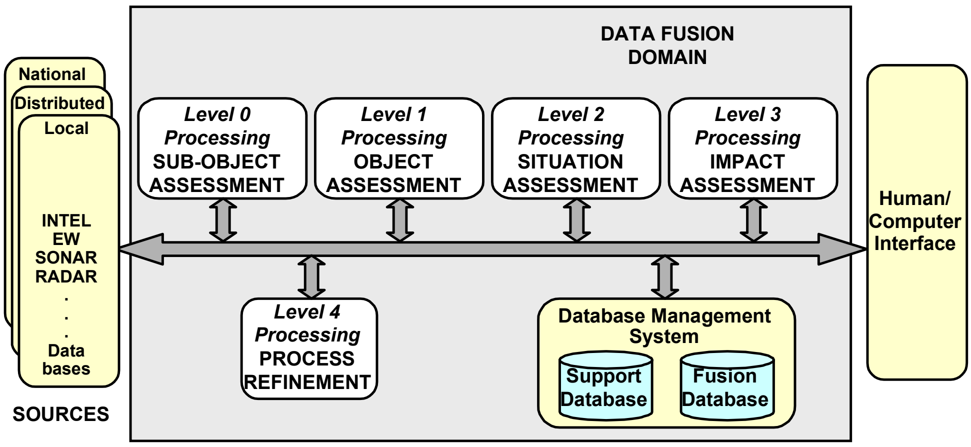


Figure 2: The United States Joint Director of Laboratories (JDL) Model of Data Fusion [8].

Data fusion is commonly described as the process of combining data or information to estimate or predict entity states [5, 6]. The functional model of data fusion introduced by the United States Joint Directors of Laboratories in 1987 [7], and revised in 1998 [8], has gained widespread acceptance within the data fusion community. It represents data fusion in terms of the five functional levels of *signal refinement* (Level 0), *object refinement* (Level 1), *situation refinement* (Level 2), *impact refinement* (Level 3) and *process*

[2] The DSTO constitutes part of the civilian sector of the Australian Defence Organisation. Its role within Defence is to ensure the expert, impartial and innovative application of science and technology to the defence of Australia and its national interests. It accomplishes this via a combination of means including prototype capability development, project support for the acquisition of defence materiel, and long range research.

refinement (Level 4) (refer to Fig. 2). It is noted that Level 1 is also often referred to as *sensor fusion* and Levels 2 and 3 jointly as *information fusion*.

Data mining, which subsumes machine learning and data visualisation, may be regarded as the process of determining patterns, trends, relationships and associations in large data sets that are not explicitly mentioned in the raw data. Thus, in contrast to data fusion as defined above, which is in broad terms a "supervised" process relying on models (eg target motion models, ontologies, inference rules, plans), data mining is fundamentally an "unsupervised" process which discovers models. Furthermore, unlike data fusion, no model of data mining has been developed to the knowledge of the authors.

Despite the distinction between data fusion and data mining that these definitions lead to, there is a strong interdependence between the two processes. Little research appears to have focussed on their integration however. One exception to this is the integrated model that Waltz proposes in [9], which is in keeping with the principles of knowledge discovery in databases (KDD)[3] [10, 11]. In his model, depicted in Fig. 3, sensor data is fed into both the data fusion and data mining modules for separate processing. While the data are analysed by the data fusion module to support decision making in the application domain, they are also passed to the data mining module for data warehousing. Selected data from the warehouse are then mined to hypothesise models that classify the entities represented by the data or the relationships between those entities; these hypothesised models then pass through a validation stage, and if validated are then evaluated and interpreted. Finally, the discovered models are passed to Level 0, 1 or 2 of the fusion model as appropriate to enhance the overall data fusion process.

While this is a fair model of integrated data fusion and data mining, it has several shortcomings. First, it is not apparent why it does not entertain the possibility of discovering models that may feed into Level 3 of the data fusion process. Second, it fails to allow established models in the data fusion process to assist in the data mining process. Finally, it has not attempted to view the integration of data fusion and data mining holistically, but instead has simply coupled the two processes together. In the remainder of this section, a strategic framework is outlined for dealing with the integration problem from a holistic perspective.

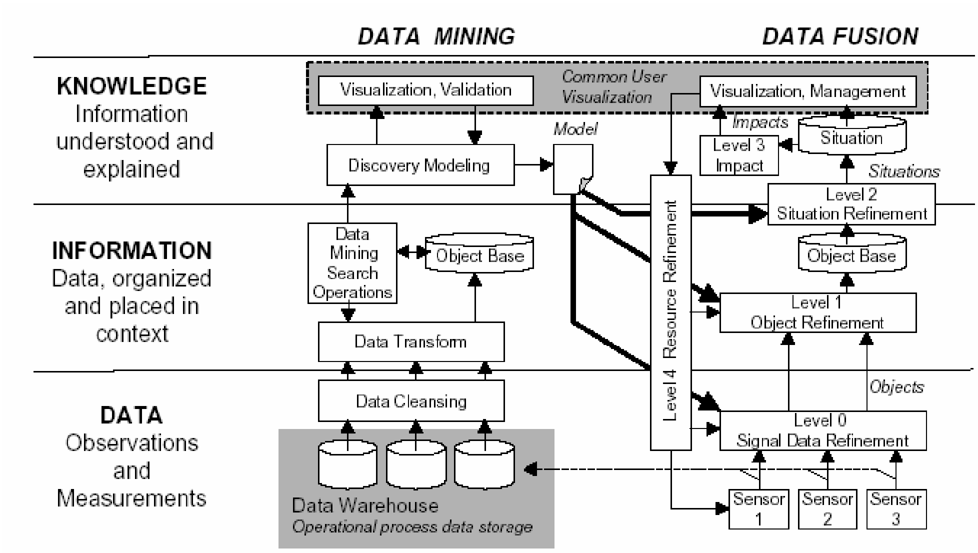


Figure 3: Waltz' Integrated Model of Data Mining and Data Fusion [9].

[3] The process of KDD involves a number of steps including data warehousing, target data selection, cleaning and pre-processing, transformation and reduction (summarisation), data mining, model selection (or combination), evaluation and interpretation, and finally consolidation and use of the extracted "knowledge" [11].

In contrast to the definition of data fusion presented above, Lambert [12] defines data fusion as the process of utilising one or more data sources over time to assemble a representation of aspects of interest in an environment. He also offers a deconstruction of the US JDL model of data fusion described earlier in the section, in which Level 0 is included in Level 1, and Level 4 is absorbed into each of Levels 1, 2 and 3 (refer to Fig. 4). Under this deconstructed JDL account, denoted hereafter by λJDL, Level 1 is about the identification of objects from their properties, Level 2 is about the identification of relations between these objects, and Level 3 is about the identification of the effects of these relationships between the objects. Articulated in more detail [12]:

- *Object refinement* is the process of utilising one or more data sources over time to assemble a representation of objects of interest in an environment. An *object assessment* is a stored representation of objects obtained through object refinement;
- *Situation refinement* is the process of utilising one or more data sources over time to assemble a representation of relations between objects of interest in an environment. A *situation assessment* is a stored representation of relations between objects obtained through situation refinement; and
- *Impact refinement* is the process of utilising one or more data sources over time to assemble a representation of effects of situations in an environment, relative to one's intentions. An *impact assessment* is a stored representation of effects of situations obtained through impact refinement.

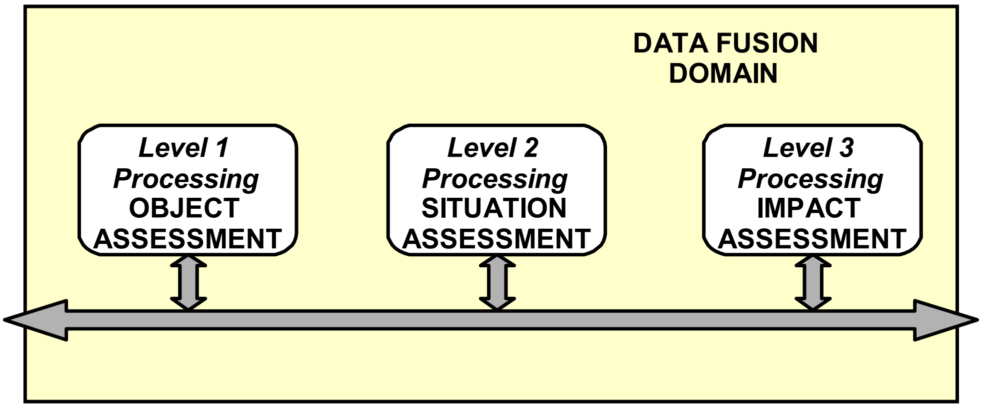


Figure 4: The λJDL Model of Data Fusion [12].

Fundamentally, the λJDL model of data fusion is not at odds with the JDL model, but proves to be far more convenient and flexible; this point will be expanded upon later in Section 5. Perhaps more interestingly, the λJDL description of the refinement processes at each level not only allows for the combination of data or information as in the JDL model, but the notion of "utilising one or more data sources over time to assemble a representation" also accommodates the use of the data or information at each level for developing or enhancing models. Thus, under the λJDL model of data fusion, data mining may be regarded as an integrated sub-process. There are a number of benefits in adopting this view. First, the data mining process acquires more structure through the inheritance of the notion of "level" from the data fusion model. Second, while the traditional data fusion and data mining processes may be distinguished within the λJDL model if required, there is no compulsion to do so. As such, more effort may be directed towards the appropriate utilisation of the data or information to meet the decision-maker's needs without having to be concerned about whether the processing is "fusion or mining". Third, analogous concepts shared by traditional data fusion and data mining such as "data alignment (data fusion)" and "data cleansing and transformation (data mining)", "multi-sensor tracking and track fusion (data fusion)" and "multi-source information retrieval (data mining)", and "information extraction (both)" for example may be either legitimately identified or classified as specific instances of general data fusion concepts. Finally, challenges common to both processes such as knowledge representation, and data or information visualisation for example may be dealt with in unison. For all of these reasons, this is the viewpoint that is taken in this paper.

3.0 TARGET BEHAVIOUR EXTRACTION FOR WIDE AREA SURVEILLANCE

As highlighted in the introduction, wide area surveillance of the sea-air gap to the north of Australia plays a key role in Australia's information capability strategy. Wide area surveillance in the Australian context is currently afforded by a suite of surveillance assets, primarily 2- and 3-dimensional ground-based microwave radars, and three over-the-horizon radars (two of which constitute the Jindalee Operational Radar Network (JORN) that became operational several months ago); refer to Fig. 5 for a stylised depiction of the radar coverage regions. These will be augmented in the coming years by an early warning and control capability in the form of a fleet of Wedgetail AEW&C platforms, and a Link 16 capability. The wide area surveillance picture (WASP), formed through the fusion of data from these sensors and additional sources, supports a number of C^4ISR activities[4] namely real-time surveillance, fighter control and situation and threat refinement at the tactical level, and mission planning and intelligence gathering at the theatre level. Typically, compilation of the WASP and these subsequent C^4ISR activities are performed manually by air defence operators. To allow the operators to focus more on decision making, and less on data and information integration, it is desirable to automate some of the data and information processing required for these activities, providing it does not lead to a loss of their situation awareness. In particular, to support automated situation and threat refinement and intelligence gathering, techniques for automatically interpreting current activities and predicting future events are required. In this section, one such technique is described which takes the approach of exploiting the contextual information in the surveillance region to "extract" the behaviours of the detected air platforms.

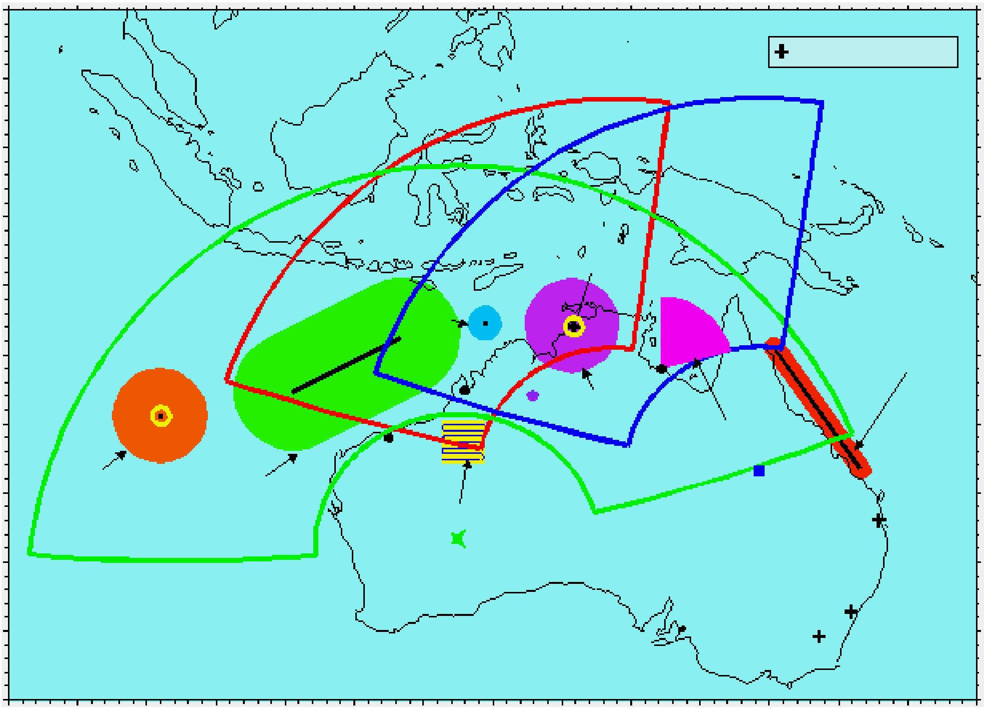


Figure 5: Stylised depiction of Australia's wide area surveillance system [16].

[4] C^4ISR stands for Command, Control, Communications, Computers, Intelligence, Surveillance & Reconnaissance.

Before outlining the methodology, it is instructive to consider how human operators accomplish the task. When human operators compile the WASP, they do not rely solely on filtered data from radars and other sensors. The track data are plotted on a geographic display which provides the operators with a wealth of contextual information to exploit for enhancing their situation awareness. This includes entities such as coastlines, air and shipping lanes, locations of towns/cities, air and shipping ports, et cetera, locations of military assets and installations, and regions such as air defence identification zones, fighter and missile engagement zones et cetera, and notional radar coverage regions. This information enables them to visually infer relations between the target tracks and the entities (including other target tracks) in order to recognise activities and events that are occurring within the surveillance region. It may also allow them to interpret the significance of those events in terms of evidence they may provide of target identity or class, as well as the threat they may pose or the impact they may have on operations of friendly and coalition forces. Finally, it may also assist them in identifying and correcting sensor misalignments, and scheduling agile sensors or cueing additional sensors. Thus, the contextual information provides support for operators at all levels of the λJDL model of data fusion. In order to introduce automation into this processing for assisting operators, algorithms are required which are capable of extracting the same types of relations or predicates between tracks and entities that the operators obtain via visual means.

To develop appropriate means for modelling and utilising the available contextual information for extracting target-entity relations, it is necessary to determine first what types of relations require automatic recognition. While no compiled list of relations can ever be truly exhaustive, those listed in Table 2, along with their auxiliary functions, have provided a firm basis from which to work:

Table 2: Entities, relations and functions used for extracting symbolic information from sensor tracks via contextual information.

ENTITIES	RELATIONS
Points, line segments, circular arcs, circles,	Is_Above_Height (or Speed)
polygons, annular sectors, and apertures	Is_Below_Height (or Speed)
	Is_Between_Heights (or Speeds)
FUNCTIONS	
Time_To_Go	Is_In or Is_On
Time_Elapsed	Is_Heading_To
Point_Of_Ingress	Is_Heading_From
Point_Of_Egress	Is_Passing
Minimum_Distance_To	Is_Within_Angle
Minimum_Time_To	Is_Within_Distance
Point_Of_Realisation	Will_Be_Within_Distance

The issue of how to model the contextual information listed at the beginning of this section[5] is somewhat dependent on the use to which the information will be put [13-17]. However, the main approach being taken to this problem by DSTO is to model the contextual information geometrically as the composition of simple geometric entities such as points, line segments, circular arcs, apertures, and circular, annular and

[5] It is noted that additional contextual information may also be available for use by air defence operators, such as the order of battle (ORBAT) and the electronic order of battle (EOB), and information about a target's weapon and speed envelopes that may be inferred from its classification or identification.

polygonal regions [16]. For example, to a first approximation, the location of an asset may be modelled as a point, an airlane may be modelled as a rectangle (along with a specified direction), the notional coverage region of a ground-based microwave radar may be modelled as a circle, the weapons envelope of a fighter aircraft may be modelled as an annular sector, and the racetrack orbit of a force multiplier may be modelled as the boundary of the union of a rectangle and two circles. Therefore, to support the recognition of target behaviours with respect to general entities, it is sufficient to develop functions for recognising them with respect to simpler entities and then to compose the functions to recognise the more complex behaviours. The recognition of target behaviours via this methodology is referred to as *target behaviour extraction*.

The methodology is explained in detail in [16] and [18]. However, it may be summarised as follows. Assume initially that the precise target state is known. Since the contextual information can be modelled geometrically, and the target state itself is geometric by nature, it is possible to use geometric criteria to establish mathematical conditions which hold if and only if the target is exhibiting the behaviour, that is if and only if the Boolean target-entity relation holds. Thus for a given relation *Rel*, the value of *Rel* (target, entity) is 1 or TRUE if the target is exhibiting the behaviour and is 0 or FALSE otherwise. For example, the relation "is_in" between a target with position (x_1, x_2) and a circle with centre (c_1, c_2) and radius R is 1 (or TRUE) if and only if the mathematical condition $(x_1 - c_1)^2 + (x_2 - c_2)^2 \leq R^2$ holds. To establish relations for compound target behaviours, the logical AND, OR and NOT connectives can be used accordingly to compose the truth values of the Boolean relations corresponding to each of the contributing simple target behaviours. To extend this approach to handle target state estimates, the target behaviour extraction algorithms can be applied to each of a specified number of random samples distributed according to the state estimate's probability density. The proportion of samples exhibiting the behaviour to the total number of samples can then be interpreted as the probability that the target is exhibiting the behaviour, given the state estimate and the model of the behaviour. Figure 6 below illustrates the methodology applied to the problem of monitoring the progress of an air target as it approaches the air defence identification zone (ADIZ) established around the coastline of an island during an air defence scenario. As time evolves, target-entity relations are used to determine if the target is heading towards the ADIZ (and the time to go until it enters the ADIZ, assuming it continues to fly at constant speed on a constant heading), if it is inside the ADIZ or if it is within a specified distance of the island. A simple tracking algorithm is used to estimate the target's state at 10 second intervals. The results for a single run are plotted in Fig. 6.

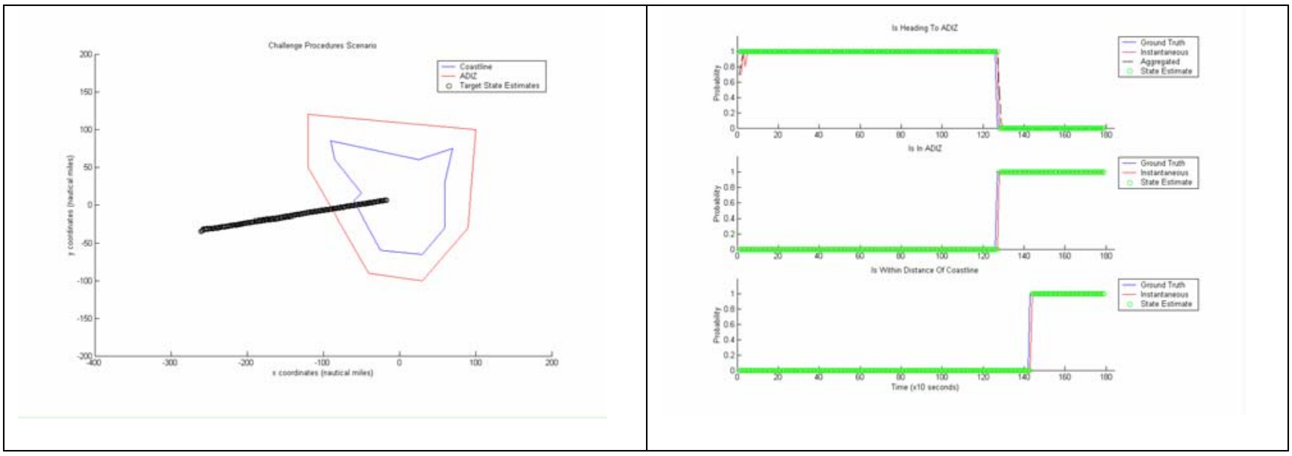


Figure 6: Target behaviour extraction methodology used to establish target-entity relations for monitoring a target's progress.

In summary, the methodology permits the behaviour of air targets to be interpreted automatically in a consistent and reasonably systematic manner that emulates the visual determination of the behaviours by air defence operators. It also has a number of desirable features such as it supports code re-use, and it provides a convenient representation of information which is flexible enough to support additional automated reasoning, whether it be by Bayesian networks, templates or intelligent software agents. Finally, the extracted information, summarised over time, may also be treated as "features" of a target track for the purposes of information retrieval and data mining (in particular clustering) to assist intelligence analysts.

4.0 INTELLIGENCE PROCESSING AND ANALYSIS

The process model employed by the intelligence community, aptly referred to as the *intelligence cycle*, comprises the five steps of *planning and direction*, *collection*, *processing*, *analysis and production*, and *dissemination*. It is evident from this model that, at a conceptual level, the business practices of the intelligence community are very similar to those employed for command and control in the defence forces. So too are the demands on intelligence analysts in regard to processing and analysing large volumes of data and information, although in the intelligence domain more emphasis is placed on knowledge discovery via data mining and information extraction.

Raw (unprocessed) intelligence data may be collected from many disparate sources via electronic and other technical means, human sources, imagery and communications for example. However, since most raw intelligence data takes the form of unstructured free text, the discussion in this paper is restricted (almost) exclusively to automated document analysis, although some of the issues raised are relevant to all forms of intelligence.

Just as the automatic processing of sensor data and the subsequent extraction of information from it is a challenging undertaking, so too is the automatic processing of documents for extracting intelligence information. However, unlike the analysis of sensor data, the automatic analysis of documents is not susceptible to treatment by numerical techniques. Thus, automated document analysis requires entirely different techniques from sensor data analysis, and must contend with the notorious difficulties that dealing with unstructured free text poses.

The general framework for intelligence information management and analysis adopted by DSTO is illustrated by the schematic in Fig. 7. With respect to this framework, there are three main challenges to be faced in introducing automation to document analysis. The first challenge lies in efficiently retrieving the documents relevant to a query. As any user of search engines on the world wide web knows, it can be extremely difficult to retrieve all relevant documents during an information retrieval query, without also retrieving a sizeable quantity of irrelevant documents. However, even assuming that individual intelligence sources, typically databases, have perfect information retrieval capabilities, often there is no connectivity between the sources, which again hampers efficient information retrieval. This lack of networking can result in the same document being retrieved multiple times because each of the sources needs to be queried individually. The second challenge is related to the processing of the retrieved documents. Information retrieval can recover documents for human analysis, but further processing of the documents for applications such as data mining and information fusion requires the unstructured information in the documents to be transformed into a well-structured form. Information extraction is a key enabling technology that can create the structured information required for these applications. However, extraction of key facts embedded within a piece of free text can prove to be problematic for a variety of reasons. For example, related facts of interest in a document may not be juxtaposed, but may instead be separated by a string of words which are extraneous to the facts, making it difficult to establish their association. In addition, references to the same fact may appear within a single document in different guises. As a case in point, dates referring to a common event may appear in either numerical or textual form, and may be written in full, abbreviated or referred to descriptively, for example 11/9/01, 11 September 2001, 11 Sept. '01, 9/11, and "the day that the twin towers collapsed". Different conventions for representing dates such as the month-day-year format preferred in the United States and the day-month-year format used in Australia may also lead to ambiguity. The third challenge is to exploit the extracted information for the purposes of data mining, either by means of unsupervised learning (clustering), link analysis (association discovery and sequence discovery) or information visualisation, and ultimately for situation and impact refinement, including the recognition of indications and warnings.

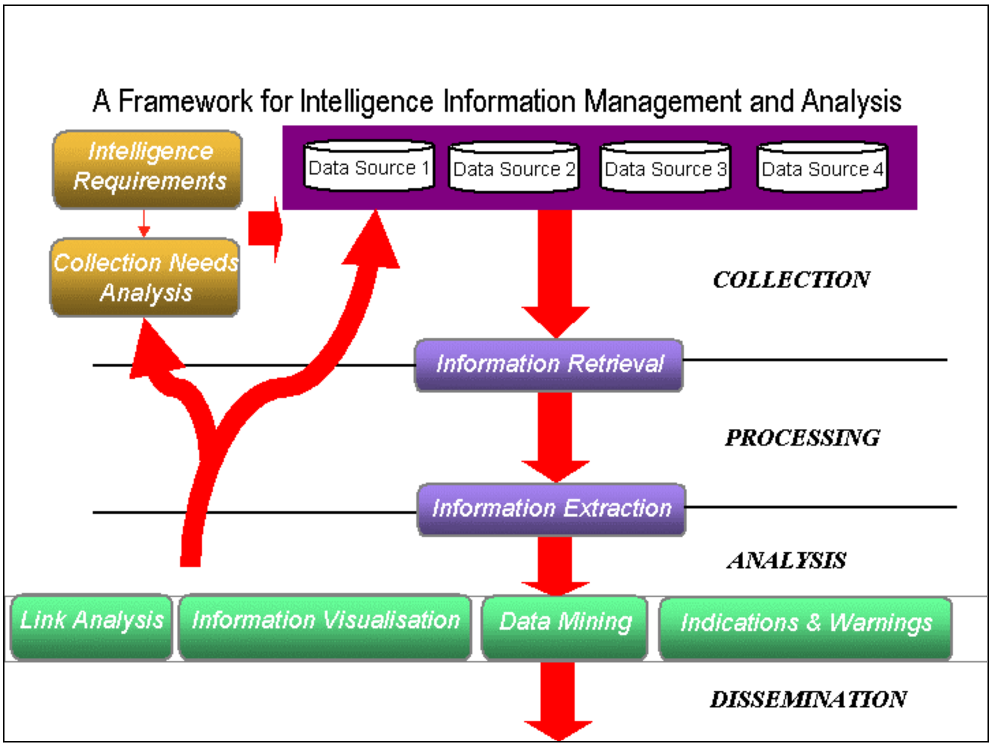


Figure 7: Schematic for intelligence information management and analysis [19].

DSTO's strategy for supporting automated text-based intelligence processing and analysis is based around these three challenges. It is evaluating a number of leading text processing and data mining commercial-off-the-shelf (COTS) products, as well as developing several in-house data mining toolkits which may assist in meeting the challenges. In the remainder of the section, a selection of these products will be profiled in brief, with particular emphasis placed on the in-house toolkits.

4.1 COTS Products[6]

Except where otherwise indicated, information on the three products below has been sourced from the internet websites of their respective manufacturers.

Autonomy™ is a product of Global Linxs that provides a software framework for automating operations on unstructured information. Some of the operations it supports include automatic categorisation, clustering, profiling, personalisation, alerting, hyperlinking and retrieval. Autonomy™ solutions are built using a modular architecture. In particular, Autonomy™ has a "portal-in-a-box" module which comprises three portals each with its own "portlets" that contain applications for performing a variety of functions.

[6] Disclaimer: The inclusion of these particular products in this paper is purely to illustrate the diversity of commercially available text processing and data mining functionality, and does not constitute their endorsement by DSTO.

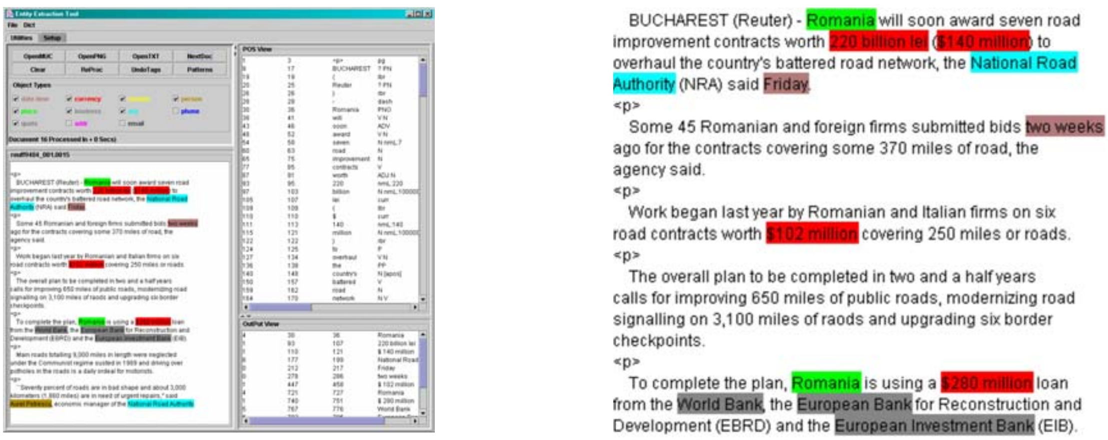
The *content portal* contains an information retrieval portlet which, like some search engines, supports concept-based retrieval and automatic summarisation of a document's content for example. However, it also features *classification* and *agent* portals. The classification portal identifies concepts in a document's contents and builds hierarchical classification schemes or categories. The agent portal performs personalisation operations using agents as concept patterns; in particular, it can use concept patterns to produce profiles to personalise information retrieval for users, recognise and alert users to highly focussed experts, and form communities of interest. For further details, refer to [20, 21].

SaffronNet™ is a product of Saffron Technology that provides the ability to process and learn about complex data in distributed multi-agent environments. It functions by establishing and maintaining a dynamic network of entities similar to a semantic network that automatically defines the context-based relationships or links between entities in the network. For further details, refer to [22].

Starlight™ is a product of Pacific Northwest National Laboratory, U.S. Department of Energy that comprises a diverse range of interactive data visualisation tools that allow the user to graphically manipulate the information under scrutiny and to display it in a number of different views. The Starlight™ information graphics fall broadly into two classes; the *non-spatial information graphics* provide spatial representations of non-spatial information such as text and numeric data, while the *inherently spatial information graphics* provide depictions of spatial information such as geospatial and CAD data. For further details, refer to [23].

4.2 In-House Products

The Fact Extractor System (FES) [24] is an information extraction product that is designed to extract "facts" from unstructured free text based on regular expression constructs. It functions on two levels. On the first level, which is commonly known as *named entity* extraction, it extracts facts of the form "who", "what", "where", and "when". On the second level, it functions by linking the basic facts extracted from the first level process into *events of interest*. The FES outputs three types of information: facts, event notifications and marked-up documents. The FES consists of a number of components; these are the fact extraction engine, the named entity extractor, a workbench for developing and testing fact extractor specifications, a batch tool for doing unsupervised information extraction, event alerting and document mark ups, and a formfiller for doing more advanced supervised information extraction. Figure 8 depicts a marked up document that has been outputted by the named entity extractor. The different entity types are distinguished by colour.



BUCHAREST (Reuter) - Romania will soon award seven road improvement contracts worth 220 billion lei ($140 million) to overhaul the country's battered road network, the National Road Authority (NRA) said Friday.
<p>
Some 45 Romanian and foreign firms submitted bids two weeks ago for the contracts covering some 370 miles of road, the agency said.
<p>
Work began last year by Romanian and Italian firms on six road contracts worth $102 million covering 250 miles or roads.
<p>
The overall plan to be completed in two and a half years calls for improving 650 miles of public roads, modernizing road signalling on 3,100 miles of raods and upgrading six border checkpoints.
<p>
To complete the plan, Romania is using a $280 million loan from the World Bank, the European Bank for Reconstruction and Development (EBRD) and the European Investment Bank (EIB).

Figure 8: Output of the named entity extractor in the form of a marked up document [24].

Finally, each set of fact extraction specifications determines a different fact extractor, so since each fact extractor functions as a software agent, different fact extractors can work cooperatively to reduce complexity for advanced information extraction. For more information, refer to [24].

The Data Mining and Visualisation Toolkit (DMVT)[7] [25] is a suite of prototype tools designed to assist in the analysis of historical positional and kinematic data produced through the tracking of dynamic entities in some problem domain, such as air targets in a wide area surveillance scenario for example. The toolkit is composed of five applications that are integrated as layers in the open source OpenMap Java toolkit from BBN Technologies [26]. Some of the strengths of the DMVT, which supports the analysis of temporal, spatial, link and track information, are its abilities to filter data for human-assisted data exploration and automatically determine normalcy patterns from the data. The DMVT features a number of customised views. For example, temporal data based on a particular activity may be displayed in a variety of modes on a grid or clock face based on "hour of the day", "day of the week" et cetera through to "month of the year" according to the user's needs, with the level of intensity of the activity for each time unit indicated via colour-coding. Similarly, target movements between given locations on a geospatial display may be indicated by colour-coded arrows, such that different colours indicate different classes of tracks and the width of the arrow indicates the volume of traffic between the locations.

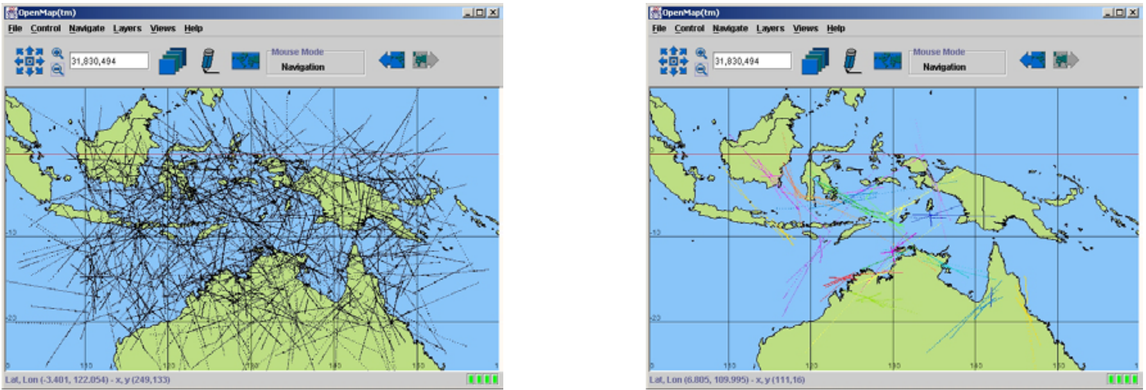


Figure 9: A set of synthetic track data before (left) and after track clustering and filtering (right) [25].

Finally, by combining the spatial view and the track clustering functionality, a set of target tracks may be clustered into classes of tracks with similar headings and close spatial separation and then viewed on the geospatial display. This technique is illustrated in Fig. 9 for a randomly generated set of synthetic air tracks (NB: no sensor model has been employed in their generation). The tracks have been clustered and clusters with a lone track have been removed to reveal potential air corridors. For more information, refer to [25].

5.0 INFORMATION FUSION FOR COMMAND & CONTROL

In Sections 3 and 4, automated data fusion and data mining systems were described for assisting defence operators and intelligence analysts in the processing and analysis of data and information. However, the important issue of human interaction with the systems was not raised. In this section, human computer interaction in information fusion is addressed in terms of DSTO's Future Operations Centre Analysis Laboratory (FOCAL) programme (FOCAL itself appears in Fig. 10). Before discussing the research being undertaken for FOCAL, it is necessary to probe more deeply into the interpretations of the λJDL model of

[7] The DMVT has been included, even though it is not a text-based information extraction product, because of its relevance to the material in Section 3.

data fusion to establish the context for the research. Note: The material for this section draws substantially from reference [12].

5.1 Interpretations of the λJDL Model of Data Fusion

The FOCAL data fusion system extends well beyond the traditional "machine sensor fusion" emphasis of the data fusion community, by including information fusion considerations involving both humans and machines. It was these considerations that led to Lambert's deconstruction of the US JDL model of data fusion in the form of the λJDL model described in Section 2. As is explained below, the λJDL model not only provides a means of integrating data fusion and data mining, but is also crucial for developing the theory which underpins the research within the FOCAL programme.

Figure 10: The Future Operations Centre Analysis Laboratory (FOCAL).

As a consequence of the FOCAL data fusion system seeking to fuse processes involving both people and machines, three distinct types of processes result: psychological processes associated with people; technological processes characteristic of machines; and integration processes facilitating interaction between the psychological and technological processes. Different interpretations of the λJDL model are obtained when different combinations of these processes are considered. Figure 11 exhibits a matrix in which the rows are the three components of the λJDL model and the columns are the three types of processes.

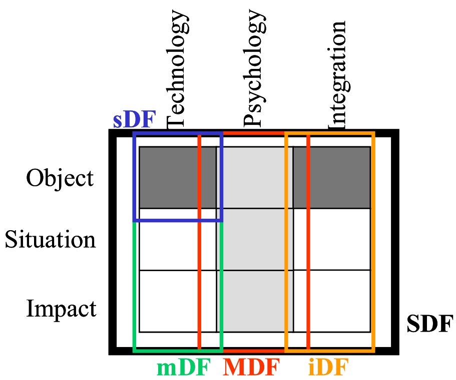


Figure 11: Interpretations of the λJDL model [12].

Five interpretations of the λJDL model are overlaid on the matrix. These interpretations are as follows:

Mental Data Fusion: Endsley [27] defines situation awareness as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". In [28], Lambert indicates that if "representation" in the definitions of the levels of the λJDL model is interpreted as "mental representation", then the object, situation and impact fusion components of *mental data fusion* (MDF) correspond to the situation awareness elements of perception, comprehension and projection respectively. In this way, mental data fusion and situation awareness may be identified.

Sensor Data Fusion: *Sensor data fusion* (sDF) refers to the technological interpretation of object refinement and corresponds to the commonly accepted definition of sensor fusion.

Machine Data Fusion: *Machine data fusion* (mDF) arises if "representation" in the definitions of the levels of the λJDL model is interpreted as "machine representation". It is concerned with technological approaches to data fusion and so refers to the technology column of the matrix in Fig. 11. Clearly, sensor data fusion is part of machine data fusion.

Interface Data Fusion: *Interface data fusion* (iDF) arises if "representation" in the definitions of the levels of the λJDL model is interpreted as "interface representation". It is concerned with human-machine integration approaches to data fusion, and so refers to the integration column of the matrix in Fig. 11.

System Data Fusion: *System data fusion* (SDF) is concerned with data fusion systems involving machines, people and interactions between the two. System data fusion refers to the entire matrix in Fig. 11. It requires a unified framework across all three components of the λJDL for handling interactions between people, interactions between machines, and interactions between people and machines.

These five interpretations of the λJDL model lead to a number of challenges for information fusion. This is expanded on in the sequel.

5.2 FOCAL Research – The Grand Challenges of Information Fusion

None of the elements of the matrix in Fig. 11 represents solved problems, and so all elements of the matrix pose challenges. However, the portions of the matrix which are unshaded pose grander challenges than the shaded portions. It is these *grand challenges*, which may be broadly categorised as either machine data fusion or system data fusion challenges, that form the basis for the research under the FOCAL programme. In [12], Lambert gives a full account of these challenges, but that level of detail is beyond the scope of this paper. Therefore, in the remainder of the section, only a brief outline of these grand challenges and a description of the related FOCAL research initiatives for each are given. For further details, refer to [12].

The Semantic Challenge: What symbols should be used and how do these symbols acquire meaning?

The semantic challenge transcends philosophical, mathematical and computational dimensions. A philosophical theory is required to conceptualise the domain of interest; a mathematical theory is required to impose structure on that conceptualisation; and a computational theory is required to bring that conceptualised structure to life. The conceptualisation suggests symbols that might be used. The (formal) mathematical theory prescribes the meaning of those symbols.

To accomplish this within the FOCAL programme, the intention is to develop formal theories to capture the semantics of selected primitive symbols [29]. The FOCAL programme needs to address a range of issues spanning multiple levels. Table 3 specifies the proposed levels, together with *some* relation symbols to be defined for each level. In developing the formal theories, the ambition is not to capture *the* meaning of time, belief, et cetera, but to formulate *a* meaning of time, belief et cetera that is sufficient for engineering data fusion systems. Ideally, the metamathematical properties of the theories will include *soundness*, *completeness* and *decidability*, but these properties are often difficult to obtain.

Table 3: Proposed levels and some relation symbols for the formal theories [12].

Social:	group, ally, enemy, neutral, own, possess, invite, offer, accept, authorise, allow
Intentional:	individual, routine, learnt, achieve, perform, succeed, fail, intend, desire, belief, expect, anticipate, sense, inform, effect, approve, disapprove, prefer
Functional:	sense, move, strike, attach, inform, operational, disrupt, neutralise, destroy
Physical:	land, sea, air, outer_space, incline, decline, number, temperature, weight, energy
Metaphysical:	exist, fragment, identity, time, before, space, connect, distance, area, volume, angle

The Epistemic Challenge: What information should be represented, and how should it be represented and processed within the machine?

While the semantic challenge for information fusion relates to the choice of symbols and what those symbols mean, the epistemic challenge involves the choice of knowledge content and the choice of a symbolic knowledge representation scheme. For analytically difficult domains in which human expertise is available, as is often the case for information fusion applications, one approach toward the epistemic challenge is to model the cognitive behaviour of people. This requires a modelling framework, a means of capturing cognitive behaviour within that framework, and a means of automating that captured cognitive behaviour within a machine. This is the approach being taken for FOCAL.

In [30], an approach for automating cognitive routines for FOCAL is outlined, based on people's explanations of what they do, *as they are doing it*, to mitigate *a posteriori* rational reconstructions. In its current form, the explanations are recorded through speech recognition before a human translation process converts them into cognitive routines executable in the ATTITUDE multi-agent reasoning system. The ATTITUDE system [31], developed by DSTO, is so named because it codes in terms of *propositional attitudes* (beliefs, desires, expectations, et cetera), which significantly eases the burden of translation from explanation to code.

The Paradigm Challenge: How should the interdependency between the sensor fusion and information fusion paradigms be managed?

The lack of an established technological information fusion approach follows from the fact that the techniques of sensor fusion do not easily scale up. As noted in [28, 32-34], there is a paradigm shift in moving from sensor fusion to information fusion, a shift from an Aristotelian world of objects with measurable properties to a Wittgensteinian world of symbolically expressed facts formed from relations between objects. The technological aspects of sensor fusion are founded on an Aristotelian paradigm based on a numerical representation, while the technological aspects of information fusion are founded on a Wittgensteinian paradigm based on a symbolic representation. The paradigm challenge arises because machine data fusion demands an interdependency between these two paradigms, and perhaps in time, a unifying paradigm.

It has become customary to think of "levels" of data fusion, despite the bus architecture in the US JDL model in Fig. 2. The notion of levels follows from a presumption that the output of sensor fusion becomes the input of information fusion. While not all inputs to information fusion derive from conventional sensor fusion, a dependency of information fusion on sensor fusion is widely recognised.

The FOCAL programme has not yet addressed the paradigm challenge, although it does support a hybrid numerical-symbolic approach, including a probabilistic inference engine within ATTITUDE [35]. Arguably, the paradigm challenge is best handled in concert with the communities processing the sensor data to ensure that the transformation from numerical sensor data to symbolic information does not lead to misinterpretations of the data. The information extraction efforts at DSTO for wide area surveillance and intelligence processing and analysis, reported on in Sections 3 and 4, support this viewpoint. It is envisioned that the information extracted from target track data and text-based information (and indeed other data sources) will be structured according to the social, physical and metaphysical formal theories arising from the semantic challenge.

The Interface Challenge: How should people be interfaced to complex symbolic information stored within machines?

As symbolic information fusion matures within the machine, a means is required of interfacing complex information encapsulated in symbolic machine representations to people interacting with those machines.

The FOCAL programme is responding to the interface challenge through a number of initiatives. The first of these initiatives is the *Virtual Battlespace* which is shown in Fig. 12. It advances beyond the typical 2-dimensional displays, which overlay target symbology or tracks (and other contextual information) on a geographic plot, by incorporating terrain, imagery, and objects of interest within a 3-dimensional stereoscopic display.

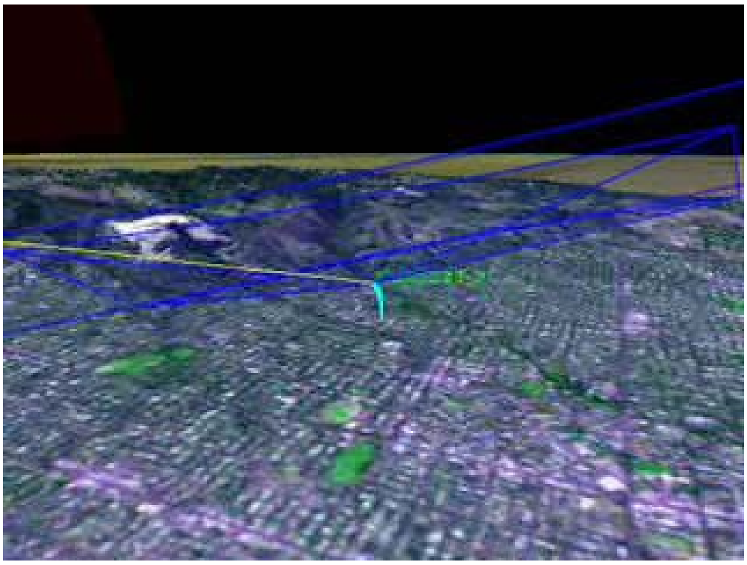


Figure 12: A snapshot of the 3-dimensional Virtual Battlespace [12].

However, while the Virtual Battlespace significantly advances the interface challenge for sensor fusion, it only marginally improves it for information fusion. The Virtual Battlespace is still not an appropriate medium for conveying information such as “Osama Bin Laden has avoided capture”. In everyday life, information of this sort is customarily received via news services, be they print, radio, television or internet based. Television news broadcasts remain the most dominant form, and are typically composed of: (a) *advisers*, such as news presenters, weather presenters, sports presenters, reporters, experts; (b) *maps and diagrams*, such as weather maps and finance charts and graphs; and (c) *video footage*, which can be live, or extracted from a library because of some relevance to news events. In [31], it is proposed that software counterparts be developed for these three components of news broadcasts. As software, the components provide *portability* and *interactivity*. Portability arises because the software can be readily distributed throughout a computer network, enabling people interacting with the system to access it on demand. Interactivity arises because, unlike the information push only mechanism of television news, the software can support an information push-pull interaction with people. This is the approach being taken in the FOCAL programme.

In FOCAL, the software counterparts of advisers, maps, diagrams and video footage are *Virtual Advisers*, the *Virtual Battlespace*, *Virtual Interaction* and *Virtual Video* respectively. The Virtual Advisers in FOCAL have two semblances, the cartoon-like Franco and the photo-realistic Jane who appear in Fig. 13. The visemes, phonemes and the emotional look of the characters, are dynamically specified by marked up text. This enables the characters to be dynamically controlled by people, scripts or the ATTITUDE software. Preliminary interactive speech dialogue with the characters has also been established.

Figure 13: FOCAL's Virtual Adviser Avatars – Franco on the left and Jane on the right [12].

The Virtual Battlespace has already been described. Virtual interaction involves the representation and visualisation of structured information, as well as the use of wands, hand gesturing and eye tracking. Finally, Virtual Video involves the automated reconstruction of 3-dimensional stereoscopic animations from formal knowledge representations of the sort alluded to in Table 3. Virtual Video combines the selection of animated models, the choice of camera perspectives, and the movie director's craft. Figure 14 illustrates a virtual video from FOCAL. Reference [36] reports some of the underlying technical issues.

Figure 14: Virtual Video in FOCAL [12].

The System Challenge: How should data fusion systems formed from combinations of people and machines be managed?

The remaining challenge – the system challenge – concerns the desire for a unified framework for system data fusion that can account for interactions between people, interactions between machines, and interactions between people and machines.

The construction of ATTITUDE agents to improve machine useability has had the side effect of producing a conceptual framework that can be applied to both people and machines. As a consequence, the response to the system challenge for FOCAL has been to treat the FOCAL data fusion system as an *agent society* composed of people and machines. The same conceptual constructs can then be applied to both people and machines. With respect to the semantic challenge, the relations of Table 3 can be applied to both people and machines. With respect to the epistemic challenge, the proposal for capturing cognitive routines is about establishing functionalist cognitive routines common to both people and machines. With respect to the interface challenge, the virtual advisers allow the machines to interface more like people.

The management of these agent societies is also under consideration within the FOCAL programme. Perugini and Lambert [37] have been investigating the organisation of agent societies for logistics operations. This has recently led to a proposed Provisional Agreement Protocol [38]. In time this is expected to develop into a basis for a system of axioms to constrain the nature of contractual interactions between combinations of human and machine agents.

6.0 CONCLUSION

The need for Australia to develop an information capability, as laid out in Australia's strategic defence policy, has been explained in terms of its strategic circumstances, and the role of the Defence Science and Technology Organisation in supporting the development of this emerging capability has been highlighted. The λJDL model of data fusion has been outlined and used to establish the theoretical underpinnings for integrating the key enabling technologies of data fusion and data mining required for the information capability. Finally, against the backdrop of this model, several of DSTO's information extraction and fusion initiatives in the areas of wide area surveillance, intelligence processing and analysis and command and control have been discussed, and the potential for synthesis of these initiatives via the grand challenges of information fusion has been articulated.

7.0 ACKNOWLEDGEMENTS

The authors would like to thank Mr. Peter den Hartog for his assistance in producing the graphics used in Figure 1.

REFERENCES

[1] Australian Federal Government (2000), *Defence 2000 – Our Future Defence Force*, The Commonwealth of Australia Defence White Paper.

[2] T.G. Feeman (2002), *Portraits of the Earth – A Mathematician Looks at Maps*, Mathematical World, Vol. 18, American Mathematical Society.

[3] Contributing Editors (2003), *SBS World Guide, 11th Edition*, Hardie Grant Publishing.

[4] Australian Federal Government (2003), *Australia's National Security – A Defence Update 2003*, The Commonwealth of Australia.

[5] D.L. Hall and J. Llinas (Eds.) (2001), *Handbook of Multisensor Data Fusion*, The Electrical Engineering and Applied Signal Processing Series, CRC Press.

[6] A.N. Steinberg and C.L. Bowman (2001), *Revisions to the JDL Data Fusion Model*, Chapter 2 of Hall and Llinas [5].

[7] F.E. White, Jr. (1988), *A Model for Data Fusion*, in the Proceedings of the 1st National Symposium on Sensor Fusion, Vol. 2.

[8] A.N. Steinberg, C.L. Bowman and F.E. White, Jr. (1998), *Revisions to the JDL Data Fusion Model*, Joint NATO/IRIS Conference, Quebec City, Quebec, 19-23 October, 1998.

[9] E. Waltz (1998), *Information Understanding: Integrating Data Fusion and Data Mining Processes*, in the Proceedings of the 1998 IEEE International Symposium on Circuits and Systems "ISCAS '98", Vol. 6, pp. 553-556, Monterey, California, 31 May – 3 June, 1998.

[10] G. Piatetsky-Shapiro and W.J. Frawley (1991), *Knowledge Discovery in Databases*, AAAI Press / The MIT Press.

[11] U.M. Fayyad (1997), Editorial, *Data Mining and Knowledge Discovery*, Vol. 1, No. 1.

[12] D.A. Lambert (2003), *Grand Challenges of Information Fusion*, Proceedings of the Sixth International Symposium on Information Fusion "Fusion 2003", pp. 213-220, Cairns, Queensland, Australia, 8-11 July, 2003.

[13] M.G. Oxenham (2000), *Automatic Air Target to Airlane Association*, Proceedings of the International Symposium on Information Fusion "Fusion 2000", Paris, France, USA, 10-13 July, 2000.

[14] H-T. Ong, M.G. Oxenham and B. Ristic (2002), *Estimating Biases in Sensor Measurements using Airlane Information*, Proceedings of the Information, Decision and Control Conference "IDC '02", pp 187–192, Adelaide, South Australia, Australia, 11-13 February, 2002.

[15] H-T. Ong, B. Ristic and M.G. Oxenham (2002), *Sensor Registration Using Airlanes*, Proceedings of the International Symposium on Information Fusion, "Fusion 2002", Annapolis, Maryland, USA, 7-11 July, 2002.

[16] M.G. Oxenham (2003), *Using Contextual Information For Extracting Air Target Behaviour From Sensor Tracks*, Proceedings of the Signal Processing, Sensor Fusion, and Target Recognition XII Conference at the 17th SPIE Annual AeroSense Symposium, 21-23 April, 2003.

[17] H-T. Ong (2003), *Sensor Registration Using Airlanes: Maximum Likelihood Solution*, Proceedings of the SPIE Conference "Signal and Data Processing of Small Targets 2003", San Diego, CA, USA, 3-8 August, 2003.

[18] M.G. Oxenham (2003), *Enhancing Situation Awareness for Air Defence via Automated Threat Analysis*, Proceedings of the Sixth International Symposium on Information Fusion "Fusion 2003", pp. 1086-1093, Cairns, Queensland, Australia, 8-11 July, 2003.

[19] R. Price (ed.) (2003), *Proceedings of the DSTO Workshop on Information Retrieval, Processing and Analysis of Disparate Data Sources*, Canberra, Australia, 11 April, 2003.

[20] http://www.digibahn.com/pdfs/PB_Autonomy_Portal_in_a_Box_0402.pdf

[21] L. Alvino (2003), *An Overview of Autonomy*, Presentation in [19].

[22] http://www.saffrontech.com/pages/products/saffronnet.html

[23] http://starlight.pnl.gov/

[24] J. Das (2003), *A DSTO Prototype Information Extraction System*, Presentation in [19].

[25] B. Williams (2003), *Advanced Data Mining and Visualisation*, Presentation in [19].

[26] http://www.bbn.com

[27] M.R. Endsley (1995). Toward a Theory of Situation Awareness in Dynamic Systems, *Human Factors*, Vol. 37 No. 1 pp. 32 – 64.

[28] D.A. Lambert (2001), *Situations for Situation Awareness*, Proceedings of the Fourth International Conference on Information Fusion "Fusion 2001", Montreal, Quebec, Canada, 7-10 August, 2001.

[29] D.A. Lambert (2003), *A Computational Metaphysics: Part I Existence*, DSTO Research Report (in preparation), Information Systems Laboratory, Defence Science and Technology Organisation.

[30] D.A. Lambert (2003), *Automating Cognitive Routines*, Proceedings of the Sixth International Symposium on Information Fusion "Fusion 2003", pp. 986-993, Cairns, Queensland, Australia, 8-11 July, 2003.

[31] D.A. Lambert (1999), *Advisers with Attitude for Situation Awareness*, Proceedings of the 1999 Workshop on Defense Applications of Signal Processing, pp. 113-118, edited by A. Lindsay, B. Moran, J. Schroeder, M. Smith and L. White, La Salle, Illinois, USA.

[32] L. Reznik and V. Kreinovich (Eds.) (2003), *Soft Computing in Measurement and Information Acquisition*, Studies in Fuzziness and Soft Computing No. 127, Springer Verlag.

[33] D.A. Lambert (2003), *An Exegesis of Data Fusion*, Chapter 6 of Reznik and Kreinovich [32].

[34] D.A. Lambert (1999), *Assessing Situations*, in Proceedings of the 1999 Conference on Information, Decision and Control "IDC '99", pp. 503-508, IEEE Press.

[35] I. Fabian, and D.A. Lambert (1998), *First-Order Bayesian Reasoning*, in Advanced Topics in Artificial Intelligence, 11th Australian Joint Conference on Artificial Intelligence "AI '98", pp. 131-142, Springer-Verlag.

[36] M. Coleman (2003), *Technical Considerations when Selecting Commercial Applications on an SGI Onyx*, DSTO Technical Report (in preparation), Information Systems Laboratory, Defence Science and Technology Organisation.

[37] D. Perugini, D.A. Lambert, L. Sterling, A. Pearce (2002), *Agents for Military Logistics Planning*, 15th European Conference on Artificial Intelligence "ECAI 2002", Lyon, France.

[38] D. Perugini, D.A. Lambert, L. Sterling, A. Pearce (2003), *Distributed Information Fusion Agents*, Proceedings of the Sixth International Symposium on Information Fusion "Fusion 2003", pp. 86-93, Cairns, Queensland, Australia, 8-11 July, 2003.

A Man-in-the-Loop Support Concept for Military Ambush Threat – Assessment Based on Reconnaissance Reports

Frank P. Lorenz and Joachim Biermann
FGAN-FKIE
Neuenahrer Strasse 20
53343 Wachtberg-Werthhoven
GERMANY

lorenz@fgan.de

ABSTRACT

On all levels of the military command hierarchy there is a general and strong demand for support by automated processing of reconnaissance reports. This is especially required for information processing and situation assessment, which aims at the inference and deduction process on the base of report information to the greatest possible extend by analysis components. Developments for computer based decision support significantly expands the capability of tactical situation display programs.

Regarding this general background, some methodical preconditions for the improvement of computer support are characterized. The central aspect is the role of the information context for an appropriate perception of the situation by the analysis of reports. This is illustrated by a specific military example of an ambush attack situation, which has the proprietary feature of disguised preparation. The example is representing an unconventional type of military operations, which are complex in prearrangement and typically clandestine in accomplishment. Their timely recognition by HUMINT information is a task, which is mandatory dependent on expertise. Our concept for the threat assessment is therefore developed on a two-stage man–in-the-loop procedure still dependant on this expertise. Its benefit lies in the ruthless incorporation of absent, but possibly significant situation context information on a trial-and-error base. The major ideas were tested by an ability assessment tool for ambush attacks. The results are discussed.

Keywords: *Knowledge-Based Information Fusion, Situation Assessment, Intelligence Preparation of the Battlefield (IPB), Command and Control, Military Scenario Simulation, Operations Other Than War (OOTW).*

1.0 INTRODUCTION

The computer assisted processing of incoming military reports is done in the first instance by theatre independent analysis. For the merging of information belonging to the same object or state of an object we need a series of regulations. Since this is not a simple one-to-one aggregation, we need to invent methods of higher level abstraction. We call these methods together with the underlying regulations the fusion process. The aggregation and fusion of information which are often incomplete, misleading and/or untimely in their content is a difficult task. These follow up processing steps can now only be accomplished on the base of background knowledge and with regards to the specific situation context. The rules for the information processing are given by a priori data. Templates are defined by recurring patterns of human behaviour and military default operations. Case studies, lessons learned and expertise establish the base for the definition of such a template for the example of a military ambush situation. The practicability of this template is illustrated with an isolated example of an ambush attack situation as a part of a comprehensive military scenario for Bosnia and Hercegovina which was developed in accordance

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

with the guidelines from [6]. The experimental analysis of this example starts with considerations of vulnerable objects and structures and explains which kind of information are the precondition for awareness using our template-based concept.

Section 2 gives a brief introduction to the information fusion topic as it is considered for the example of this paper. Fundamental constituent terms like information, situation assessment, situation context, information modelling, databases and templates are outlined. Section 3 touches one of the most important prerequisites, that military operations are determined by rules, doctrines and default behaviour. In section 4 the elements of a specific observation situation are introduced, like the scenario frame, experimental details, the report situation and some knowledge base data examples are given. Some preliminary testing results for the example of a non-conventional military ambush situation are discussed in section 5. The conclusions for this concept so far are recapitulated in section 6. Section 7 gives the references for major work used in this study.

2.0 INFORMATION FUSION

The processing model for data fusion which was developed by the Joint Directors of Laboratories (JDL) Data Fusion Subpanel (DFS) provides a framework and common reference for addressing data fusion issues and problems [31]. Though different models and concepts are existing, like [7] (Boyd Control or Observe, Orient, Decide and Act – OODA Model), [26] (Intelligence Cycle), [1] (Waterfall Model), [28] (Revised JDL Data Fusion Model), [2] (Omnibus Model) or [25] (Extended OODA Model), to name just a selection of most influencing works on this topics. A brief overview was given by [11] recently. The following roughly grouping is still valid:

- Level 1 Fusion: Fused position, and time-stamped identity estimates
- Level 2 Fusion: Hostile or friendly military situation assessment
- Level 3 Fusion: Hostile force threat assessment

These levels are not architectural aspects of computing, they focus on an inferential hierarchy as the process moves from Level 1 to Level 3. In this paper we discuss a concept to obtain intelligence products from report processing of an example study of a non-conventional military operation. From the scope of the Intelligence Cycle (IC) defined by the NATO Glossary of Terms and Definitions (AAP-6) we focus on the phase of the *Processing of Intelligence.*

This example study is a sequel to the work on knowledge-based data fusion on templates for a classical conventional military scenario [3], [4], [13], [14]. The aim of the work is, to show the benefits of comprehensive a priori situation analysis which is the base for the creation of suitable templates for timely awareness of hostile operations, facing imperfect information. Military expert knowledge is the base for the information fusion, but the systematic incorporation of missing or weak information as elastic constraints is the ultimate strength of a computer support tool, which is in accord with template-based information fusion. There we see the key for a necessary problem solving technique for threat assessment.

This is the underlying idea for the man-in-the-loop concept. The incoming semantically pre-processed information is, depending on its content, allocated either to the processing steps of the *capability oriented analysis* (in the case of a ‘static’ content) or the *ability oriented analysis* (in the case of a ‘dynamic’ type). The difference between a static or a dynamic information is explained in details in a later paragraph. After the level one processing of these information (Section I and II in Fig. 1.) we obtain the input elements for situation assessment on templates, which of course will be incomplete in terms of the whole set of possible entries to a template. If not incomplete, it will at least not cover all aspects with respect to a *Ground Truth.* In spite of that we try pattern matching for subsets of predefined templates for specific situation assessment tasks. An example for that could be e.g. a template for an ambush attack threat.

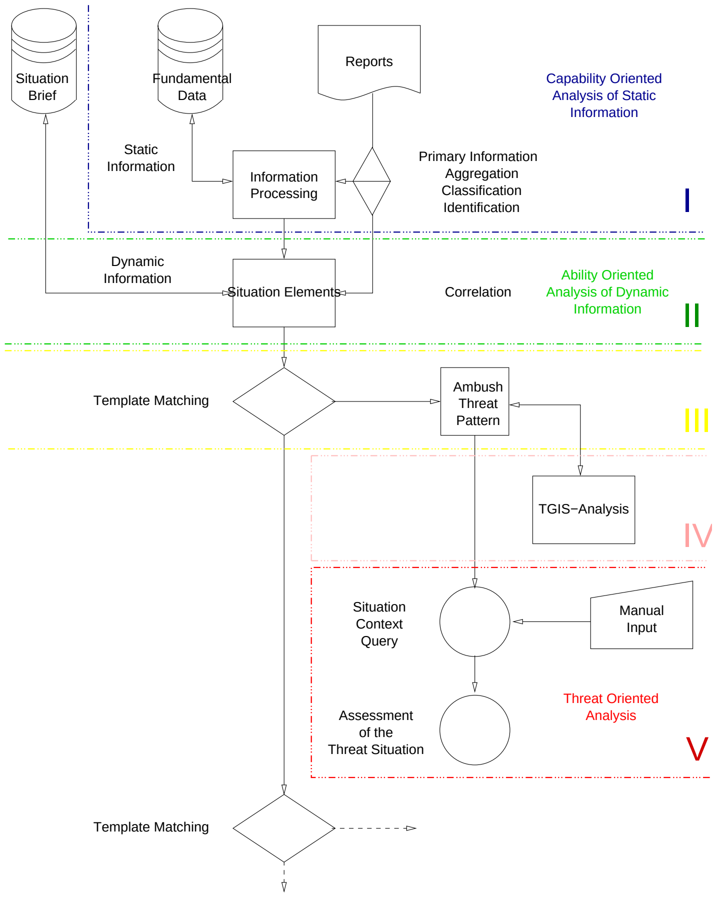


Figure 1: Scheme of the man-in-the-loop support concept.

2.1 Information in Reports

In our simplified scenario driven study we focus on incorporation of not explicitly stated situation context dependent information into the process of information fusion. Reasons for that could be e.g. low observability of objects, minor detection-discrimination capabilities or behavioural deception, to name just a few. The practically important report processing is not within the subject area of this paper. We do not address the problem of information extraction from natural language intelligence reports. [23] is a recent publication devoted to this problem.

2.2 On Fusion Concepts and Our Attempt for the Military Ambush Situation Assessment

Assessment means a given set of data for all objects (states) is correlated with a priori identified behavioural pattern for possible situations to describe the status of operations in the surveillance volume. Spatial and temporal characteristics serve to discriminate between patterns (e.g. suspicious activities). An automated data fusion method applied, would start computing degrees of fitness or associations between the set of pre-processed information from reports and a catalogue of available templates. The results of such a process gives the extend to which each template is consistent with and able to account for the received reports [30]. Exemplified, the degree of fit can be related to a degree of belief or a possibility that a certain template represents the situation. This could lead to the formulation of confidence factors as a weight assigned to particular data, the number of priority data occurrences, and so on. So far there is still no best solution formulated, but anyway, approaches aiming at fully automated problem solution diminish, whereas the communication (interaction) and interpretational skills of a human operator is accepted as necessary. Early advices to abandon any loading of elements in a template to individual situational conditions on an ad hoc base could only be checked out if, and this is our lucky position now, brute force computing can be done almost at no cost. Simulation-based analysis of this concept is ongoing.

The template for our ambush attack assessment is a summary of so called slots, specifying typical elements and characteristics of events, and typical and required relationships between assets and events. We will see later in this paper how we deal with the problem of unavailable but 'required' information. The situation assessment can be thought of as a means to estimate the enemy battle situation, what the enemy is doing (activities) and attempting to achieve (intent). We keep further in mind that the threat assessment task involves identifying the probable situation causing the observed data (information) and events. The inputs include event detections, state estimates and a set of hypotheses for evaluation. The outputs are, at least theoretically, conditional probabilities of the various hypotheses being considered in simulation [30, p. 276]. Diagnostic accuracy as well as timeliness are the dominant concerns in the assessment process.

2.3 On Modelling the Information

Military reports are usually based on HUMINT observations of the physical world. For almost every information a spatial and temporal reference (timestamp) can be obtained. This is a fundamental condition for fusion tasks like association, tracking or identification. The main entities of concern in a spatio-temporal application are the states of objects or features, their relation with space and time, and their interrelation in space and time. This leads accordingly to a representation on three axes.

- S - The spatial axis, representing the location of object states,
- T - A temporal axis for time, stamping object states and measuring of change between states.
- K - A knowledge axis, representing the classification process (and identification) of objects or features.

Any information I_i gives us a triple (S_i,T_i,K_i) defining the location S_i^k of an object O^k, giving it a timestamp T_i^k and increase the knowledge K_i^k of its state. Any shift on axis S is a change $\delta\ S_1^1$ for a given object O^1. The change in state $(S_0^1 + \delta\ S_1^1,\ T_0^1 + \delta\ T_1^1,\ K_0^1)$ for the object O^1 e.g. describes a movement of this object. Considering a proper implementation, the system will be able to answer the question if, regarding the basic knowledge about the type of object O^1, this is a valid solution within the model domain. In other words, can an object O^1 reach the location $S_2^1 = S_0^1 + \delta\ S_1^1$ within the available time $T_2^1 = T_0^1 + \delta\ T_1^1$. This is just an example for a regular query to the dataset. If we consider a change in state for an object O* along all three axes, we are forced to formulated queries in terms of higher abstraction [9].

The modelling of movement of the low level combat elements is a decisive process in every high resolution combat simulation system. Especially if the system is able to conduct automated route planning, the cost of movement and an estimation of the remaining distance from the regarded node to the target terrain has to be taken into account. Route planning should not solely be confined to time consumption along a linear distance to reach targets, but additionally to the three-dimensional path length including slopes, concealment and cover. This will substantially support the C²-authority in developing suitable tactical behaviour. The geo-referenced processing of available information as a base for a complete and systematic Intelligence Preparation of the Battlefield (IPB) is self-evident [18], [27]. This includes the integration of the so called *Ground Truth* (realistic reproduction of scenario relevant aspects of the theatre of war), this type of situation context includes e.g. weapon engagement zones, supposed mine fields or impact studies of local weather conditions.

The already mentioned separation in static and dynamic information (Fig. 1.) is explained by the addressing of the processed information either to the set of base (fundamental) data in case of a ‘static type’ of information or alternatively in case of a ‘dynamic’ information as an element to the database of the perceived situation, the situation brief [17]. In this sense we explain an association of ‘static’ information, like strength, composition, logistics, technical characteristics of equipment, weapons and others to a capability oriented analysis. In opposition to that we treat information on movements and activities as preconditions for follow-up activities, and therefore we consider a different type of analysis, the ability analysis (Fig. 1.) as appropriate.

2.4 The Template Method

Templating is a logic-based pattern recognition technique used in data fusion applications since the 1970s [30]. The pattern matching concept has basically no limitations in complexity. This is valid for both, the observational data and the logical relationship used to define a pattern. It is a feasible way to combine algorithmic approaches, figure-of-merit approaches and knowledge-based heuristic methods. The setup of a template consists on necessary (mandatory) and sufficient conditions, with acceptance and rejections thresholds and the components that describe or make up the object.

Independent from the acceptance of the concept of knowledge-based information fusion by templates as a possible solution to the fusion problem, one might argue, that the existence of templates demands nothing else than formally correct information. In such a case, vague information or even randomly incorporated elements would contradict the application of templates. And to make things even worse, how to incorporate elements of surprise in an evaluation process for the fitness of an information regarding a specific template? If we would be able to treat elements of surprise as alternatives to more rational (doctrinal) behaviour, the loophole from this dilemma would simply be the procedural option generation and evaluation of alternative actions. This is one of the key aspects of our chosen attempt.

All elements or slots of the specific template for this example are given in detail by [17]. It was i.e. build on expertise given in the works of [8], [10], [12], and [29].

2.5 Factors of the Situation Context

Given any terrain setting of an observation area and receiving e.g. a report stating that four military trucks loaded with soldiers and obviously some boxed equipment left their home barracks, driving with (non-doctrinal) fast speed, could be alarming for a military observer. On the other hand without any supplementary context, despite the cars went on roads and their assumed driving direction, the task of assessment of a conflict or even threat situation cannot be carried out. A lot more of situation dependent background knowledge, the situation context, is necessary. But usually this knowledge will not be complete, will not be granular enough or just impossible to be fed into a computer database.

Situation context can partially be accessed via databases. GIS server e.g. provide terrain data (topography, hydrography, soils, vegetation, infrastructure, etc.). Climatical data can be digitally accessed like metereological or socio-political data. Though there still remains quite some situation related context, like for the ambush attack example the actual state of a conflict or relations of alliances, the interaction with the operator remains indispensable.

2.6 Databases Contents

The knowledge bases necessary for the report and information processing should contain knowledge on

- Entities, events, attributes and relationships (Ontologies)
- Spatial relationships descriptions (Pattern)

- Temporal behaviour descriptions (Scripts)
- Decision rules
- Tactics, doctrines
- Own-Forces status
- Vulnerability.

3.0 ON RULES, DOCTRINES, MILITARY DEFAULT BEHAVIOUR

Military actions are sequences of operations with strong dependencies. Operating sequences are interweaved by doctrines and rules. Adopting this fact, we can make the assumption that specific military operations could possibly be infered from antecedent behaviour or acts.

To depict the consequences of this standard behaviour, we give some examples for the principles which are contained in a movement order (which is incorporated in our example and e.g. [16], [24]). The order contains i.a. the column gap, time gap, road gap, march sequence, average speed, length of column, route order, march route and movement table (Tab. 1.). All these settings remain fixed for an untroubled march until the release point will be reached by the last unit of the column. A march column is under centralized command, is heading in one defined direction and units never cross each others course or overtake. A march column never splits. All these are examples for doctrines.

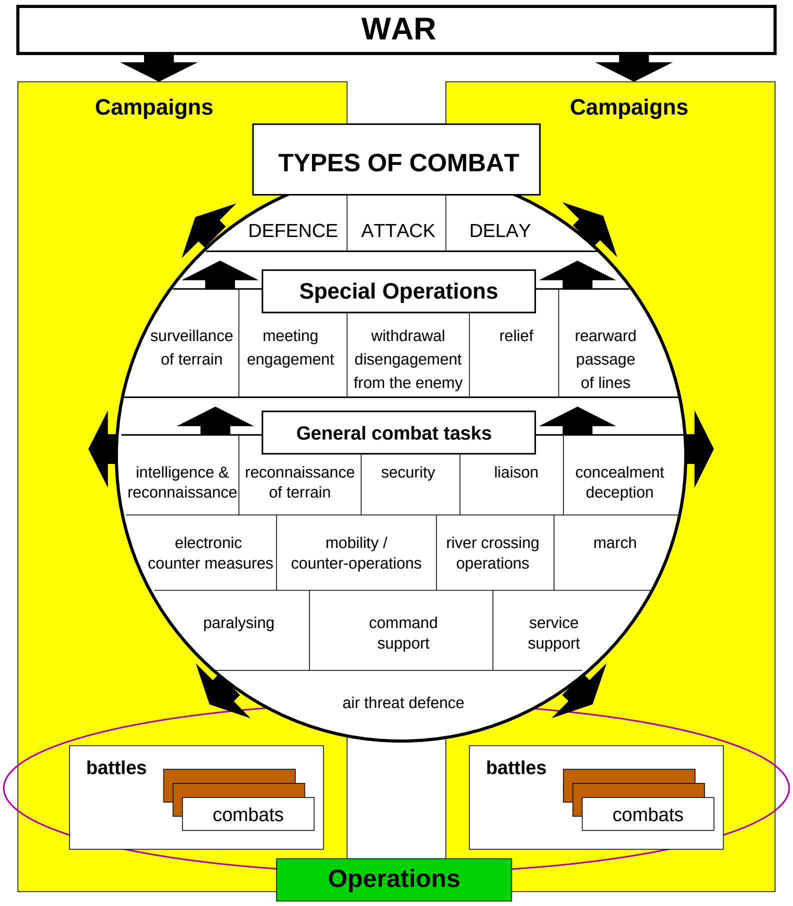


Figure 2: War as a structured system of separable interweaved operations, (from [20]).

Table 1: Extract from movement tables (from [20]).

Level	Motor Units	Distance [m]	Length of March Column [km]	Pass Time [min]			Remarks
				20 km/h	30 km/h	40 km/h	
Battery, Company	15	50	1	3-4	2-3	1-2	
		100	2	6-8	4-6	3-4	
Battalion	100	50	13	40-50	25-35	20-30	-5 March Units - Unit Distance 2-3 km
		100	20	60-80	40-60	30-40	
Brigade	900		120 (2x60)	3 hours	2 hours	1,5 hours	2 March Routes for 3 March Serials each

4.0 THE EXAMPLE

The developed concept for information fusion is basically independent from a specific military scenario. To illustrate it as model for further discussion we took the example of a military ambush attack on a deploying march column as a separable non-conventional military action from a comprehensive military scenario.

4.1 The Scenario

The political situation of the Republic of Bosnia and Hercegovina (BiH) (approx. 42° – 45° N, 16° – 19° E), as the scenario battlefield is shown in Fig. 3. This simulation scenario offers the possibility of describing tactical and operational intentions as concepts of operation on the level of individual missions of subordinate combat within a higher command level.

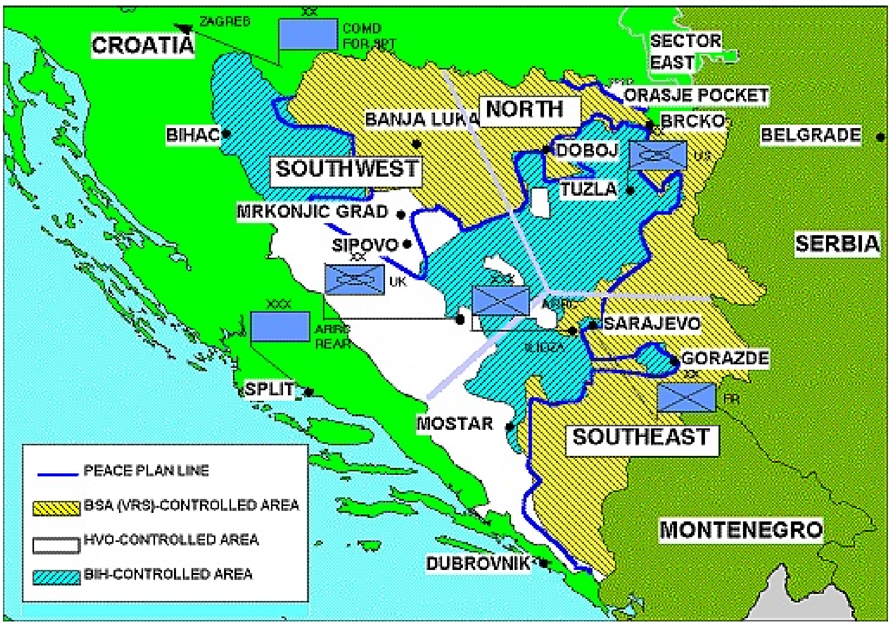


Figure 3: The Republic of Bosnia and Hercegovina (BiH) as theatre for this simulation example.

The two entities are the Federation of Bosnia-Hercegovina (the Moslem-Croat alliance), and the Serb Republic of Bosnia-Hercegovina (Republika Srpska). Since the declaration of independence of Slowenia and Croatia (6/25/1991) the complexity and diversity of conflicts within the region of former Yugoslavia is hardly to understand. This is still valid for the evaluation of possible future operational aims of war parties. Irrational and emotional based actions of military and political leaders contradict western logic. Animosity e.g. can solely be driven by emotions of revenge or regionally focused struggle for power. This can be explained by the geographical conditioned sectionalism, the unrelated peoples and religions, and the forced heteronomy.

4.2 Experimental Details

The Dayton Peace Accord (11/22-25/1995) not only defined the new frontiers, but also fixed the amount and allocation of remaining weapons and military personal of the former war factions within this territory. Therefore a global a priori overview can be accessed. The initial dislocation of all engaged forces can be seen as given information, due to the regulations fixed. The rival armies are separated by a demilitarised zone of about four kilometres width (Peace Plan or Inter-Entity Boundary Line (IEBL), Fig. 3.).

4.3 Report Situation

The message acquisition situation can be considered as adequate for the problem, due to a large amount of international groups and their observers, e.g. the International Police Task Force (IPTF), the United Nations High Commission on Refugees (UNHCR) or the United Nations Mine Action Centre (UNMAC). Fig. 4. shows some statistical information about the report collection of phase I of the scenario. The upper chart shows the absolute number of reports relative to the delay of the information. We see, e.g. 26 instantaneous information, respectively about 20 % of the information with individual timestamp, which are without any time delay between observation and report. The lower chart shows the report delay for each individual message. Several colours per bar indicate the report of more than one information with individual timestamp (up to four, message no. 63). Important for the testing of the concept is, that for the whole military report stream in our scenario the average information ‘age’ is marginally less than 15 Min. This should be kept in mind for the analysis tasks and possible expectations on the efficiency of a support concept exclusively dependent on HUMINT reports. The reports used for our example are a subset of this Phase I.

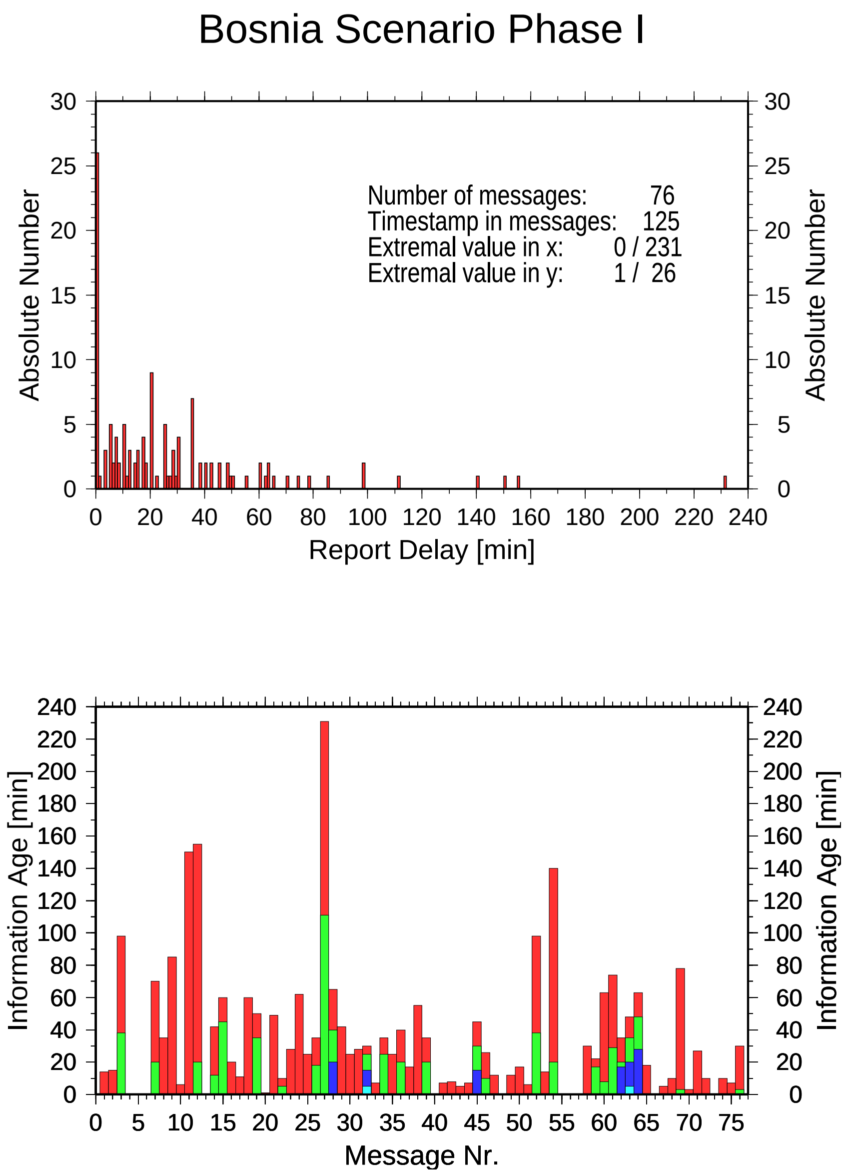


Figure 4: Charts displaying time delays between observations and reports.

4.4 Database Examples

The ambush attack scenario under examination can be described as follows: A place with high priority for its strategic importance is the city of Doboj in the central North of the country (Fig. 6.). The cities control is the guarantee for the access to the so called Oszren pocket, a hilly area south of the Spreča river still belonging to the Republika Srpska (RS). The Serb Republic Army (VRS) has some major installations in the city of Banja Luka about 70 km from the IEBL close to Doboj. A major road leads via Teslič, about 10 km west from the IEBL to the centre of actions Doboj, heading there from south-western direction. Therefore it is an important task for the Bosniak Army (ABiH) to get this road under safety catch as soon as possible. A ABiH brigade being based about 90 km south at Zenica is going to overtake this mission. This is done under the traditional policy of centralized command and control, which implies, that in case of an alarm or mobilization, the first command is already given a priori. This is military default behaviour, and therefore gives a chance to be calculable.

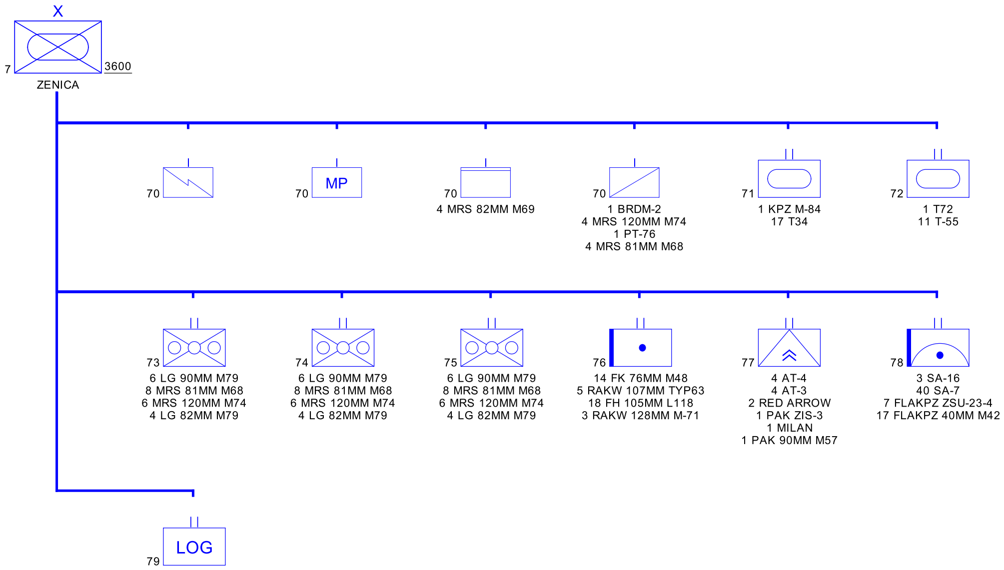


Figure 5: Organizational chart of the mechanized infantry brigade sending out the march column.

One problem for the brigade commander is, that despite the major road along the Spreča valley, which is unfortunately already under artillery fire by the VRS, only one secondary road exists leading directly (and fast) to the ordered area of operation. This road cuts the mountainous area close and along to the IEBL. There is no accessible road crossing this borderline along a distance of about 15 km for this region. The zone of the closest distance between two roads, one on the side of the RS and the other on ABiH territory, lies about six kilometre south of Teslič. The distance is less than five kilometre. After mobilization, the march column needs about four hours to get on the march route. To reach the mentioned zone of maximum convergence of the two roads, another three and a half hours are necessary (Fig. 6.).

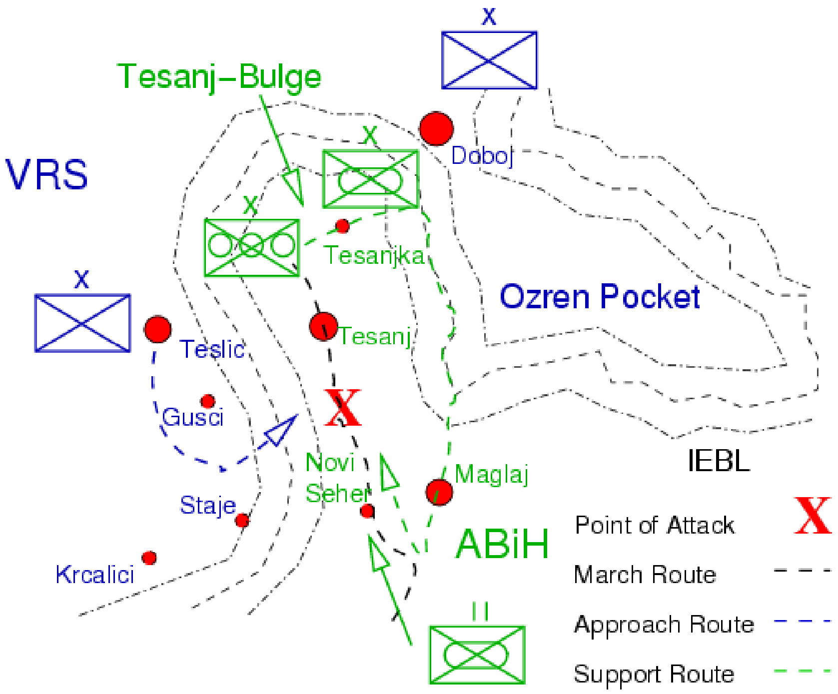


Figure 6: The situation of the military ambush attack.

Figure 6. shows the situation along the approximately last 25 km for the march column. The factions of the Army of the Serb Republic (VRS – blue colour) and the Bosniak Army (ABiH – green colour) aim at the defence (VRS), respectively attack (ABiH) of the city of Doboj (Fig. 2.). The relocation movements by the ABiH to the northern part of the Tesanj-Bulge can only be deferred by the weaker VRS. The latter was surprise attacked by the ABiH and needs time for preparing its over all defence.

5.0 SIMULATION AND FIRST RESULTS

Our testing of the concept starts with the following conditions. The commander of the brigade has to order his two tank battalion (71 and 72, Fig. 5.) to a release point at Tesanjka (Fig. 6.). The tanks are crucial equipment for the operational tasks of the Bosniak Army in the follow-up conflict. By setting up the movement table for the march, he is keeping in mind that the march units of these two tank battalions will be the most vulnerable elements within the column, in case the adversary succeeds to clear up the Bosniak intentions. About this time the Bosnian Army receives a report stating that four VRS military trucks loaded with an unknown number of soldiers and obviously some boxed equipment left their home barracks in Teslič, driving with (non-doctrinal) fast speed southwards. The same trucks return empty about three quarters of an hour later.

In the absence of any further situation context , the analysis of this information will end up pretty fast and the outcome will be pretty poor. The support tool may add four trucks to the knowledge base and an uncertain number of soldiers driving around to the situation brief database. But even with the sparse information from section 4, any brigade commander responsible for giving the order to a march column in that direction may be aware at least of the information of relocation movements in conspicuous fast speed. Analysing the situation, we face, that all information on dislocation of forces and equipment used in this ambush attack situation are common knowledge for both sides. In addition the geographical setting is well known to the factions. Even uncovering the most probable first action mission of the tank battalion is no difficult task, all the rest are simple algorithms, like standard Travelling-Salesman solver utilizing the facts from section 3. Just considering so far only the capabilities of the potentially war factions [21]. Nevertheless we assumed quite some context information in addition, which in fact cannot be neglected in carrying out real simulations of this kind. Some examples are i.e. the peace plan line or information on ethnic separation. Tests with a so far implemented support tool give a result of a necessary time consumption of 1.5 hours for the advance of a platoon between the two roads, equipped with light mortars, anti tank weapons, grenades and mines. Not taking into account a possible support of this operation by irregular local forces, even the regular tactical reserve troops might be able to carry out a surprise attack against the Bosniak tank battalion in time. The remaining time of almost one hour will be sufficient to set an ambush trap targeting the ABiH tank battalion. This is exactly what happens in the scenario. Variations in parameter settings and incorporation of randomly generated ‘report’ information allow the designing of a marching order with the same tool, which avoids to set any unit under added risk.

We can sum this up with the conclusion that the knowledge-based fusion of messages is an important task for the Intelligence Preparation of the Battlefield (IPB). The introduced method is capable for dealing with queries like, is the location S^* at the time $T^* + \delta T^*$ still save as a marching route. Or, is it possible for the enemy to reach a specific place with any special equipment in a given amount of time, just having a course idea on starting and release point.

6.0 CONCLUSION

The impulse for this work were questions like, can a template based information fusion support be given for highly complex non-conventional military operations which suffer by a couple of limitations in a priori setup and restrictions to the useful information available, which came up following the work in [16]. The answer to this topic is not only positive, far more it seems, that the template based problem solving strategy may be the only one dealing with this class of 'ill-posed' problems. The major constraint is to rely on the man-in-the-loop. Without dialogue and interaction no machine support seems to be feasible. We see in our of course still rudimentary conception especially benefits for operations in Low Intensity Conflicts (LIC [5], [22]). They are characterised by manageable dimensions in space and time. The creation and setup of templates for other examples of (non-conventional) military operations will be an important step for further succession.

7.0 REFERENCES

[1] M. Bedworth, *Probability Moderation for Multilevel Information Processing*, DRA Technical Report DRA/CIS(SEI)/651/8 M94.AS03BPO32/1, 1994.

[2] M. Bedworth and J. O'Brien, *The Omnibus Model: A New Model of Data Fusion?*, Proceeding IEEE AES Systems Magazine, pp. 30-36, 2000.

[3] J. Biermann, *Das Experimentalsystem HADES – Methodische Ansätze zur wissensbasierten Fusion strukturierter Informationen,* FKIE-Bericht Nr. 21, Wachtberg, 2000.

[4] J. Biermann, *HADES – A Knowledge-Based System for Message Interpretation and Situation Determination,* in: A. Pasqual del Pobil, J. Mira, M. Ali (Eds.), Tasks and Methods in Applied Artificial Intelligence, 11th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA-98-AIE, Lecture Notes in Artificial Intelligence 1416, Springer Verlag Berlin, 1998.

[5] P. Billing, *Eskalation und Deeskalation internationaler Konflikte. Ein Konfliktmodell auf der Grundlage der empirischen Auswertung von 288 internationalen Konflikten seit 1945*. 364 S., Frankfurt a.M., 1992.

[6] Bundesministerium der Verteidigung, Fü H III 2, *Rahmenvorgaben zur Erarbeitung von Standardszenaren für die nichttechnische Studienarbeit im Heer*, AZ 31-02-03/VS-NfD, 31+13 S., Bonn, 2002.

[7] J. Boyd, *A Discourse on Winning and Losing*, Maxwell AFB Lecture, 1987.

[8] V. Černov and S. Tuaev, *Skala cojuznica ili Ešče raz o zasade v gorach, (Der Fels als Verbündeter oder noch einmal über den Hinterhalt im Gebirge).* Bundessprachenamt - Referat SM 7, Auftrags-Nr.: F 0756, aus: Armejskij sbornik, Nr. 11/1999 /6/, pp. 25-27, 1999.

[9] B. El-Geresy and Ch. Jones, *Models and queries in a spatio-temporal GIS*, in: Innovations in GIS 7, GIS and Geocomputation, ed. by P. Atkinson and D. Martin. Taylor & Francis, London, pp. 27-39, 2000.

[10] A. Folscheid, *Terrorismus und Guerilla: Versuch einer Begriffsklärung*. Ethik Internationaler Konflikte (EIKon), Diskussionspaper 02/01, Saarbrücken, 9 S., 2001.

[11] A. Gad and M. Farooq, *Data Fusion Architecture for Maritime Surveillance*. Proceedings of the 5th Int. Conf. on Information Fusion, Fusion'2002, Annapolis, MD, pp. 448-455, 2002.

[12] R. Guth, *Checklisten für den Ersten Angriff: Handbuch für die Schutz- und Kriminalpolizei*. Richard Boorberg Verlag, Stuttgart, München, 150 S., 2000.

[13] T. Hoffmann, *Ein Reduktionssystem zur Klassifikation von Feindbeobachtungsmeldungen,* FFM-Bericht Nr. 481, 1997.

[14] T. Hoffmann, *Musterbasierte Informationsfusion in HADES,* FKIE-Bericht Nr. 7, Wachtberg, 1999.

[15] C. Karg, *Ein Verfahren zur Synthese von Konvoibewegungen,* FKIE-Bericht Nr. 10, Wachtberg, 1999.

[16] F. P. Lorenz and J. Biermann, *Knowledge-Based Fusion of Formets: Discussion of an Example,* Proceedings of the 5th Int. Conf. on Information Fusion, Fusion'2002, Annapolis, MD, pp. 374-379, 2002.

[17] F. P. Lorenz, *Ein Ansatz zur rechnerbasierten Unterstützung bei der Analyse einer Hinterhaltsbedrohung aus Meldungsinformationen und Situationskontext,* FKIE-Bericht Nr. 62, Wachtberg, 2003.

[18] NC3A Geo Officer Draft Proposal, *Raster Map Configuration Concept for SFOR GIS Database*, Information Systems Division, NATO C3 Agency -NATO UNCLASSIFIED-, 2+4 S., The Hague, March 1999.

[19] D. Noble, *Template-Based Data Fusion for Situation Assessment*, Tri-Service Data Fusion Symposium, John Hopkins Univ., pp. 156-161, 1987.

[20] R. Oestmann, *Dazu befehle ich ...!,* Handbuch für militärische Führer, 3. Auflage. Walhalla Fachverlag, Regensburg Berlin, 2000.

[21] V. Pötzsch, *Entwicklung und Implementierung von Modulen für die Prozesse Bewegung, Abnutzung und Aufklärung und Entwurf eines Rahmens für Führungsmodule für das Gefechtssimulationssystem COSIMAC*, Diplomarbeit, ID 14/97, Fakultät für Informatik, Univ. der Bundeswehr München, Neubiberg, Dezember 1997.

[22] M. Rauchensteiner and J. Ch. Allmeyer-Beck, *Formen des Krieges. Vom Mittelalter zum "Low-Intensity-Conflict"*, Verlag Styria, Graz Wien Köln, 208 S., 1991.

[23] U. Schade, *Ontologie-Entwicklung für Heeresanwendungen*. FKIE-Bericht Nr. 57, 41 S., Wachtberg, 2003.

[24] L. Schuberth und J. Biermann, *Marschkolonnensynthese in HADES ohne verlässliche Straßenkarten*, FKIE-Bericht Nr. 35, 66 S., Wachtberg, 2001.

[25] E. Shahbazian, D. Blodgett and P. Labbe, *The Extended OODA Model for Data Fusion Systems*, Fusion'2001, pp. FrB1-19-25, Aug. 2001.

[26] A. Shulsky, *Silent Warfare: Understanding the World of Intelligence*, Brassey's, 1993.

[27] OTL d.R. Spitzer, *Auf neuem Terrain: MilGeo-Beratung im Heer*, PzH II-01, pp. 3-4, 2001.

[28] A. Steinberg, C. Bowman, and F. White, *Revisions to the JDL Data Fusion Model*, SPIE, Vol. 3719, pp. 430-441, 1999.

[29] I. N. Vorob'ev, *Tactica boevych grupp,(Eine Taktik für Einsatztrupps)*. Streitkräfteamt Dez. Sprachendienst, Auftrags-Nr.: 368/01, Freigabe-Nr.: F 3236, aus: Voennaja Mysl', RF, pp. 27-30, 2001.

[30] E. Waltz and J. Llinas*, Multisensor Data Fusion,* Artech House, Boston London, 464 p., 1990.

[31] F. White*, A Model for Data Fusion,* Proc. Of the 1st Int. Symposium on Sensor Fusion, 1988.

Representation and Recognition of Uncertain Enemy Policies Using Statistical Models

Robert Suzić
Tel: +46 (0) 8 55 50 32 17
Swedish Defence Research Agency (FOI)
SE-172 90 Stockholm
SWEDEN

Email: robert.suzic@foi.se ; rsu@nada.kth.se

ABSTRACT

In this work we extend from the single agent to the on-line multi-agent stochastic policy recognition problem using a network structure. By using knowledge of agents' interrelations we can create a policy structure that is compatible with that of a hostile military organisation. Using this approach we make use of existing knowledge about the military organisation and thereby strongly reduce the size of the hypothesis space. In this way we are able to bring down the problem complexity to a level that is tractable. Also, by using statistical models in policy recognition we are able to deal with uncertainty in a consistent way. This means that we have achieved improved policy recognition robustness.

We have developed a ***proof of concept*** *Bayesian Network model. For the information fusion purpose, we show with our model that it is possible to integrate the pre-processed uncertain dynamical sensor data such as the enemy position and combine this knowledge with terrain data and uncertain a priori knowledge such as the doctrine knowledge to infer multi-agent policy in a robust and statistically sound manner.*

1.0 INTRODUCTION

Dealing with uncertain information in a complex and in some cases chaotic environment is the difficult task for military commanders. A set of methods which improve the process of collecting and reasoning about uncertain information is called *information fusion*. The goal of information fusion is to describe a particular state of the world of interest by making best possible use of all available information. In military applications this can apply to anything from the position and type of hostile forces to an enemy's plans and intentions. This paper shows how our knowledge about the enemy can be represented and his policies recognised. We choose Bayesian Networks (BN) as the method for representation of our knowledge about the enemy, designated *static knowledge*, and Dynamic Bayesian Networks (DBN) for inference based on sensor data, designated *dynamic knowledge*. Those problems, as many others, must of course be addressed in the operational system. In order to focus on policy recognition, in this work we do not deal with the classical identification and association problems.

The stochastic nature of policies is derived from the fact that we do not have full knowledge about the enemy and his actions. Instead, policies are represented as discrete probability density functions on different modelling levels. This means that military commanders should not only pay attention to the policy with the highest probability but to all policies that have a significant probability to occur.

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

In this paper we will present the dynamical stochastic policy recognition problem. By dynamical we mean that policy can change over time and as soon as new observations arrive. The automatic process of policy recognition is performed for each new observation. The process is on-line and involves many agents (units) acting together. Thus we characterize the task as *on-line multi-agent stochastic policy recognition.*

Situation awareness plays a decisive role in achieving information and next-turn decision superiority on the battlefield. "Battle space is an abstract notion that includes not only the physical geography of a conflict but also the plans, goals, resources and activities of all combatants prior to, and during, a battle and during the activities leading to the battle" [1]. It is generally difficult to derive conclusions about the enemy's intentions from a chaotic, uncertain and complex environment, see [2]. Military commanders have to be able to act fast. In order to achieve agility military commanders need to have good situation awareness. On-line policy recognition gives users, military commanders in this case, hints about what the enemy is going to do next, provided relevant sensor information and a priori knowledge about the enemy is present. The next step the military commander has to make is to investigate if the warning received from the system is serious. If the system correctly deduces that an attacking policy is the most likely to happen or that the probability is high that the military commander's own force has been discovered, then situation awareness, i.e., the ability to respond in time, improves and the security level increases by using the on-line multi-agent stochastic policy recognition method.

This work has been performed within the Information Fusion strategic kernel project at the Swedish Defence Research Agency (FOI) and the Royal Institute of Technology (KTH) in Sweden.

2.0 POLICY RECOGNITION AND INFORMATION FUSION

2.1 Aim

On-line multi-agent stochastic policy recognition aims to detect which policies an agent or group of agents are executing by observing the agents' actions and by using *a priori* knowledge about the agents in a noisy environment.

The method chosen for the representation of this task is Bayesian inference using dynamic Bayesian nets and other statistical models. The inference is intended to derive belief measures for enemy plans. Thus, the paper provides yet another example, see also [3, 4, 5, 6], of how Bayesian Networks (also known as probabilistic networks or belief networks) can be used to support situation awareness.

In everyday life people recognise intentions of other people, and in some cases intentions of animals. Here is a burglary example: How can we recognise that a thief or a number of thieves are trying to rob our neighbour's house? You can be fairly certain that a burglary is taking place if you see that some person is trying to get in through the window of your neighbour's house. But what if in the next moment some member of your neighbour's family opens the window to the "burglar"? This means that the probability of a burglary is lowered and you update your belief dynamically (on-line). The hypothesis "an inappropriate visit" gets more support. The probability density function over the hypothesis space changes. This is a main feature in BN modelling, illustrated by similar examples in [10].

In military applications the issue is how to recognise certain military behaviours of the enemy. Using the movement pattern, speed, distance, weather, maneuverability distance to presumptive target, etc., it might be possible to fuse the acquired knowledge about the enemy and use it in policy recognition. The advantage would be that military commanders, having better knowledge about the enemy's intentions, will be able to act earlier. The ability to act preventively increases as well. In this paper we claim that by using our knowledge about hostile force doctrines and fusing this knowledge with sensor information we can recognise certain military behaviours of the hostile force. Our aim is to demonstrate the statement above in a proof-of-concept model. This work uses a DBN model to represent our beliefs about hostile force units.

2.2 Modeling Approach

As mentioned in the introduction, in this work we do not deal with the classical identification and association problems. A further delimitation is that sensor reports are assumed to be pre-processed so that the input to our model are positions of the enemy represented as discrete probability density functions. In the present work we create a model of a hostile tank company. The model is hierarchical, corresponding to an hierarchical policy structure. The company consists of three tank platoons, each platoon containing three tanks. For each level there is a certain set of *policies* that are invoked by the higher level. The simplest policies, their *atoms*, consist only of a set of *actions*. In this example the simplest policy is the tank (group) policy. More complex policies consist of other policies, also referred to as *sub-policies*, or a mixture of sub-policies and actions. Higher level policies invoke lower level policies down to their action atoms.

The connection between different policy levels is modeled using conditional probabilities which represent uncertain causal relationships. A policy of agent i on level k depends on the policy that a superior agent is executing on level $k + 1$ and on the state of agent i in the previous time step. Single agent stochastic policy recognition is represented by an Abstract Hidden Markov Model (AHMM) in [7]. In Kevin Murphy's Ph.D. work [8] an AHMM is described as a DBN. In this work we extend the on-line single agent stochastic policy recognition problem to the corresponding on-line multi-agent stochastic policy recognition problem by using a network structure that allows many agents. By using knowledge of agents' (hostile units') interrelations we create a policy structure that is compatible with that of a hostile military organisation. Using this approach we both make use of existing knowledge about the military organisation and strongly reduce the size of the hypothesis space. In this way we are able to bring down the problem complexity to a level that is tractable.

There are several advantages with this kind of representation. One of them is that one can build policy hierarchies modularly, in a way that corresponds to military organisational hierarchy. Using this kind of approach we can facilitate:

- the process of knowledge reuse,
- the process of updating knowledge,
- verification and validation processes.

When observing actions of an agent (hostile unit) we face several uncertainties. One kind of these uncertainties are the stochastic outcomes of actions, see [7, 9]. An agent intends to carry out some actions to achieve his goals but he is in some cases prevented from using these actions. In some situations, the effects of the agent's actions therefore do not correspond to the goal state. E.g., in some cases one tank unit is planning to take a hill but during some part of the plan execution it is understood that this is not possible because of bad maneuverability. In our model we do not care if the agent is rational. Instead we model and analyse if some policies that the agent is executing are more likely than others. This modelling approach respects terrain data, a list of the agent's possible targets and our knowledge about the agent's behaviour. Other obstacles in the policy recognition problem are uncertain observations and incomplete knowledge about the agent and his behaviour.

In order to be able to handle those different types of uncertainties we use Bayesian Networks as the modelling technique. See [8, 10, 11] for an introduction to BN and DBN. A Bayesian Network is a set of uncertain causal connections between variables of interest. The connections are modeled as conditional probability distributions. A BN uses a distributed representation of the current state. Thus we can also represent the probability density function in a distributed way. There are three types of variables in our type of BN:

1) hypothesis variables,
2) information or evidence variables,
3) mediating variables.

In our case, the hypothesis variables are the agent's policies. Hypothesis variables are either impossible or too costly to observe. Information or evidence variables represent our observation models, e. g. the uncertain position of the enemy. Mediating variables are introduced for special purposes. In this case we use mediating variables to model terrain restrictions and the agent's doctrines.

The multi-agent policy hierarchy in this paper consists of the policy of the company at the top level, the platoon policies at the next level and the tank policies. Our BN modelling approach is that the company policy causes change in platoon policies and platoon policies cause change in group (tank) policies. One reason for using this modelling approach is that this model follows military hierarchy; commanders give orders to their subordinates who are superior at the next level. The other reason is to minimize complexity, see also [7, 13]. The ability to represent uncertain causal relationships is one of the main advantages of Bayesian Networks. A BN can only represent a snapshot of the situation, whereas a DBN allows new evidence to arrive while taking into account old variable values.

Evidence nodes propagate new information through the network by updating variables of the DBN.

2.3 Model Description

Our model was built by using both a Bayesian Network and a Dynamic Bayesian Network representation. The variables in the BN graph are represented by nodes while uncertain causal relationships are represented by arcs. The description of uncertain causal relationships is performed by using conditional dependence probabilities between the variables. By using BN we are able to express our subjective belief. As an example, we can state that one type of unit formation has lower probability to occur when maneuverability is bad, see also "The Movement-Analysis Challenge Problem" in [1].

In our model we make a distinction between hypothesis nodes (policies nodes, discovery node), mediating nodes (restriction nodes such as the terrain, doctrinal nodes such as formation) and observation nodes (sensor nodes).

Policy hierarchy is represented by the BN. Policy for each agent (hostile unit) is represented as a BN node. The simplest policy is on tank (group) level, $k = 0$.

Tank i's policy variable, $\pi_{o,i}$, has the following discrete states:
$\pi_{o,i}$:= <agent i is moving in the direction of our own force, agent i is moving in the direction opposite to our own force, agent i is moving in neutral direction>
On the next level we have the policy of the tank platoon i at level $k = 1$:
$\pi_{1,i}$:= <attack, defence, reconnaissance, march>
Finally on the top level we have the policy of the tank company at level $k = 2$:
$\pi_{2,i}$:= <frontal attack, frontal attack and flange attack, defence, delay battle, march>

In Figure 1 we show a Bayesian network representing the policy hierarchy model of a hostile company.

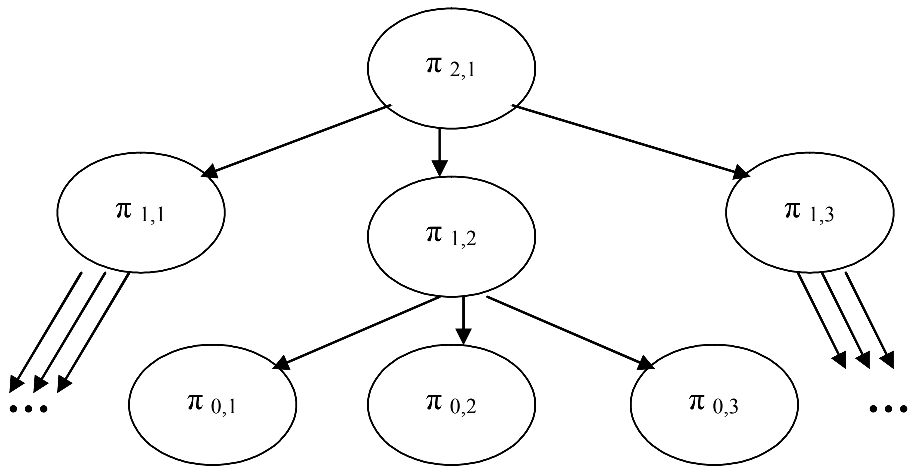


Figure 1: Policy hierarchy.

In this model we combine terrain representation with uncertainty about hostile force position. The output of this fusion is represented in a BN node called maneuverability. We find it important to model information about the terrain in the policy recognition problem. In particular environments, particular policies are assumed to be more probable than in some other environments. In our DBN model the representation of environment is limited to terrain, visibility, and the possibility to find cover. The variables formation, distance to presumptive target, and direction of guns are used to connect observations to different policies. We call those nodes strictly doctrinal. The hostile company model is implemented in MATLAB using K. Murphy's BN package [8,16]. The DBN representation of a hostile company is visualised in Figure 2.

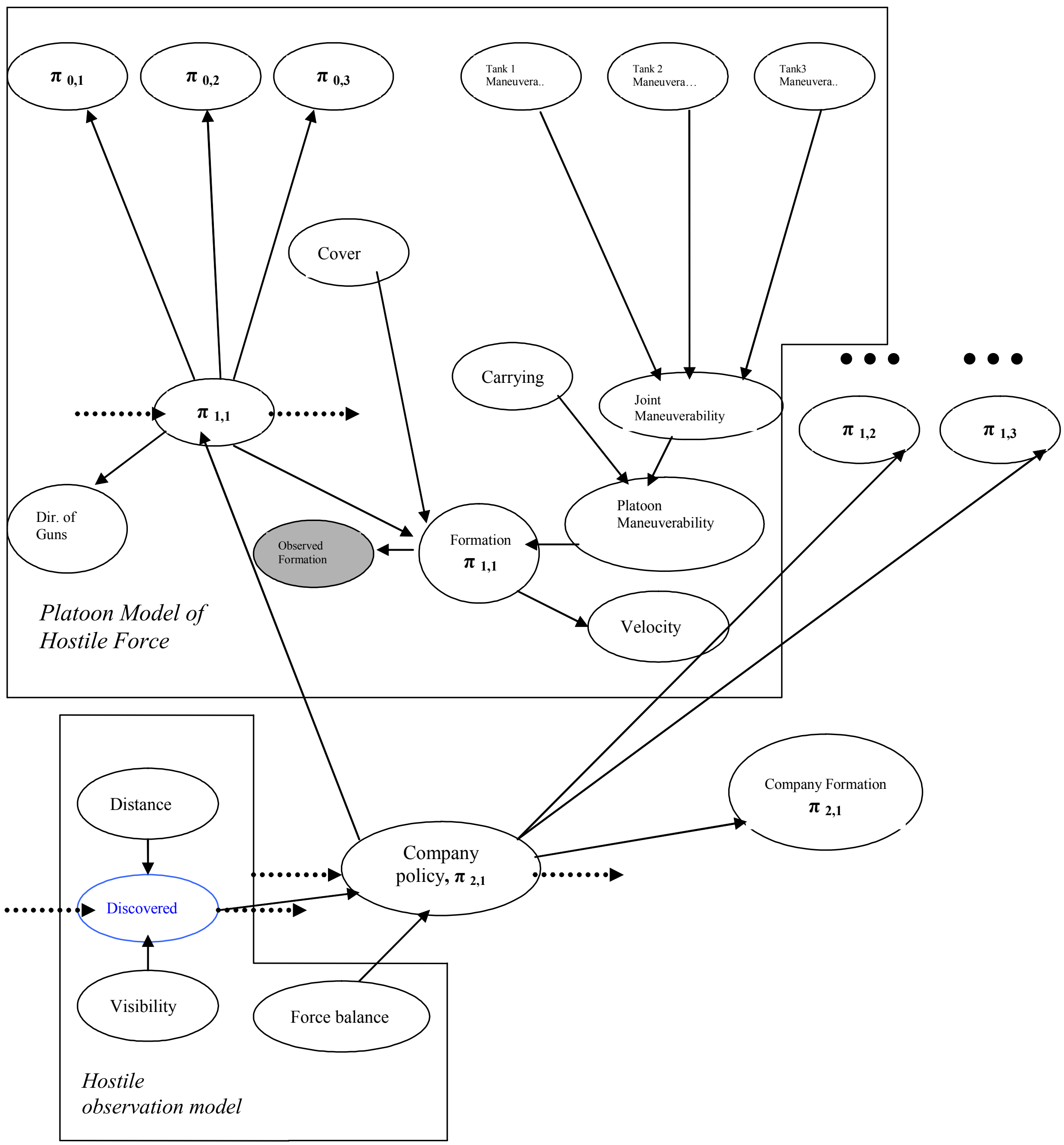


Figure 2: Bayesian Network representing a hostile company.

The whole network contains 57 nodes. The purpose of the network is to make qualified estimation of the opponent's behaviour based on observations, knowledge about opponent's doctrines as well as data from the terrain. If the whole terrain would be represented as Bayesian nodes then the problem of high computational complexity would appear. To represent the terrain in this model we use a fragmented representation of it. Such nodes as *Tank Maneuverability*, *Carrying* and *Cover* represent terrain. The probability distribution of the value *Tank Maneuverability* depends on many factors. For instance, the tank maneuverability depends on slopes in the surroundings and if there are any significant obstacles such as large water areas. It depends also on if the tank is on road or not. In this implementation we only use this fact when representing tank manoeuvrability. The carrying capacity depends on the terrain's capability to carry heavy loads such as tanks. The *Cover* depends, in our implementation, on vegetation. In a pre-processing stage, before the simulation is started, we performed a classification of the terrain for each pixel of the map that we had read into MATLAB. That is, the classification of each pixel is based on whether it represents a road or not. In the same manner we performed classification on vegetation of the terrain. As the result we get data that represents different aspects of the terrain.

We use MATLAB functions which take an estimate of the tank position as input and make a weighted calculation reading the terrain data that corresponds to the tank position and its close surrounding. The weighting function should correspond to the uncertainty of the observation. The greater the uncertainty about enemy tank position, the more data representing the surroundings should be taken into concern. If an observation is pretty certain then weighting should be high on data representing an estimate of the tank position. In our example we use only rectangular distributions over a set of pixels. The output of the functions are probability distributions. Those values are entered as soft evidence in Bayesian nodes. The example of such nodes are *Cover* and *Tank Maneuverability*. The variable *Tank Manoeuvrability* has two states, Good and Bad. A MATLAB function assigns the probabilities over variable *Tank Maneuverability* states for each observation.

To facilitate the defining of our a priori knowledge of the variable *Platoon Maneuverability* we introduce a mediating node designated as *Joint Maneuverability.* Consequently, by using the mediating node the variable *Platoon Maneuverability* gets two parents instead of four. Thus we define distributions that depend on two parents instead of four. This method is called *divorcing*, see also [10] p. 52.

One of the key variables that reveal enemy behaviour and intentions is the enemy formation. It is represented as a Bayesian node in the network. According to doctrine manuals, when the enemy has the intention to attack it usually attacks in battle line formation. When transporting to a certain destination the enemy transports in march formation. There are many reasons why the enemy may not use a certain type of formation. One of the factors that have influence on building a formation is the environment. When the platoon maneuverability is bad the enemy will not have the opportunity to attack in battle line and the probability of the formation type battle triangle increases. The probability that the enemy is performing reconnaissance increases when the enemy is moving in a stepped formation. As for all Bayesian variables in the network, we define a priori knowledge about the variable *Formation* as the probability density function **P**(*Formation* | *Platoon Policy, Cover, Platoon Maneuverability*). In Table 1 we describe the distribution that we have implemented. We found the variable *Formation* important because it models the connection between doctrines and environment in a statistical and therefore soft manner.

Table 1: A Priori Distribution of the Formation variable

P (Formation | Platoon Policy, Cover, Platoon Maneuverability)						
Policy Platoon	**Cover**	**Maneuverability**	Formation value = Lead	Formation value = Battle Line	Formation value = Stepped Formation	Formation value = Battle Triangle
1)Attack	**Good**	**Good**	**0.05**	**0.75**	**0.15**	**0.05**
2)Defend	Good	Good	0.05	0.60	0.3	0.05
3)Reconnaissance	Good	Good	0.35	0.2	0.25	0.2
4)March	Good	Good	0.9	0.01	0.01	0.08
5)Attack	Bad	Good	0.01	0.79	0.1	0.1
6)Defend	Bad	Good	0.01	0.53	0.41	0.05
7)Reconnaissance	Bad	Good	0.3	0.3	0.35	0.05
8)March	Bad	Good	0.7	0.2	0.08	0.02
9)Attack	**Good**	**Bad**	**0.08**	**0.5**	**0.25**	**0.17**
10)Defend	Good	Bad	0.08	0.55	0.35	0.02
11)Reconnaissance	Good	Bad	0.3	0.13	0.47	0.1
12)March	Good	Bad	0.75	0.01	0.04	0.2
13)Attack	Bad	Bad	0.08	0.3	0.2	0.42
14)Defend	Bad	Bad	0.02	0.2	0.5	0.28
15)Reconnaissance	Bad	Bad	0.2	0.38	0.4	0.02
16)March	Bad	Bad	0.56	0.08	0.12	0.24

In the first row where *Platoon Policy* = Attack, *Cover* = Good and *Manoeuvrability* = Good, the probability for *Formation* = Battle Line is 75 % while the hypothesis *Formation* = Battle triangle has only 5 % support. However, if the maneuverability is bad the enemy may not be able to hold the battle line formation. Therefore if *Maneuverability* = Bad then the triangle formation gets greater support, 17 %, see row nine. This is easy to explain by the fact that the enemy will not have good coordination ability and enough free space to attack in the battle line formation when manoeuvrability is bad.

In the same manner we define knowledge that the most probable formation is Lead when the platoon is performing march. The lead formation is also under influence of the terrain and the cover. Therefore probability for *Formation* = Lead decreases from 90 % to 75 % when maneuverability changes from Good to Bad. If *Maneuverability* = Bad and *Cover* = Bad this probability would decrease further to 56 %.

A natural question that arises is: Why these numbers? The simple answer is that we model this distribution to make a distinction regarding which hypothesis is more probable. The purpose was not to define, and we doubt that is possible in this application area, how probable a certain hypothesis is. The numbers can differ. E. g. in the previous example, row three in Table 1, the a priori support for the hypothesis Battle Line could be redefined and set to 70 %. At the same time the support for the remaining hypotheses would increase. However, the qualitative representation should remain the same considering the knowledge gained from doctrine manuals. That means that the shape of the probability density function should remain the same. In this example, row three, the support for the hypothesis Battle Line should be higher than the support for any other hypothesis. However, how much higher this support should be is an open issue. This kind of approach is part of the Bayesian way to solve statistical problems. Instead of expressing for which confidence interval the certain hypotheses is valid, we compare hypotheses as in this example. The output in this case is the distribution of the different hypothesis variables such as policies for platoon, company policy and our guess if the enemy has discovered us. In the same manner as we modeled the input data, attention is focused on which hypotheses are the most probable to occur given the a priori knowledge and sensor reports. Second priority, but also of importance, is their quantitative representation. E. g. we show in section three tables that represent the most probable and the least probable hypotheses for different time steps.

According to our model, the variable *Observed Formation* depends on the actual formation.

Because of the environment, uncertain observations and enemy's coordination problems we are not always able to observe his formation pattern. It is the rule rather than the exception that the enemy's formations do not follow the same geometrical properties as described in the doctrine manuals. Therefore we introduced a heuristic function in MATLAB that takes the estimates of the tank positions as inputs and as output returns the distribution of the observed formation's values. The result is entered as soft evidence in the variable *Observed Formation*. By Bayes rule it has influence on the value *Formation*.

The node *Force Balance* represents status on the battle field. If the enemy is stronger the probability that he will attack is higher than if he is weaker.

The company policy is also under strong influence of discovery. To trigger attack behaviour the enemy has first to discover our forces. Therefore we also build a limited model that represents hostile force ability to observe us. This model is a part of the network. We assume that the discovery depends only on the distance from the enemy to our forces and visibility.

Another aspect of the network representation is that some node values are under influence of the node values of the previous time step. The BN gets dynamic properties. This kind of the network is denoted as DBN. The topology of the network remains the same but the node values vary with the time. The distribution of the variable is conditioned by its parent(s) and the node values of the previous time step.

The list of nodes that are time dependent is the following: *Platoon policies*, *Company policy* and *Discovered*. The reason for introducing those variables as time dependent is purely logical. If enemy has discovered us in one time step, $t - 1$, it implies that the probability that he rediscovers us, finds us again, is higher in the next time step, t. The similar property follows company and platoon policy.

We define a conditional probability density function for the *Discovered* node in Table 2.

Table 2: A priori Distribution of the variable Discover

P(Discovered (t)| Discovered(t-1), Distance(t),Visibility(t))				
Discovered (t-1)	**Distance(t)**	**Visibility(t)**	Discovered = Yes	Discovered = No
Yes	Near	Good	0.99	0.01
No	Near	Good	0.8	0.2
Yes	Neutral	Good	0.9	0.1
No	Neutral	Good	0.65	0.35
Yes	Far away	Good	0.8	0.2
No	Far away	Good	0.4	0.6
Yes	Near	Bad	0.9	0.1
No	Near	Bad	0.7	0.3
Yes	Neutral	Bad	0.8	0.2
No	Neutral	Bad	0.4	0.6
Yes	Far away	Bad	0.6	0.4
No	Far away	Bad	0.15	0.8

In the order to achieve tactical superiority on the battlefield, tanks maneuver very often. That implies for our modeling approach, that we do not connect nodes representing tank policies over time. We connect platoon policies and company policy over time because of higher inertia. If the whole company is attacking at one time step there is significant probability that the company will continue to execute its policy Attack in the next time step.

To variable *Company formation* we assigned abstract type values. We found it difficult to make automatic formation recognition in a manner that the doctrine manual describe formation patterns for company formation. Instead we defined values for the company formation as follows: Lead, Together, Spread, and Very Spread. Again, we use a MATLAB function that reads the estimates of all enemy tank's positions and assigns a value to *Company formation*.

3.0 SCENARIOS, SIMULATION AND RESULTS

We have visualised the movement of a hostile company unit. The hostile company unit consists of three platoons. Each platoon consists of three tanks. There are two scenarios that we use for this simulation. The first one is an attacking scenario and the second one is a movement, march, scenario. The first phase of the both scenarios is the same. In the beginning, the tanks are marching in a neutral direction. The company is not spread over a vast area and the distance to our forces is large. MATLAB functions take as input enemy's tank positions, position of our forces and terrain data. The outputs of those functions are the probability distributions over the variables *Cover, Distance, Tank Policy,* (Platoon) *Formation, Company Formation* and *Tank Maneuverability*. This operation is performed for each tank. The uncertainty about the observations takes into account how the information of the pixels of interest should be weighted. In our example we used a simple approximation of the rectangular distribution over each position.

In time step two observations about current positions and direction of the enemy arrive, see Figure 3. We performed the computation on-line and the results for this step are documented in Table 3.

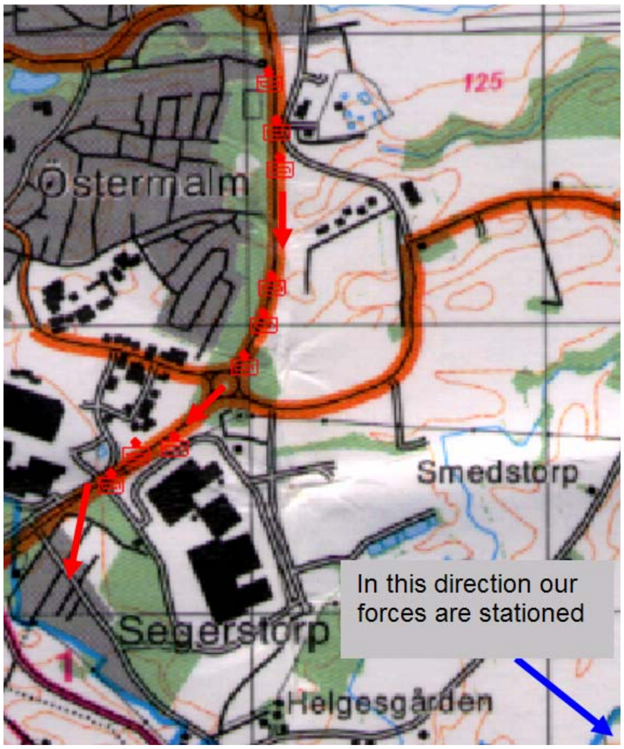


Figure 3: The enemy company is approaching at time step two.

Table 3: Company, platoon policies and values

Most probable and least probable policies at time step two (Figure 3)		
	Most probable states ***(Probability %)***	**Least probable** ***(Probability %)***
Company Policy	March *(47 %)* Flange attack *(21 %)*	Defence *(3 %)*
Platoon One Policy	March *(80 %)*	Defence *(5 %)*
Platoon Two Policy	March *(70 %)*	Defence *(7 %)* Attack *(8 %)*
Platoon Three Policy	Attack *(90 %)*	Defence *(0.5 %)* March *(1 %)*

The probability that the enemy company has discovered us is 33 %. The most probable state of platoon three, according to Table 3, is Attack. The explanation is that this platoon is approaching in the direction towards us. However, the most probable state of the company policy is March. This is achieved by weighting with other nodes including the two other platoon policies. It is usually difficult to infer intentions of a single tank if this unit is not put in a greater context such as platoon or company.

Figure 4: The enemy company is passing by.

After some time observations are received. The enemy begins to approach and then passes by, see Figure 4. The movement and formation pattern indicates that enemy has not discovered us although the distance is short. Thus, the most probable hypothesis is that the hostile force is performing march with probability 98 %. The most probable platoon two policy is March and is 97 % in this case. But for platoon three there is a probability of 44 % that the platoon will attack us and only 27 % probability that this platoon is marching.

The probability that the enemy company has discovered us in this time step (time step = 6) is only 7 %.

3.1 Attacking Scenario

At this time step we divided the scenario in two. The first one is the attacking scenario.

The enemy forces move toward us in two line formations from two different directions.

Figure 5: The enemy approaches in attacking formation.

The first and the third platoon change direction to us and begin approaching. The second platoon continues to move in the same, neutral, direction and follows the road, see Figure 5. We obtain the following results and show them in Table 4.

Table 4: The results for time step seven

Most probable and least probable policies at time step seven in attacking scenario (see Figure 7)		
	Most probable states ***(Probability %)***	**Least probable** ***(Probability%)***
Company Policy	Flange attack *(48 %)* Delay Battle *(25 %)* Frontal *(23 %)*	Defence *(0.5 %)* March *(3 %)*
Platoon One Policy	Attack *(96 %)* Reconnaissance *(3%)*	March *(0.3 %)* Defence *(0.7 %)*
Platoon Two Policy	March *(92 %)* Reconnaissance *(6 %)*	Attack *(1,4 %)* Defence *(1,6)*
Platoon Three Policy	Attack *(61%)* Reconnaissance *(25 %)* Defence *(9 %)*	March *(5 %)*

The formation of the company is much spread out and the most probable type of attack is flange attack combined with frontal attack. On platoon level the platoon two seems to still perform march while the other two platoons have attacking policy as the most likely behaviour.

The probability that we have been discovered is increased to 51 %.

Figure 6: The company movement towards us at time step seven.

In the next time step, eight, observations arrive. The movement pattern is towards us and the formations of all the platoons are battle lines. The probability of a frontal attack increases to 54 % and the probability of a combined frontal and flange attack decreases to 43 %. This means that the probability that the company will attack is extremely high. If we add probabilities of both attacking policies it is 97 %. The probability that we have been discovered increases further to 67 %.

3.2 Marching Scenario

At time step six the enemy will continue the march without performing any attacking movements, see Figure 5. The tank positions are the same as in the attacking scenario but the direction is different. The enemy force will follow the road.

Therefore the probabilities of each hypothesis will be different.

Table 5: Marching enemy

Most probable and least probable policies at time step seven in marching scenario (see Figure 5)		
	Most probable states ***(Probability %)***	**Least probable** ***(Probability %)***
Company Policy	March *(90 %)* Delay Battle *(6 %)*	Attack *(0.1 %)*
Platoon One Policy	March *(70 %)* Reconnaissance *(23%)*	Attack *(0.09 %)*
Platoon Two Policy	March *(97 %)*	Attack *(0,04 %)*
Platoon Three Policy	March *(96 %)*	Attack *(0,08%)*

On both company and platoon levels there is strong support for the hypothesis value March, see Table 5. The probability of attack is low. There is only weak support for the hypothesis Delay Battle. The probability that we have been discovered drops to 10 %.

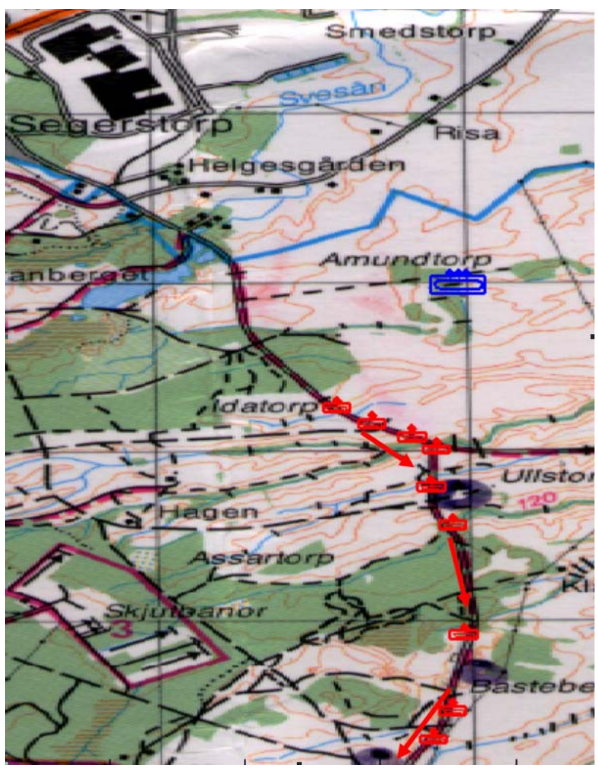


Figure 7: The company movement running away at time step eight marching scenario.

The support for the hypothesis March will increase when the enemy company continues to march in the direction opposite to our position. In Figure 7, we give an example of how the enemy force is continuing its march and how this hypothesis gets even more support. The probability for the company performing march is increased to 98 % while the probability that we have been discovered drops further to 7 %.

4.0 CONTRIBUTIONS AND CONCLUSIONS

One of the difficult tasks in plan recognition is the interaction between the terrain and recognition processes, see [12]. To decrease complexity we do not use terrain representation by Bayesian nodes. The terrain is instead represented in different data sets. When new observations arrive the MATLAB functions combine them with the relevant terrain data and supply terrain nodes with soft evidence. This information is propagated through the network as the statistical inference process.

The policies do not have stopping conditions in the same manner as in the policy recognition problem where single agents are moving through different rooms, see [7]. In [7], there are distinct sets of policies for each room. The stopping conditions are the walls of the rooms. In the military domain it is usually hard to find such boundaries, but you can connect certain terrain types and certain military policies.

By using statistical models in plan recognition we are able to deal with uncertainty in a consistent way. This means that we have achieved improved policy recognition robustness. In the paper [14], a correlation formula between observations and objects is presented but the drawback is that the plan recognition presented does not take different abstraction levels into account.

We state that our model also has a vague qualitative interpretation. The structure of nodes and arcs explains the model in a qualitative way, see [15]. Also, the results may be interpreted comparing hypotheses instead of expressing how probable they are. DBN are used in this implementation by representing our knowledge in a fragmented manner. By using this kind of approach we obtain a better overview. The knowledge is transparent and the black box concept is avoided. Our model is still incomplete in the sense that we do not incorporate the association and identification problems.

One of the main challenges in this work was to model and extend the problem structure from previously studied single-agent policy recognition to multi-agent policy recognition that is applicable in military information fusion. This was achieved by combing several modelling approaches. The first issue was to model agents and interrelations between different agents on different abstraction levels in the DBN model. By using DBN we also connect policies over time. The second issue was to represent the relevant terrain for the DBN model in a fragmented and dynamical manner for each new observation. Finally, we implemented functions for recognition of the physical relations between the agents such as formations and enemy tank policies. Our simulation results show that it is possible to recognise certain multi-agent stochastic policies on-line.

5.0 ACKNOWLEDGMENTS

I would like to express my thankfulness to: Stefan Arnborg (KTH), Per Svensson (FOI), Magnus Sparf (FOI), Claes Henrikson (FM), Ronnie Johansson (KTH), Joel Brynielsson (KTH), Klas Wallenius (SAAB), Karim Oukbir (KTH), Jesper Fredriksson (KTH), Pontus Svensson (FOI), Hedvig Sidenblad (FOI), Fredrik Samulsson (KTH/FOI) and Mateusz Moszkowicz (SU).

6.0 REFERENCES

[1] Paul Cohen, Robert Schrag, Eric Jones, Adam Pease, Albert Lin, Barbara Starr, David Gunning and Murray Burke, "The DARPA High-Performance Knowledge Bases Project", AI Magazine, Winter 1998, p. 30.

[2] Pearce, A. R. and Heinze, C. A. and Goss, S. "Enabling perception for plan recognition in multi-agent air-mission simulations". Proc. Fourth International Conference on Multi-Agent Systems (ICMASS2000), pp 427-428, 2000.

[3] K.B. Laskey, S.M. Mahoney and E. Wright "Hypothesis Management in Situation-Specific Network Construction", Proc. Seventeenth Conference on Uncertainty in Artificial Intelligence, 2001.

[4] S. Mahoney and K.B. Laskey "Constructing Situation Specific Networks", Proc. Fourteenth Conference on Uncertainty in Artificial Intelligence, 1998.

[5] G. Laskey, K. Laskey "Combat Identification with Bayesian Networks", Proc. Command and Control Research and Technology Symposium 2002.

[6] A.J. Pfeffer "Probabilistic Reasoning for Complex Systems" (Ch. 8.2, p. 257), Stanford University, March 1999, USA.

[7] Bui, H. H., Venkatesh, S., and West, G. "Policy recognition in the Abstract Hidden Markov Model", Journal of Artificial Intelligence Research, 17: 451-499, 2002.

[8] Kevin P. Murphy, "Dynamic Bayesian Networks: Representation, Inference and Learning", University of California, Berkeley, 2002, USA.

[9] Jim Blythe, "Decision-Theoretic Planning", AI Magazine, Summer 1999.

[10] Finn V. Jensen, "An introduction to Bayesian Networks", UCL press, London, ISBN: 1-85728-332-5, 1996.

[11] Bruce D'Ambrosio, "Inference in Bayesian Networks", AI Magazine, p. 37-54.

[12] Robert P. Goldman, Christopher W. Geib and Christopher A. Miller, "A New Model of Plan Recognition", Proc. Fifteenth Conference on Uncertainty in Artificial Intelligence, 1999.

[13] Terran Lane and Leslie Pack Kaelbling, "Toward Hierarchical Decomposition for Planning in Uncertain Environments" , MIT Artificial Intelligence Laboratory, May 2000, USA.

[14] Frank Mulder, Frans Voorbraak, "A formal description of tactical plan recognition", Information Fusion 4 (2003) 47-61.

[15] Jon Doyle and Richmond , "Background to Qualitative Decision Theory", AI Magazine, American Association for Artificial Intelligence, Summer 1999, USA.

[16] Kevin Murphy, "Bayes Net Toolbox for Matlab" , *Last updated on 29 May 2003*, http://www.ai.mit.edu/~murphyk/Software/BNT/bnt.html

Information Evaluation: Discussion about STANAG 2022 Recommendations

Laurence Cholvy
ONERA Centre de Toulouse
BP 4025, 2 avenue Edouard Belin
31055 Toulouse
FRANCE

cholvy@cert.fr

Vincent Nimier
ONERA Centre de Châtillon
BP 72, 29 avenue de la Division Leclerc
92322 Châtillon
FRANCE

nimier@onera.fr

1.0 INTRODUCTION

An important step in the intelligence gathering process is the fusion of information provided by several sources. The objective of this process is to build an up-to-date and correct view of the current situation with the overall available information in order to make adequate decisions. Moreover, to succeed in this process, it is important to associate with each available information, some attributes like the number of the sources that support it, their reliability, the degree of truth of the information etc. For the moment, these attributes are managed by the human operator when he fuses information provided by the different sources. However, there is no real methodology to do this in a formal manner. And, when relaying this fusion process to a machine, we need to develop formal definitions and algorithms to manage these attributes in addition to fusing information. Furthermore, in a context of interoperability where different systems exchange information, common definitions of these attributes have to be shared. The Standardization Agreements (STANAG) 2022 of North Atlantic Treaty Organization (NATO) defines a framework for such common definitions.

The purpose of this paper is first to analyze the STANAG 2022 recommendations about information evaluation and then to set a first step in the definition of a formal and non ambiguous system for evaluating information. Indeed, as it will be shown, the present recommendations, written in natural language are rather ambiguous and imprecise and are open to discussion. This paper is organized as follows. Section 2 presents a review of STANAG 2022 recommendations and points out to the main notions that underline these recommendations. Section 3 analyzes the different assumption underlying the evaluation proposed in the STANAG. Section 4, an example of symbolical formalism for the fusion of information is presented. In this example, we consider some of the assumption of STANAG 2002. Finally, section 5 is devoted to a discussion.

2.0 REVIEW OF STANAG 2022 RECOMMENDATIONS

The Annex to STANAG 2022, Edition 8 ([1]) explicitly mentions that the aim of information evaluation is to indicate the degree of confidence that may be placed in any item of information which has been obtained for intelligence. (...) This is achieved by adopting an alphanumeric system of rating which combines a measurement of the reliability of the source of information with a measurement of the credibility of that information when examined in the light of existing knowledge.

Examining the whole text leads us to point out that the two main concepts in this evaluation system are the reliability of the sources and the credibility of the information. These concepts are defined in the STANAG 2022 recommendations, as follows:

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

Reliability of the source is designated by a letter between A and F signifying various degrees of confidence as indicated below.

- a source is evaluated A if it is completely reliable. It refers to a tried and trusted source which can be depended upon with confidence.
- a source is evaluated B if it is usually reliable. It refers to a source which has been successfully used in the past but for which there is still some element of doubt in particular cases.
- a source is evaluated C if it is fairly reliable. It refers to a source which has occasionally been used in the past and upon which some degree of confidence can be based.
- a source is evaluated D if it is not usually reliable. It refers to a source which has been used in the past but has proved more often than not unreliable.
- a source is evaluated E if it is unreliable. It refers to a source which has been used in the past and has proved unworthy of any confidence.
- a source is evaluated F if its reliability cannot be judged. It refers to a source which has not been used in the past.

Credibility of information is designated by a number between 1 and 6 signifying varying degrees of confidence as indicated below.

- If it can be stated with certainty that the reported information originates from another source than the already existing information on the same subject, then it is classified as "confirmed by other sources" and rated 1.
- If the independence of the source of any item of information cannot be guaranteed, but if, from the quantity and quality of previous reports, its likelihood is nevertheless regarded as sufficiently established, then the information should be classified as ``probably true" and given a rating of 2.
- If, despite there being insufficient confimation to establish any higher degree of likelihood, a freshyly reported item of information does not conflict with the previously reported behaviour pattern of the target, the item may be classified as ``possibly true" and given a rating of 3.
- An item of information which tends to conflict with the previously reported or established behaviour pattern of an intelligence target should be classified as ``doubtful" and given a rating of 4.
- An item of information which positively contradicts previously reported information or conflicts with the established behaviour pattern of an intelligence target in a marked degree should be classified as ``improbable" and given a rating of 5.
- An item of information is given a rating of 6 if its truth cannot be judged.

The previous definitions of information evaluation can be criticized. Indeed, since they are given in natural language, they are quite imprecise and ambiguous and a lot of points are open to discussion. For exemple, if we consider the credibility of information, it seems that the rating defined according to the recommendations does not describe a unique property. For instance, how should we qualify an item of information supported by several sources of information which is also conflictual with some already registered information? According to STANAG definitions, this item should be given a rating of 1 and should also be given a rating of 5.

Furthermore, according to the recommendations, a rating of 6 should be given to an item whose truth cannot be judged. This suppposes that the other ratings (1...5) concern the evaluation of the truth of information. If so, the rating of 1 should be given to a true information. But, as it is defined, the rating of 1 is given to an item supported by at least two sources. This is questionable since, the different sources (even if they agree) may emit false information.

However, even if the previous recommendations can be criticized, we can see that they present three basic concepts that present a cornerstone in evaluation system. These concept are the following:

- the reliability of a source
- the number of independent sources that support an information
- the fact that the information tends to conflict with some available information.

It is clear that these three concepts are independent. and that they have to be included in all quotation systems, whether these systems are automatic or not.

3.0 ANALYSIS OF THE DIFFERENT EVALUATION CONCEPTS

As it was noticed above three basic concepts are underlying the operational notion of quotation.

Nevertheless, these concepts are used, in the STANAG, in an operational framework which is not necessarily usable in an automatic process. For this reason, we try in this section to give a more comprehensive and formal definitions of these concepts with the aim of using them in an automatic process of fusion.

- ***Reliability***

In mathematical logic, the reliability of a source is defined in a binary way as follows: *an information source is (totally) reliable if and only if the information it delivers are true in the real world.* For instance, a sensor which measures the temperature is totally reliable if and only if the temperature it indicates is the correct one; a human expert is totally reliable if and only if any information (opinion, conjecture etc) he gives is true.

According to the recommendations of STANAG 2022, , the reliability of a source is not a binary notion but a graded one and is defined in reference to its use in the past. It can be measured for example, as the ratio of the number of times the source gave a true information to the number of times it gave information. However, this definition does not take into account the actual environment of use of the information source. For instance, even if it is known to be reliable, an infra-red sensor loose reliability when it rains. So we have to take into account the condition in which the source is used. Some practical consideration of the use of this notion in a numerical context can be find in [7], [8].

However, it must be noticed that there is no consensus yet on a formal definition of reliability. For instance, we can read in the APJ 2.1 "*every piece of information produced by an impeccable source is not necessarily correct*". If "impeccable" intends to mean "reliable", this sentence is contradictory with the definition given previously. Here, it implies that the reliability of a source is not defined by its ability to deliver truth and that even a reliable source can be wrong. But it can be wrong not because it is not sincere but because its model of discernment is maybe not precise enough to distinguish true from false.

- ***Independence***

This notion have to be a little bit enlightened. The classical statistical definition of the independence of two events is given by the rule:

two events A and B are independent iff $P(A,B)=P(A)P(B)$

The STANAG highlights the independence of sources to confirm information. Indeed this independence is important if we want to improve our confidence in a decision when many sources agree on the same decision. However this condition on independence is not sufficient. Let us consider n sources given a decision about an hypothesis H_i. Our confidence that this hypothesis is true given the fact that the n sources agree on the fact that this hypothesis is true is given by the probability: $P(H_i/d_1=i, \ldots,d_n=i)$ this probability may be written in terms of elementary probability of each source by the formula:

$$P(H_i / d_1 = i,\ldots, d_n = i) = \frac{P(d_1 = i,\ldots, d_n = i/H_i)P(H_i)}{P(d_1 = i,\ldots, d_n = i)}$$

For the sake of simplicity if we suppose that elementary sources are independent, then the above formula may be written as:

$$P(H_i / d_1 = i,\ldots, d_n = i) = \frac{\prod_j P(d_j = i/H_i)P(H_i)}{\prod_j P(d_j = i)}$$

Furthermore if we suppose that each source has the same performance, it is easy to see that this probability is higher than the probability that the hypothesis is true given only one source if and only if each of the source is informative:

$$P(d_j=i/H_i)\hat{}P(d_j=i)$$

This means that the probability of giving a right decision for a given hypothesis is higher than the probability of giving the same decision no matter what is the hypothesis.

- ***Conflict***

In this part we address the concept of conflict with a focus on it use in a symbolical.

In mathematical logic, conflict is defined by the notion of contradiction. *Some pieces of information, i.e, some formulas, are conflictual if and only if, they are contradictory (i.e, there exists no possible world in which they are all true).*For instance, “the object is a plane”, “the object is an helicopter” are two conflictual pieces of information (given the background knowledge that no object can be both a plane and an helicopter). “its speed is 600kmh” “its speed is 625kmh” are also are two conflictual pieces of information (since obviously, not both of them can be true).

However, it seems that the STANAG recommendations want to make a distinction between information which are conflictual and information which tend to be in conflict.

The notion of conflict may be related to the notion of distance between information. In logic an information is modelised by a propositional variable. Then the conflict between tow information may be modelised by a distance which is null if the propositional formulas are not in conflict and different of zero if they are.

In the following we give some definition and property of the notion of distance in the context of logic for dealing with the credibility given in the STANAG.

definition

Let $\mathcal{L}$ a propositional language on a finite alphabet of propositional variables $\mathcal{P}$. An interpretation is a mapping from $\mathcal{P}$ toward $\{0,1\}$. The set of all the interpretation is write W. An interpretation I is a model of a formula if and only if this formula is true for this interpretation. Let w define a formula and mod(w) is the set of model of w. That is $mod(\varphi)=\{I\in \mathbf{W} / I \models \varphi\}$.

Let us now define a distance between two interpretations. This is a function d : W×W →N such that:

$$d(I,J) = d(J,I)$$

$$d(I,J) = 0 \text{ if } I=J$$

This notion helps us to naturally define a distance between an interpretation and a propositional variable by:

$$d(I,\varphi) = \min_{J \models \varphi} d(I,J)$$

Now it is possible to extend this definition to the distance between an interpretation and a set of propositional variables. For this, we have to introduce an operator for the combinatory of the elementary distances. The choice of this operator is outside the scope of this paper and we only focus here on two operator: sum and max. With these operators the distances are given by:

$$d_s(I,\psi) = \sum_{i=1}^{n} d(I,\varphi_i) \text{ for the sum operator}$$

$$d_M(I,\psi) = \operatorname*{Max}_{i=1,\dots,n} d(I,\varphi_i) \text{ for the max operator}$$

With $\psi = \{\varphi_1, \dots, \varphi_n\}$

In the symbolical formalism, the simplest distance that it is possible to define between two interpretations is the drastic distance. This distance gives 0 if the two interpretations are equal and 1 if not.

$$d_D(I,J) = \begin{cases} 0 & \text{si } I = J \\ 1 & \text{sinon} \end{cases}$$

In this case the distance between an interpretation and a set of logical proposition is also equal to 0 or 1.

Given w and w' two models of a formula, the number of propositional letter whose valuation differs from w to w' is called the hamming distance.

4.0 EXAMPLE FOR A FORMAL MODEL OF EVALUATION FOR SYMBOLICAL INFORMATION

In this paragraph, we give some ideas on how an evaluation as close as possible to the one recommended by the STANAG may be use when dealing with symbolical information The fusion process we illustrate on an example takes into account the number of the sources that support an information, their reliability. However it does not take into account a graded notion of conflict.

- ***Classical fusion operator***

We recall here some definitions introduced by Konieczny and Pino-Pérez [2], [3] in the context of fusion of database.

Let db1...dbn be n sets of information and Poss be a set of models. Konieczny and Pino-Pérez defined a majority merging operator, denoted Δ such that the set of models of the information source which is obtained from merging db1... dbn with this operator, $Mod((\Delta([db_1,...,db_n])))$, is semantically characterized by:

$$Mod(\Delta([db_1,...,db_n])) = \min_{\leq[db_1...db_n]}(Poss)$$

where ≤[db1…dbn] is a total pre-order on Poss defined by :

$$w \leq_{db_1...db_n} w' \text{ iff } \mathrm{d}(\mathrm{w},[\mathrm{db}_1...\mathrm{db}_\mathrm{n}]) \leq \mathrm{d}(\mathrm{w}',[\mathrm{db}_1...\mathrm{db}_\mathrm{n}])$$

with

$$d(w,[db_1...db_n]) = \sum_{i=1}^{n} \min_{w' \in Mod(db_i)} d(w,w')$$

where d(w,w') is the Hamming distance. In other words, when merging db1...dbn with the operator Δ, the result is semantically characterized by the models in Poss which are minimal according to the pre-order $\leq_{[db_1,...,db_n]}$.

- ***Fusion operator with reliability of the sources***

Here, we to extend this operator by taking a weighted sum instead of the sum.

We assume now that any db_i is associated with a weight, i.e, an integer denoted $r(db_i)$ representing its degree of reliability. For instance, a source whose reliability is A (resp, B, C, D) will be weighted by 5 (resp, 4, 3, 2, 1).

We define Δ_{WS} by :

$$Mod(\Delta_{WS}([db_1,...,db_n])) = \min_{\leq^{WS}_{[db_1...db_n]}}(Poss)$$

$\leq^{WS}_{[db_1...db_n]}$ is a total pre-order on Poss defined by:

$$w \leq^{WS}_{[db_1...db_n]} w' \text{ iff } d_{WS}(w,[db_1...db_n]) \leq d_{WS}(w',[db_1...db_n])$$

with

$$d_{WS}(w,[db_1...db_n]) = \sum_{i=1}^{n} \min_{w' \in Mod(db_i)} d(w,w').r(db_i)$$

i.e

$$d_{WS}(w,[db_1...db_n]) = \sum_{i=1}^{n} r(db_i). \min_{w' \in Mod(db_i)} d(w,w')$$

- ***Example 1***

We consider that the problem consists in identifying a flying object. This object can be a plane (p) or an helicopter (h) or a missile (m). Assume we have five information sources called One, Two, Three, Four and Five whose reliability are respectively A, B, C, D and A. We consider the following flow of information:

- One emits $p \vee h$
- Two emits $h \vee m$
- Three emits p
- Four emits h
- Five emits p.

The questions that raise are: is the object a plane? is the object an helicopter ? is the object a missile ?

By the previous definitions, we can define:

$$
\begin{aligned}
&db_1 = \{p \vee h\} \text{ and } r(db_1) = 5 \\
&db_2 = \{h \vee m\} \text{ and } r(db_2) = 4 \\
&db_3 = \{p\} \text{ and } r(db_3) = 3 \\
&db_4 = \{h\} \text{ and } r(db_4) = 2 \\
&db_5 = \{p\} \text{ and } r(db_5) = 5
\end{aligned}
$$

The three possible worlds are:

$$
\begin{aligned}
&w_1 = \{p, \neg h, \neg m\} \text{ in which the object is a plane,} \\
&w_2 = \{\neg p, h, \neg m\} \text{ in which the object is an helicopter,} \\
&w_3 = \{\neg p, \neg h, m\} \text{ in which the object is a missile}
\end{aligned}
$$

Then we can compute the following distances:

$$
\begin{aligned}
&d_{WS}(w_1, [db_1 ... db_5]) = 12 \\
&d_{WS}(w_2, [db_1 ... db_5]) = 16 \\
&d_{WS}(w_3, [db_1 ... db_5]) = 30 \\
&\text{then } Mod((\Delta([db_1, ..., db_5]))) = w_1 = \{p, \neg h, \neg m\}
\end{aligned}
$$

It means that, given the information transmitted until now by the different sources and given their reliabilities, the most plausible answer to the query “what is the object?” is “the object is a plane”.

- ***Example 2***

Now, let us consider that the previous information sources have different degrees of reliability, for instance:

$$
\begin{aligned}
&r(db_1) = 2 \\
&r(db_2) = 5
\end{aligned}
$$

$$r(db_3) = 2$$
$$r(db_4) = 5$$
$$r(db_5) = 2$$

And suppose that the five sources send the same information as before.

Then the distances are now the following:

$$d_{WS}(w_1, [db_1 \ldots db_5]) = 20$$
$$d_{WS}(w_2, [db_1 \ldots db_5]) = 8$$
$$d_{WS}(w_3, [db_1 \ldots db_5]) = 16$$

which give as a result: “it is an helicopter”.

This example shows the impact of the reliability of the sources in the result of the decision: changing reliability degrees of the sources may lead to different results.

5.0 DISCUSSION

This work intended to formalize some informal recommendations about information evaluation in information fusion. Some of the informal notions underlying the recommendations have been given a formal interpretation, even if, as it has been shown, no consensus exist yet on these definitions and more work is needed. Based on this formalization, we have suggested to implement the fusion process by a weighted majority because this fusion operator takes into account the number of the sources and their reliability. If the number of the sources had not been so important, we could have chosen a method taking into account the reliability of the sources only (see [4], [5], [6] for instance).

However, some of the notions underlying the STANAG recommendations have not yet been taken into account.

For instance, in classical literature, we made no difference between ``being conflictual'' and ``tending to be conflictual''. This last notion assumes that there is a scale for defining conflicts. Generally, this scale is binary and the formal notion which represents the conflict is the logical inconsistency. Extension to graded conflict may be modeled thank to the definition of distance between interpretation but further work have to be made to go closer to the STANAG recommendation.

Another point which has been left aside is the total ignorance about the reliability of a given source, which was a case foreseen in the recommendations. In the present formalization, it seems difficult to represent that. Indeed, which number can be attached to a source whose reliability is not known? That is an open question.

6.0 REFERENCES

[1] Annex to STANAG 2022 (Edition 8). North Atlantic Treaty Organization (NATO) Information Handling Services, DODSID, Issue DW9705.

[2] S. Konieczny and R. Pino-Pérez. On the logic of merging. In Principles of Knowledge Representation and Reasoning (KR'98), pages 488--498, 1998.

[3] S. Konieczny and R. Pino-Pérez. Merging with Integrity Constraints In Proceedings of ESCQARU'99.

[4] C. Baral and S. Kraus and J. Minker and V.S. Subrahmanian. Combining multiple knowledge bases. IEEE Trans. on Knowledge and Data Engineering, 3(2), 1991.

[5] L. Cholvy. Proving theorems in a multi-sources environment. Proceedings of IJCAI, 1993.

[6] C. Liau. A conservative Approach to Distributed Belief Fusion. Proceedings of FUSION, 2000.

[7] V. Nimier « Supervised Multisensor tracking algorithm », FUSION-98, International Conference on Multisource-Multisensor Information Fusion, Las Vegas, USA, July 6-9, 1998, pp 149-156.

[8] V. Nimier, « Introducing Contectual Information in Multisensor tracking Algorithms » Advances in intelligent Computing, Springer, 1994.

Using A Priori Databases for Identity Estimation through Evidential Reasoning in Realistic Scenarios

Dr. Pierre Valin
Lockheed Martin Canada
R&D Department
6111 Royalmount Ave
Montréal, Québec H4P 1K6
CANADA

E-mail: Pierre.Valin@lmco.com

Dr. Eloi Bossé
Defence Research & Development Canada
(DRDC) Valcartier
2459 Pie-XI Blvd North
Val-Bélair, Québec G3J 1X5
CANADA

E-mail: Eloi.Bosse@drdc-rddc.gc.ca

INTRODUCTION AND SCOPE

Canadian defence companies and Government Research and Development (R&D) laboratories have long ago recognized the necessity to develop comprehensive a priori databases containing all the possible attributes that can be inferred by measurements coming from a given sensor suite. In order to maintain this document at a NATO unclassified level, a small portion of an existing (consisting of more than 2200 platforms) database is presented, which nevertheless contains all the salient features needed for refining the identity (ID) of any target by the fusion of sensor information. In addition, only the information gathered from unclassified sources such as Jane's and Periscope is presented. This a priori Platform DataBase (PDB) contains all the possible naval and air targets, military or commercial, that can be encountered in realistic scenarios, and all the attributes that can be measured by any sensor belonging to any own-platform of the Canadian Forces (CF), ensuring its possible common use throughout the CF. Also presented and explained are all the attributes and all the correlations between platforms that are appropriate to Situation and Threat Assessment (STA or higher-level fusion), and which are present in the larger database. This paper focuses on only one own-platform of the CF in relevant scenarios, the maritime surveillance aircraft CP-140 Aurora (a Canadian version of the US's P3-C with S2-B avionics) in its present operational status, and also with an anticipated upgraded sensor suite. Validation and benchmarking of the chosen evidential reasoning scheme for identity estimation, is performed through several simulated scenarios that make use of DRDC-Valcartier Concept Analysis and Simulation Environment for Automatic Target Tracking and Identification (CASE-ATTI) sensor module. Two missions have received special consideration because they make full use of all the CP-140 sensors: 1) Maritime Air Area Operations (MAAO) involving a mix of commercial merchant ships and enemy line combatant ships performing exercises, and 2) Direct Fleet Support (DFS) involving fleets of Canadian and American ships on a NATO mission, that are imaged with the Aurora's imaging sensors.

1.0 ATTRIBUTE INFORMATION FROM SENSORS

The present and foreseen CP-140 sensors can be divided into three broad classes depending on the type of input they provide to the MSDF function:

1) *Identification* sensors, namely special-purpose "intelligent" sensors, reporting distilled ID information such as Electronic Support Measures (ESM), Identification Friend or Foe (IFF response), and datalink (Link-11 ID information)

2) *Imaging* sensors, namely the existing Forward Looking InfraRed (FLIR) and the upcoming Spotlight Synthetic Aperture Radar (SSAR) upgrade to the present radar. The SSAR will have

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

4 modes: 2 for Land Use (the Land Spotlight and StripMap) and 2 for Ships (Sea Spotlight and Range Doppler Profiling)

3) *Tracking* sensors (radar when not SSAR mode, Link-11 positional information)

The set of databases needed for the identification of platforms can be decomposed into a Platform DataBase (PDB), an Emitter Name List (ENL) and a Geopolitical List (GPL):

1) the PDB used for identification purposes spans all possible targets that can be met in the most important Aurora missions and contains all the possible attributes for each platform
2) the ENL includes all emitters (navigation, targeting, etc.) corresponding to each platform of the PDB
3) the GPL provides the allegiance of various countries and thus can evolve as the geo-political context of missions changes.

The attributes coming from sensors that are catalogued in the PDB can be split into three groups:

1) Kinematical attributes
2) Geometrical attributes
3) Identification attributes

These are further detailed in the sub-sections below [1,2].

1.1 Kinematical Attributes

Kinematical attributes can be estimated through tracking in the positional estimation function of Multi-Sensor Data Fusion (MSDF), and through reports from IFF and datalink. A minimum list contains:

- maximum acceleration, both tangential (if available for aircraft, it denotes the likely presence of afterburners) and centrifugal (with higher g values likely denoting a fighter aircraft)
- maximum platform speed (the quoted value is relevant to static ambient atmosphere and must be interpreted, or fuzzified, to account for air currents, particularly when the aircraft can reach the jet stream)
- minimum platform speed, e.g. a null value for an aircraft denotes the a Short Take-Off & Vertical Landing (STOL) capability, or a helicopter.
- maximum altitude that a platform may reach can serve as a bound for altitude reported by the IFF or which can be deduced from a tracker in 3-D from sensor reports in 2-D and assumed flight characteristics. For a submarine, this corresponds to the maximum depth that can be achieved (a negative quantity).
- cruising speed

The first 4 serve as bounds to discriminate between possible target IDs, while the fourth can suggest a list of the most plausible IDs.

1.2 Geometrical Attributes

Geometrical attributes can be estimated by algorithms which post-process imaging information from sensors such as FLIR or Electro-Optics (EO) and SSAR.

Classifiers that perform such imagery post-processing can be thought of as Image Support Modules (ISM) performing much the same functionality as the ESM does for the analysis of electromagnetic signals. These ISMs need

- the three geometrical dimensions of height, width and length (for FLIR and EO), or at least their ratios if range information is not provided by the tracker that cued the imaging sensors, or if no laser ranging is possible, and
- the Radar Cross Section (RCS) of the platform seen from the front, the side and the top (for the SSAR and radiometric radar).

In addition, the distribution of relevant features may be needed for classifiers, but may be considered part of the algorithms that generate plausible IDs. Such features can be superstructure locations (for the SSAR) or hot point distribution (for the FLIR).

1.3 Identification Attributes

Identification attributes can be directly given by the ESM, or as outputs of the FLIR and SSAR ISM at various stages of their classification. The ESM requires an exhaustive list of all the emitters that are carried by each platform. In addition certain acoustic features leading to possible submarine or ship ID can be listed, such as propulsion type, number of propellers, number of blades, number of cylinders.

However the ISMs are usually designed [3] to not only attempt at providing the best single ID possible but also to estimate confidence in higher levels of an appropriate taxonomy tree, since often only general characteristics application to several platforms of a given "type" can be identified. This taxonomy tree is usually derived from some standard, either STANAG 4420 or MIL-STD 2525 (A or B), and can have many levels of branching, e.g. a "domain" includes several "types", which include many "classes", etc., etc., eventually reaching the leaves of the tree providing the specific platform ID or singleton.

Taxonomy trees can be implemented as an extension of the PDB. In the example shown in Figure 1 below, a number of K attributes (8 are specifically indicated) are linked to N elements (7 are specifically indicated) of the platform library including 5 elements of library taxonomy. The taxonomy is a tree of few levels where the branches are also attributes.

Often, especially for STA applications, it is enough to know the type of platform rather than its specific ID. Naval battlegroup compositions, air attack formations, and the content of convoys over land, follow general well-known patterns, and need not necessarily be detailed to the platform level to ascertain lethality and/or intent.

The usefulness of the extra taxonomy elements will be shown in the next section, in the context of a chosen evidential reasoning scheme [4].

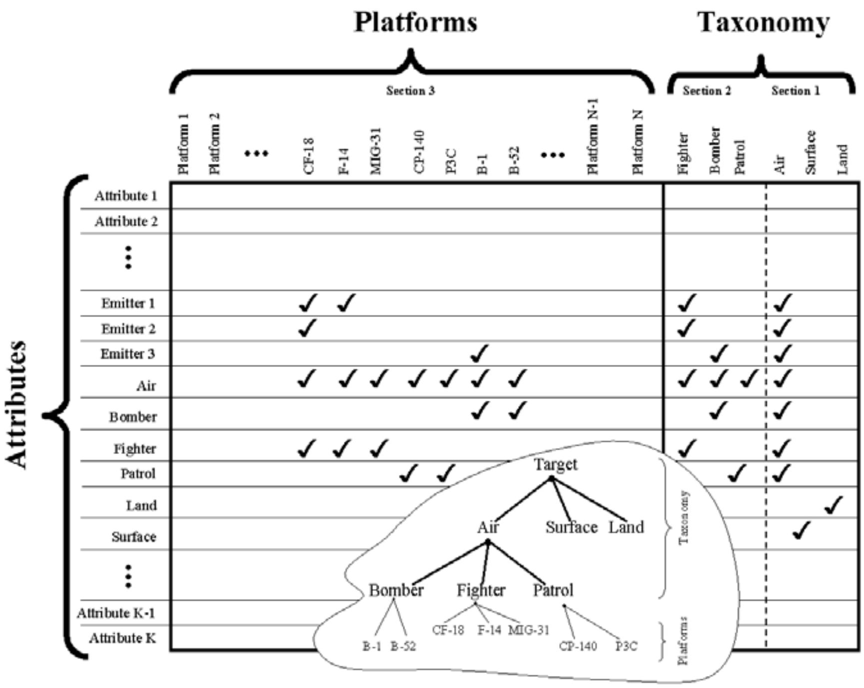


Figure 1: Example of a taxonomy tree structure for the a priori PDB.

In an alternative "transposed" description, the tabular form of the PDB can be thought of as rows being indexed by single platforms (singletons), and with columns representing measured attributes and the various levels of the taxonomy. The actual implementation is a matter of choice. Such a representation is shown in Table 1 below for a sample of Russian ships (later to be used in the MAAO scenario), with the emphasis only on the subtype taxonomy and the complete emitter list (each number in the list corresponds to a NATO designation in the ENL). It should be observed that the emitter list is exhaustive and contains many emitters common to several platforms. This immediately implies that many fusion steps of ESM reports must be performed in order to reach a proper identification to the lowest level of the taxonomy tree, namely the desired singleton.

Table 1: An example of an alphabetical list of Russian ships

#	Platform name	Type	List of emitters
242	BEREZINA AOR	TANKER	44,64,47,45,83,46,103,109
249	DUBNA AOR	TANKER	47
237	GORYA MHO	MINE SWEEPER	65,274,46,275,109
219	GRISHA I FFL	FRIGATE	44,105,103,104,47,109,106,83
220	GRISHA II FFL	FRIGATE	44,105,103,104,47,109,106
221	GRISHA IV FFL	FRIGATE	44,105,47,45,83,46,103,104,109,106
232	IVAN SUSANIN AGB	ICEBREAKER	44,56,64
234	IVAN-ROGOV-ALEKSANDR LPD	LANDING SHIP	93,89,103,101,68,46,45,65,64,62
235	IVAN-ROGOV-MITROFAN LPD	LANDING SHIP	93,89,103,101,68,46,45,65,64,105
213	KARA-AZOV CG	CRUISER	78,84,62,64,85,45,92,68,46,93,104,103
214	KARA-KERCH CG	CRUISER	78,84,62,64,47,85,45,68,46,93,104,103
215	KIROV-ADM-USHAKOV CGN	CRUISER	77,89,90,65,67,92,84,80,45,71,46,93,101,106
224	KRIVAK IB FFG	FRIGATE	63,69,67,45,68,103
225	KRIVAK II	FRIGATE	62,69,67,45,103
226	KRIVAK IIIA FF	FRIGATE	62,69,66,45,71,46,103,101
212	KUZNETSOV CV	AIRCRAFT	95,63,65,97,301,99,100,307

240	MATKA PH	PATROL BOAT	303,327,46,103,101,109
216	MOSKVA CHG	CRUISER	77,84,62,47,85,83,103,104
241	MURAVEY PH	PATROL BOAT	66,46,103,109
227	NANUCHKA I FFG	FRIGATE	128,333,83,45,104,109,334
228	NANUCHKA III FFG	FRIGATE	128,333,45,104,109,334,46
239	NATYA II MS	MINE SWEEPER	47,86,87,109,103
229	NEUSTRASHIMY FFG	FRIGATE	63,65,335,71,106
246	OSKOL AR	SUPPORT VESSEL	47
230	PARCHIM II FFL	FRIGATE	335,338,46
247	PRIMORYE AGI	SUPPORT VESSEL	64
217	UDALOY AND KULAKOV DDG	DESTROYER	69,97,65,67,91,46,71,101,106,89,93
218	UDALOY II DDG	DESTROYER	69,97,63,65,128,91,71,93,131

For convenience land targets are usually treated separately, since their taxonomy tree is much more complex.

2.0 EVIDENTIAL REASONING ON THE A PRIORI PDB

The question then arises as to which reasoning framework to use for compounding, or fusing, successive sensor declaration about a given attribute. The best choice seems to be the Dempster-Shafer (DS) formalism because it can:

- process incomplete information, implying that ignorance should be a concept defined mathematically rigourously (ignorance corresponds to the complete PDB in DS reasoning)
- not require a priori information, sometimes impossible to gather for a given mission, which rules out Bayes reasoning (Bayesian would have to split the ignorance equally across all other platforms, which can lead to large computational loads)
- handle conflicts between contact/track, implying that conflict should be defined a concept mathematically (conflict is calculated as the sum of BPAs with null set intersection in DS)
- have a real-time method, which means that DS truncation is essential (rules to that effect have been empirically determined and benchmarked)
- present the operator with the best ID, i.e. give preference to singletons, then the next best thing, i.e. doublets, then triplets, etc., or equivalently…
- have the possibility of computing beliefs in higher nodes of the tree, as sums of BPAs of the underlying branched structure
- have the possibility of providing several different functions as a decision aid, which can be, in the DS scheme, the plausibility or the expected utility interval, for example.

A sensor reporting a given attribute, then declares an ID proposition containing all the singletons having that attribute, with a Basic Probability Assignment (BPA or mass) reflecting its confidence in the attribute determination (which can be quite complex if the attributes are fuzzified).

Identity estimation is performed by the fusion of existing ID information from previous repeated fusion (therefore a long list, possibly truncated in number), with the new (possibly complicated) declaration from any of the sensors or ISMs. The DS combination method can done one of two ways

1) either using the original Orthogonal Sum (OS) which calculates conflict and redistributes it to the whole set of propositions [1,2],
2) or using a Hierarchical OS (HOS) which redistributes possible conflict by assigning the mass value to the first node of the tree that resolves the conflict [4].

The standard OS reads [1,2], for set of proposition A_i (coming from the sensor) and B_j (from previous fusion steps) yielding the set of proposition C_k

$$m(C_k) = \sum_{i,j \ni} \sum_{A_i \cap B_j \equiv C_k} \frac{m(A_i)m(B_j)}{1-K} \qquad K = \sum_{A_i \cap B_j \equiv \Phi} m(A_i)m(B_j)$$

with K being the conflict arising between the sets A_i and B_j (whenever the intersection is the null set).

In the case of the HOS the algorithm is explained below [4], when each of A and B has the form of bitstreams given by Figure 1 (checkmarks being 1, otherwise entries are zero), where the operator "==" refers to the standard equality comparison operator, $\cup$ and $\cap$ refer to the logical OR and AND operations applied on bitstream, the $\oplus$ symbol stands for the concatenation procedure applied on a section of bitstream, and $\varnothing$ refers to the null set. Given two propositions described by the bitstreams A and B:

HOS : IS A.section[3] $\cap$ B.section[3] == $\varnothing$?

No: $C = A$.section[1,2,3] $\cap$ B.section[1,2,3]

Yes: IS A.section[2] $\cap$ B.section[2] == $\varnothing$?

No: C = {(A.section[3] $\cup$ B.section[3])} $\oplus$ {A.section[1,2] $\cap$ B.section[1,2]}

Yes: IS A.section[1] $\cap$ B.section[1] == $\varnothing$?

No: C = {A.section[2,3] $\cup$ B.section[2,3]} $\oplus$ {A.section[1] $\cap$ B.section[1]}

Yes: C = Ignorance.

The OS is regularly used in the MAAO and DFS scenarios to be discussed later, with an appropriate truncation scheme to keep the potentially NP-hard problem CPU-tractable. An example of such truncation rules and the related parameters is shown below [5].

1) All combined propositions with BPA > MAX_BPM are kept
2) All combined propositions with BPA < MIN_BPM are discarded
3) If the number of retained propositions in step 1 is smaller than MAX_NUM, it retains, by decreasing BPA, propositions of length 1
4) If the number of retained propositions in step 3 is smaller than MAX_NUM, it does the same things with propositions of length 2
5) Repeat a similar procedure for propositions of length 3
6) If the number of retained propositions is still smaller than MAX_NUM, it retains propositions by decreasing BPA regardless of length.

Typical values for the parameters are MAX_BPM = 0.05 and MIN_BPM = 0.01 for MAX_NUM = 8, and some simple empirical rules relating these parameters can be found, for example the following simple inversely proportional formula is suggestive of a possible simplification [5]:

MAX_BPM = 1 / (2.5 * MAX_NUM)

MIN_BPM = 1 / (12.5 * MAX_NUM)

In addition, to prevent the algorithm from being unable to recover from a series of countermeasures, a small lower limit has to be imposed for the ignorance at all times.

3.0 HIGHER LEVEL INFORMATION FOR THE A PRIORI PDB

The Levels 2 and 3 databases, i.e. the STA/RM DB should contain all the platform parameters relevant for STA as well as RM, i.e., since missiles (number and detailed properties) on enemy ships are relevant for STA, while the same information on possible own-platforms is relevant for RM. In a Network Centric Warfare (NCW) context, the lethality of enemy platforms in the red force is important for STA, and the lethality of cooperating Participating Units (PUs) is relevant for RM within the blue force. A non-exhaustive list contains elements pertaining to [6]:

1) Platform and Mission:
    - displacement,
    - number of operational copies of the platform,
    - list of hull numbers & names (if it can be provided for ID in harbors, air fields, etc.),
    - range of deployment,
    - platform type (with amplification),
    - role for the mission, and
    - crew (for full operation)
2) Armament (type and number of examples present on platform, both HW as well as humans for mission deployment):
    - Surface-to-Surface Missiles (SSMs), including submarine launched missiles
    - Surface-to-Air Missiles (SAMs), including submarine launched missiles
    - Close-In Weapon Systems (CIWSs),
    - Air-to-Surface Missiles (ASMs),
    - Air-to-Air Missiles (AAMs),
    - CIWSs,
    - conventional bombs,
    - troop complement (number of special force for assault, landing or parachuting),
    - lethality,
    - guns, and
    - torpedo tubes
3) Sensors (mostly passive, in order to estimate probability of own-platform detection, excluding the radar already in PDB):
    - Infra Red Search and Track (IRST),
    - sonars (e.g., Hull-Mounted Sonar, towed-array, sonobuoy, tethered sonar), and
    - imaging sensors (e.g., EO, FLIR, SAR)

4) Air Platforms on Deck (for surface ships):
   - Number of helicopters (on any line combatant ship)
   - Number of aircraft (on aircraft carrier)

4.0 TYPICAL DS RESULTS USING THE A PRIORI PDB INFORMATION

Results from the two complex scenarios that use DRDC-Valcartier's Concept Analysis and Simulation Environment for Automatic Target Tracking and Identification (CASE-ATTI) sensor module are presented in this section [1,2]:

1) Maritime Air Area Operations (MAAO) involving a mix of commercial merchant ships and enemy line combatant ships performing exercises, and
2) Direct Fleet Support (DFS) involving fleets of Canadian and American ships on a NATO mission, that are imaged with the Aurora's imaging sensors (only the SSAR results are presented here).

For both the MAAO and DFS scenarios, the DS OS is used for the ID fusion. In both scenarios, SSAR imagery is used to attempt to help the classification. A SSAR simulator provided by DRDC-O was used to produce the imagery for the appropriate acquisition parameters. The two scenarios were intentionally built to test

1) the limits of the association mechanism for ESM reports (with inter-ship distances at the theoretical limit of discernability), leading to occasional miss-associations, and
2) the performance of the SSAR ISM, by choosing ships whose imagery can be misconstrued. This is achieved by having the scenario contain ships of unusual length for their types, which fools the Bayesian classifier, which uses length distribution to detect ship type.

The section will demonstrate that the DS scheme is robust under countermeasures (i.e. wrong ESM associations, either by deliberate spoofing, or by algorithmic error). Also this section will present a toy scenario illustrating the HOS method for the taxonomy tree previously shown in Figure 1.

4.1 OS Results for Real-Life MAAO and DFS Scenarios

At appropriate times in the MAAO scenario, several ESM contacts are received for each hostile vessel, one such contact being incorrect for one platform (chosen arbitrarily to be the Udaloy destroyer), in order to test the robustness of the DS evidential reasoning algorithm under countermeasures. This discrepancy will prompt the use of SSAR imaging of the Udaloy and members of its convoy. As soon as the operator has imaged the Udaloy, he/she will then in short order image two other ships in the Russian convoy, namely the Kara cruiser and the Mirka frigate with roughly the same acquisition parameters, since the surveillance aircraft's motion over such a short period of time is not very significant. To create the imagery of enemy ships, a simulator from DRDC-Ottawa was used with permission. The imagery thus generated is unclassified, and its interpretation by the SAR ISM is reproduced below in Figure 2 (removal of artifacts by thresholding between the top 2 rows of imagery, and centerline detection by a Hough transform are clearly visible). The category (merchant vs. line combatant) is always properly achieved by the Neural Net category classifier, and the type ID by the Bayes classifier is always correct. More information on the hierarchical SSAR classifier can be found in reference [3].

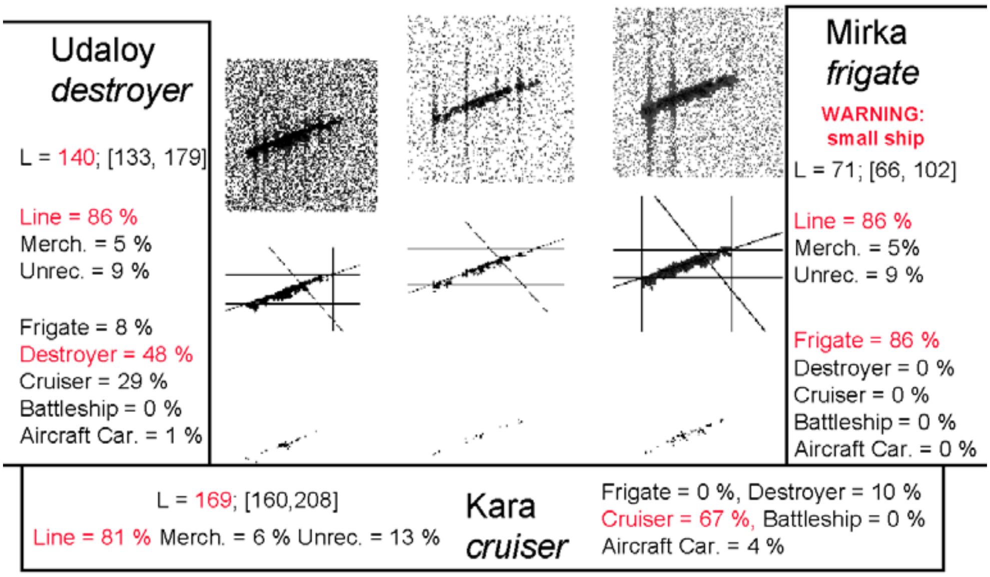


Figure 2: SAR imagery and ISM performance for Russian ships in MAAO scenario.

The PDB contains many Russian ships with emitters common to the 3 ships in the MAAO scenario, as was shown in Table 1 (not all Russian ships are in that Table, e.g. the Mirka is not listed), hence many ESM reports are necessary to achieve correct ID. The ESM reports are chosen at random amongst the possible list given in Table 1 when no countermeasures are present. The most complicated DS reasoning occurs for the Udaloy II as mentioned previously and is shown in Figure 3 below. In this Monte-Carlo case, no emitter report was detected during this time by the ESM, that could distinguish the 2 versions of the Udaloy in the PDB (triangles in blue are discriminating emitter reports and a SSAR ISM fusion confirming an Udaloy triplet of same RCS signature).

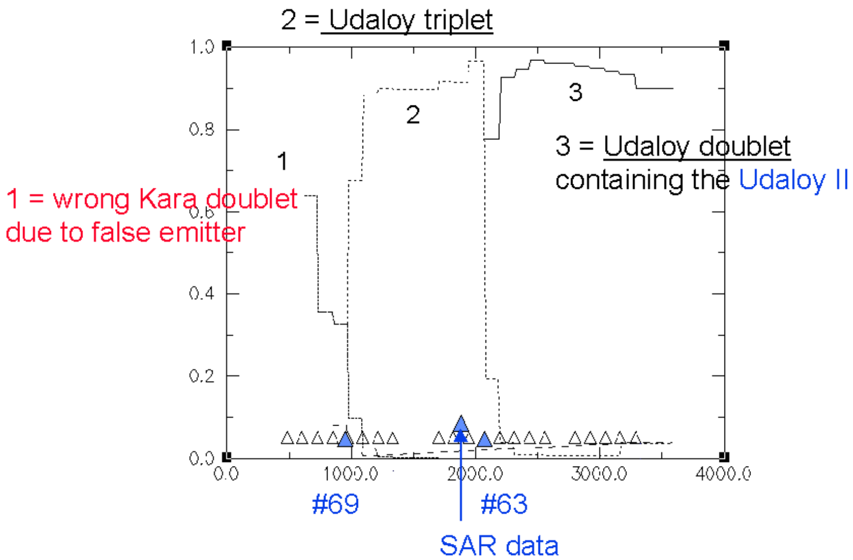


Figure 3: ID for the Udaloy II after countermeasures and with SAR ISM fusion.

If one concentrates on American ships in the DFS scenario, the resulting SAR imagery and ISM interpretation achieved is depicted in Figure 4 below.

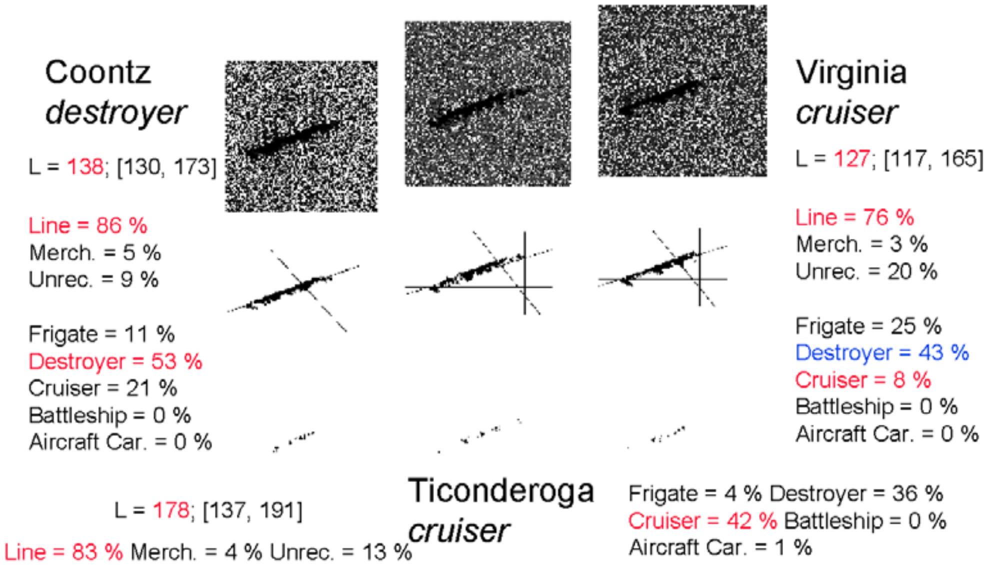


Figure 4: SAR imagery and ISM performance for American ships in DFS scenario.

In this case, it should be noted that the SAR ISM incorrectly identifies the Virginia cruiser as a destroyer because its length is small for a cruiser, (blue indicates a mistake by the SAR ISM, red a successful ID). If the imagery is done sufficiently early in the scenario, the ESM reports will eventually confirm the Virginia (rather than a destroyer such as the Spruance), otherwise it may not (depending on the Monte-Carlo run).

4.2 HOS Results for a Toy Scenario

To isolate the effect of ID fusion in the presence of a hierarchical taxonomy tree, and the functioning of the HOS, the toy scenario is much less complicated than the MAAO or DFS scenarios, and it incorporates simple trajectories, but provides conflicting ID information as time increases. Whereas the MAAO and DFS scenarios involve ships seen from a maritime surveillance aircraft, this toy scenario involves air targets seen from a Canadian Patrol Frigate of the HALIFAX (or City) class. Since the PDB contains all possible platforms, either application uses the same PDB.

In the toy scenario, we will focus on one AIR target track for which the 3 levels of taxonomy shown in Figure 1 apply. In order to provide conflict, one adds a dummy emitter which, as Figure 1 shows, is only found on bomber planes in our database. Thus the ESM will report fighter emitters most of the time, but once in a while, it will report a bomber emitter for that same plane. The other sensors simulated do not report any conflicting information (unlike the SAR in the DFS scenario).

The simulated results are as follows and are presented in Figure 5 which schematises the results of the best ID proposition, the one with the largest BPA (or mass), found for this target in the simulation case described below at critical fusion times [4]:

1) at time t_1 a report is received from the radar sensor reporting only kinematics. This, using a fuzzy logic test, leads to the statement Air with a BPA of 80%.
2) at time t_2, a first ESM report is received. It reports one of the emitters for a fighter (with a BPA of 80%) let us call it E_1. At this point this emitter is sufficient to correlate the information with the database and we know that this emitter is on a fighter without knowing which one.
3) at time t_3 another fighter emitter, E_2, is reported giving the full identification of the CF-18 which is the only one with both emitters on board.

4) at time t_4, the conflicting report is received reporting an emitter E_3 belonging to bombers only. We see that, at this point, there is identity degradation once the report is fused with the previous information but not completely.

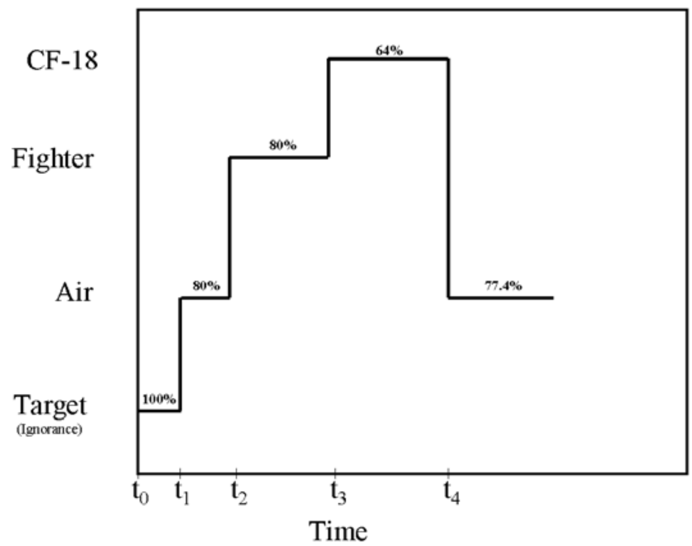


Figure 5: Taxonomy level of the target best ID proposition as a function of the time.

The following Table 2 summarises the results for all identities by showing the BPA of each proposition as a function of the time of fusion:

Table 2: Taxonomy level of the target best ID proposition as a function of the time

Proposition	t_1	t_2	t_3	t_4
CF-18	0%	0%	64%	12.8%
Air-Fighter	0%	80%	32%	6.4%
Air	80%	16%	3.2%	77.4%
Bomber	0%	0%	0%	3.2%
Ignorance	20%	4%	0.8%	0.2%

Even in the presence of conflict, the identity of the platform after the conflicting report remained partly correct. We observe that the belief in the AIR statement (sum of all BPAs of propositions that are sub-sets of air) remained after t_4 with 99.8%. This is in contrast to the OS method, which would have reduced the BPA of the Air proposition to approximately 24%. Here also the probability of the platform being a Fighter is diminished and a new Bomber statement appeared with lower confidence. The next ESM report, if not conflicting, will bring the correct identity back. This treatment of conflict, is probably more adapted to what an operator would like to see in the presence of conflicting evidence, namely assigning BPA to the first non-conflicting branch of the tree, rather than renormalizing all the fused propositions in the presence of a large conflict.

5.0 CONCLUSIONS

This paper has presented the design of a comprehensive a priori PDB for MSDF and STA applications, and examples of its successful use though the DS evidential reasoning scheme, under extreme conditions, e.g. involving countermeasures, badly interpreted imagery, wrongful associations for large target density, etc. Several versions of the DS method and several scenarios were presented in order to present a large spectrum of realistic applications. The realism of the scenarios was provided by simulators from DRDC-V for radar (and ESM) and from DRDC-O for SSAR imagery.

Furthermore, this DS-based ID determination has been demonstrated with success in current sea trials which have been underway as part of the Command Decision Aid Technology (COMDAT) for the HALIFAX class frigates.

6.0 ACKNOWLEDGEMENTS

The authors would like to thank all the co-authors of the various references below, in particular former Lockheed Martin Canada research scientists Frédéric Lesage for implementation of the HOS rule on a toy scenario, and Alexandre Jouan (presently a DRDC-V Defence Scientist) for the implementation of the OS rule to the complex scenarios.

7.0 REFERENCES

Only the most recent references are shown below. Earlier references (there are many before 1999) can be found within these.

[1] Valin, P., Jouan, A., and Bossé, E. (1999), Fusion of imaging and non-imaging sensor information for airborne surveillance, Proc. SPIE Aerosense Conf. 3719 on Sensor Fusion: Architectures, Algorithms, and Application III, Orlando, 7-9 April 1999, pp. 126-138.

[2] Jouan, A., Valin, P., and Bossé, E. (1999), Testbed for Fusion of Imaging and Non-Imaging Sensor Attributes in Airborne Surveillance Missions, FUSION 99, 6-8 July 1999, Sunnyvale, CA, Vol. 2, pp. 823-830.

[3] Valin, P., Tessier, Y., and Jouan, A. (1999) Hierarchical Ship Classifier for Airborne Synthetic Aperture Radar (SAR) Images, Proc. 33rd Asilomar Conference on Signal, Systems and Computer, 24-27 October, Pacific Grove, pp. 1230-1234.

[4] Simard, M.A, Valin, P., and Lesage, F. (2000), The ID box: a Multisource Attribute Data Fusion Function for Target Identification, Proc. SPIE Aerosense Conf. 4051 on Sensor Fusion: Architectures, Algorithms and Applications IV, Orlando, April 24-28 2000, pp. 180-190.

[5] Valin, P., and Boily, D. (2000), Truncated Dempster-Shafer Optimization and Benchmarking, Proc. SPIE Aerosense Conf. 4051 on Sensor Fusion: Architectures, Algorithms and Applications IV, Orlando, April 24-28 2000, pp. 237-246.

[6] Valin, P. (2001), Reasoning Frameworks for Fusion of Imaging and Non-imaging Sensor Information, in Proceedings of the Workshop on Estimation, Tracking and Fusion: A Tribute to Yaakov Bar-Shalom, Naval Postgraduate School, Monterey, CA, May 17 2001, pp. 269-282.

Analysing an Identity Information Fusion Algorithm Based on Evidence Theory

Anne-Laure Jousselme, Éloi Bossé, Alexandre Jouan
Defence Research and Development Canada – Valcartier
2459 Pie-XI Blvd North
Val-Bélair, Quebec G3J 1X5
CANADA

E-mail: Eloi.Bosse,Anne-Laure.Jousselme,Alexandre.Jouang@drdc-rddc.gc.ca

ABSTRACT

In this paper, we analyse an identification algorithm in the evidence theory framework. The identification algorithm is composed of four main steps: (1) sensor reports are transformed into initial Basic Probability Assignments, (2) the successive BPAs are combined through Dempster's rule, (3) the resulting BPAs are approximated to avoid algorithm explosion, and (4) in parallel to step (3) a decision is taken on the identification/classification of an object from a database based on the maximum of pignistic probability criterion. The identification algorithm is applied to a Direct Fleet Support scenario where ESM reports are fused to identify six targets among a possibility of 142 in the database. As a basis for the analysis, we observe the behaviour of (1) the pignistic probability of the singletons of the database, used as decision rule, (2) the distance between a BPA and a "solution" (ground truth), (3) the distance between an approximated BPA and a non-approximated one, and (4) the non-specificity of the BPA.

Keywords: *Target Identification, Evidence Theory, Distance, Non-specificity.*

1.0 INTRODUCTION

The need for timely and accurate processing of large amounts of uncertain and possibly incomplete data from multiple dissimilar sources is felt in many industrial and defence contexts. Most of the time, the fusion of information coming from the multiple sources is being manually performed by the operators/users. This process of manually and mentally re-plotting information by the staff and the commander is very laborious, complex, time consuming and prone to error. Furthermore, the amount and complexity of information now available has made this type of data fusion impractical and the situation is worsening as new surveillance sources become available. Mental and manual data fusion must be replaced by automated data fusion wherever it makes sense and is possible to assist the operators in coping with the ever-increasing flow and complexity of information, in their task of tracking and identifying multiple targets.

This paper focuses on the fusion of identity information that can handle organic and non-organic, local and remote types of information characterized by different accuracy and timeliness. The theory of evidence (Dempster-Shafer) has been proposed [1,2] as a promising avenue in combining information coming from different sources in the particular objective of target identification. However, one major inconvenient of the combination rule used in this theory (Dempster's rule) lies on the exponential increase of the number of

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

propositions (focal elements). This number becomes rapidly unmanageable causing a serious problem for real time applications.

To avoid this problem, some approximation algorithms [3] have been investigated that aim at reducing the number of focal elements according to different criteria such as the maximum of focal elements allowed, the minimum mass for the focal elements to be kept, etc.

In previous papers [3,4], we proposed an error measure based on the distance between the approximated function and the original one (without approximation). This distance provides a means to quantify the quality of the approximation.

In this paper, we propose an additional measure used for comparison of approximation algorithms based on some measures of uncertainty as described in [5], especially the non-specificity. A new representation using error bars whose the size depends on these additional measures, combined with the distance between both belief functions proposes an easy way to analyze the efficiency of approximation rules. Indeed, in addition of quantifying the distance between both belief functions, different uncertainty measures can be represented, quantifying thus in a different manner the loss of information induced by the approximation. We present some results based on simulations for target identification scenarios, and show how a good approximation rule can significantly reduce the algorithm processing time. In section 2 the problem of target identification is introduced, as well as the evidence theory. Then, the target identification process consisting of four main steps is described: *initial Basic Probability Assignment construction*, *combination of BPAs*, *propositions management* and *decision*. In section 3, some measures to analyse the identification algorithm are proposed and the application is illustrated by a scenario in section 4, where some interesting behaviours are shown.

2.0 TARGET IDENTIFICATION WITH EVIDENTIAL THEORY

2.1 The Target Identification Problem

Each platform (surface, sub-surface, air and land) that can potentially be detected and identified by the sensors or the information sources of a military surveillance system is listed in a Platform Data Base (PDB). Each of them is described in term of the parameters that are measured by the sensors. The PDB also contains two additional knowledge sources: the Geo-Political Listing (GPL) and the Emitter Name Listing (ENL). The GPL lists the attribute data that are assessed by the COMINT sensor for each country or organization of the world while the ENL includes the name and class of all the radio emitter sources that can be detected by the ELINT sensor. Note that several platforms are enumerated in different variants differing mostly by their emitter list, corresponding to sequential platform upgrades while in some cases, the emitter list may be so specific as to correspond to an unique platform rather than a variant of a specific class.

In the Direct Fleet Support (DFS) scenario presented in the section 4.1, a CP-140 surveillance aircraft equipped with a radar antenna (for target tracking and SAR imaging) and ESM antennas is sent for a reconnaissance mission over a fleet of US ships. The information to be captured is made of ESM reports, SAR imagery, and kinetic data. This information can be retrieved sometimes quite directly such as in the case of ESM reports that consist of an emitter list with some confidence level about the accuracy of the list that reflects the confidence in its electromagnetic spectral fit. Kinematic information is however more complex in nature and is generated from the data provided occasionally by the target tracker. Firstly, each physical quantity has a different dimension (speed, acceleration) and an accurate determination is not necessarily needed for fusion. Indeed, it is convenient to bin the attribute "speed" into fuzzy classes like "very fast",

"fast", "average", "slow" and "very slow" (separately for air and surface targets). Similar binning for acceleration could range from "very large *g*" to "very small *g*". Membership in each class is a measure of how well the measured value fits into the descriptor as described below in the next section. Further, speed or acceleration reports must be fused only if they involve a significant change from past historical behaviour in that track. The reason is two-fold: firstly no single sensor must attempt to repeatedly fuse identical ID declarations otherwise the hypothesis that sensor reports are statistically independent is violated, and secondly the benefits of the fusion of multiple sensors are lost when one sensor dominates the reports. This is clearly the case if positional fusion reports the same value of speed for hours at intervals of a few seconds!

Furthermore, a measured value of speed (or acceleration) only indicates that the target is capable of that speed but not to correspond to either V_MAXI (or V_MINI) of the PDB (nor the maximal ACC of the PDB). It is a reasonable working hypothesis to fuzzify the value reported by the tracker into adjacent "bins" to account for the target being at, say only 80% of its optimal speed (a "very fast" target can occasionally travel "fast"), or travelling with a strong tailwind (a "fast" target can occasionally appear as "very fast"). Finally the concept of binning can be generalized to continuous membership functions of a fuzzy set as outlined in the next section. Similarly, the image interpretation module for the SAR imager can generate a nearly infinite set of declarations from a single given image. Care must be taken to preserve as much independence as possible between the declarations and certainly to try preventing any conflict. Such independence can be achieved to a reasonable extent if different features are extracted from the image in different steps or if totally different mathematical algorithms are used in each step.

The Dempster-Shafer (DS) theory of evidence offers a powerful approach to manage the uncertainties within the problem of target identity. DS theory is particularly suited for our application because it requires no a priori information, can resolve conflicts (present in hostile environments due to countermeasures), and can assign a mathematical meaning to ignorance (which is the result of some of the chosen algorithms).

However, traditional DS has the major inconvenience of being an NP-hard problem. As various evidences are combined over time, DS combination rules will have a tendency to generate more and more propositions (*i.e.* focal elements) which in turn will have to be combined with new input evidences. Since this problem increases exponentially, the number of retained solutions must be limited (see section 2.2.4).

2.2 Review of the Theory of Evidence

The theory of evidence or Dempster-Shafer's theory [6,7] has been shown to be a good tool for representing and combining pieces of uncertain information.

Let Θ be the frame of discernment, *i.e.* the finite set of N mutually exclusive and exhaustive hypotheses, $\Theta = \{1,2,\ldots,N\}$. The power set of Θ, 2^Θ is the set the 2^N subsets of Θ, $2^\Theta = \{\varnothing, 1, \ldots, N, (1,2), (1,3), \ldots, (N\text{-}1,N), (1,2,3), \ldots \Theta\}$, where $\varnothing$ denotes the empty set.

2.2.1 Basic Probability Assignment

A Basic Probability Assignment (or mass function) is a function m from 2^Θ to $[0,1]$ which satisfies the following conditions:

$$\sum_{A \subseteq \Theta} m(A) = 1 \qquad (1)$$
$$m(\varnothing) = 0$$

$m(A)$ is called Basic Probability Number (BPN), or simply mass. It represents our confidence in the fact that "all we know is that the object belongs to A". In other words, $m(A)$ is a measure of the belief attributed exactly to A, and to none of the subsets of A. The elements of 2^Θ that have a non-zero mass are called focal elements.

Given a BPA m, two functions from 2^Θ to $[0,1]$ are defined: A belief function Bel, and a plausibility function Pl such that

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B) \text{ and } \mathrm{Pl}(A) = \sum_{A \cap B \neq \varnothing} m(B) \tag{2}$$

It can also be stated that $\mathrm{Pl}(A)=1-\mathrm{Bel}(A)$, where A is the complement of A. Bel(A) measures the total belief that the object is in A, whereas Pl(A) measures the total belief that can move into A. The functions m, Bel and Pl are one to one corresponding, so it's equivalent to talk about one of them, or also about the corresponding body of evidence.

2.2.2 Combination

Let m_1 and m_2 be two BPAs. The new BPA resulting from their combination is given by the Dempster's rule of combination:

$$m(A) = (m_1 \oplus m_2)(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) m_2(C) \tag{3}$$

K is called the conflict factor and is defined by

$$K = \sum_{B \cap C = \varnothing} m_1(B) m_2(C) \tag{4}$$

K measures the degree of conflict between m_1 and m_2: $K = 0$ corresponds to the absence of conflict, whereas $K = 1$ implies a complete contradiction between m_1 and m_2. Indeed, $K = 0$ if and only if no empty set is created when m_1 and m_2 are combined. On the other hand, $K = 1$ if and only if all the sets resulting from this combination are empty.

2.2.3 Decision

A BPA is distributed among different elements of 2^Θ. However, the observed object must be identified among the elements of Θ. Hence, the current BPA must be transformed so that a value can be assigned to each element of the frame of discernment. The identified object will be the one with the highest value. Different transformations exist on which a decision can be made. A popular one is the pignistic transformation proposed by Smets [8] as basis for decision in the evidential theory framework. The decision rule based on a BPA m is then

$$\theta^* = Arg\left[\max_{\theta \in \Theta} BetP(\theta, m)\right] \text{ and } BetP(\theta, m) = \sum_{\theta \in A \subseteq \Theta} \frac{m(A)}{|A|} \tag{5}$$

θ^* is thus the identified object.

This decision presents the main advantage compared to the maximum of plausibility that it takes into account the cardinality of the focal elements. Hence, when the algorithm is unable to converge towards a singleton, the pignistic probability does not tend towards 1, contrary to the plausibility. This will be illustrated in section 4.

2.2.4 Approximation

One major drawback in the combination rules of the evidential theory, is the exponential growing of the number of focal elements. Indeed, if the frame of discernment contains N elements, 2^N subsets can be created by the combination rules during the fusion process. Unless if N is very small, the number of new focal elements becomes rapidly unmanageable, and we often talk about "algorithm's explosion". To avoid this kind of "explosion", some methods are available which prune some useless focal elements. Among the proposed alternatives we retain two promising ones:

k-*l*-*x* **approximation** This algorithm for approximation of a BPA has been first proposed in [9]. It involves three parameters k the minimum number of focal elements to be kept, l the maximum number of focal elements to be kept and x the maximum threshold on the sum of the lost masses. It can be summarized as follows:

1) Select the k focal elements with highest masses;
2) While the sum of their masses is less than 1-x, and while their number is less than l, add the next focal element with highest mass.

$D1$ **approximation** This algorithm has been first presented in [10] and is summarized here. k is also the desired number of focal elements.

1) Select the k-1 focal elements with highest masses;
2) Distribute the other masses among the selected focal elements.

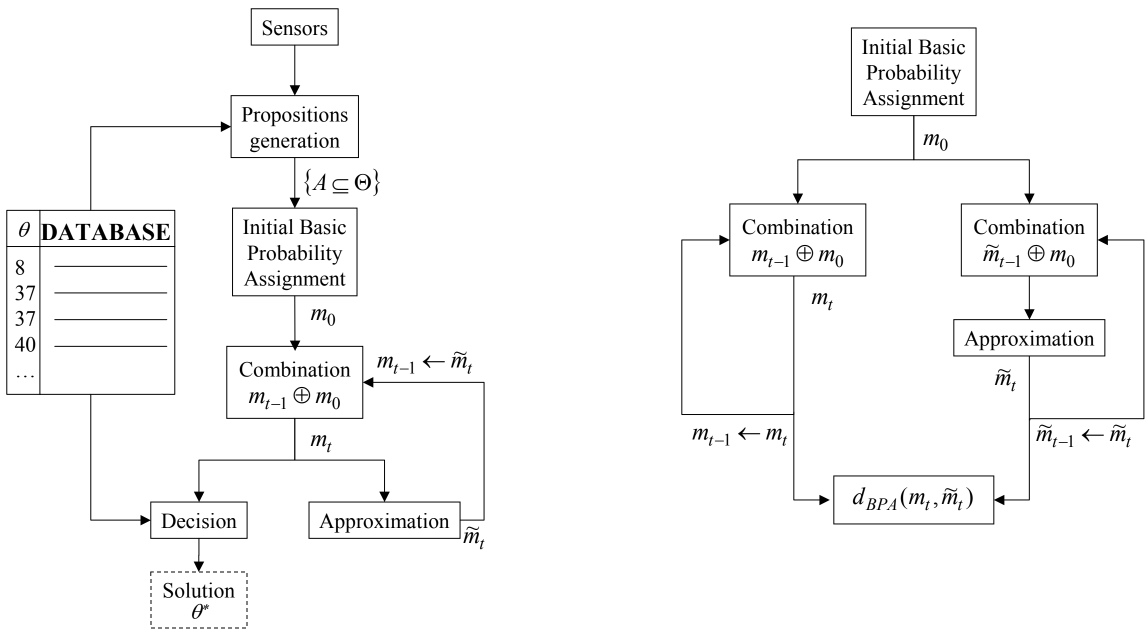
The distribution of the masses of the prunes focal elements lies on successive iterations and accounts for the relations between these focal elements and the remaining ones (objects in common, etc.). After the last iteration, the remaining mass is affected to the frame of discernment, which is then the k^{th} focal element.

The complete algorithm of masses distribution will not be detailed here, but can be found in [10].

2.3 The Identification Algorithm

A typical identification algorithm is shown on figure 1-(a). Such an algorithm is most of the time included in a general fusion process enclosing time data alignment, data association process, refinement of the database, decision process, etc.

From the information provided by sensor sources and by the use of a priori information (database), a new proposition is built. Then, based on this proposition, the Basic Probability Assignment takes into account some uncertainty or vagueness. Let us call m_0, the new incoming BPA. The core of the fusion process is the combination of m_0 and the BPA at the previous time, m_{t-1}. The resulting BPA at time t, m_t, is then the support for decision making. Using different criteria, the best candidate for identification is selected from the database. On the other hand, m_t must be combined to a new incoming BPA and thus becomes m_{t-1}. However, this step must be preceded by a proposition management step, where m_t is approximated. Indeed, the combination process being based on intersections of sets, the number of focal elements increases exponentially, and becomes rapidly unmanageable. This latter step appears as a crucial one as it can influence the whole identification process.



(a)- Typical Identification Algorithm **(b)- Approximation quality.**

Figure 1: Identification process and analysis.

3.0 ANALYSIS OF THE IDENTIFICATION ALGORITHM

In order to analyze the identification algorithm, we propose two main measures: (1) a distance between two BPAs, that can quantify the quality of an approximation rule as well as show the convergence of the identification algorithm; (2) a measure of non-specificity that can precise the quality of the convergence of identification algorithm.

3.1 Distance between two BPAs

Let m_1 and m_2 be two BPAs defined on the same frame of discernment Θ, then the distance between m_1 and m_2 is defined as [11]:

$$d_{BPA}(m_1,m_2) = \sqrt{\frac{1}{2}\langle (m_1 - m_2),(m_1 - m_2)\rangle} \quad \text{with} \quad \langle m_1, m_2 \rangle = \sum_{A \subseteq \Theta} \sum_{B \subseteq \Theta} m_1(A) m_2(B) \frac{|A \cap B|}{|A \cup B|} \tag{6}$$

This distance can be used for two distinct purposes in the analysis:

3.1.1 Distance to the Solution

In this case, $m_1 = m_t$ and $m_2 = m_{GT}$, where m_{GT} is the BPA based on the ground truth, the expected solution, and defined by:

$$m_{GT}(A) = 1 \tag{7}$$

A, being any subset of Θ. Although A is expected to be a singleton (see section 4.2), it can be a larger set of indiscernible objects for example, as it will be illustrated in section 4.3.

3.1.2 Approximation Quality

In this case, $m_1 = m_t$ the current BPA issued from combination rule (see figure 1-(a)) and $m_2 = \tilde{m}_t$ the approximated BPA of m_t. On figure 1-(b), the analysis of the approximation in the identification process is detailed. After the new incoming BPA m_0 has been combined with the older one $m_{t\text{-}1}$ giving m_t, the latter is approximated to guarantee a reasonable number of focal elements. The distance between the original BPA (*i.e.* without approximation) and the approximated one is then computed, and different approximation algorithms can then be compared, for example *k-l-x* or *D*1 algorithms introduced in section 2.2.4.

3.2 Measure of Non-Specificity

A BPA represents two types of uncertainty: non-specificity and conflict [5]. In this paper, we consider only the non-specificity, as a complementary measure of the quality of the identification algorithm. The non-specificity of a BPA m on Θ is defined as [12]:

$$NS(m) = \sum_{A \subseteq \Theta} m(A) \log_2(|A|) \tag{8}$$

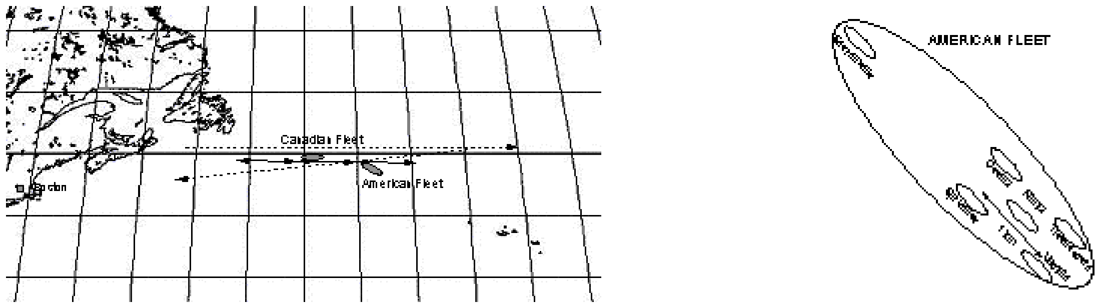
where |.| denotes the cardinality. The expression is widely accepted as the only one measure of non-specificity in evidence theory, because it is the only function satisfying the five properties of (1) additivity (for non-interactive BPAs), (2) subadditivity (for interactive BPAs), (3) normalization, (4) symmetry and (5) branching [5,13]. In particular, $NS(m) = 0$ if and only if m is a probability distribution on Θ (all focal elements are singletons), and especially if $m(\{\theta\}) = 1$ for some singleton of Θ[1] 1. This behaviour will be illustrated in sections 4.2 and 4.3.

4.0 ILLUSTRATION OF THE TARGET IDENTIFICATION PROCESS

4.1 Scenario

We consider the American Direct Fleet Support scenario whose complete description can be found in [14]. The location is 1,000 km due east of Greenwood in the mid-Atlantic where several CPFs and Iroquois class ships are heading towards Europe, eventually to enter the Mediterranean and pass through the Suez canal for support of NATO forces off Irak. Aircraft speed is close to the most economical cruising speed at 170 m/s (roughly 610 km/h or 330 knots) at an altitude of 7.62 km (25,000 ft). The scenario length is 3 hours as indicated by the 3 double arrows each covering 610 km (Figure 2-(a)). The Aurora passes 20 km south of a first group of Canadian ships heading due East just after the first hour and 100 km north of a second group of U.S. ships with SE heading towards the Islands just after the second hour. The flight pattern was chosen by the Aurora pilot so that the SAR need not be used to identify the Canadian contingent but far enough to be able to image the American flotilla.

[1] An additional measure of entropy on a BPA such that NS(m) = 0 will solve this ambiguity. Indeed, the entropy of this BPA will be 1 if and only if it exists θ such that m({θ}) = 1.



(a) Ships and aircraft in DFS operations **(b) American flotilla**

Figure 2: Direct Fleet Support scenario.

Table 1: The 6 targets of the American flotilla and their main features

Index	Name	Type	Emitter list
8	COONTZ	SURMILI	7 8 13 16 18 33 34 35 57
17	VIRGINIA	SURMILI	7 8 13 15 16 31 32 53 54 57
37	TICONDEGORA	SURMILI	7 8 13 32 53 54 57 110 112
39 1	SPRUANCE	SURMILI	8 14 18 31 32 43 53 57 114 115 119 12
55	NIMITZ	SURMILI	7 8 16 17 54 57 115 117 121 122 124 125 126 127
41	SACRAMENTO	SURMILI	7 13 18 33 121 130

The flotilla of US ships is composed with 6 surface targets named Coontz, Virginia, Ticondegora, Spruance, Nimitz and Sacramento (Figure 2-(b)). The American ships are in formation at 26 knots with longitudinal spacing between of 1 km and its transverse spacing half that (as seen in the figure below after two hours of the scenario when the Aurora flies by). Again the support ship lags behind a bit at 25 knots. It needs not be part of the convoy at this time because no threats have been anticipated from intelligence reports, even if it is within range of some possible threats from advance bases of hostile intent. In the DFS scenario, the American fleet is far enough to be imaged by the SAR but the geometrical considerations have degraded the ESM's contribution to the ID of the American fleet. Indeed, this time, the typical angular separation between ships can be as little as 0.5 km at 100 km or 0.005 radians which is quite comparable to the intrinsic bearing accuracy of the radar. In other words, even after Kalman filtering, the tracks are expected to be angularly very close and the changing aspects of the American fleet (heading at 45 degrees), as viewed by the CP-140, can cause several crossings in bearing between the tracks corresponding to each American ship.

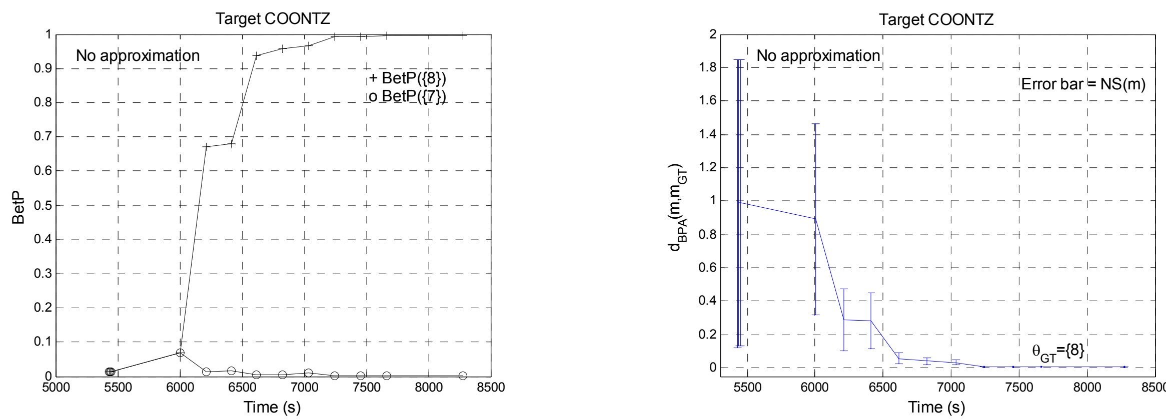
One thus expects a larger amount of false associations of ESM-to-track than in the Canadian fleet case, due to the combined effect of large (classified) ESM bearing accuracy and intrinsic filtered radar track accuracy. In order to follow the ID evolution, Table 1 shows the emitter list for the American fleet. Again the emitters carried by the platforms have many common elements and emitters are selected at random from run to run. From each ESM report an initial BPA m_0^{ESM} is built:

$$m_o^{ESM}(A_i) = 0.8 \quad \text{and} \quad m_o^{ESM}(\Theta) = 0.2 \tag{9}$$

where $A_i = \{\theta \in \Theta |\ \theta \text{ holds emitter } i\}$. $m_0^{ESM}(A_i)$ is a confidence level about the accuracy of the list that reflects the confidence in the electromagnetic spectral fit. In this scenario, only ESM reports are fused. The results presented below follow those presented in [4] and [3].

4.2 Convergence to the "Good" Solution

The solution is represented by a BPA m_{GT} such that $m(\theta_{GT})$=1, where θ_{GT} is the ground truth of the index of the observed target. Hence, in the studied scenario, $\theta_{GT} \in \{8;17;37;39;55;41\}$. In addition, we consider the non-specificity as the size of the error bars. Figures 3-(a) and 3-(b) show the good convergence of the identification algorithm towards singleton {8}, which corresponds to target COONTZ in the database. The evolution of the pignistic probability of this singleton is represented on figure 3-(a). The decision here is very clear and singleton {8} appears as the only one candidate for the target observed. The second in importance of pignistic probability is singleton {7} and is also represented on the figure for comparison. Figure 3-(b) shows the convergence of the algorithm towards the solution $m_{GT}(\{8\})$=1. Besides the fact that there is clearly only one possibility for the observed target, this solution is the good one. The error bars precise the non-specificity of the combined BPAs during combination time. Indeed, $NS(m)$ decreases rapidly towards 0 as the algorithm converges. This supports the fact that the object of convergence is a singleton.



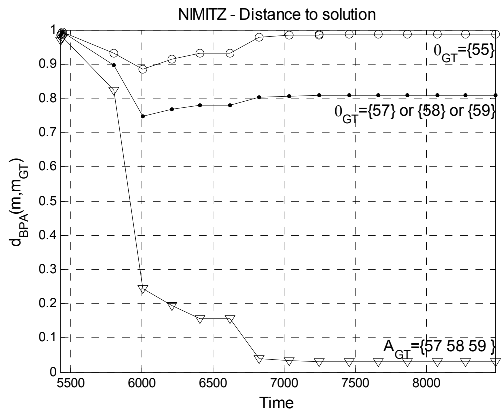
(a) Pignistic probability of singleton {8} without approximation

(b) Distance to the solution $m_{GT}(\{8\})$=1; Error bars are non-specificity.

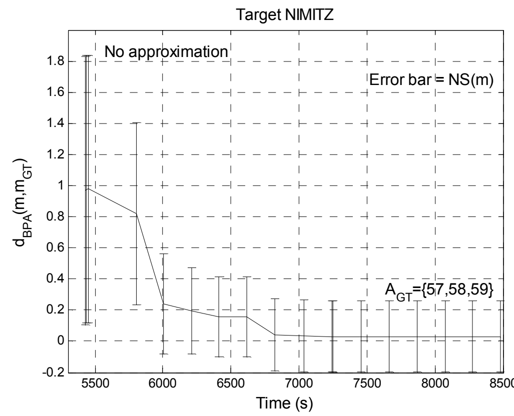
Figure 3: Correct identification of target COONTZ.

4.3 Convergence to the "Best" Solution

The behaviour of the identification algorithm can however differ even if it correctly identifies the target. The identification of target NIMITZ is represented on figures 4-(a) and 4-(b) and illustrates a case where the algorithm is unable to converge towards a singleton. On figure 4-(a) the convergence to three different solutions is shown: $m_{GT}(\{55\})$=1, corresponding to the target of the scenario but giving poor results in terms of the distance to the solution, $m_{GT}(\{57\})$=1, corresponding to another singleton approached closer than the good solution by the algorithm, and $m_{GT}(\{57;58;59\})$=1 corresponding to a subset of the frame of discernment, subset towards which the algorithm converges. This behaviour illustrates a case of indiscernible objects by the available sensors. Hence, solution should not be θ_{GT}={57}, but rather A_{GT}={57;58;59}, these three targets differing only by their name, an unavailable feature.



(a) Distance to different solutions.



(b) Distance to $m_{GT}(\{57; 58; 59\})=1$ and nonspecificity for error bars.

Figure 4: Convergence of identification algorithm for target NIMITZ.

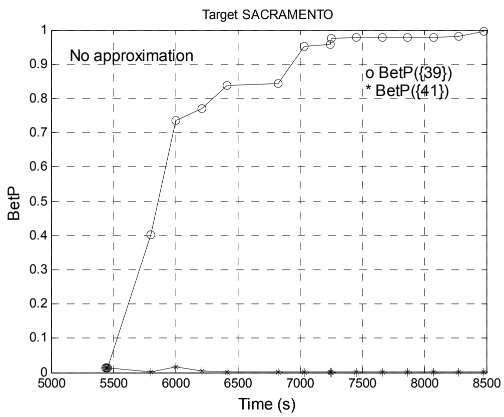
4.4 Convergence to a "Wrong" Solution

Due to the fleet configuration (see figure 2-(b)), some miss-identifications are directly issued from miss-associations. Indeed, some targets are mixed up. Figures 4-(a) to 4-(d) show the convergence of identification algorithm for targets SACRAMENTO and SPRUANCE is shown through the evolution of both BetP and d_{BPA} for different solutions. From these results, it appears that target SACRAMENTO is identified as target SPRUANCE, and target SPRUANCE is identified as target COONTZ. In this study, we assumed a correct association for all the targets and then processed each target separately. Hence, miss-association leads to miss-identification. Although the problem of association should be solved prior to identification, the result of identification could be helpful in a refinement of the association process. This is the purpose of a current research.

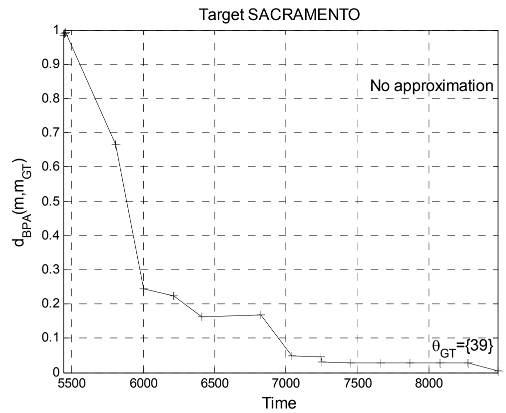
4.5 Impact of the Approximation Rule

The approximation rule aims at reducing the number of focal elements for the next BPA to be combined. The effect of such a rule is illustrated on figures 6-(a) and 6-(b) for targets COONTZ and VIRGINIA. In order to consider only the number of focal elements preserved by the approximation rule, we use the two following algorithms: k-l-x-n ($k = n$, $l = n$, $x = 0$) and $D1$-n ($k = n$-1). n is the number of focal elements to be kept, an integer set either to 8 or 10 in this paper. For evaluating the quality of the approximation rules we use the distance between two BPAs (equation (6)) following the scheme described in section 3.1.2 and on figure 2.

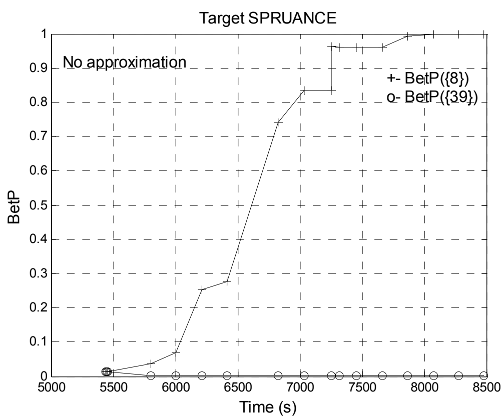
The comparisons of $D1$ and k-l-x algorithms for the six targets of the scenario are shown on figures 7-(a) and 7-(b), with a number of focal elements set to 8. Following these results the two algorithms are equivalent, both leading in finality to very low distances to the original BPA (*i.e.* without approximation), and this although a previous analysis on Monte-Carlo simulations identified $D1$ algorithm as the best one [3, 10].



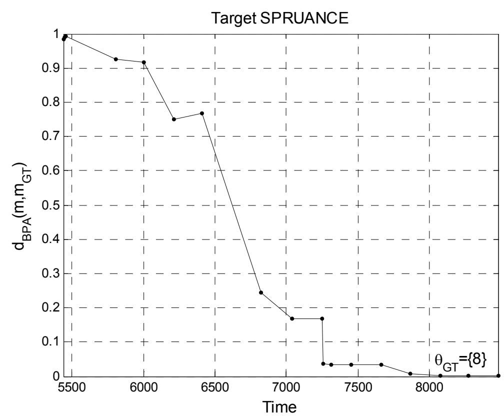
(a) Pignistic probability of singletons {41} and {39} for target SACRAMENTO.



(b) Distance to solution $m_{GT}(\{39\})=1$ for target SACRAMENTO.

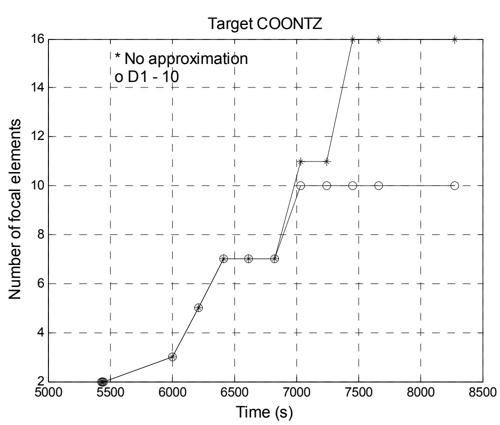


(c) Pignistic probability of singletons {39} and {8} for target SPRUANCE.



(d) Distance to solution $m_{GT}(\{8\})=1$ for target SPRUANCE.

Figure 5: Convergence to wrong solutions due to errors of association.



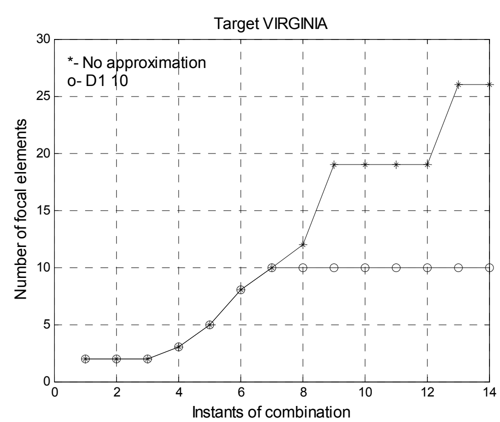


Figure 6: Impact of the approximation rule on the number of focal elements – Targets COONTZ and VIRGINIA - Algorithm $D1$ with $k=10$.

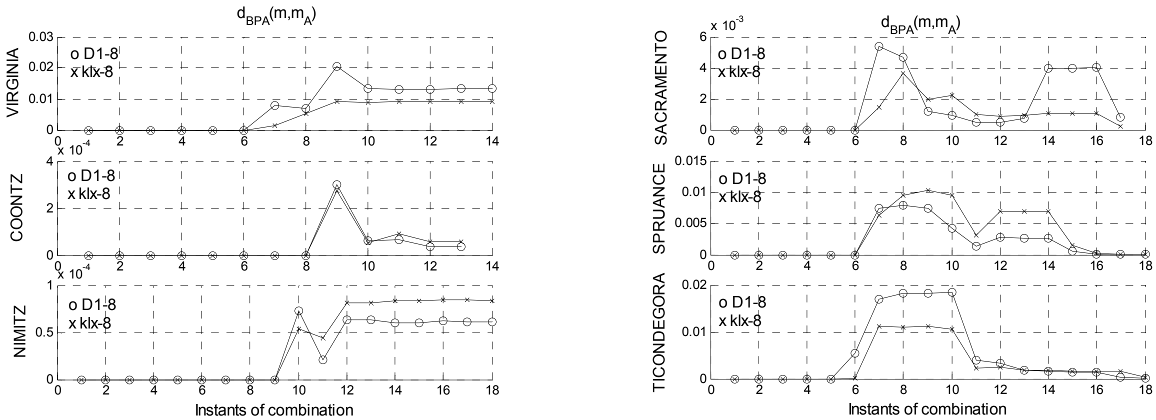


Figure 7: Comparison of approximation algorithms *klx* -8
and *D1*-8 based on the distance to the original BPA.

5.0 CONCLUSION

In this paper, we analyzed an identification algorithm based on evidence theory, through a simulation scenario.

As a basis for the analysis, we used

1) BetP($\{\theta\}$) the pignistic probability of the singletons of the database, used as decision rule. When this value tends to 1, the corresponding singleton is selected to be the solution, *i.e.* the observed target.
2) $d_{BPA}(m_t,m_{GT})$ the distance between a BPA and a "solution", to analyse the convergence of the identification algorithm. m_{GT} can represent the ground truth (expected solution) but also any other subset of the database towards which the algorithm converges. An exhaustive of the closest BPA reached by the algorithm could be accepted as a new decision rule.
3) $dBPA(m_t,\tilde{m}_t)$ the distance between two Basic Probability Assignments to quantify the performance of different approximation algorithms.
4) $NS(m_t)$ the non-specificity of the BPA, a complementary measure of the distance to the solution which clarify the convergence of the identification algorithm. In particular, when $NS(mt)$ tends to 0 besides the fact that $d_{BPA}(m_t;m_{GT})$ tends to 0 means a convergence towards a singleton.

Target COONTZ is the best-identified target as singleton {8} with no ambiguity. In two cases (targets SPRUANCE and SACRAMENTO), probable errors of association lead to wrong identifications, *i.e.* a convergence towards one of the targets of the scenario, but not the good one. The identification of target NIMITZ highlights a lack at the database level (hence of the frame of discernment). Indeed, according to the available features provided by the sensors on the platform, some targets are indiscernible and should be gathered as a single element.

REFERENCES

[1] E. Bossé and J. Roy, "Fusion of identity declarations from dissimilar sources using the Dempster- Shafer theory," *Society of Photo-Optical Instrumentation Engineers*, vol. 36, pp. 648–656, mar. 1997.

[2] E. Bossé and M.-A. Simard, "Managing Evidential Reasoning for Identity Information Fusion," *Society of Photo-Optical Instrumentaion Engineers*, vol. 37, pp. 391–400, feb. 1998.

[3] A.-L. Jousselme, D. Grenier, and E. Bossé, "Analyzing approximation algorithms in the theory of evidence," in *Sensor Fusion: Architecture, Algorithms and Applications VI* (P. of SPIE, ed.), vol. 4731, (Orlando, FL), pp. 65–74, 2002.

[4] A.-L. Jousselme, E. Bossé, and D. Grenier, "More results on a metric to measure the performance of evidential reasoning for identity information fusion," in *4th International Conference on Information Fusion*, vol. I, pp. TuB3–3–TuB3–8, 2001.

[5] G. J. Klir and M. J. Wierman, *Uncertainty-Based Information*, vol. 15 of *Studies in fuzziness and Soft Computing*. Heidelberg, New York: Physica-Verlag, 2 ed., 1999.

[6] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[7] A. Dempster, "Upper and Lower Probablilities Induced by Multivalued Mapping," *Ann. Math. Statist.*, vol. 38, pp. 325–339, 1967.

[8] P. Smets, "Constructing the pignistic probability function in a context of uncertainty," *Uncertainty in Artificial Intelligence*, vol. 5, pp. 29–39, 1990. Elsevier Science Publishers.

[9] B. Tessem, "Approximations for efficient computation in the theory of evidence," *Artificial Intelligence*, vol. 61, pp. 315–329, June 1993.

[10] M. Bauer, "Approximation Algorithms and Decision Making in the Dempster-Shafer Theory of Evidence-An Empirical study," *International Journal of Approximate Reasoning*, vol. 17, no. 2-3, pp. 217–237, 1997.

[11] A.-L. Jousselme, D. Grenier, and E. Bossé, "A new Distance Between two Bodies of Evidence," *Journal of Information Fusion*, vol. 2, pp. 91–101, June 2001.

[12] D. Didier and H. Prade, "A note on measures of specificity for fuzzy sets," *International Journal of General Systems*, vol. 10, pp. 279–283, 1985.

[13] A. Ramer, "Uniqueness of information measure in the theory of evidence," *Fuzzy Sets and Systems*, vol. 24, no. 2, pp. 183–196, 1987.

[14] A. Jouan, P. Valin, and E. Bossé, "Testbed for Fusion of Imaging and Non-Imaging Sensor Attributes in Airborne Surveillance Missions," in *FUSION 99*, vol. 2, (Sunnyvale, CA), pp. 823–830, July 1999.

A Bayesian Network for Combat Identification

Ing Sicco Pier van Gosliga, MSc
Ir Henk Jansen
TNO Physics and Electronics Laboratory
P.O. Box 96864
2509 JG The Hague
THE NETHERLANDS

E-Mail: vangosliga@fel.tno.nl / h.jansen@fel.tno.nl

SUMMARY

This paper reports the results of an investigation at TNO Physics and Electronics Laboratory for the Royal Dutch Navy on how a Bayesian network can be used to introduce more transparency in decision aid as to the level of confidence of the information that is used. To assess the feasibility of a Bayesian approach to new decision support system concepts, we have chosen to use the identification process of air contacts aboard navy frigates as a case study. The identification process is a time intensive and mind consuming process that is critical to anticipate an air attack. Wrong decisions may have fatal consequences and for example: identifying a neutral aircraft as an hostile aircraft may cost the lives of many people and may cause undesired political instability. Vice versa, mistakenly taking a hostile aircraft for a friendly one will give opponents the tactical advantages of a surprise attack. Typically a list of identification criteria is used by air defense personnel to discriminate hostile aircraft from neutral and friendly aircraft. These predefined criteria may change from mission to mission but will always follow a strict scheme of "if-then-else" clauses. Although identification involves reasoning with uncertain information, current procedures do not make this uncertainty explicit. Embedding Bayesian inference techniques in decision support systems would enable us to reason with uncertainty in a scientifically sound and consistent manner. Bayesian networks can express the likelihood of a hypothesis such as the identity of air contact being hostile as an explicit value, even when information about a contact is uncertain and incomplete. Making this uncertainty explicit, enables navy personnel to know how much confidence it should have the probability of hypotheses that are based on it.

BAYES THEOREM

Aircraft contacts, defined by tracks, are processed by automated information systems and presented to navy personnel that evaluate each track's identity and threat by interpreting its observed characteristics. Typical characteristics are [1][2]: altitude, speed and maximum speed, sudden manoeuvres, flight in formation, country of origin, adherence to air lane and adherence to air traffic control orders (ACO). There are multiple methods in use to determine the most likely identity for a track: electronic support measures (ESM), identification friend or foe (IFF), cross-told identification and visual identification. Each of these methods will deliver an identity for the track. Obviously air defence personnel are facing a problem when two methods suggest a different identity to a track. In a broader context this problem is commonly referred to as the Information Fusion problem. An approach to deal with such inconsistencies is using Bayesian statistics to derive the most likely identity.

The Bayesian theorem is based on the assumption that the probability of an event is dependent on the presence of other events, if causal relations exist between them. The basic mathematical rules for probabilistic reasoning are the product rule, the sum rule, the theorem of total probability and Bayes

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

theorem (where Bayes rule is actually a consequence of the product rule). Since especially the last rule forms the foundation of all Bayesian systems it will be discussed briefly. We will use ‘P(h | e)’ as a formal notation for the probability (i.e., the degree of belief) of hypothesis ‘h’ being true, given circumstantial evidence ‘e’. The probability of an event is expressed as a value scaling from 0 to 1, where the value 1 (or 100%) says that the event certainly takes place or is believed to take place and 0 (or 0%) definitely not.

$$p(h \mid e) \times p(e) = p(e \mid h) \times p(h)$$

Figure 1: Bayes theorem.

The left-hand term, ‘P(h | e)’, is called the posterior probability. It represents the probability of hypothesis ‘h’ after considering the effect of evidence ‘e’. The second term, ‘p(e)’, returns the chance of observing ‘e’. The term ‘p(e | h)’ is called the likelihood, and it stands for the probability of circumstantial evidence ‘e’ assuming the hypothesis ‘h’ is true. The last term, ‘p(h)’, is the prior probability of event ‘h’ reflecting our belief in ‘h’ without any additional information (i.e. without evidence ‘e’).

$$p(h \mid e_1 , e_2) = \frac{p(h \mid e_2) \times p(e_1 \mid h , e_2)}{p(e_1 \mid e_2)}$$

Figure 2: An alternative (extended) notation of Bayes theorem, where ‘e_2’ is additional evidence.

This paper discusses how to embed domain knowledge into a Bayesian network. A Bayesian network is a graphical model that represents the joint probability distribution for a large set of variables. Graphical models are graphs in which nodes represent variables, and the arcs represent the relations between them. We make distinction between two types of variables: observed variables and hypotheses. Being a graph structure, a Bayesian network can be manipulated with a large variety of mathematical methods from graph theory. Furthermore several belief-updating algorithms based on the Bayesian theorem, have been developed for such networks. Hence, the name: Bayesian belief network. Given that algorithms based on the Bayesian principle are mathematically consistent [3] (as opposed to e.g. Fuzzy Logic, Dempster-Shafer and uncertainty factors) we have chosen to use a Bayesian network structure to approach the combat identification problem.

Observed variables are those variables that have been assigned a value. The hypotheses are variables whose values are derived from the observed variables, but of which the actual values are unknown. For hypotheses an approximated value can be derived given the values of observed variables (e.g. ‘$\max[\, p(h \mid e_1 , e_2)\,]$’), based on a priori knowledge (e.g. ‘$p(e_1 \mid h , e_2)$’). This paper will not discuss how the Bayesian theorem is used to derive this information. People who are interested in the underlying algorithms be may be interested in the following literature: [4] [5]. The focus of this paper lies on how a domain can be modelled into a Bayesian network and uses the combat identification process as an example. Designing a Bayesian network takes three phases, which will be described now.

PHASE 1: DEFINING VARIABLES AS HYPOTHESES

Since the nodes in a Bayesian network form the basis of the graph, we need to define the variables first. Typically a set of variables is needed that represent facts that can be observed. The variables that cannot be observed will be treated as hypotheses. Common criteria for identifying tracks have been mentioned above, before we discuss how these criteria relate to each other we will discus how they can be used as variables in a Bayesian network model. The variables can be interpreted as representing so-called events. A variable may in theory have any number of states. A variable may, for example, be the speed of an aircraft (300 m/h, Mach 2 etc.) or the identity of a track (neutral, hostile or friendly). Important is that a variable may only be in exactly one state at a time. An aircraft cannot fly 300m/h and 600 m/h at the same time; neither can a track be hostile and neutral at the same time. The state may be unknown to us. When the state of a variable is unknown, the known states of other variables and prior knowledge can give us an indication of what the unknown state may be.

The variable "track identity" is the central hypotheses. In the case discussed in this paper, its value is unknown. To assess the most likely state of "track identity" observations are necessary. Prior knowledge alone is inadequate to reach the required level of certainty. Such prior knowledge may be the known ratio of friendly planes to neutral and hostile planes in our neighbourhood. This prior knowledge is as dynamic as the surrounding world is. The mentioned criteria that are used for combat identification will be embedded as variables in our model.

On decision level, the set of feasible actions is limited and therefore discrete. For this reason we will strictly use discrete variables. Continuous variables (e.g. speed and altitude) will be discretized. There are also practical reasons why discrete variables are preferred. To support the decision-making process the likelihood of each state must be calculated for all unobserved variables. The total probability of all these states should sum up to 1. When the set of values for a variable would be very large, the probability for a value will be close to zero. Such small numbers introduce numerical problems and unnecessarily increase the computational complexity of the problem. The smaller the range, the quicker the algorithm is. There are Bayesian algorithms that can be used on continuous variables. However these algorithms use restrictions on the format of probability density functions (e.g. Gaussian distribution functions) or do not use a complete search of the solution space (e.g. Monte Carlo simulation). People who are interested in using continuous variables in Bayesian network are advised to read the following literature: [6].

$$\text{speed} \in \{0..k, k..2k, \ldots, s_{max} - k \ldots s_{max}\}$$

$$\text{altitude} \in \{0..a, a..b, b..c\}$$

Figure 3: Instead of continuous, discretized versions will be used for speed and altitude.

If we want to use continuous variables, they should be converted to discrete variables. To illustrate this process, we will discuss the variable speed that we will use for the speed of a track. First of all, we need to limit the range of this variable. We assume that a track will not fly faster then a certain speed (say s_{max}) and we won't accept speeds below zero: 0 < speed < smax. We can split this range in a number of equal sized parts, e.g. speed$\in\{0..k, k..2_k, \ldots, s_{max} - k \ldots s_{max}\}$. This gives us a discrete version of the continuous variable speed. Obviously the size of the set affects the accuracy of the model. Rather than using a high resolution, it is more effective to choose a range for speed that discriminates various possible identities of a track. Since civil planes in general fly within a certain speed range. Military jets may exceed the speed of civil planes, but not necessarily (e.g. military helicopters fly at lower speeds). A similar process is used for the variable altitude.

$$\text{adherence\_to\_ACO} \in \{\text{true, false}\}$$
$$\text{adherence\_to\_airlanes} \in \{\text{true, false}\}$$
$$\text{sudden\_manoeuvres} \in \{\text{true, false}\}$$
$$\text{flight\_in\_formation} \in \{\text{true, false}\}$$
$$\text{identification\_manoeuvre} \in \{\text{correct, incorrect}\}$$

Figure 4: The variables that will be used to describe a track's behaviour.

Commercial planes are obliged to follow civil air lanes. These lanes are static and commonly known. The recorded history of a track can be used to determine whether the plane was adherent to an air lane.

When an airliner has been outside its air lane, it might be an airliner in trouble or a hostile aircraft in disguise. Whether a plane is coherent to an air lane can be expressed as a fact (true or false), alternatively the distance to the mean of an air lane can be used to express the degree of belief of coherence. The same approach can be used for the other behavioural characteristics: adherence to air traffic control orders, sudden manoeuvres and flight in formation. An instructed manoeuvre is performed by friendly forces for identification purposes. Since new instructions are given on a daily basis and are restricted knowledge, performing a correct manoeuvre is considered to be a reasonable hard criterion to positively identify friendly forces. We will only use “correct” and “incorrect” in our model.

$$\begin{gathered}\text{ID\_by\_ESM} \in \{\text{neutral, hostile, friendly}\}\\ \text{IFF\_mode} \in \{\text{mode 3/ac, mode 4, other}\}\\ \text{ID\_by\_visual} \in \{\text{neutral, hostile, friendly}\}\\ \text{ID\_by\_intel} \in \{\text{neutral, hostile, friendly}\}\\ \text{ID\_by\_FAC} \in \{\text{neutral, hostile, friendly}\}\\ \text{ID\_by\_ally} \in \{\text{neutral, hostile, friendly}\}\end{gathered}$$

Figure 5: Various methods can be used to derive an identity for a track. However, the results of these methods may differ. This set of discrete variables will be used to hold their values.

Electronic warfare support measures (ESM) forms the division of electronic warfare involving action taken to search for, intercept, identify and locate radiated electromagnetic energy for the purpose of immediate threat recognition. It provides a source of information required for immediate decisions involving electronic countermeasures, electronic counter-countermeasures and other tactical actions such as avoidance, targeting and homing. ESM is limited as a method of identification since it relies on goniometry to relate sensed radiation to a specific track. Therefore its readings are less reliable in dense air traffic, because it will be uncertain which sensor reading relates to what track. One might consider adding an extra variable to describe the density of air traffic around the assessed track, when using EMS as a criterion for identification. Characteristic patterns of electromagnetic radiation from known planes can be measured effectively and used as a reference. The availability of usable reference material affects the reliability. Another automated way to determine a track's identity is using IFF, “identification friend or foe”. This system was developed as a military system to discriminate hostile forces from friendly forces. All IFF systems use a “challenge and response” system. In civil systems the ATC Beacon system will “challenge” an airborne transponder. The challenge will be for a specific mode and wait for a response. The aircraft’s transponder will determine which mode was used and reply with an appropriate response. Modes are determined by the timing between the challenge pulses. IFF has 4 modes:

- The first mode originates from the original military system that was used in military air traffic control to determine what type of aircraft is answering or what type of mission it is on.
- The second mode is a newer system that replaced the first, also only for military use. This mode requests the tail number to identify a particular aircraft.
- The third mode is in use for civilian systems. It knows optional additional modes:
  - Sub-mode A is used internationally to identify airliners for air traffic control.
  - Sub-mode C altitude encoding for air traffic control (normally used in conjunction with 3/A).
  - Sub-mode is WIC war identification code for military use.
- Finally mode 4 is the military system that is in use nowadays. Instead of modes 1 and 2, this mode is encrypted and therefore more secure. It uses NATO specific crypto-secure IFF broadcast messages and is considered to be the most reliable criteria to determine the identity of a track.

IFF is a discrete variable. The introduction of mode 3 and mode 4, made the first two modes obsolete. It is unlikely that they will be used in wartime. Airliners normally use mode 3 in conjunction with sub-modes

A and C. Air traffic control relies on airliners to use these modes, therefore we will only accept mode 3 when both sub-modes are used. This leaves us with NATO secure mode 4 and the value "other reply" for all replies that are not covered by both states. Note that the value "other reply" is not the same as "no reply observed". Normal visual information and visual images derived from the optical monitoring of the electromagnetic spectrum from ultraviolet through far infrared can be used effectively to assess the identity of tracks. The visual identification process is traditionally performed by human agents. Automated systems match observed images to data and patterns stored in a database. The reliability is dependent on the quality of the image and the visibility of distinguishing features. When another party (e.g. a forward air controller, an ally or intelligence) has identified a track, the reliability of this identity is dependent on the source. Since multiple sources may provide a cross-told identity for a single track. Therefore the values for information source are not mutually exclusive and a separate variable for each source required.

PHASE 2: DESIGNING A CAUSAL NETWORK

Bayesian networks or belief networks are directed graphical models. Directed graphical models take into account the directionality of the arcs that link the nodes together and are not allowed to have directed cycles. This has several advantages of which the most important is that one can regard an arc from A to B as indicating that event A "causes" event B. Thus the arc direction is determined by chronological order of events. Using causal relations as a guide to construct the graph structure, leads to a reasonably intuitive model of its domain. In addition causal structures can be learned by fitting the strength of the causal relation to data. This can be convenient in problem areas were domain modelling is problematic.

Although the direction of each causal relation must point into one single direction, it does not mean that the state of a child will not affect the likelihood of its parent's states. The likelihood of events that form the circumstantial evidence of hypotheses may on their turn depend on other criteria. For example: for a track the adherence to an air lane depends on its position and course, while the probability of that track having a certain position may depend on the weather conditions. It is forbidden to use bi-directional arcs. In practice this means that causal cycles like the chicken-and-eggs problem are not allowed. On occasions where the directionality of a causal relation is not clear it is hard to construct a network. It is not allowed to have $A \rightarrow B$ and $B \rightarrow A$ in the same network, because cycles are prohibited. Normally, $A \rightarrow B$ and $B \rightarrow A$ should be redundant to each other. Because Bayes product rule is symmetrical, $[P(A \mid B) \times (B) \Leftrightarrow P(B \mid A) \times P(A)]$, there should be no functional difference. If it is not possible to translate $A \rightarrow B$ into $B \rightarrow A$ and vice versa, the model is semantically incorrect. The problem is likely to be caused by the fact that two separate events are represented by one and the same variable: $A \rightarrow B$ and $B \rightarrow A'$, where A' is in fact a different instance than A. Another cause may be that A and A' are referring to the same instance, but at a different moment in time. Our model is centred on the identity of a track. In the previous section we introduced a set variables based on criteria that are normally used for combat identification. We will now bind these variables together in a network structure based on the causal relations between them. Some of the causal relations are strengthened in the presence of other events. For these events extra variables will be introduced.

Figure 6: On the left a decision tree is shown for the identity of a track. On the right a graph is seen that shows the structure of the causal relations between altitude, speed and identity.

Both models in figure 6 describe the relation between altitude, speed and identity. The left model describes how the identity of a track is *determined* by altitude and speed, where the right model describes how the speed and altitude are *influenced* by identity. What is the exact difference? If we consider all nodes as variables describing events, the timing of these events defines the direction of the arc. When a track has a certain identity, say hostile, it is likely to fly low at a high speed. The track was already hostile before it decided to fly low. It did not become hostile because it was flying at a high speed or a low altitude. A counterexample: a civil plane can fly low, when it is nearing an airport. The airport did not suddenly appear because the plane was flying low; of course it was already there. Since the airport is near, the pilot is triggered to fly low.

identity $\Rightarrow$ altitude $\wedge$ speed
near airport $\Rightarrow$ altitude
near airport $\in$ {near, not present}

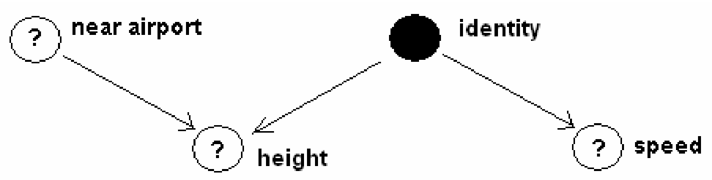


Figure 7: The graph on the right shows how knowing the state of 'identity' blocks the relation between 'speed' and 'altitude'. On left a textual representation of the same causal relations.

The direction of the arcs guides the belief update algorithm through the network. One way of looking at a Bayesian network is that observed events trigger other events that are directly related to it to update the likelihood of their states. These updated events will trigger related events on their behalf, and so on. In that way belief updates will propagate though network away from the observed events. This is why the direction of the arcs is that important. Cycles would create an endless loop for such belief updates, exactly as it does in the chicken-and-egg problem. The directionality guarantees that the belief update will stop at certain points in the network. Without going into further detail on how Bayesian inference algorithms work, the small network in figure 7 illustrates how belief updates are “blocked” by directed arcs. The figure illustrates the following relations: “a hostile aircraft is likely to fly at a low altitude to avoid radar contact” and “a civil plain is not likely to fly that low, unless it is landing”. Suppose we only know the *altitude* of a plane, then the *altitude* gives us a clue of the *identity* of a plane, hence the most probable *speed*. When we do know the *identity*, but do not know the altitude. Knowing the *speed* will not help us to estimate the *altitude*, since there is no direct relation between them, only via *identity*. The knowledge of *identity* blocks the relation. Knowing identity dictates the likelihood of *altitude*. Only knowing whether the near is near an airport force us to reassess the probabilities of *altitude*. If *track near airport* is unknown, the prior probability of *track near airport* is used to derive the most likely state of *altitude*.

identity $\Rightarrow$ made sudden-manoeuvres
heavy turbulence $\Rightarrow$ made sudden-manoeuvre
identity $\Rightarrow$ flight in formation
identity $\Rightarrow$ adherence to air lane
turbulence $\in$ {true, false}

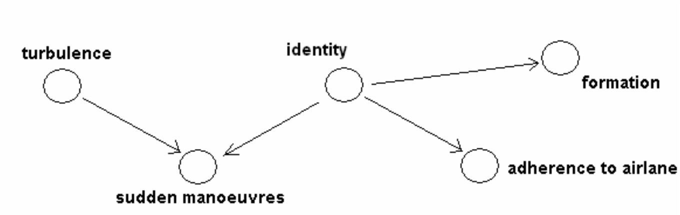


Figure 8: Graph structure of the causal relations between 'identity' and flight characteristics.

A track’s identity can be recognised by several behaviours that are typical for civil or military aircraft. Civil tracks follow flight lanes and keep a steady course and speed. Heavy turbulence may cause sudden manoeuvres as well, but will naturally not affect the identity of a track. Military tracks normally fly in formation and may use sudden course changes on their path. Carriers are unable to make fighter-jet manoeuvres.

$$\text{identity} \Rightarrow \text{ESM}$$
$$\text{identity} \Rightarrow \text{IFF}$$
$$\text{tracks close by} \Rightarrow \text{IFF}$$
$$\text{tracks close by} \Rightarrow \text{ESM}$$
$$\text{tracks close by} \in \{\text{true, false}\}$$

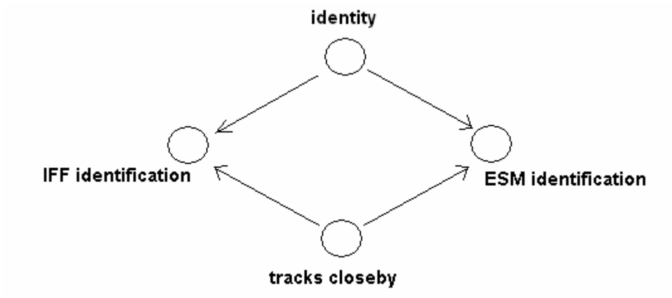


Figure 9: On the right a graph structure showing the causal relation between identity and identification methods IFF and ESM. Note the double dependencies.

Various technical aids are available to discriminate tracks. Established methods are Electronic Support Measures and the NATO secure IFF protocol. ESM identifies tracks by using pattern matching techniques on electromagnetic energy emitted by the track. Radiated energy is related to known characteristics. ESM is less reliable when multiple tracks are flying close to each other, because on such occasions it will be hard to pinpoint the source of the radiation. IFF has this same problem. Figure 9 shows the causal relations between these methods and a track's identity. Together with the three causal relations for the cross-told identities (allies, intelligence and forward air controllers) the relations in this section make up a Bayesian network as shown in figure 10. Note that this network does not represent a decision tree of the identification process. In a decision tree, the arcs would be the other way around. The network is a causal model centred on the identity of a track, describing a variety of factors that are symptoms of it.

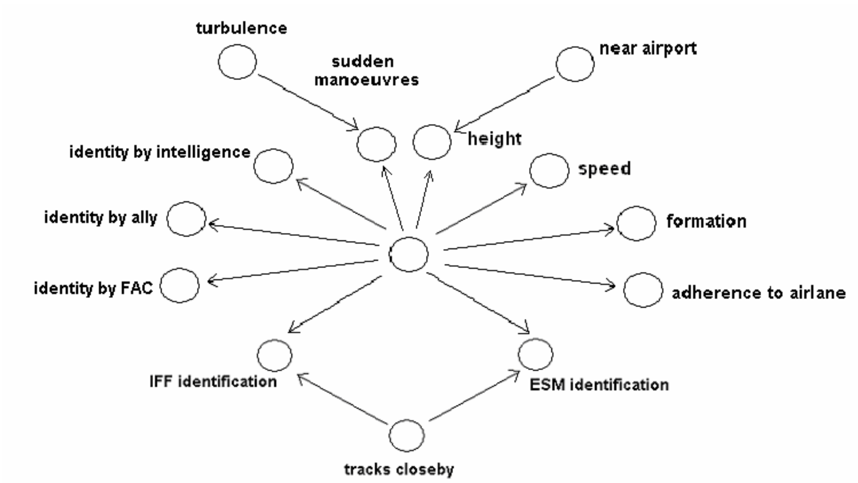


Figure 10: This graph structure gives an overview of all causal relations described in this paper.

PHASE 3: DEFINING THE PRIOR PROBABILITIES

Bayesian inference relies on probability distributions. Defining these distribution is a challenging process, since many probabilities are not known precisely, if not impossible to know exactly. A probability function describes the probability of a variable having a certain value, given the values of other variables to which a causal relation exists. A probability function over discrete variables can be structured as a probability table. In the previous section the following variables and correlations were introduced to describe the relations between the identity, speed and altitude of an aircraft.

altitude $\in$ {0..a, a..b, b..c}
identity $\in$ {hostile, friend, neutral}
speed$\in$ {0..k, k..2k,… ,s_{max} - k …s_{max}}
near airport $\in$ {near, not present}
identity $\Rightarrow$ altitude
identity $\Rightarrow$ speed
near airport $\Rightarrow$ altitude

p(airport = present) = 50%
p(airport = not present) = 50%
p(identity = hostile) = $33^1/_3$%
p(identity = neutral) = $33^1/_3$%
p(identity = friendly) = $33^1/_3$%

Figure 11: On the left a set of variables is seen and how they relate to each other. Two variables have no dependency on another variable (i.e. not present on the right-hand side of a relation). For both variables, the prior-probability is show on the right.

Table 1: this table holds all probabilities for the function P(altitude | identity, airport)

		altitude		
identity	**airport**	0..a	a..b	b..c
neutral	not present	10%	80%	10%
neutral	present	70%	20%	20%
hostile	not present	$33^1/_3$%	$33^1/_3$%	$33^1/_3$%
hostile	present	$33^1/_3$%	$33^1/_3$%	$33^1/_3$%
friendly	not present	$33^1/_3$%	$33^1/_3$%	$33^1/_3$%
friendly	present	$33^1/_3$%	$33^1/_3$%	$33^1/_3$%

For the part of network that encapsulates these variables (seen in figure 7) the following probability functions need to be defined: '*P(altitude | identity, airport)*', '*P(speed | identity)*', '*P(airport) and P(identity)*'. That is one table for each variable. The number of relations that affect a variable defines the number of dimensions for the table that describes its prior probabilities. The variables *airport near* and *identity* are independent of other events. If the likelihood of all states of a variable *v* are equal ($p(v{=}s_1) == p(v{=}s_2) == p(v{=}s_3)$ *etc.*), this means that the prior probabilities of *v* will not play a discriminating role of influencing the probabilities of the states of variables related to it. Basically this means that we do not want to assume anything on the presence of an airport when we do not know whether this is the case. Neither we want to assume anything on the identity. We can safely assume that the position of a track is always known. Given a database of all airports we could use the position of the track to determine whether an airport is near or not. In that case *position* is an observed variable, the state of variable *airport* will not have to be estimated. We could also use the position of a track to say something of the likelihood of a track having a certain identity. When we know how many friendly aircraft are in the area, we may have an estimate of the number of neutral and hostile planes in the area as well. This information could be used to dynamically generate prior probabilities, based on actual knowledge of the neighbourhood. The variable identity will remain unobserved, but if nothing is known about the context of a track except its location we could say something about the probability of that track being hostile. Naturally since the numbers of friendly in the area shifts over time, so do the probability tables of *identity*.

A method to determine the likelihood of each value for each context is learning from recorded data. Statistic methods can be used to estimate the chance of value based on the frequency of occurrence in the data for a certain context [7]. For many applications it is unlikely that training can learn all possible scenarios, because all possible states of the context will have to be available in the data. This is especially the case for combat identification. Even if this requirement is met, it is still not certain that recorded data is representative for future missions because of the ever shifting paradigm in dynamic environments. It is possible however to train subparts of the network which function within a relatively static system.

BENEFITS

By recognizing causal relations between criteria may be weak or strong depending on their context, it is better to include this knowledge in the decision-making process. Various tools can aid this process, but not all of them will take into account the dynamics of causality. Probabilistic models are an attempt to model causality in a less rigid way than rule-based systems. A Bayesian-based system can evaluate the identity of a track based on incomplete information, because it is able to estimate the values of the unobserved variables by using the prior knowledge and relate this to known observations. Not relying on complete information makes the system more reliant. When not all the information can be available at once, the system can determine the probability of posterior events and anticipate on the most likely before it has complete knowledge of what is to come. For the identification case this means, that all tracks can be given a likely identity even if certain information such a IFF or ESM is not yet available. Bayesian networks have the benefit to be able to determine the priority of missing information given the current context, without being vulnerable to tunnel vision. Because of this the information gathering process can be directed efficiently. Furthermore, a Bayesian-based system knows the weight (relevance) of all used criteria. For each hypothesis the most important reason can be given why it is valid or not. Rather than listing all reasons, it can highlight the most important one for the current situation. In a time-critical decision-making process, information overflow is a widely addressed problem. Context aware information filters have been proposed as a solution to ease the stress on decision makers. A Bayesian-based system can be used as a context aware information filter. The ability of Bayesian methods to prioritise information on relevance makes an effective tool for information presentation and gathering.

Knowledge in a Bayesian network is scalable. Adding new variables does not affect the prior knowledge. Adding causal relations, only affects the prior knowledge of the variables that are directly related to it. In a rule-based system the set of rules affects the outcome of the inference mechanism. For some mechanism also the order in which the rules are addressed is of importance. If rules are added to the system, one has to be aware that each extra rule can alter the functioning of the whole rule-set. To prevent undesirable effects, complete knowledge of the rule-set is necessary. Because of this, maintenance on rule-based system is a elaborate process and will never be a routine operation. Compared to rule-based systems, Bayesian-based systems can be scaled up easier.

CONCERNS

Probabilistic reasoning itself will not be the main challenge for the technical implementation of a Bayesian-based decision support system. Algorithms have been extensively described in scientific papers and implementations of these algorithms are available as ‘of-the-shelf’ software components from a broad range of parties. Most effort will be required for the construction of a sound model and the human factors of reasoning with uncertainties, presenting the output of the reasoning algorithm.

Special interest should be given to modelling and maintaining the knowledge embedded in the network. Since Bayesian systems use a model-based approach, a sound understanding of the domain is required. In the past expert systems have been usually based on heuristic knowledge. For model-based systems it is favourable not to use heuristic rules, when they are based on a weak correlation and not founded on understanding of the underlying mechanisms. Typically heuristics are simple diagnostic rules that draw a strict conclusion based on observed symptoms, while these symptoms may not be directly related to the cause. Such rules are out of place in a causal network, however rules that are founded on causal relations can be translated and will fit in [8]. In this paper the variables are presented as events. Although an event is related to a moment in time, this property is not exploited in the architecture of the described causal model. The presented model takes only into account the *occurrence* of events. However, knowing *when* and in *what order* events occur gives us a strong indication on the cause of events. Research in the field of probabilistic modelling and reasoning currently concentrates on embedding the dynamic aspects of time into Bayesian belief networks. Although these aspects have not been discussed in this paper, the authors would strongly advice engineers involved in causal modelling to be aware of the crucial role of time and make use of the latest developments in the field of dynamic Bayesian networks.

In this paper probabilities are represented as percentages. Research exposes that people find it hard to interpret hypotheses in such a notation [9][10]. Care should be given on how to represent probabilities. An alternative notation is the frequency notation, which expresses the probability of an event as the frequency of occurrence of one single event. A frequency-based notation seems to work better then percentages, but is not applicable for all variables and domains. Suppose a frequency based notation is used for the Air Defence case, how should someone act on a computer generated sentence: “On 1 in 3 times under these circumstances that plane would be hostile.”? A graphical representation avoids the textual interpretation problems. For the combat identification case a triangular representation was used (as seen in figure 12). An equilateral triangle can visualize three related dimensions such as hostile, friendly and neutral. Each corner of the triangle represents the upper boundary of a dimension. Because the values of a probability variable are mutually exclusive, a triangle can visualize each event that has three possible states. In this way hostile targets will be positioned in another corner then the neutral tracks. The tracks that have an unknown identity (i.e. not enough is known to discriminate between them) will be placed in the middle of the triangle. One advantage of this visualization is that hostile tracks are visually discriminated from other tracks, and found on a fixed location. For displaying certainties on a topological display, colour codes can be used to add an extra layer of information on a map. When all possible states of an event are represented by a colour; the mixture of these colours represents the identity of a track. This method may have some disadvantages, like incompatibility with the standard NATO colour codes and symbols. Although some possible ways of visualizing uncertainty have been examined, it has not been the focus of our research at TNO Physics and Electronics Laboratory. However, since research indicates that people have difficulties in both reasoning with and interpreting uncertainties [10], it is important for the acceptance of decision aid systems that human factors are given full attention in the engineering phase. The authors encourage research in this field.

Finally the problem remains what to do when a track seems to be hostile. This however is not mainly – or at all– an engineering issue but a military and political one. When is a track hostile enough to act? Release criteria have been unaddressed in this paper, although they are closely related and triggered by identification criteria. The benefit of using decision aid is to support the decision makers by delivering plausible and solid conclusions on the identity of each track during the dynamic process of air combat. By making uncertainty explicit, de decision process is getting more transparent.

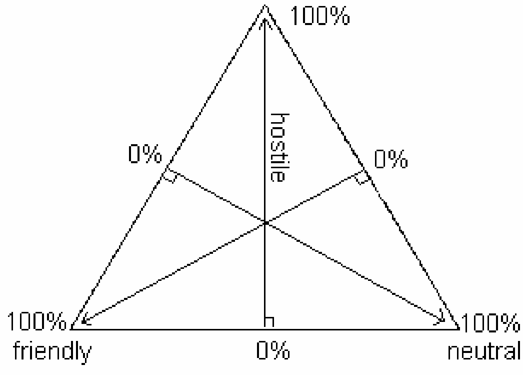


Figure 12: A triangular representation of a tracks identity.

CONCLUSION

By incorporating Bayesian techniques into the decision loop to assist in the identification process, time and energy can be saved. The model presented in this paper is merely a sketch of the actual Bayesian model for air defence identification and serves merely to illustrate our point of view on how such a model should be designed. The availability of software implementations of multiple Bayesian algorithms makes the technical implementation relatively easy. In general, designing an adequate and sound causal model of application domain will be the greatest challenge. Intimate knowledge of Bayesian statistics will not be a necessity, but planners do require a thorough understanding of the subject. Concerns remain however on how to present the results to navy personal and how the knowledge in the network should be kept up to date. The authors expect that further research may result in a better way to incorporate time aspects in Bayesian models and better ways to present the outcomes to navy personnel.

REFERENCES

[1] NATO Military Agency for Standardisation, “STANAG 4162 Technical Characteristics of the NIS Identification Data Combining Process”, March 2001 (NATO RESTRICTED)

[2] NATO, Military Agency for Standardisation, “STANAG 4162 Annex A: Source type description, the standardized requirement”, March 2001 (NATO RESTRICTED)

[3] C. P. Robert, Springer “The Bayesian Choice”, 11:507-418, 2001

[4] N. J. Nilsson, Morgan Kaufmann “Artificial Intelligence, A New Synthesis”, 19:317-342, 1998

[5] F. V. Jensen, Springer Verlag “Bayesian Networks and Decision Graphs (Statistics for Engineering and Information Science)”, 2001

[6] R. Schachter, C. Kenley, Management Science “Gaussian Influence Diagrams”, 35:527-550, 1989

[7] D. Heckerman, Microsoft Research Technical Report “A Tutorial on Learning with Bayesian Networks”, March 1995

[8] M. Korver, P.J.F. Lucas, Medical Informatics 18(3): “Converting a Rule-Based Expert System into a Belief Network”, 1993

[9] G. Gigerenzer, P. Sedlmeier, Journal of Experimental Psychology “Teaching Bayesian Reasoning in less than two Hours”, September 2001

[10] G. Gigerenzer, U. Hoffrage, Psychological Review “How to improve Bayesian Reasoning without instruction: Frequency Formats”, September 2001

Multi-Sensor Kinematic and Attribute Tracking Using a Bayesian Belief Network Framework

Mark L Krieg
Intelligence, Surveillance and Reconnaissance Division
Defence Science and Technology Organisation, Australia
PO Box 1500, Edinburgh, SA 5114
Australia

E-mail: mark.krieg@dsto.defence.gov.au

SUMMARY

Situational awareness plays a major role in many military and civilian operations. Apart from the identity, type, location and dynamics of targets of interest, the situational picture may also provide other target information, such as weapons state, fuel status and intent. Many legacy systems incorporate an automatic tracking capability, with identification, situational assessment and decision-making being left to the operators. The automation of many of these functions is the focus of much research and development.

A necessary prerequisite for updating the state of a target is the correct association of measurements or other information to the track. The ability of Bayesian belief networks (BBNs) to model the uncertain relationships between continuous and discrete variables make them excellent candidates for incorporating both kinematic and attribute information in the association process. A BBN model for a single scan data association problem is presented and used to develop a global nearest neighbours solution using both kinematic and attribute information. Monte Carlo simulations demonstrate the benefit of using attribute information in the association process.

1 INTRODUCTION

Situational awareness plays a major role in many military and civilian operations. The automated situational picture traditionally displayed the position, speed and heading of targets within a region of interest, with the operator manually adding identity and other information. However, modern multi-source algorithms have the potential to automaticially update additional target information or attributes, such as identity, class, category, weapons state, fuel status, threat level and intent, using the kinematic and attribute information from a variety of disparate sensors and sources. The Australian Defence Organisation is currently acquiring a number of capabilities, such as the Airborne Early Warning and Control, Air Defence Ground Environment and Joint Strike Fighter that will, or are expected to, provide some level of automation in this area.

The most critical and difficult problem in state estimation or tracking problems is that of data association. The use of attribute information in data association has the potential to provide greater discrimination and reduce association ambiguities. This paper addresses the joint kinematic and attribute data association problem, using a common probabilistic framework defined for both kinematic and attribute data using Bayesian Belief Networks (BBNs). This framework accomodates the uncertainty inherent in all sensor data, which is ignored in simple attribute gating techniques.

The paper is organised as follows. Following this introduction, an overview of relevant previous work is presented in Section 2, and a brief tutorial on BBNs is provided in Section 3. In Section 4, the joint kinematic

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

and attribute data association algorithm is developed, with the results of the *Monte Carlo* analysis in Section 4, and conclusions in Section 6.

2 BACKGROUND

Tracking the dynamic or kinematic state of a target is well documented [1]. However, the dynamic behaviour of a target and information from other sensors and sources, such as secondary surveillance radar (SSR), electronic surveillance (ES), tactical data links and intelligence sources, can be used to estimate other attributes of a target, such as its identity, class, threat level and intent. Two key challenges are to successfully combine the disparate information or data, and to handle the uncertainty or errors in the data. Bayesian and Dempster-Shafer are two approaches that are able to accommodate these challenges. Bayesian techniques use Bayesian probability theory to handle uncertainty, whereas Dempster-Shafer introduces the concepts of support and plausibility [2, 3]. Bayesian probability theory provides a common frame of reference for the kinematic and attribute variables, and has previously been exploited for joint tracking and classification [4, 5].

Bayesian belief networks (BBNs) are causal belief networks that model the dependencies between variables or network nodes using Bayesian probability [6]. The ability of BBNs to model both continuous and discrete variables, and the relationships between such variables, make them excellent candidates for developing Bayesian approaches to attribute tracking and fusion systems. Korpisaari and Saarinen [7] proposed a generalised Bayesian network structure with a many-dimensional association vector for applying attribute data in a Joint Probabilistic Data Association (JPDA) context. With Hautaniemi [8], these authors recognised how the dependencies between attributes generally produce non-singly connected networks for attributes that evolve over time. Chang and Fung [9] used a BBN to incorporate attribute information into a multiple hypothesis tracking algorithm. Krieg [10] showed how BBNs may be used as a basis for deriving the Kalman filter tracking algorithm and a joint kinematic and attribute nearest neighbours data association algorithm. He also developed an attribute estimation algorithm, which he applied to the target tracking and recognition application. In [11], he used BBNs to develop a joint tracking and classification solution.

3 OVERVIEW OF BAYESIAN BELIEF NETWORKS

A Bayesian belief network (BBN) is a causal network that is represented by a directed acyclic graph, where the nodes represent stochastic variables and the directed links or arcs represent the direct causal influences between the variables [12]. Bayesian calculus, in particular predetermined prior and conditional probabilities, is used to determine the posterior node probabilities from the data or evidence applied to the network. Although the network provides a complete probabilistic model of all the variables in a specific domain, its design is based on local interactions between variables that directly influence one another. Therefore, it is only necessary to determine the conditional probabilities for each variable [6]; a conceptually simpler task than defining the entire joint probability.

The *belief* or posterior probability of the state of each variable in a BBN is dynamically calculated from the static conditional probabilities and the propagation of evidence. When applied to the network, evidence is propagated in the direction of the causal links by π messages and against the direction of the links by λ messages. On receipt of a message from another node, a node will update its belief and propagate new messages to each node directly connected to it, thereby ensuring the evidence is propagated throughout the network. The belief at any node is the normalised product of the λ and π information at that node.

For simplicity, the following description of evidence propagation in BBNs is restricted to tree network structures. Unless stated otherwise, a node or variable is denoted by an uppercase alphabetic character, and its value is denoted by its lowercase equivalent.

Consider the propagation of evidence through the BBN in Figure 1. The evidence influencing the proposition $X = x$ is separated into two disjoint subsets, namely that introduced to X through the arcs between X and its

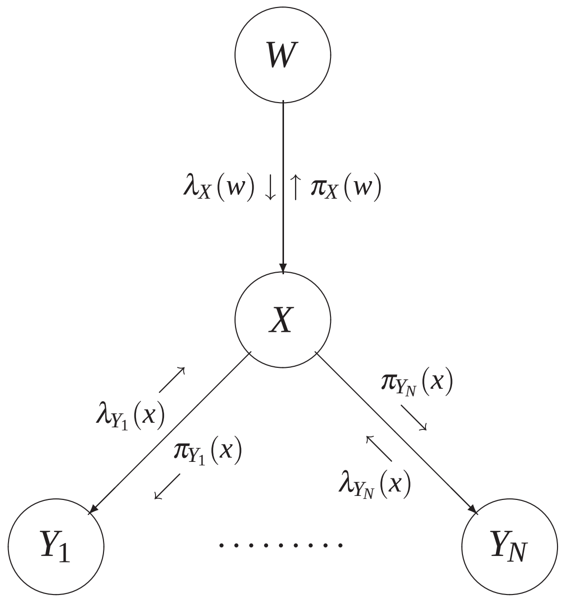


Figure 1. Evidence propagation in a BBN

children, $Y_1, \ldots, Y_N$, and that introduced through the arc between X and its parent, W. The former is aggregated into the likelihood $\lambda(x)$, representing the *diagnostic* or *retrospective* support for the proposition $X = x$, and the latter the prior probability $\pi(x)$, representing the *causal* or *predictive* support for $X = x$. The belief of $X = x$ is then simply

$$\mathrm{Bel}(x) = \alpha\,\pi(x)\lambda(x), \tag{1}$$

where α is a normalising constant.

For simplicity, and without loss of generality, it is assumed that all evidence is introduced into the BBN through leaf nodes. Then the retrospective support provided by the evidence on each leaf node Y_i, $i = 1, \ldots, N$, is reflected by the message $\lambda(y_i)$, which takes the value $\delta_{y_i y_i'}$, if instantiated with y_i', or some other value representing the probability that $Y_i = y_i$. This support is then propagated to X as

$$\lambda_{Y_i}(x) = \sum_{y_i} \mathrm{p}\,(y_i \mid x)\,\lambda(y_i), \tag{2}$$

and combined with the support from the other children of X using

$$\lambda(x) = \prod_{i=1}^{N} \lambda_{Y_i}(x). \tag{3}$$

The retrospective support is then propagated to W according to

$$\lambda_X(w) = \sum_{x} \mathrm{p}\,(x|w)\,\lambda(x). \tag{4}$$

The π messages propagate predictive support in the direction of the links. The support provided by the root node W is simply the prior probability of W, that is, $\pi(w) = \mathrm{p}\,(w)$, and this is propagated to X using

$$\pi(x) = \sum_{w} \mathrm{p}\,(x \mid w)\,\pi(w). \tag{5}$$

The propagation continues from X to the children of X using

$$\pi_{Y_i}(x) = \alpha \prod_{k=1\backslash i}^{N} \lambda_{Y_k}(x)\pi(x) = \alpha \frac{\mathrm{Bel}(x)}{\lambda_{Y_i}(x)}, \tag{6}$$

and

$$\pi(y_i) = \sum_x \mathrm{p}\,(y_i \mid x)\,\pi_{Y_i}(x). \tag{7}$$

The belief at all nodes may now be calculated.

4 JOINT KINEMATIC AND ATTRIBUTE FUSION

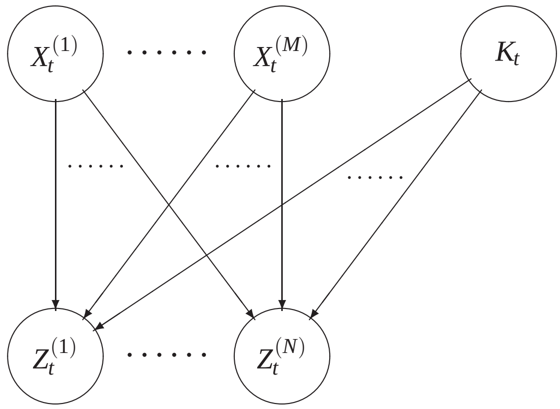


Figure 2. BBN for associating measurements to tracks

Consider the problem of associating N measurements at some time t to M tracks or other entities. Denoting the measurements as $z_t^{(n)}$, $n = 1, 2, \ldots, N$, and the track states as $x_t^{(m)}$, $m = 1, 2, \ldots, M$, and introducing an $N \times 1$ association vector, $k_t = (k_t^{(1)}, k_t^{(2)}, \ldots, k_t^{(N)})^\top$, $k_t^{(i)} \in \{1, 2, \ldots, M\}$, where the value of the i^{th} element indicates the track to which the i^{th} measurement is associated, the problem becomes one of estimating the value of k_t. This may be represented by the BBN in Figure 2. Both the measurements and track states may contain both attribute and kinematic information, and the corresponding probabilities represent the joint probabilities of all the attribute and kinematic elements.

The BBN of Figure 2 is not singly connected, that is, it contains loops in the underlying Markov network. Therefore, the clustering approach [6, 13] is used here to produce the tree structure of Figure 3, where the clustered variable $A_t = \{X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(M)}, K_t\}$. The conditional probabilities are now given as

$$\mathrm{p}\left(z_t^{(n)} \mid A_t\right) = \mathrm{p}\left(z_t^{(n)} \mid k_t, x_t^{(k_t^{(n)})}\right), \qquad n = 1, 2, \ldots, N. \tag{8}$$

From (1), the posterior probability, or belief, of A_t is given by

$$\mathrm{Bel}(A_t) = \alpha\lambda(A_t)\pi(A_t). \tag{9}$$

The predictive support is simply the prior probability of A_t, namely

$$\pi(A_t) = \mathrm{p}\,(A_t) = \mathrm{p}\,(k_t) \prod_{m=1}^{M} \mathrm{p}\left(x_t^{(m)}\right), \tag{10}$$

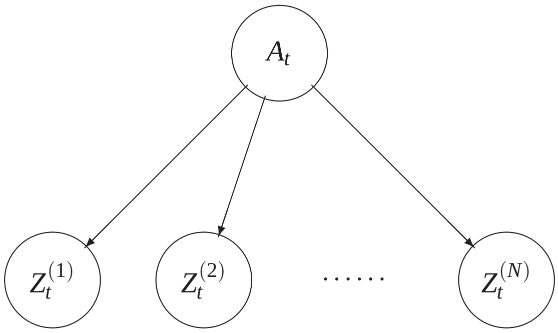


Figure 3. Cluster BBN for associating N measurements to M tracks

assuming that the track prior probabilites are independent, which is valid if it is assumed that a measurement can only be associated with a single track.

Denoting the known value of $z_t^{(n)}$ as $z_t^{(n)\prime}$,

$$\lambda\left(z_t^{(n)}\right) = \delta_{z_t^{(n)} z_t^{(n)\prime}}, \tag{11}$$

where $\delta_{z_t^{(n)} z_t^{(n)\prime}}$ equals 1 when $z_t^{(n)} = z_t^{(n)\prime}$ and 0 otherwise. Using (2) and (8), the retrospective support to A_t from the n^{th} measurement becomes

$$\lambda_{Z_t^{(n)}}(A_t) = \mathtt{p}\left(z_t^{(n)\prime} \mid k_t, x_t^{\left(k_t^{(n)}\right)}\right), \tag{12}$$

and, from (3), the support for A_t from all measurements is

$$\lambda(A_t) = \prod_{n=1}^{N} \mathtt{p}\left(z_t^{(n)\prime} \mid k_t, x_t^{\left(k_t^{(n)}\right)}\right). \tag{13}$$

Invoking (9) with (10) and (13) gives

$$\mathrm{Bel}(A_t) = \alpha \mathtt{p}(k_t) \prod_{m=1}^{M} \mathtt{p}\left(x_t^{(m)}\right) \prod_{n=1}^{N} \mathtt{p}\left(z_t^{(n)\prime} \mid k_t, x_t^{\left(k_t^{(n)}\right)}\right). \tag{14}$$

The belief of k_t may now be found by marginalising A_t, that is,

$$\mathrm{Bel}(k_t) = \alpha \mathtt{p}(k_t) \sum_{x_t^{(1)}} \sum_{x_t^{(2)}} \cdots \sum_{x_t^{(M)}} \prod_{m=1}^{M} \mathtt{p}\left(x_t^{(m)}\right) \prod_{n=1}^{N} \mathtt{p}\left(z_t^{(n)\prime} \mid k_t, x_t^{\left(k_t^{(n)}\right)}\right), \tag{15}$$

which may be rewritten

$$\mathrm{Bel}(k_t) = \alpha \mathtt{p}(k_t) \prod_{m:m\in k_t} \sum_{x_t^{(m)}} \mathtt{p}\left(x_t^{(m)}\right) \prod_{n:k_t^{(n)}=m} \mathtt{p}\left(z_t^{(n)\prime} \mid x_t^{(m)}\right), \tag{16}$$

where α is a normalising constant. Using the most probable value as the best estimate for the association vector, α may be ignored, as the solution is simply

$$\hat{k}_t = \arg\max_{k_t} \mathtt{p}(k_t) \prod_{m:m\in k_t} \sum_{x_t^{(m)}} \mathtt{p}\left(x_t^{(m)}\right) \prod_{n:k_t^{(n)}=m} \mathtt{p}\left(z_t^{(n)\prime} \mid x_t^{(m)}\right). \tag{17}$$

Of particular interest is the prior distribution of k_t. If all permutations of the association vector are equiprobable, then the solution under linear and Gaussian assumptions is equivalent to the nearest neighbours association algorithm for both kinematic and attribute data. However, zeroing the prior probability of all the permutations of the association vector that associate more than one measurement from any sensor to one same target constrains the solution and the result is a global nearest neighbours algorithm for both kinematics and attributes.

5 RESULTS

The association algorithm of Section 4, with the constraint that each target can only produce one measurement from each sensor, is applied to the tracking and identification problem for crossing targets. Probabilistic gating is used to eliminate low probability associations. Measurements are simulated from two sensors, namely a secondary surveillance radar (SSR) returning range, azimuth and identity, and an Electronic Surveillance (ES) sensor returning azimuth and platform class, which is related to identity. Kinematic and identity updates are performed by the Extended Kalman Filter (EKF) and the identity update algorithm of [10], respectively. Two cases are considered, one with association based only on kinematic data, and the other using both kinematic and attribute data for association.

Figures 4 and 5 show the track locations for a single run for kinematic only association and joint attribute and kinematic association, respectively. The track positions are indicated by solid lines, and the ground truth by dotted lines. The SSR position measurements from each target are shown, where the '$+$' represents those from one target, and the '$\circ$' those from the other target.

Frequent track swaps, as illustrated by Figure 4, are observed when only kinematic data is used for data association. The use of both kinematic and attribute data overcomes this problem. This is reflected in the assocation accuracy results presented in Table 1, where the percentage of measurements from each target that are associated with each track is presented for both the SSR and ES sensors. Perfect association would result in all measurements from target 1 associated with track 1, and all measurements from target 2 associated with track 2. The results for both sensors reflect the difficulty of associating reports to closely spaced entities using only kinematic data. The additional discrimination introduced by the attribute data provides the observed improvement in performance of the joint association over the kinematic only association.

This example demonstrates the advantage of using both attribute and kinematic data for data association. This is equally applicable to other data fusion applications, where other attribute information may be involved.

Table 1. Association accuracies for kinematic and joint kinematic and attribute association

Sensor	Target	Track	Assoc. Accuracies (%)	
			Kinematic	Joint
SSR	1	1	86	99
		2	13	0·05
	2	1	13	0·05
		2	85	98
ES	1	1	58	71
		2	41	27
	2	1	42	28
		2	57	71

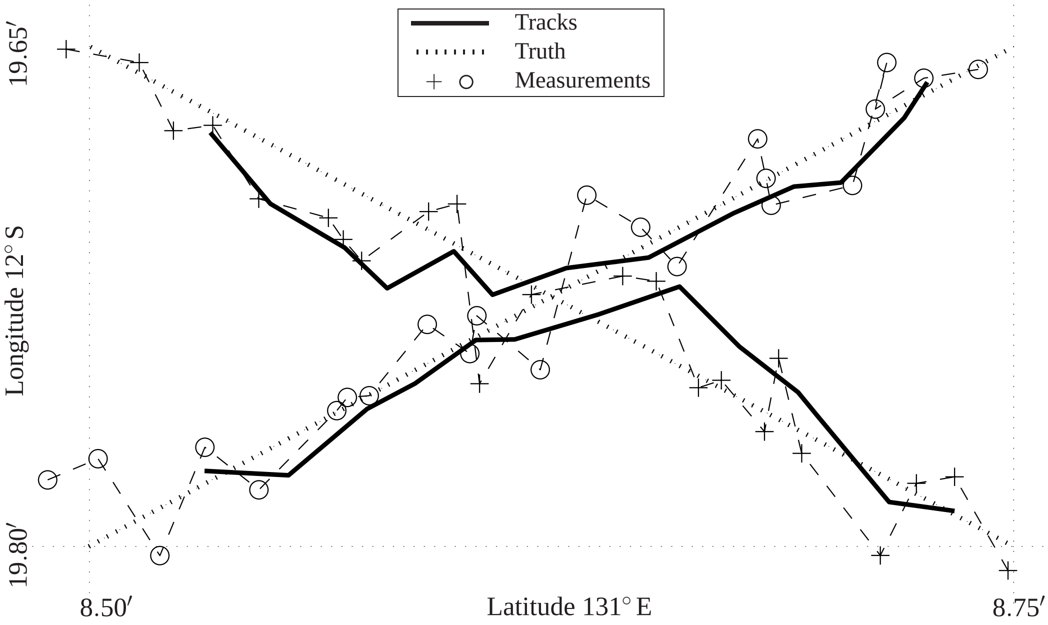


Figure 4. Track location using kinematic only association

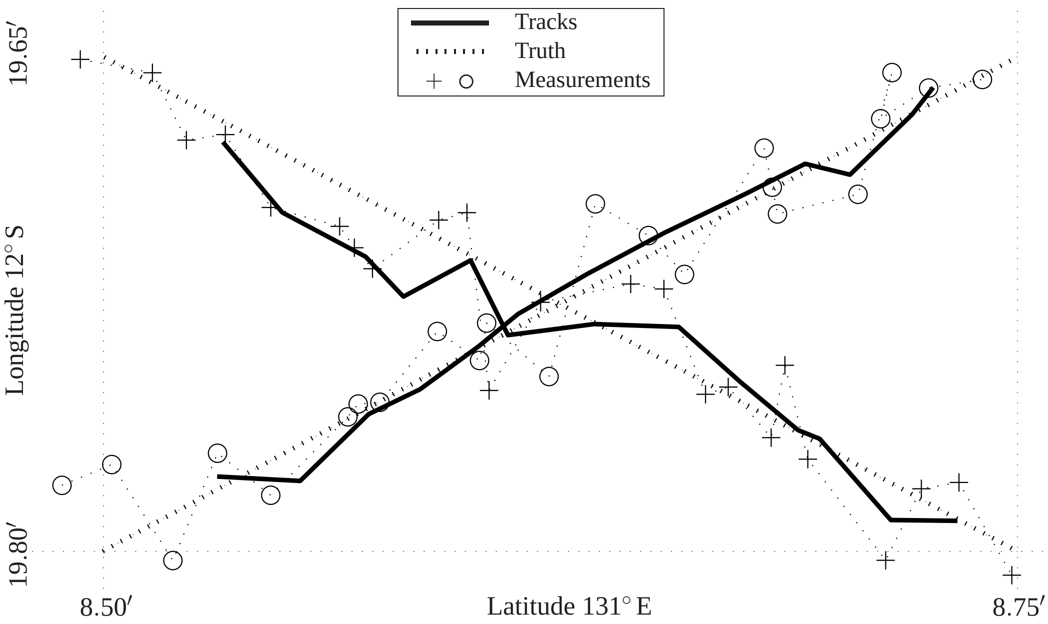


Figure 5. Track location using joint kinematic and attribute association

6 CONCLUSIONS

Accurate data association is a necessary requisite for the fusion of multi-source multi-target information. A joint kinematic and attribute data association algorithm that incorporates data uncertainty has been developed using a BBN framework. Significant improvement in association accuracy over kinematic only association has been demonstrated through simulation.

REFERENCES

[1] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood MA: Artech House, 1999.

[2] P. Bogler, "Shafer-Dempster reasoning with applications to multisensor target identification systems," *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-17, no. 6, pp. 968–77, 1987.

[3] M.-A. Simard, J. Couture, and É. Bossé, "Data fusion of multiple sensors attribute information for target identity estimation using a Dempster-Shafer evidential combination algorithm," in *Proc. SPIE*, vol. 2759, pp. 577–88, 1996.

[4] S. Challa and G. W. Pulford, "Joint target tracking and classification," in *Proc. 5th Int. Conf. on Radar Systems (Radar '99)*, (Brest, France), 1999.

[5] N. Gordon, S. Maskell, and T. Kirubarajan, "Efficient particle filters for joint tracking and classification," in *Signal and Data Processing of Small Targets 2002*, vol. 4728 of *Proceedings of the SPIE*, pp. 439–49, 2002.

[6] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco CA: Morgan Kaufmann, 1988.

[7] P. Korpisaari and J. Saarinen, "Bayesian networks for target identification and attribute fusion with JPDA," in *Proc. 2nd Int. Conf. Information Fusion*, vol. II, pp. 763–9, 1999.

[8] S. Hautaniemi, P. Korpisaari, and J. Saarinen, "Target identification with Bayesian networks," in *Proc. SPIE*, vol. 4051, pp. 55–66, 2000.

[9] K.-C. Chang and R. Fung, "Target identification with Bayesian networks in a multiple hypothesis tracking system," *Optical Engineering*, vol. 36, no. 3, pp. 684–91, 1997.

[10] M. L. Krieg, "A Bayesian belief network approach to multi-sensor kinimatic and attribute tracking," in *Proc. Information, Decision and Control (IDC2002)*, (Adelaide, Australia), Feb. 2002.

[11] M. L. Krieg, "Joint multi-sensor kinematic and attribute tracking using bayesian belief networks," in *Proc. Int. Conf. Information Fusion (Fusion 2003)*, (Cairns Australia), July 2003.

[12] F. Jensen, *An Introduction to Bayesian Networks*. London UK: UCL Press, 1996.

[13] M. L. Krieg, "A tutorial on Bayesian belief networks," Tech. Rep. DSTO–TN–0403, DSTO Australia, 2001.

Set classification of military targets

Mr. Reinert Korsnes / Mr. Bjørn Jervell Hansen
Norwegian Defense Research Establishment (FFI)
Division for electronics
Postboks 25, NO-2027 Kjeller
NORWAY
Email:{reinert.korsnes | bjorn-jervell.hansen}@ffi.no

August 29, 2003

Abstract

The present study shows by example the potential amount of information available in a set of observations of targets where there are known relations between these targets. Known relations between objects significantly reduces the set of possible explanations behind a set of observations. The application here is classification of military targets. Cost functions and interaction with decision makers extend the feasibility of the present approach meaningfully to treat many observations and possible targets behind these.

1 Multi-object state-estimation

Classification and tracking of military targets are today often made individually for each object. An approach also utilizing observations of other objects can provide better results. Mahler [1997], Kreinovich [1997] and many others elaborate, within the framework of random sets, on classification of many objects based on a variety of data sources. Theory of random sets provides a direct approach to treat a variety of observations of many targets. This work also takes a direct approach to classify a set of objects.

The focus of this work is by example quantitatively to illustrate the potential amount of information available in the whole set of observations of actual targets. This approach utilizes knowledge about relations between possible targets and their time varying parameters using computer based tools. The present study in general demonstrates the idea of searching through the set of possible explanations (state space) for a given set of observations for the purpose of object tracking and classification.

This work consists of two main parts. The first part illustrates the potential of a full search through a small state space (sections 2, 3 and 4). The second part discusses search in state spaces which are too large to search throughout (section 5).

2 Limited example of information extraction from a set of observations

Appendix A gives an example where a restricted and known set $\mathbf{T} = \{a, b, c, d\}$ of military targets are confined within an area. An observation (or report) $O = a$ means that an observer believes he has observed a (instead of b, c or d) at a given location and time. Let X denote the true origin of the observation O. If $X = a$ then the observation $O = a$ is correct. Otherwise it is false.

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

Assume, for simplicity, that each observation has a given uncertainty as follows: A single observation is, in our example, correct with 58 percent probability and the probability to give the report b, c or d is each 14 percent (in total 100 percent). So with standard probabilistic notation, $P(O = a|X = a) = 0.58$ and $P(O = b|X = a) = 0.14$ etc. We assume symmetry with respect to the objects, so $P(O = b|X = b) = 0.58$ and $P(O = a|X = b) = 0.14$ etc. Equation 1 summarizes this information where $x, y \in \mathbf{T}$ and $1 \leq i \leq 4$.

$$P(O_i = x|X_i = y) = \begin{cases} 0.58 & \text{if } x = y \\ 0.14 & \text{otherwise} \end{cases} \tag{1}$$

Given the observations O_1, O_2, O_3 and O_4 with real origin X_1, X_2, X_3 and X_4 respectively. Assume that each set of possible values for X_i (i.e. each *state*) has the same initial probability.

Assume we want to estimate/guess what is the real origin X_1 behind the observation $O_1 = a$. The Bayes rule provides the probability that the real target X_i is a:

$$P(X_i = a|O_i = a) = \frac{P(O_i = a|X_i = a)P(X_i = a)}{P(O_i = a)} \tag{2}$$

where $P(O_i = a) = \sum_{s \in \mathbf{T}} P(O_i = a|X_i = s)P(X_i = s) = 0.58 \cdot 0.25 + 3 \cdot 0.14 \cdot 0.25 = 0.25$. This gives $P(X_i = a|O_i = a) = 0.58$. The probability for $X_i = b$ similarly is:

$$P(X_i = b|O = a) = \frac{P(O_i = a|X_i = b)P(X_i = b)}{P(O_i = a)} \tag{3}$$

which has numerical value 0.14. Each estimate above only depends on one single observation.

If the observations are related, the observations O_2, O_3 and O_4 contain information about X_1. I.e. $P(X_1|O_1 = a, O_2, O_3, O_4)$ is in general different from $P(X_1|O_1 = a)$. We may, for example, be sure that all observations have different objects as origin. This can be the case if the observations for example are simultaneous or the objects are geographically far from each other. The real targets (objects) are in this case all different (provided reliable position estimates for the observations). Spatial location is just one special example of a temporal parameter for an object, so this situation is much more general than for simultaneous observations of geographically distributed targets or where the targets cannot move quickly enough to be observed several times. It can often be easy to find out that two objects, X_1 and X_2, are different even if it is hard exactly to classify the objects. For example, one may easily see that there are two different people on two different unclear photos. However, it is in general much harder to identify the people more exactly. One may only observe that one person is higher than the other or have different clothes the same afternoon. One may have a cellular phone while the other has not etc.

One may in general assume a known (general) relation $\prec$ on the set of possible targets and that this is observable. See the illustration of Figure 1. The actual relations may be static (given for a set of objects) or it can be situation dependent. The relations must in all cases, to be utilized, be part of (or given by) the set of observations.

Assume, for example, that for any possible target X_i and X_j we know if $X_i \prec X_j$ is true or not. $\prec$ may simply mean 'is larger/higher/longer than', 'is different from', 'in front of', 'earlier than', 'faster than' or it can be based on more complex geometric aspect ratios, properties of radar return, emittance etc. $X_i \prec X_j$ may also mean for example that 'X_i wears glasses while X_j does not' from which one can derive that $X_i \neq X_j$.

Assume one is able to observe if the relation $X_i \prec X_j$ is true or not for any target X_i and X_j even if one do not know the exact value of X_i and X_j. We show below that this can include significant information for classification of targets.

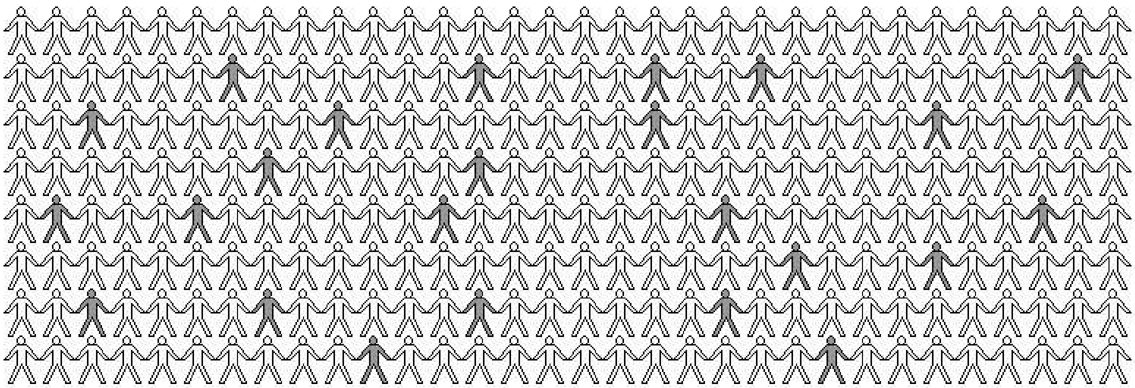


Figure 1: Known relations between objects significantly restricts the number of candidate explanations (sets of classifications) behind a set of observations. Here each person illustrates an explanation for a given set of observations. If there are no restrictions, then four possible objects and four observations gives $4^4 = 256$ possible explanations. If any two observations come from two different targets (for example spatially distributed simultaneous observations), then there are only $4! = 24$ possible candidate explanations. This gives a ratio of 0.094. This ratio decreases fast with increased number of possible targets/observations.

If we assume all objects are different, and denote this 'relation' by $\mathcal{R}$, we get by simple numerical simulations (see appendix B) that

$$P(X_1 = a|O_1 = a, O_2 = b, O_3 = c, O_4 = d, \mathcal{R}) = 0.80 \tag{4}$$

which is considerably different from $P(X_1 = a|O_1 = a) = 0.58$. If we have ground truth data for an object, for example $X_2 = b$, we get even more certainty:

$$P(X_1 = a|O_1 = a, O_2 = b, O_3 = c, O_4 = d, X_2 = b, \mathcal{R}) = 0.88 \tag{5}$$

And if we also have the ground truth data $X_3 = c$ we get:

$$P(X_1 = a|O_1 = a, O_2 = b, O_3 = c, O_4 = d, X_2 = b, X_3 = c, \mathcal{R}) = 0.94 \tag{6}$$

If the set of observations are not consistent, this will increase the uncertainty. Assume we know that all observed targets are different, but two observations give the same independent classification such as for example $O_3 = O_4 = c$. We then get that

$$P(X_1 = a|O_1 = a, O_2 = b, O_3 = c, O_4 = c, \mathcal{R}) = 0.67 \tag{7}$$

The intuitive explanation here is that the inconsistency (i.e. that $O_3 = O_4$) tells that at least one of the observations O_3 and O_4 is wrong and therefore a may be hidden behind one of these observations.

If a is given by the observations to be two places at the same time, this further increases the confusion:

$$P(X_1 = a|O_1 = a, O_2 = b, O_3 = c, O_4 = a, \mathcal{R}) = 0.50 \tag{8}$$

And if a is given to be 3 places at the same time, the uncertainty even increase:

$$P(X_1 = a|O_1 = a, O_2 = b, O_3 = a, O_4 = a, \mathcal{R}) = 0.33 \tag{9}$$

If a is initially classified to be at 4 places at the same time, then $P(X_1 = a|O_1 = a, O_2 = a, O_3 = a, O_4 = a, \mathcal{R}) = 0.25$. These estimates conform to common sense since inconsistent observations have to include errors.

These estimated probabilities can be used in a cost assessment as shown in appendix A.2. The cost assessment might be of help to a decision maker making a decision of how to react to these observations.

3 Generalizations of limited example

3.1 Increased and reduced number of observations

The limited example of Section 2 assumes the number of observations to be the same as the number of targets. This section shows a way to generalize this restriction allowing the number of targets to be different from the number of observations. The purpose of this generalization is to show that the amount of available information - the number of observations - is important with respect to probability estimates like the ones in equations 4 - 9.

Assume $\mathcal{R}_1$ represents the relation that the objects generating the observations $O_1, \ldots, O_4$ are different (as in Section 2), while the objects behind O_1 and O_5 are the same (i.e. $X_1 = X_5$). A similar simple numerical simulation as in Section 2, provides the following estimate:

$$P(X_1 = a|O_1 = a, O_2 = b, O_3 = c, O_4 = d, O_5 = a, \mathcal{R}_1) = 0.95 \tag{10}$$

This shows that the additional observation O_5, and the known relationship that $X_1 = X_5$, gives an improved confidence in the classification of X_1 (cf Equation 4).

Reduction of the set of observations also has effect on the quality of classification in the present approach. Let $\mathcal{R}_2$ denote that the objects generating the observations O_1, O_2, O_3 are different. We then get that

$$P(X_1 = a|O_1 = a, O_2 = b, O_3 = c, \mathcal{R}_2) = 0.69 \tag{11}$$

which is a considerable reduction of confidence in the classification of X_1 (cf Equation 4).

3.2 Classifying a set of not precisely known targets within a restricted area

This section illustrates a generalization of the classification problem of Section 2 which assumed a small number of known targets within a restricted area. Assume now that the actual four targets are members of a set $\Omega = \{a_1, a_2, \ldots, a_{11}\}$ of 11 classifications and that the quality of observations are the same as in the above examples. Equation 12 defines (similar to Equation 1) the quality of the observations:

$$P(O_i = x|X_i = y) = \begin{cases} 0.30 & \text{if } x = y \\ 0.07 & \text{otherwise} \end{cases} \tag{12}$$

where both x and y are an element of Ω and $1 \leq i \leq 4$. This conforms to the quality of observations assumed in Section 2 (Equation 1).

If there are no known relations between the targets, then the probability of $X_1 = a_1$ given the observation $O_1 = a_1$ is $P(X_1 = a_1|O_1 = a_1) = 0.30$ (cf Equation 2).

Assume, as above, four observations $O_i = a_i$, $i = 1, 2, 3, 4$ and that all objects are different ($\mathcal{R}$). We get in this case that (still via simple simulations):

$$P(X_1 = a_1|O_i = a_i, \mathcal{R}) = 0.32 \qquad i = 1, 2, 3, 4 \tag{13}$$

where $\mathcal{R}$ as above denotes that all objects $X_1,\ldots,X_4$ are different. As opposed to in the classification problem of section 2 (see equation 4), the information ($\mathcal{R}$) that the targets are different do not significantly decrease the uncertainty of classification (note from above that $P(X_1 = a_1|O_1 = a_1) = 0.30$).

Assume now, as an illustration, that the relative sizes of the 4 targets constitutes a strong linear relation $\prec$ on the set Ω of 4 targets. It might e.g. be known that object 1 is smaller than object 2, which is smaller than object 3 etc. (i.e. as if the elements of Ω were logically organized along a line).

This means, mathematically, that for any two different objects X, Y in Ω, either $X \prec Y$ or $Y \prec X$. It also means that $X \prec Y$ and $Y \prec Z$ gives that $X \prec Z$ (and it is never true that $X \prec X$). If we know, for example, that X_1 is 'less than' another observed object X_2 (i.e. $X_1 \prec X_2$), then we get:

$$P(X_1 = a_1|X_1 \prec X_2) = 0.56 \tag{14}$$

Similarly we have that

$$P(X_1 = a_1|X_1 \prec X_2 \prec X_3) = 0.76 \tag{15}$$

and if all the four observed objects are strongly linearly related:

$$P(X_1 = a_1|X_1 \prec X_2 \prec X_3 \prec X_4) = 0.90 \tag{16}$$

4 Initial probabilities

The above statistical approach assumes a homogeneous distribution of the initial probabilities over the set of possible states. This is a natural choice when there is no information giving preference to special alternatives. The use of the relation $\mathcal{R}$ above can be looked at as defining an initial probability where the states not conforming to the relation has probability zero (and the other states each having the same initial probability). A certain relation, in other words, restricts the search space or concentrates the probability to the subspace conforming to the relation. Other types of initial distributions can have similar effects on the classifications similar to those above.

5 Treatment of many objects

5.1 Restricting the search

The examples above assumes, for a large number of objects, heavy computing and the method is therefore not directly practical for large problems. A main challenge is therefore meaningfully to search through only a smaller part of the state space or to search through it in a cost effective manner and eventually to stop processing when sufficient results are obtained. An algorithm can use knowledge about physical restrictions and information from intelligence in order to reduce the search space. Situation and threat assessments can also provide information for restricting the search space for finding sufficient information for actual decisions. This approach for efficient search requires flexible user interactions.

5.2 Ordered and restricted search

Assume a set $\mathbf{o} = \{O_1, O_2, \ldots, O_M\}$ of M independent observations of singular members of a large set $\{a_1, a_2 \ldots, a_N\}$ of N targets within a restricted area (say $N \simeq 20$). X_i denotes the true value behind the observation O_i and an observer with incomplete information about it, may treat X_i as a random variable.

Assume that each state $\mathbf{x} = (X_1, \ldots, X_M)$, has an associated "interest to know". I.e. there is a payoff (or saved cost) associated with information about the values of X_i. One may therefore want to estimate the probability distribution for x or at least obtain partly knowledge about its probability distribution.

An efficient search after estimates of the probability of "interesting" states, can utilize an initial probability distribution on the state space. The search can in this way first look through parts of the search space with the highest expected payoff. Let $\mathbf{A}$ and $\mathbf{B}$ be two different parts of the search space. Given that $I(\mathbf{x} \in \mathbf{A})$ and $I(\mathbf{x} \in \mathbf{B})$ are the payoffs to know if $\mathbf{x} \in \mathbf{A}$ and $\mathbf{x} \in \mathbf{B}$, respectively, and that $\mathbf{x} \in \mathbf{A}$ and $\mathbf{x} \in \mathbf{B}$ similarly has initial (unconditional) probabilities $P(\mathbf{x} \in \mathbf{A})$ and $P(\mathbf{x} \in \mathbf{B})$. If a decision component must make a priority between calculating $P(\mathbf{x} \in \mathbf{A}|\mathbf{o})$ or $P(\mathbf{x} \in \mathbf{B}|\mathbf{o})$ (for example within a given time), then estimating $P(\mathbf{x} \in \mathbf{A}|\mathbf{o})$ would be a natural choice if the expected payoff $C(\mathbf{A}) = I(\mathbf{x} \in \mathbf{A}) \cdot P(\mathbf{x} \in \mathbf{A})$ is significantly larger than $C(\mathbf{B}) = I(\mathbf{x} \in \mathbf{B}) \cdot P(\mathbf{x} \in \mathbf{B})$. If the payoff from estimates for $\mathbf{x} \in \mathbf{B}$ is expected to be much larger than the cost of waiting for the estimates, then a natural choice would be to wait for estimates for $\mathbf{x} \in \mathbf{B}$. If $I(\mathbf{x} \in A) = 0$ or $P(\mathbf{x} \in \mathbf{A}) = 0$, then the part A may not be interesting to search through.

References

V. Kreinovich. Random sets unify, explain, and aid known uncertainty methods in expert systems. In J. Goutsias, R.P.S. Mahler, and T. Nguyen, editors, *Random Sets - Theory and Application*, pages 321–345. Springer-Verlag, New York, 1997.

R.P.S. Mahler. Random sets in information fusion - and overview. In J. Goutsias, R.P.S. Mahler, and T. Nguyen, editors, *Random Sets - Theory and Application*, pages 129–164. Springer-Verlag, New York, 1997.

Appendix

A Scenario

A.1 Intelligence operation within a restricted area

The following scenario illustrates the use of the present approach to help establishing a situation picture. An unknown (potentially hostile) group of four boats (a, b, c, d) is observed entering a Norwegian fjord. Intelligence and general situation assessment results in that a team is set up with the mission to monitor the boats. The monitoring is to be conducted in a discrete/hidden manner to prevent revealing the level of monitoring, hence helicopters, radars etc. cannot be used. Revealing, for an enemy, the level of monitoring and response patterns is regarded to increase to vulnerability in the given situation and it will also increase the need for costly change in security related investments.

There are several hypotheses for why the boats have entered the fjord:

H1: They are regular tourist boats.

H2: At least one boat (boat a) is testing our intelligence, level of alert and response patterns.

H3: The boat a is for sabotaging a power plant in the fjord.

The intelligence team is assumed to report if the power plant is at risk (early enough to protect this possible target).

The outlet of the fjord is well observed, meaning that it is known how many boats have entered the fjord and that any boats coming into the fjord will be spotted. The boats can be distinguished from each other via different characteristics (different sizes, different types of boats, etc.). Further, one can easily find out if two observations of boats are for different boats or not. Each observation is given a confidence as in Equation 1. The mission for the team is (in a hidden way) to monitor all four boats as well as possible while evaluating the risk for sabotage.

A.2 Cost Assessment

The power plant is assumed to have a (relative) value of $I_{\text{power plant}} = 100$. The cost of revealing the intelligence/monitoring level is assumed to be less than the limit $S = 5$. Only one of the boats (boat a) is assumed to be large enough to be equipped for attacking the power plant. If the boat a enters a "red zone" close to the power plant, it is given a probability p_{attack} to attack the power plant. Assume, according to Section 2 above, the observations $\mathbf{o} = \{O_1 = a, O_2 = b, O_3 = c, O_4 = d\}$. So if a boat X_1 is observed within the "red zone", and the classification only includes use of observation O_1, the expected cost of the choice of not taking action would be:

$$C_{\text{no action}} = p_{\text{attack}} \cdot P(X_1 = a|O_1 = a) \cdot I_{\text{power plant}} \tag{17}$$

For $p_{\text{attack}} = 0.10$, $P(X_1 = a|O_1 = a) = 0.58$, and $I_{\text{power plant}} = 100$ this gives $C = 5.8$ which is close to the cost of revealing the intelligence. Hence the individual classification estimate $P(X_1 = a|O_1 = a) = 0.58$ does not provide a clear conclusion (decision). The similar expected cost, taking into account the whole set of observations, would be

$$\begin{aligned} C'_{\text{no action}} = p_{\text{attack}} \cdot \\ P(X_1 = a|\mathbf{o}, \mathcal{R}) \cdot I_{\text{power plant}} \end{aligned} \tag{18}$$

giving, according to Equation 4, $C = 8$. If there were ground truth data for X_2 and X_3 (cf Equation 6), then we would get $C = 9.4$.

Assume (still) that boat a is known to be hostile. If there is no significant cost associated with checking up a regular tourist boat, then the expected cost of taking action similarly would be less than

$$C_{\text{action}} = S \cdot P(X_1 = a|O_1 = a) = 2.9 \tag{19}$$

and

$$C'_{\text{action}} = S \cdot P(X_1 = a|\mathbf{o}, \mathcal{R}) = 4.0. \tag{20}$$

If there were a high cost associated with the possibility that X_1 was a regular tourist boat, and b, c and d were confirmed to be tourist boats, then there would be an additional cost proportional to $P(X_1 \neq a|O_1 = a) = 0.42$, $P(X_1 \neq a|o, \mathcal{R}) = 0.20$ or alternatively $P(X_1 \neq a|o, X_2 = b, X_3 = c, \mathcal{R}) = 0.06$. These alternative estimates for the probability for $X_1 = a$ could in this case give large differences in cost estimates.

B Numerical simulation algorithm

This section describes the simulations providing the estimates of classification probabilities used in this paper. The following points outline the algorithm:

1. A set of all possible realities explaining the original observation vector (explanations) is generated. The size of this set depends on the size of the original observation vector (the number

of observations), the number of possible classifications of the objects generating the original observation vector, and on the possible restrictions. E.g.: with an original observation vector of 5 single observations, 4 possible classifications, and no restrictions, the set of explanations is a matrix of size $5^4 \times 5$. With an original observation vector of size 4, 4 possible classifications and the restriction that the 4 items in the observation vector originate from different objects, the size of the matrix will be $4! \times 4$.

2. For each explanation, 1000 observation vectors of the same size as the original observation vector are generated. Each of these observation vectors are generated randomly based on the sensor likelihood, i.e.: for each entry in the explanation, a random number is drawn and compared to the likelihood. The outcome of this comparison decides what observation is generated by this particular explanation.

3. The fraction of these 1000 observation vectors equal to the original observation vector, represents how well the possible explanation explains the original observation.

This part of the algorithm is repeated N times, and the mean fraction for each possible explanation is recorded as the final result.

Consider, as an example, an original observation vector $\mathbf{oov} = \begin{bmatrix} a & b & c \end{bmatrix}$, where the possible classifications are a, b and c. With the restriction that the 3 items in **oov** originate from different objects, the matrix of explanations will look like this:

$$\begin{bmatrix} a & b & c \\ a & c & b \\ b & a & c \\ b & c & a \\ c & a & b \\ c & b & a \end{bmatrix}$$

The algorithm then could produce the following result:

possible explanation	result	fraction
$[\,a\ b\ c\,]$	750	0.75
$[\,a\ c\ b\,]$	130	0.13
$[\,b\ a\ c\,]$	40	0.04
$[\,b\ c\ a\,]$	20	0.02
$[\,c\ a\ b\,]$	20	0.02
$[\,c\ b\ a\,]$	40	0.04

This result gives the estimate

$$\begin{aligned} & P(X_1 = a | O_1 = a, O_2 = b, O_3 = c, \mathcal{R}_3) = \\ & \quad P(X_1 = a, X_2 = b, X_3 = c | O_1 = a, O_2 = b, O_3 = c, \mathcal{R}_3) + \\ & \quad P(X_1 = a, X_2 = c, X_3 = b | O_1 = a, O_2 = b, O_3 = c, \mathcal{R}_3) = \\ & \quad 0.75 + 0.13 = 0.88 \end{aligned}$$

where $\mathcal{R}_3$ represents the restriction that X_1, X_2 and X_3 are different.

Ontological Approach to Military Knowledge Modeling and Management

Anne-Claire Boury-Brisset
Defence R&D Canada – Valcartier
2459 Pie-XI North
Val-Belair, QC G3J 1X5
CANADA
Phone: 1 (418) 844 4000, Fax: 1 (418) 844 4538
Email: Anne-Claire.Boury-Brisset@drdc-rddc.gc.ca

ABSTRACT

Ontologies have received increasing interest in the computer science community and their benefits are now recognized for military applications. In addition to explicitly encode a shared understanding of some domain, ontologies can be used in a wide range of applications including natural language processing, intelligent search engines, information retrieval, or as a means to facilitate interoperability between heterogeneous knowledge sources at a high level of abstraction. Our objective is to contribute to the building of ontological models in the information fusion domain and to exploit these models to provide enhanced knowledge management assistance to military commanders. In the paper, we first review the roles of ontologies for military applications. Then, we describe the characteristics of high-level information fusion processes and address methodological aspects related to ontological engineering. This leads us to derive an ontological approach to situation and threat assessment domain modeling. Finally, we present ontology-based knowledge management services that could be exploited in support of high-level information fusion processes.

1.0 INTRODUCTION

In military environments, commanders and their staff are overwhelmed by the increasing amount of information available to them. Large volumes of information and data from heterogeneous information sources have to be integrated and interpreted in order to gain situational awareness. Both structured and unstructured data sources have to be managed. In the first case, different data models and formats have to be considered. In the latter, content extraction from unstructured data (e.g. military messages containing free text) using natural language processing techniques are required. Furthermore, the techniques being developed for data fusion and resource management in Decision Support Systems for Command and Control are becoming increasingly more sophisticated, particularly through the incorporation of methods for high-level reasoning processes.

Providing commanders with intelligent information fusion systems or advanced decision aids requires a good understanding of the processes involved, their information requirements, and the development of formal domain models upon which reasoning processes can be based (e.g. knowledge-based systems, intelligent agents). In this context, we aim at contributing to the building of ontological models that could be used to automatically support information fusion processes. Furthermore, these models would enhance knowledge

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

management services for information fusion and would provide commanders with a gain in situational awareness.

Ontologies have received increasing interest in the computer science community and their benefits are now recognized for military applications as they provide a foundation for the representation of domain knowledge. They explicitly encode a shared understanding of a domain that can be communicated between people and application programs. Gruber [6] defines an ontology as « *an explicit specification of a shared conceptualization* ». The specification of an ontology comprises a vocabulary of terms, each with a definition specifying its meaning. Ontologies range from controlled vocabularies to highly expressive domain models [12]: integrated data dictionaries designed for human understanding, taxonomies organizing concepts of a domain into inheritance hierarchies, structured data models suitable for data management, and finally highly expressive computational ontologies. In addition to explicitly encode a shared understanding of some domain, ontologies can be exploited in a wide range of applications including natural language processing, intelligent search engines, information retrieval, or as a means to facilitate interoperability between heterogeneous knowledge sources at a high level of abstraction.

Ontologies can play different roles for knowledge modeling and management in military applications. First, ontological models constitute a language that facilitates communication between a community of agents: communication between humans, interaction between humans and computers, and communication between computer applications, software agents or multiagent systems. From a knowledge modeling perspective, ontologies provide domain models representing a shared understanding of some domain that can be reused across a wide range of applications in that domain. The development of larger and more complex intelligent systems in knowledge-intensive domains make impractical to develop knowledge bases from scratch. Thus, knowledge engineering promotes the design of libraries of reusable components that include both ontologies and problem-solving methods. Consequently, the design of ontologies is considered as a pre-requisite to the development of large knowledge bases (e.g. HPKB DARPA program). These models provide a support for the development of data fusion systems or decision support systems for command and control by system engineers and knowledge engineers. Predefined ontologies can then be exploited to provide more advanced knowledge management services.

Most significant modeling efforts in the military domain have been devoted to planning activities. Common representation of plans has been a subject of interest for a long time. Among the work reported in the literature, some aim at providing plan models (e.g. O-PLAN, SPAR, PLANET), whereas other focus on the representation of courses of action (COA), i.e. outlines of plans, for example, the Disciple-COA project from the DARPA HPKB and RKF programs. Bowman's work [4] extends the COA ontology to represent the military concept of Center of Gravity used at the strategic level. In the context of military data models, the NATO LC2IEDM (Land C2 Information Exchange Data Model) is also of great value because it stores the core data needed to describe information to be exchanged between C2 systems with respect to the battlefield domain. Few researches have been devoted to analyze high-level data fusion processes from an ontological perspective. W. Johnson and his colleague [7] present an ontological analysis for situation and threat assessment and describe the different types of relations between objects of the domain. M. Kokar describes a formalization of situation awareness of in [2]. As ontologies aim at building reusable components, these models should be analyzed in order to contribute to the representation of high-level information fusion domain knowledge.

Our objective is to contribute to the building of ontological models in the information fusion domain and to exploit these models to provide enhanced knowledge management assistance to military commanders. The remainder of the paper is organized as follows. First, we describe the characteristics of high-level information fusion processes. Then, we address methodological aspects related to ontological engineering. This leads us to derive an ontological approach to situation and threat assessment domain modeling. Finally, we present ontology-based knowledge management services that could be exploited in support of high-level information fusion processes.

2.0 CHARACTERISTICS OF HIGH-LEVEL FUSION PROCESSES

Data fusion is the process of combining data from multiple sources to refine state estimates and predictions [14]. In a perspective to provide a unifying terminology and a shared understanding of the domain within the Data Fusion community, efforts conducted by the Joint Directors of Laboratories (JDL) Data Fusion Working Group have resulted in the refinement of a process model for data fusion [14]. The JDL model is the most widely accepted model of the data fusion process. It is divided into four levels (the sub-object data assessment being excluded). The definitions are as follows.

Level 1 – Object assessment: Estimation and prediction of entity states on the basis of observation-to-track association, continuous state estimation and discrete state estimation. The product of this level consists of tracks, i.e. hypothesized entities with their type and identity.

Level 2 – Situation assessment: Estimation and prediction of relations among entities, to include force structure and cross force relations, communications and perceptual influences, physical context, etc. This involves associating tracks into aggregations. The state of the aggregate is represented as a network of diverse relations among its elements.

Level 3 – Impact assessment (or Threat assessment): Estimation and prediction of effects on situation of planned or estimated actions by the participants, to include interactions between action plans of multiple players. The product is the impact estimate of the assessed situation, i.e. the outcome of various plans as they interact with one another and with the environment.

Level 4 – Process refinement: Metalevel process consisting of adaptive data acquisition and processing to support mission objectives.

In the same perspective, J. Roy [13] attempts to synthesize the main notions put forward by data fusion and situation awareness models into a situation analysis model. He defines Situation Analysis (SA) as *a process, the examination of a situation, its elements, and their relations, to provide and maintain a product, i.e. a state of Situation Awareness (SAW), for the decision maker*. The proposed SA model captures the representation of various elements of the situation as well as how they relate to create a meaningful synthesis, i.e., a comprehension of the situation. It includes: situation element acquisition, data alignment and association, situation element perception refinement, situation element contextual analysis, situation element interpretation, situation classification and recognition, situation assessment, situation element projection, impact assessment, situation watch, and process refinement. By describing the different functions involved in the SA process, the model provides a basis for the building of domain models and problem-solving methods related to the processes.

New efforts within the information fusion community aim at formalizing the concepts related to high-level information fusion with the objective to improve human understanding across the data fusion community (e.g. defence researchers and system developers), and ultimately facilitate communication between distributed fusion systems. In this context, an ontology for high-level data fusion captures the main concepts and relationships between concepts at these levels. The ontology should specify both physical and non-physical entities involved in level 2 (situation assessment) and level 3 (threat assessment) information fusion processes.

High-level fusion processes are complex processes involving many elements and interactions among a wide variety of battlefield components. In particular, a large number of data items are included and used in the processes. High-level data fusion processes have the following properties:

- They emphasize on symbolic reasoning rather than numeric reasoning;
- Hierarchical reasoning is required due to vertical organization of military entities and the multiple levels of abstraction nature of the reasoning process. Objects involved in the data fusion process are organized in multiple levels of abstraction where objects at one level are linked to objects at an adjacent level (e.g. entities aggregation).
- Reasoning in context is performed: data are analyzed with respect to the evolving situation including environmental considerations.

Concepts involved in high-level data fusion processes have specific characteristics:

- They manipulate both concrete and abstract entities;
- Multiple types of dynamic and static domain knowledge have to be processed;
- There exist numerous constituency-dependency relationships among objects as well as events and activities of interest;
- Reasoning in context exploits domain feature databases to facilitate hypothesis management. Thus, it requires a data representation that supports efficient spatial and semantic search.

Before presenting some of the relevant concepts to be incorporated in a baseline ontology for information fusion, we provide some guidelines for construction of ontologies.

3.0 ONTOLOGY CONSTRUCTION METHODOLOGY

The development of ontologies is a modeling activity that is complex and time-consuming. Therefore, methodologies have emerged based on experiences gained in the construction of large ontologies. A survey of these methodologies is presented in [8]. These aim at making the development of ontologies more an engineering process rather than an art. The main stages that can be derived from these methodologies consist of the following:

- Definition of the requirements for the ontology (purpose and scope);
- Building informal specification of concepts (taxonomy);
- Formally represent the concepts and axioms in a language (ontology);
- Evaluation of the ontology.

The purpose of an ontology defines its intended use, and the scope of an ontology delimitates the world to be modeled. The next step in the process consists of identifying the most important concepts in the domain, build a lexicon for these terms, and derive a comprehensive taxonomy of terms of the domain. The use of a mixed top-down and bottom-up approach to ontology development is recommended. The top-down mode may extend the definition of concepts from an existing upper-level ontology, i.e. establish links to upper-level categories that have already been defined within large ontologies (e.g. CYC) or relevant military models (e.g. NATO LC2IEDM data model). The bottom-up approach adds more specific concepts from additional reference sources (e.g. glossaries, terminology or domain databases, etc.).

The semantics of concepts in the ontology is specified through their definition, their properties, relations with other concepts, and eventually axioms that formally specify definitions and constraints of terms in the domain. Usually, an ontology is decomposed into subdomains organized into different hierarchies of concepts. Top-level concepts being at the top of class hierarchies are sometimes called microtheories, (e.g. Military equipment).

An important aspect in the ontology development process is to explicitly establish relationships that exist between concepts. Some of the relations that can be defined between concepts are:

- Relations that link a concept with more specific concepts (*is-a/subsume* relation);
- Relations that link a complex object to its constituents (*part-of/contains* relation);
- Any variety of relations that should be specified. These relations include for example causal, functional dependencies, or temporal relations.

A thorough analysis of the processes to be supported by the ontology should be performed as well as an analysis of the relevant data/knowledge sources involved in the processes. Ontological engineering could be seen as the domain modeling component of a global knowledge engineering method for the design of intelligent systems (knowledge-based systems, knowledge management systems, knowledge portals, etc.). For example, the CommonKads method for knowledge engineering and management incorporates the ontology building process in its knowledge model. IDEF is a method widely used within the military community designed to model the decisions, actions, and activities of an organization or system that incorporates an ontology capture method (IDEF5). In some cases such as knowledge-intensive military environments, a Cognitive Work Analysis (CWA) is performed to analyze the cognitive processes involved, and provide a functional decomposition of the processes and their information requirements. Such analysis should help derive and/or refine the concepts of the ontology, and better determine the scope of the ontology.

Moreover, the development of ontologies should be an incremental process, validated by subject matter experts at each stage of the process, and should maximize subsequent reuse and extensibility.

In the case where the ontology is exploited as a language to facilitate access to heterogeneous sources, a mapping has to be provided between the concepts of the ontology and the meta-models of the different data sources (description of the content of the sources, i.e. data models, or meta-data).

The whole ontological engineering process is presented in figure 1.

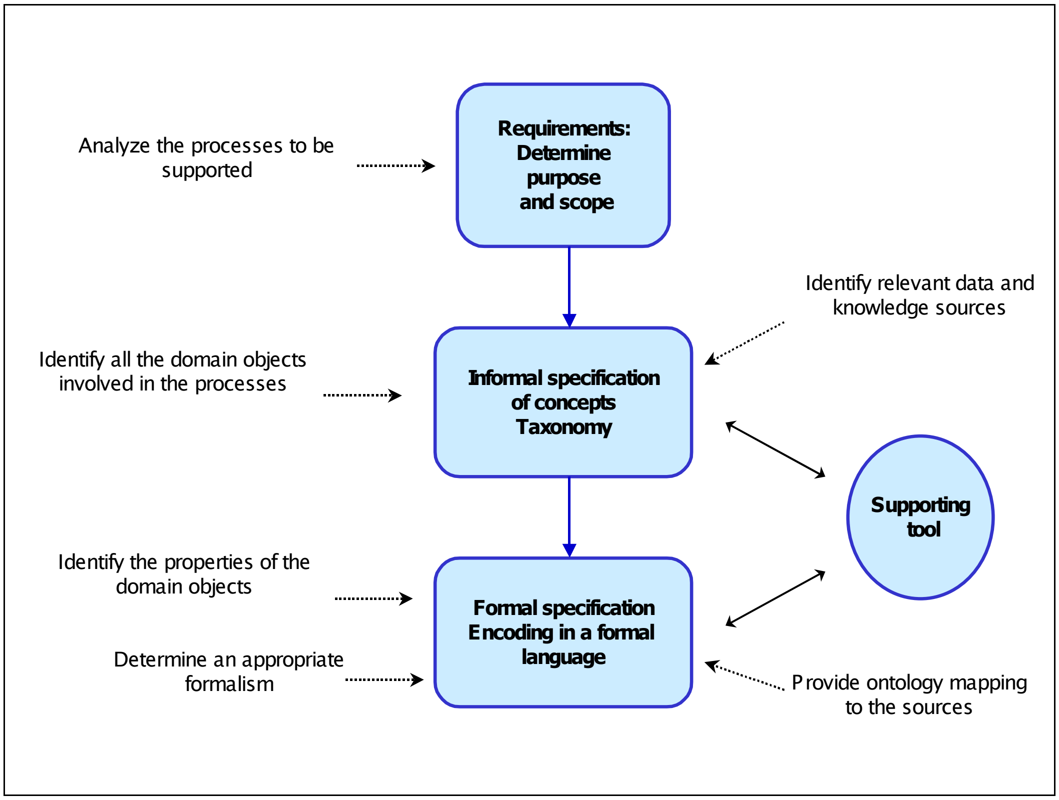


Figure 1: Ontology development process.

An ontology representation language has to be chosen to encode the ontology. The degree of formality of the ontology is mainly determined by the purpose of the ontology. If the ontology is a framework for communication among people, the representation can be informal, but if the ontology is to be used by software tools or intelligent agents to support automated tasks, then a more formal representation is required [16]. Different formalisms and knowledge representation languages have been proposed to describe ontologies. Some are limited to describing concepts, attributes, and relations and resemble conceptual models in databases or object-oriented models (ex. UML class models). UML has recently been extended to become a suitable candidate to support ontological engineering. Other formalisms use knowledge representation paradigms such as first-order logic, frame-based or description logic. New developments within the Internet community and the semantic Web have led to new ontology languages proposals such as DAML+OIL [5,11], or OWL (Ontology Web Language) [10]. An appropriate ontology language for high-level information fusion modeling would allow the formal specification of information fusion concepts, their properties and the different types of relationships that exist between concepts. This specification should serve as a basis for automated reasoning.

Ontological engineering requires environments to facilitate the process of building and maintaining ontologies, especially when the number of concepts and relationships become large. Such tools provide functions to edit concepts, browse the ontology, etc. They should facilitate the integration of existing ontologies, and the translation from the original formalism to some standard (e.g. RDF Schema, DAML+OIL, or OWL).

4.0 ONTOLOGICAL APPROACH FOR HIGH-LEVEL FUSION KNOWLEDGE MODELING

Ontological engineering for high-level fusion can be performed by analyzing Level 2 and 3 fusion processes in order to characterize the most important concepts that are part of these processes, and derive a specification for these concepts. This process constitutes a problem-oriented approach to ontology creation.

4.1 Situation Assessment

While level 1 data fusion deals with concrete entities such as emitters, platforms, low-level military units, level 2 and 3 are more concerned with abstract entities (e.g. event, intent, or goal).

Situation Assessment attempts to answer the questions: who is out there? Where are they and what are they doing? How are they situated? Which kinds of units are formed?

The output of Level 1, i.e., information about individual objects, their position, movement and identity, is aggregated into a composite tactical picture at Level 2. This concerns the situation assessment issue that leads to a more symbolic representation of the environment and the relationships among the entities and the events in it, to produce a higher-level of statement of what is really happening. Situation assessment focuses on relational information to determine the meaning of a collection of entities (into combat units and weapon systems), by aggregating the objects by location and type. The location attribute helps determine if an entity is part of an aggregation. Knowledge of the types of units helps determine what kind of actions, operations and strategies are possible. Hostile behaviour patterns help to establish whether the situation corresponds to a normal activity, threatening conditions, a suspicious activity, etc.

Environmental information (terrain, surrounding media, hydrology, weather, sea-state, underwater conditions) is taken into account in the situation assessment process. Thus, models of terrain, oceanography and meteorology have to be provided.

Figure 2 illustrates some important elements of Level 2 fusion that should be further refined and specified in the ontology. Relationships are not specified here, but typical relationships between entities include constituency-dependency, causal, temporal relations or geometrical proximity. In particular, temporal and spatial relationships should be carefully specified. Moreover, functional relationships that group entities carrying out the same functions have to be specified.

Ontologies of time and space are particularly important in military domains because reasoning in a dynamic world requires a formal means to describe spatial and temporal entities. For example, the order and sequence in time in which the entities are observed are crucial for their interpretation. Moreover, the analysis of movements of entities gives some indications on their intention (offensive or defensive).

Different time structures can be considered, such as point structure or interval structure. For the latter, a set of relations on temporal intervals proposed by Allen consist of the following: equals, before, meets, overlaps, starts-by, contains, ended-by. A space ontology concerns the representation of points or regions objects and relations amongst them. In a dynamic world, objects are located in time and space (notion of motion). Thus, representations of time and space have to be combined.

Different ontological models of time and space are described in [15]. Standardization efforts are also reported within the DAML community.

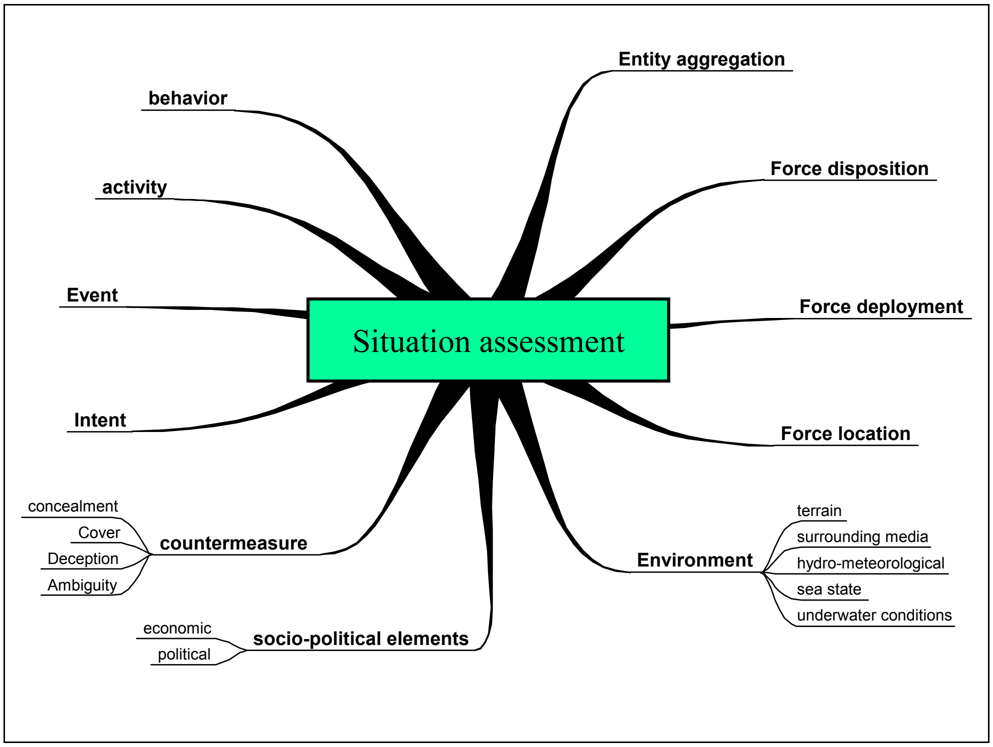


Figure 2: Elements of Situation Assessment.

4.2 Threat Assessment

The highest level of data fusion is the threat assessment level, or impact assessment (i.e. level 3), that projects the current situation into the future and infers about the impact of the assessed situation, the friendly and enemy vulnerability, the force capabilities, and determine levels of danger. Threat assessment attempts to answer questions such as: what is the overall force level? What is their readiness? How effective are they? What are their intentions? What are their goals? What are their strategies? What can we do about it? What will be the outcome of engagements?

Threat assessment aims at determining engagement outcomes as well as assessing an enemy's intent based on knowledge about enemy doctrine, level of training, political environment, etc. The focus is on intent, capability and opportunity [9]. These elements are depicted in Figure 3.

A major element of threat assessment is the prediction and evaluation of the enemy most probable courses of action. Thus, plan or courses of action (outlines of plans) models from the military domain that encapsulate important planning concepts (e.g. goal, action, event) are of interest here.

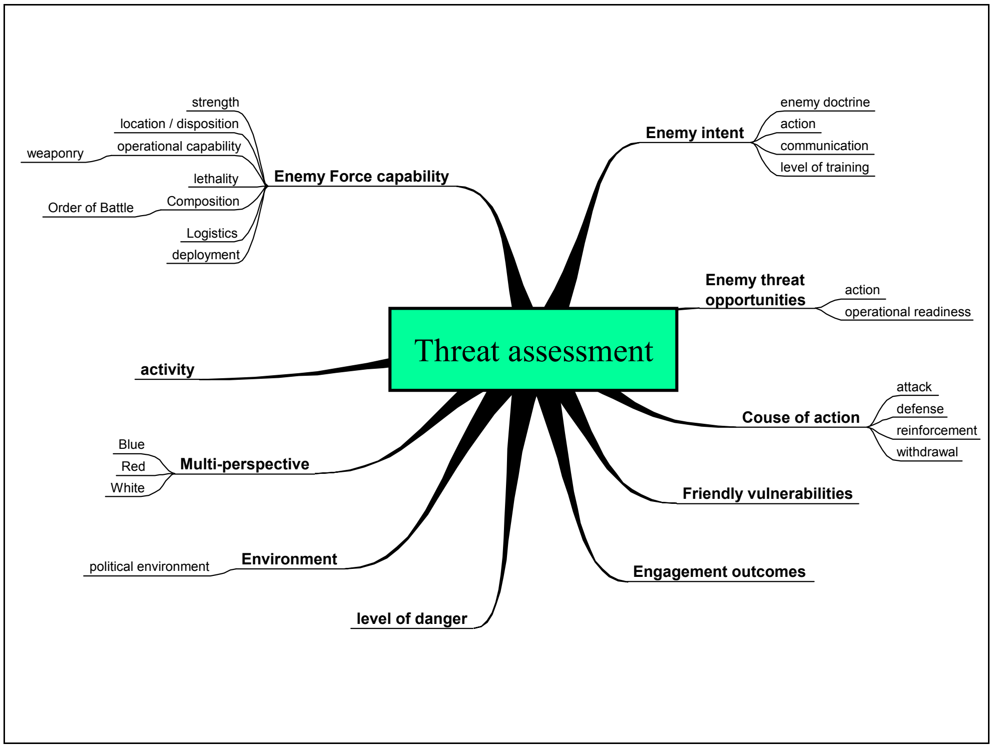


Figure 3: Elements of Threat Assessment.

5.0 ONTOLOGY-BASED KNOWLEDGE MANAGEMENT SERVICES

Complex military environments require knowledge management techniques and tools to deal with information overload, to exploit available data from different sources in multiple formats, to provide effective access to information, to analyze large data sets using intelligent data mining techniques, and to facilitate content extraction from unstructured data using artificial intelligence and natural language processing techniques, etc.

Ontological models can be exploited to provide more advanced knowledge management services. First, ontologies facilitate interoperability between heterogeneous knowledge sources by providing a knowledge-level description of a domain that can be mapped to heterogeneous data or information sources. Secondly, in military environments where an increasing volume of both structured and unstructured information has to be digested (e.g. the intelligence domain), ontologies can be exploited to index semi-structured information sources, to help extract semantic content from unstructured data, and to provide more advanced knowledge management services such as categorization of documents or ontology-based search engines.

5.1 Ontologies for Information Integration

Level 2 and 3 information fusion processes require a significant amount of a priori database information to support the components analyses. Information sources supporting information fusion processes are in different forms and formats. Major categories of databases required for level 2 and 3 data fusion are provided in [17]. Representational techniques to support data fusion processes are usually spatial and object-oriented [1].

However, high-level information fusion processes also exploit military doctrine, procedures, or lessons learned, usually expressed in semi-structured textual format.

The KNOWMES project (KNOWledge Management and Exploitation Server) conducted at DRDC-Valcartier aims at providing an ontology-based knowledge server that exploits a priori databases containing heterogeneous information in support of Situation/Threat Assessment and Resource Management (STA/RM) [3]. The proposed server relies on a three-tiered architecture that includes an ontology layer (cf. figure 4). The client tier provides various knowledge management functions available to the user, such as search and retrieval, knowledge base browsing, and handles user interactions. The middleware layer implements the knowledge management services. It includes an ontology server to manage and exploit ontologies, as well as an ontology mapper that links the ontology to the sources (through their models). The data tier concerns the heterogeneous information sources, namely an object-oriented database, a GIS database, and an XML repository. The object-oriented database contains a priori data to support the different fusion levels (e.g. descriptive information about military equipment, object behaviour patterns, formation, communication scheme, weapon and sensor capabilities). It has been chosen due to its capability to organize objects and relationships between objects [1], and to manage them effectively. The GIS database contains geographical and topological information such as terrain elevation, roads, buildings, or commercial corridors. The XML database contains documents of relevance for high-level fusion processes, such as doctrine, standard operational procedures, or rules of engagement.

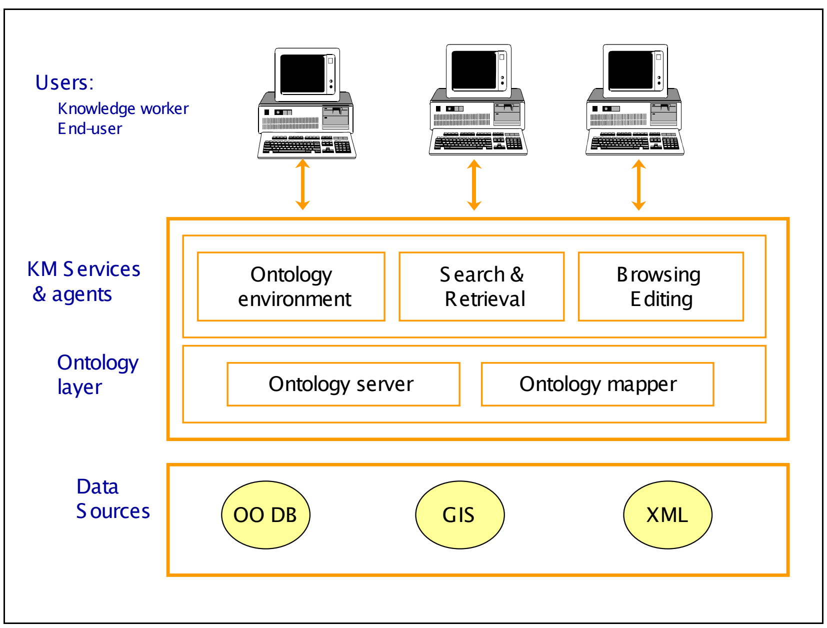


Figure 4: Architecture of KNOWMES.

5.2 Advanced Search Engines

The exploitation of ontologies for query-answering is twofold:

1) query formulation,

2) query interpretation and processing.

First, an ontology can serve as a language to express queries. Using a common terminology, users can formulate queries without worrying about exact names of the entities in the data sources nor the structure of the underlying data sources. An ontology browser may be available to present the ontology structure to the user who can then select the concepts of interest. A common terminology facilitates also a unified access to multiple sources. A federated search mechanism has to reformulate user queries (expressed using the ontology) in the query languages of the different sources (e.g. SQL statement for relational databases).

The ontology itself can be exploited to improve query-answering mechanisms. Ontologies can be exploited to enhance search engines mechanisms by analyzing relationships between concepts and thus providing more relevant results (e.g. remove noisy results or redundancies, enable search refinement). The relationships linking concepts (i.e. abstraction and composition) can be used to extend or refine queries. For example, a search for information about *helicopter* will derive a search for information for *CH-146* given that *CH-146* is a type of *helicopter*. So, documents about *CH-146* will be retrieved even if they do not contain the word *helicopter*. More complex relations and their properties could be exploited to process queries and draw inferences leading to more relevant results. Moreover, ontology-based search engines should also take into account synonyms of terms that are incorporated in the ontology. Then, the search engine looks for the terms specified in the query as well as their synonyms. Acronyms should also be considered when their definition is available in the ontology. These mechanisms are incorporated in a prototype under development to enhance search capabilities within a lessons learned management system for the Canadian Forces.

Finally, retrieving information by exploitating ontological models and by taking into account the context of a current situation should lead to more relevant and accurate results for the decision maker.

5.3 Ontologies for Document Categorization

Military organizations have to deal with an increasing number of documents coming from different sources and in various formats. These documents have to be screened, analyzed and categorized in order to interpret their content and gain situation awareness. They should be categorized according to their content to enable efficient storage and retrieval A better categorization and management of information would facilitate correlation of information from different sources, avoid information redundancy, improve access to relevant information, and thus better support decision-making processes.

Ontologies, or the taxonomy aspect of ontologies, can be exploited as a support for automatic document categorization. On one hand, they organize concepts in a hierarchical structure that can be utilized for categorization (e.g categories of Yahoo). On the other hand, they provide the semantics of a domain that can be exploited to improve traditional classification methods based on statistics. Providing a formal model of a domain through ontologies, or classification of terms through taxonomies has often been identified of potential utility to support information extraction from texts or for automated document indexing. For example, WordNet, a large electronic lexical database publicly available may be used to support information extraction or query formulation. But it is considered not suitable for processing texts in specific domains.

Natural language processing techniques supported by a domain ontology can be used to extract semantic meaning from unstructured text and provide semantic indexes from the ontology. An ontology-based semantic analysis consists of analyzing candidate concepts resulting from the statistical analysis from a semantic perspective by exploiting a domain ontology, in order to restrict the document descriptors to the attributes that semantically characterize the text (e.g. remove poorly meaningful words). At each level of the ontology structure, specific semantic expressions are attached to concepts, to guide the semantic processing. A semantic search engine is used to search for semantic similar expressions in documents being processed.

This allows the system to refine the semantic analysis of the document and thus provides a fine-grained documents categorization.

Some research projects exploiting ontologies have been undertaken at DRDC-Valcartier to organize military knowledge bases and search for information more effectively. In particular, Lessons Learned knowledge warehouses contain relevant experiential knowledge reported from previous missions (observations, issues, recommendations) expressed in natural language with many acronyms and abbreviations. If these knowledge sources are adequately indexed and organized along a taxonomy of the domain, they can be exploited by military officers to effectively retrieve information when assessing a situation or when planning a new mission. Moreover, we are conducted some experiments in order to exploit both statistical methods and semantics relying on an ontology for automatic document classification in the terrorism domain to help military officers facing information overload.

6.0 CONCLUSION

In this paper, we have presented an ontological approach to military knowledge modeling and management, in particular for high-level information fusion. We have also illustrated how these models could provide enhanced knowledge management assistance to military commanders. As ontologies promote knowledge reuse and sharing, we should benefit from previous models built in related domains.

Beside the KNOWMES project related to knowledge management in support of information fusion, several research projects conducted at DRDC-Valcartier aim at exploiting ontological models. This includes the building of a situational awareness knowledge portal, the management of lessons learned, and the design of an automatic document analyzer and classifier. The work presented herein and the other ongoing initiatives should benefit from each other.

Finally, there are some ontological engineering issues that have not been addressed here, for example, the maintenance and exploitation of extensible ontologies, and the integration/mapping of several ontologies within different military environments. These aspects are important and should be analyzed further.

7.0 REFERENCES

[1] R. T. Antony, Principles of Data Fusion Automation, Artech House, 1995.

[2] K. Baclawski, M. Kokar, M. Malczewski, Formalization of Situation Awareness, in *Proceedings of OOPSLA 2002*.

[3] A.-C. Boury-Brisset, Ontology-based Approach for Information Fusion, in *Proceedings of the Sixth International Conference on Information Fusion*, Cairns, Australia, July 2003.

[4] M. Bowman, A.M. Lopez, Jr., and G. Tecuci, Ontology Development for Military Applications. In *Proceedings of the SouthEastern Regional ACM Conference*, Atlanta, GA, March 16-17, 2001.

[5] DAML, DARPA Agent Markup Language Web Site, http://www.daml.org.

[6] T. Gruber, A translation Approach to Portable Ontology Specifications, *Knowledge Acquisition*, 5(2), pp. 199-220, 1993.

[7] W. Johnson, I. D. Hall, From Kinematics to Symbolics for Situation and Threat Assessment, in *Proceedings of Information, Decision and Control 99*, Adelaide, Australia, 1999.

[8] D. Jones, T. Bench-Capon, P. Visser, Methodologies for Ontology Development, in *Proceedings of IT&KNOWS (Information Technology and Knowledge Systems) of the 15th IFIP World Computer Congress*, Budapest, Hungary, 1998

[9] D. L. Hall, J. Llinas, *An Introduction to Multisensor Data Fusion*, in Proceedings of the IEEE, Vol. 85, No. 1, 1997.

[10] J. Heflin, R. Volz, J. Dale, Requirements for a Web Ontology Language, Working draft of the W3C Ontology Working Group, http://www.w3c.org/TR/webont-req.

[11] I. Horrocks, DAML+OIL: a Description for the Semantic Web, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2001.

[12] D. L. McGuinness, Ontologies Come of Age, In Dieter Fensel, Jim Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. MIT Press, 2002.

[13] J. Roy, From Data Fusion to Situation Analysis, *Proceedings of Fusion'2001 International Conference*, Montréal, Canada, 2001.

[14] N. Steinberg, C. L. Bowman, and F. E. White, Revision to the JDL data fusion model, Joint NATO/IRIS Conference, Quebec City, October 1998.

[15] O. Stock, *Temporal and spatial reasoning*, Kluwer Academic Publishers, 1997.

[16] M. Uschold, M. Gruninger, Ontologies : principles, methods and applications, *Knowledge Engineering Review*, 11(2), 1996, pp. 93-155.

[17] E. Walz, J. Llinas, *Multisensor Data Fusion*, Artech House, 1990.

Ontological Approach to the Representation of Military Knowledge

Zdenka Winklerová
Military Academy in Brno
Command and Staff Faculty
612 00 Brno Kounicova 44
CZECH REPUBLIC

Email: zdenka.winklerova@vabo.cz

ABSTRACT

The answers to commanders' priority intelligence requirements are typically based on a speedy, partial analysis of the information available. The research findings have made a significant progress in certain areas of data interpretation for ***intelligence analyses*** *which are also referred to as* ***sensor fusion, data fusion and information fusion****. The accuracy and speed of the fusion can be significantly improved by utilizing the knowledge processed at the appropriate level.*

In this document, the general problem of information provision for decision making is addressed. The most appropriate solution for implementation of a data fusion system, the-multi-agent architecture, is suggested. One multi-agent application related to a specific kind of knowledge, the social knowledge, is outlined, presenting some experience about usability of ontological approach to the representation of military knowledge.

Keywords: *data fusion, multi-agent architecture, social knowledge, ontology, background knowledge, pragmatic meaning of information*

1.0 INTRODUCTION

1.1 Command and Control Assessment Challenges

For the purposes of Code of Best Practice (COBP) [1], the term Command and Control (C2) is intended to be an umbrella term that encompasses the concepts, issues, organisations, activities, processes, and systems associated with the NATO definition of C2. Analyst will increasingly be called upon to provide insights into non traditional operations, Operations Other Than War (OOTW), and to work in a new conceptual dimension in order to examine the impact of new information-related capabilities coupled with new ways of organising and operating.

1.2 Principles

Three often cited principles of conventional warfare include the need for unity of command, the importance of hierarchical decision making, and the criticality of achieving surprise in operations. The COBP [1] has proposed alternative principles for OOTW. It cites the need for unity of purpose, consensus in decision making, and transparency of operations.

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

1.3 Analysis

The analysis of C2 remains among the most challenging. In addition, the analyses of OOTW often require consideration of individual behaviour. This has led to the application of "softer" analytic approaches, e.g. extensive reliance of expert elicitations.

1.4 Decision Making Process

It is important to have a proper representation of the decision making process to represent information operation effects. Representation of the decision making process, however, remains difficult because of the difficulty in representing human performance, command styles and organisational relationships. As the first step, it seems suitable to address the general problem of information provision for decision making. The question is, what kind of information the observers should communicate to the decision makers? If we have an improvement in decision making through the provision of information, then we consider the information to be of *value*. Improvement of decision can be viewed in the following ways:

1) Improvement in the decision outcome, i.e. an alternative is selected resulting in a better outcome.
2) Improvement in the decision process. Here a decision is considered as improved, not (necessarily) because a new and better alternative is selected, but because the obtained information enables the selection procedure to become more logical, resulting either in a quicker selection or a selection with greater ease and confidence.
3) Through appropriate provisions of information: An improvement may result if decisions can be made delegable to other lower levels of decision makers.

1.5 Information

Thus, the information should be represented and handled as a commodity. Information should be considered as a resource that can be collected, processed, and disseminated. Information is rarely valuable in its original form. It usually has to be processed in some way. Typical processing requirements include filtering, correlation, aggregation, disaggregation, and *fusion of information*. These processes can be accomplished by either manual or automated means. The ability, or inability, to properly process information, and the time it takes, can have a direct impact on the combat operational outcome.

1.6 Data and Information Fusion Task

At the first level of information processing, we like to use data fusion methods for *object refinement*. In other words, based on the multi-sensor data, we have got our best estimate of the identity of the object or objects out there in the given distance.

Next, we are in need of estimating the position of the objects and constituting entities called *situations*.

Once the situation has been determined, we want to know what the objects are doing, and we are trying to predict their future movement. We call this level of information processing *contextual understanding* and it also involves behavioural questions related to our goal of estimating human intents.

Thus, the objective of Data Fusion (DF) Task, and/or Information Fusion Task, is making decisions on the basis of distributed data sources accessible through a system (organisation). An objective of a DF System is to combine data from many different sources to make decisions.

According to the commonly accepted JDL[1] view [2], Data and/or Information Fusion is a multilevel process comprising several types of tasks:

[1] Joint Directors of Laboratory Data Fusion Model was formulated in 1992

- *Level 0 – Data Producing* – corresponds to the fusion of sensor signals to produce data specifying semantically understandable and interpretable attributes of objects.
- *Level 1 – Object Refinement.* This level is called *Data Fusion*, it aims at processing the above data to make decisions with regard to classes of the objects in question, i.e. the classes of the states of the objects.
- *Level 2 – Situations*. This level is normally called *Information Fusion* and its goal is to assess a *situation* constituted by the set of the above mentioned objects considered as a single whole called *a system* hereafter.
- *Level 3 – Contextual Understanding*. Information Fusion of this level corresponds to an *Impact Assessment*, which means, for example, adversary intent assessment produced on the basis of the situation development prediction.
- *Level 4 – The Feedback* – assumes calculation of a feedback, like planning resource usage, sensor management, etc.
- *Level 5 – Situation Management* – the upper level, involves human activity and situation management.

1.7 Communication

Any DF, or Information Fusion System, is inherently distributed. Information has a specific source, and that source is usually not the end user of the information. A requirement exists, therefore, to move information from one place to another in the operational environment. Communication systems of all forms exist to accomplish this movement.

The classical theory[2] differentiates three conceptual levels of communication:

1) *Syntactic level* which is concerned with the rules for building up sentences. At the syntactic level, we solve a technical problem: how accurately the symbols used in communications can be transmitted.
2) *Semantic level* which examines the meaning of signs in relation to the represented objects or actions. At the semantic level, we solve a representation problem: how intelligibly do the transmitted signs represent the intended message, and, how precisely the transmitted symbols convey the desired meaning.
3) *Pragmatic level* which features how the senders and receivers evaluate and understand the meaning including psychological impact, action consequences, etc. At the pragmatic level, we solve an efficiency problem: how efficiently the received message influences the behaviour of the receiver, or, more precisely, how effectively the received meaning affects the conduct in the desired way.

There is a close relation between these levels and the semiotic distinctions:

1) *Syntax* and the forms of language
2) *Semantics* and the meanings of language
3) *Pragmatics* and the use or function of the language

The first two levels of information processing, according to the JDL schema, produce estimates about the states of single (physical) objects. Estimates about more abstract entities called "situations", contextual understanding, feedback calculations, situation management and human intent estimating, are, in general,

[2] Shannon's Mathematical Theory of Communication and Warren Weaver's three levels of problems in communication (1949)

supposed to take knowledge from human experts as a line of reasoning that accurately reflects the way a military analyst would look at something.

Thus, according to the JDL schema, we are concerned:

1) with syntactic aspects of information at the Level 0
2) with syntactic and semantic aspects of information at the Level 1 and higher
3) with syntactic, semantic, and also pragmatic aspects of information at the Level 2 and higher

1.8 Pragmatic and Contextual Meaning of Information

The classical theory mentioned above is concerned with the syntactic level only, it is not at all concerned with the pragmatic aspects of information.

Attempts to raise the theory to the semantic level have been made by J. L. Austin in his work about *speech act theory*.[3] A need for the development of a pragmatic theory of information still exists.

2.0 MULTI-AGENT ARCHITECTURE OF DATA FUSION

For the design and implementation of a DF system, the most advantageous strategy is the strategy using a multi-level hierarchy of classifiers. The source based classifiers make decisions on the basis of data of particular sources followed by the meta-level decision making based on combining the source-level decisions. The advantages of such a scheme are

1) decrease of the data source information exchange
2) simplicity of data source classifier fusion even if they use data of different representation structure, certainty, accuracy, etc.
3) possibility to use mathematically sound mechanisms for combining decisions of multiple classifiers.

In some applications, this strategy is the only applicable one. For example, a group of applications in which the data are private and the data holders do not want to share the data, but agree to share decisions produced on the basis of such data sources.

2.1 Peculiarities of Data and Data Processing in DF Task

The first problem is the development of the shared thesaurus providing for *mono semantic understanding* of the terminology used in formal specification of domain entities. According to the modern understanding, it is necessary to use a meta-model of data and knowledge presented in terms of *ontology* and shared by all entities of the DF System. The structure of ontology used is depicted in Fig.1. It comprises the domain, problem ontology together with task ontology and also with the application ontology. The application ontology comprises two types of components of the DF system: a part of application ontology that is shared by all components of DF system and parts of the application ontology that are private for particular data source. It seems reasonable, at least in theory, to have a unified top-level ontology for large communities of users. Domain ontology and task ontology describe the vocabulary related to a generic domain and/or a generic task activity by specializing the terms introduced in the top-

[3] John Langshaw Austin's 1955 lectures were published posthumously under the title of *How to Do Things with Words* [7]. The basic idea is concerned with the provision of a measure of semantic information content of simple declarative sentences in a defined language system. Knowledge Communication Meta Language (KQML) was one of the earliest attempts to construct a communication language based on the speech act theory. The Agent Communication Language (ACL) represents a version of KQML which is more precisely defined. There is a standard version of ACL written in XML and sponsored by FIPA.

level ontology. Application ontology describes concepts depending both on a particular domain and task, which are often specializations of both the related ontologies.[4] These concepts often correspond to the roles played by domain entities while performing a certain activity.

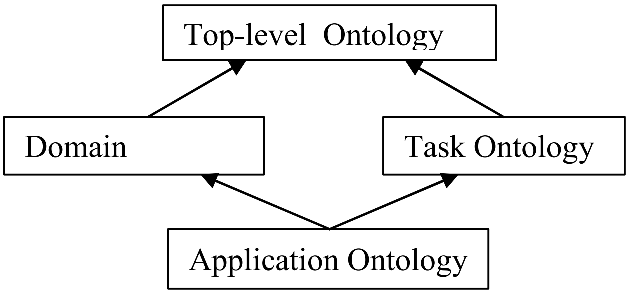


Figure 1: The Structure of Ontology.

The second problem corresponds to the *entity identification problem*. The data specifying an object may be represented in different data sources. This is why each local data source only partially specifies the above mentioned objects. Its complete specification is made up of data fragments distributed through the data sources. Therefore, to form a complete object specification, a mechanism to identify such fragments is needed. It should be noticed that some fragments of data associated with an object can be absent in a number of sources.

2.2 Combining Decisions of Multiple Classifiers

In the most DF tasks, decision is understood as classification of an entity (e.g. object, state of an object, situation), that is, assigning the entity a class label from a fixed set. Each local data source is associated with a single or several classifiers. Each of them produces classification of an entity using only local data or local data fragments and then transmits the classification produced to meta-level, where these classifications are combined by meta-level classifier(s) using an appropriate method.

2.3 Implementation of a DF System

To implement the above described conception, multi-agent architecture is the most appropriate solution for the desired DF system.

The developed architecture of DF software consists of two kinds of components:

1) the component responsible for the design of source based parts of the desired DF system
2) the component supporting iterative and interactive designs of the meta-level part of the desired DF system

Both of them present allocation of a particular task through agents. The source based component includes data source managing agent and local classification agents of a DF system. The meta-level component includes an agent-classifier of a meta-level which receives decisions from local source based agents of the DF system, and via the semantic processing of input messages produces a top-level decision. The selection

[4] It may be important to make clear the difference between an ontology and a knowledge base. Ontology is a particular knowledge base, describing facts assumed to be always true by a community of users. Within a generic knowledge base, we can therefore distinguish two components: the ontology, containing state-independent information, and the core knowledge base, containing state-dependent information.

of scenario depends on the input and inner state of the agent. In turn, inner agent's state depends on the prehistory of the agent's operations. This prehistory is reflected in the state of agent's knowledge base.

Design and implementation issues of such a DF system are proposed in [3].[5]

3.0 SOCIAL KNOWLEDGE IN MULTI-AGENT SYSTEM

Any DF system deals with distributed data sources and executes distributed or decentralized data processing. Decision about decisions from local sources is produced by meta-level agent. The base-level and the meta-level agents interact during solving a DF task. Both the idea of advancing decision to reasoning and the usability of the multi-agent system for Information Fusion seem suitable.

A multi-agent system (MAS) usually consists of a set of autonomous units capable of:

1) independent operations aimed at meeting their *local goals*
2) cooperative actions contributing jointly to the *global goal* shared across the community.

The agents' abilities to communicate, mutually coordinate their actions, cooperate and share the global goals determine the level of their integration oriented behaviour. These abilities depend mainly on the extent and quality of knowledge available to the agents. Knowledge, a true piece of evidence in which the agent believes, can either

1) guide the agent's autonomous local decision making processes, aimed e.g. at providing an expertise or search in the agent's database; this is what we call agent's problem solving knowledge, or
2) express the other agents' behavioural patterns, their capabilities, working loads, experience, commitments, knowledge describing conversations or negotiation scenarios, which we will refer to later as *social knowledge*.

Let us outline one application related to this specific kind of knowledge.

3.1 CPlanT Multi-Agent System

This part will describe the CPlanT multi-agent system that has been implemented for planning the OOTW coalition.[6] [4] The further specified principles and ideas have been tested and implemented on humanitarian relief (HR) operations which are a subset of the OOTW types of operations. A hypothetical humanitarian scenario has been designed and implemented for this purpose.[7]

Unlike in classical war operations, where the technology of control is strictly hierarchical, OOTW are very likely to be based on the cooperation of a number of different, vaguely organized groups of people, such as non-governmental organizations (NGOs) providing humanitarian aid, but also of army troops and official governmental initiatives.

Unlike hierarchical approach, collaborative approach to the operation planning allows greater deal of flexibility and dynamics in grouping optimal parties playing active roles in the operation. The main reason

[5] This research was supported by grant of AFRL/IF-European Office of Aerospace Research and Development (Project 1993P)

[6] This project was supported by AFOSR/European Office of Aerospace Research and Development under contract number F61775-00-WE043.

[7] The scenario has been encoded in the XML files and the computational model of the scenario has been implemented in Allegro Common Lisp. While the inter-agent communication is FIPA compliant, each agent in CPlanT is a stand-alone application, and the agents communicate via TCP/IP connection.

why we can hardly plan operations involving different NGOs by a central authority results from their reluctance to provide information about their intentions, goals and resources. Actual information may become unavailable also due to unreliable communication channels. It may happen that a collaborative entity gets cut off the communication links for a certain period of time and the rest of the community still wishes to be able to form/execute plans relying upon the missing player. For this reason, each participating entity should be able to maintain approximate model of the collaborating members of the coalition.

For the above described OOTW, *the multi-agent community* denotes the whole collection of participating agents. A *coalition* is a set of agents who agreed to fulfil a single, well specified goal. An agent may participate in multiple coalitions. A process of *coalition planning* will be understood as an agent's reasoning about possible candidate for membership in the planned coalition. *Coalition formation* is the process of reaching an agreement among candidates for membership in a planed coalition. Unlike in classical coalition formation problems, the quality of coalition in OOTW is not given only by the function of different utilities, such as a task completion time or the number of collaborating agents, but also by the amount of private information that the agents had to reveal.

There are three levels of agent's knowledge representation suggested:

1) *Public Knowledge* is shared within the entire multi-agent community. This class of knowledge is freely accessible within a community. As public knowledge we understand agents name, type of the organization the agent represents, general objective of the agent's activity, country where the agent is registered, agent's human-human contact (fax number).
2) *Alliance Accessible Knowledge* is shared within a specific alliance. We do not assume the knowledge to be shared within the overlapping alliances. Members of an alliance will primarily share information about free availability of their resources and respective positions. This resource oriented type of knowledge may be further distinguished as material resources, human resources and transport resources.
3) *Private Knowledge* is owned and administered by the agent himself. The agent, provided he finds it useful, may communicate some pieces of the private information upon a request. As an instance of private knowledge we can consider mainly specific status of the resources the agent administers, future plans and allocations of resources, his past performance, agent's collaboration preferences and restrictions, and agent's planning and scheduling algorithms.

3.2 Acquaintance Model

Models containing social knowledge are usually called *acquaintance models*. An acquaintance model is a knowledge based model of agent's mutual awareness that summarizes the agent's knowledge about his collaborators and about suitable communication and negotiation scenarios. An acquaintance model is meant to maintain permanent, semi-permanent and non-permanent information about other agent's services, knowledge, statuses, about potential negotiation scenarios, delegation principles etc. It is required that the acquaintance model will also contain certain knowledge about its own knowledge, status and intended activities which is usually referred to as agent's self-knowledge. The corresponding part of this knowledge can be accessible to the collaborating agents and they maintain it in an identical form.

A specific methodology for organization and administration of agent's mutual awareness is known as the *tri-base acquaintance model (3bA)*. The most important virtue of the 3bA is the absence of a central element. If an agent in the community dies or gets overloaded, the system is expected to reorganize itself in order to solve its tasks anyway.

3.3 Knowledge Structures of the Acquaintance Model

Within the tri-base model, each agent maintains three knowledge bases where all the relevant information about the rest of the community is stored.

1) *Co-operator Base (CB)* maintains permanent information of co-operating agents, i.e. their addresses, communication languages, their predefined responsibility. This type of knowledge is expected to be changed rather rarely.

2) *Task Base (TB)* stores in its *Problem Section (PRS)* general problem solving knowledge – information about possible decompositions of the tasks to be solved by the agent, and in its *Plan Section (PLS)* the agent maintains the actual and most up-to-date plans on how to carry out those tasks.

3) *State Base (SB)* has two parts, the *Agent Section (AS)* and the *Task Section (TS).*

The agent *A* stores in his AS all relevant information characterizing the present state of the relevant part of the system, e.g. the current load of cooperating agents. This part of the state base is updated frequently and informs the agent who is busy, who is available for collaboration and makes the agent possible to evaluate what conditions hold at present. A sophisticated agent can include here very complex knowledge including knowledge about himself.

A slightly redefined tri-base model has been designed for planning peacekeeping operations. The co-operator base, as the community belief-base, stores public information about all the community members. The task base keeps possible coalition with respect to a particular task. The state-base has been split into two bases: social-belief base, where non-permanent information about the alliance members is stored, and the self-belief base, where the alliance accessible information of the agent himself is stored and offered to the other alliance members. All three types of knowledge – private knowledge, alliance accessible knowledge and public knowledge may be stored in a single knowledge structure.

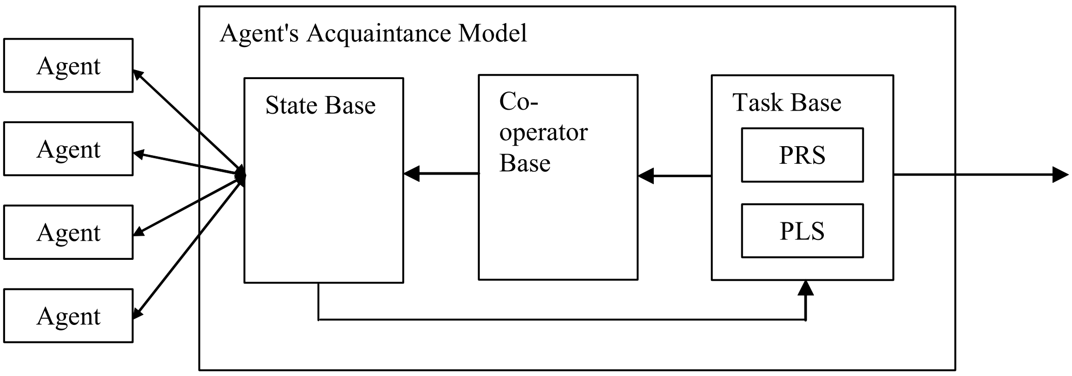


Figure 2: Tri-Base Acquaintance Model.

3.4 Generation of Plans

Suppose that the agent *A* is in charge of a task *T*. The agent can either

1) use an existing plan stored in the plan section PLS of his task base, or

2) generate a new plan using his own knowledge and inference mechanisms.

In the latter case the agent *A* takes problem knowledge found in the problem section PRS of his task base. There can exist several rules in PRS the conditions of which are met. A good choice has to be supported by appropriate techniques of conflict resolution. The plan with the highest *trust* is viewed as the actual plan. Whenever the contents of the agent section of the state base gets updated, the *trust* is to be recomputed and each of the rules is to be re-evaluated. This kind of activity makes the plan the most up-to-date.

3.5 Knowledge Improvement

Besides knowledge maintenance, which keeps the knowledge in the state base up-to-date, the 3bA concept allows knowledge to be improved as well. There are two ways how to implement permanent knowledge improvement:

3.5.1 Object Level Knowledge Improvement

Object level knowledge improvement is based on agent's ability and responsibility to optimise, reorganize, deduce new pieces of knowledge and improve the knowledge stored. Object level knowledge improvement is primarily implemented by machine learning techniques that allow the agent mainly to find specific patterns of inter-agent communication, produce generalization, etc. Alternatively, the agent may be equipped with meta-reasoning (reasoning about other agents) capabilities and with explicit knowledge how the pieces of state base knowledge may be put together and new knowledge formed.

3.5.2 Meta-Level Knowledge Improvement

Meta-level knowledge improvement is not carried out directly by the agent owning the knowledge. Improvement and knowledge revision is provided by an independent meta-agent, who observes activities of the community, collects relevant pieces of information and tries to draw certain assumptions about the individual agent's behaviour. Meta-agent can meta-reason and learn how to enhance the community efficiency and he is able to provide advice to the agents.

In the next part it will be explained how the concept of meta-reasoning can contribute to improvement of the 3bA knowledge.

3.6 Meta-Reasoning and Reflectivity in Multi-Agent System

Let us consider a computational system capable of certain class of decision making, for example, language translation. The system will carry out computation in order to perform behaviour that will meet its designed objective. Such a computation is regarded as primary reasoning (object level reasoning). If we require this computational system to be reflective, it needs to be able to reason about itself.

A reflective system consists of

1) object component (object agent)
2) reflective component (meta-agent(s))

Reflective reasoning allows the system

1) to analyse and learn from its past course of decision making (*learning*)
2) to detect inconsistence in manipulated knowledge (*reality check*)
3) to suggest efficiency improvements in the respective problem solving (*adaptation*).

A reflective multi-agent system should contain either a single meta-agent, or a collection of meta-agents who are capable of reasoning about other agents who carry out the primary decision making:

1) about their knowledge
2) about their reasoning process.

The meta-agents individually or collectively constitute a *reflective component* of the multi-agent system. The multi-agent system performs reflective reasoning, reasoning about itself, by *meta-reasoning* carried out by meta-agents. Meta-reasoning in a 3bA model can be considered as a form of *social reasoning*.

If we understand meta-reasoning as to be reasoning about reasoning, meta-reasoning is not only useful for implementation of reflection. The idea of meta-reasoning extends the concept of the 3bA model twofold:

1) *Assumed Belief.* An agent can update a record in his state base not only by being told. Due to the content of his state base, the agent can reason about other agents to analyse their behaviour and to predict future course of actions. This capability will upgrade agent's social intelligence.
2) *Patterns of Community Interactions*. Sometimes it is impossible to detect interesting patterns of community interactions from the single agent's point of view. This is true mainly due to the fact that the agents have usually their organizational roles and cannot monitor or even understand the whole of the community.

4.0 ONTOLOGICAL APPROACH TO KNOWLEDGE REPRESENTATION

As a basis for cooperation, the capability of cooperative agents' actions assumes the communication across the multi-agent community. The communication among different agents is therefore an important aspect of multi-agent systems.

Agents do not have a direct access to each other but they can request services by sending messages. For the agents, the act of communication denotes the activity of sending some information from a sender to a set of (intended) receivers. An advantage of this approach is that one can get loosely coupled open systems that only use message passing as a vehicle for collaboration. The use of ontologies in message exchange communication gives meaning to the contents of messages sent between the agents.

4.1 Message Content Ontology

Message Content Ontology is a technology to support inter-agent communication by providing a definition of the world on which an agent can ground his beliefs and actions, as well as by providing terms that can be used in communication. Thus, a message content ontology helps agents to describe facts, beliefs, hypotheses and predications about a domain.

4.2 Agent Communication

The most appropriate model of agent communication seems to be the abstract communication model of FIPA[8] derived from the speech act theory. In this model, communication occurs through the exchange of asynchronous messages corresponding to communicative acts. The Agent Communication Language (ACL) defines format of these messages. Within a message, elements in the world are defined in a domain ontology. A content language expression is used to represent the content of the message. Finally, a speech act as the agent's intention to describe or alter the world is wrapped around.

[8] FIPA, The Foundation of Intelligent Physical Agents, is a non-profit organisation aimed at producing standards for the interoperation of heterogeneous software agents.

(See the example given [9])

In order to preserve agent autonomy as much as possible, the FIPA communication model is based only on the speech act as the communication idea. For the agents to be able to reason about the effects of their communications, ACL messages could be inserted into proper Agent Interaction Protocol (AIP) which describes communication patterns as allowed sequences between agents and the constraints on the content of those messages.[10] [5]

4.3 Background Knowledge as Context

This type of knowledge has been called *real world knowledge*. What is usually meant is the knowledge the conversation participants might deduce that the others had before, or independently of, a particular conversation, by *virtue of a membership in a community*. Each community implies certain types of knowledge which might be shared with other members and which the listener must deduce in the course of the interaction.

(See an example given [11])

4.4 Ontology as Background Knowledge

Ontology might be built which could quickly reveal the extent to which a human readers make inferences to gain an understanding of a message context. Those inferences are often based on background knowledge. Various forms of knowledge representation have been proposed to model this background information.

Implementation of the background knowledge as ontology in appropriate knowledge structure should significantly decrease the traffic across the agent community.

---

[9] The ACL message shows an example in which agent Peter informs agent John that today it is raining. Domain ontology is 'Weather-ontology', language is English, content language expression is 'today it is raining', and speech act is 'INFORM':
<fipa-message act="INFORM">
<agent-identifier>
<sender>
<.name id=Peter@host1:8888/JADE/>
</sender>
<receiver>
<.name id=John@host1:8888/JADE/>
</receiver>
</agent-identifier>
<content>today it is raining</content>
<language>English</language>
<ontology>Weather-ontology</ontology>
<.conversation-id>Peter-John253781</conversation-id>
</fipa-message>

[10] CPlanT multi agent system for planning humanitarian relief operations uses classical negotiation algorithms such as Contract Net Protocol (CNP)

[11] The following example demonstrates how the cultural knowledge is the basis for the inference:
A: Come over next week for lunch.
B: It is Ramadan.
If *A* and *B* are Muslims then *A* will probably infer that *B*'s reply means *No*

CONCLUSIONS

In this document, the general problem of information provision for decision making process is dealt with, focusing on the typical processing requirements including fusion of information. Since any data or fusion information system deals with distributed data sources, the multi-agent paradigm was mentioned as a challenging framework for solving very complex tasks in a distributed way. The CPlanT multi-agent system was outlined as being intended for the OOTW coalition with collaborative approach to operation planning.

In virtue of the fact that the quality of coordination, cooperation and functional integration in multi-agent systems depends strongly upon the knowledge explored, a specific part of knowledge called *social knowledge* was described, as well as the models containing this kind of knowledge, acquaintance models, enabling to perform 'reasoning about reasoning' that allows agent to analyse behaviour and to predict future course of actions as a form of social reasoning.

Because the capability of cooperative agents' actions assumes the communication across the multi-agent community, the message content ontology was mentioned as a technology supporting inter-agent communication, as well as the background knowledge the implementation of which as ontology structure allows message context understanding.

It may be important to suggest that ontologies traditionally reside within the field of knowledge organization. The *sharing of ontologies* between diverse communities allows them to compare their own information structures with other communities that share a common terminology and semantics. *Information Flow* provides a foundation for the sharing of ontologies in a distributed setting. For more detailed discussion, see [6].

Successful and effective communication is heavily reliant on the capability of agents to communicate not only data and information, but knowledge as well, and calls for information sharing on all levels – syntactic, semantic and pragmatic. The pragmatic aspect of communication among agents focuses on two basic spheres:

1) the knowledge which agent to address, and how to locate that agent in question
2) the knowledge how to initiate and maintain the communication.

REFERENCES

[1] NATO *Code of Best Practice for Command and Control Assessments*, SAS-039 Symposium, NATO C3 Agency, 2002.

[2] Blash, E.: Tutorial TD2: *Fundamentals of Information Fusion and Application*. in Proceedings of Information Fusion 2002, Annapolis, MD, USA, 2002.

[3] Gorodetski, V., Karsayev, O., and Samoilov, V.: *Multi-agent Data Fusion Systems: Design and Implementation Issues*. in Proceedings of the 10th International Conference on Telecommunication Systems - Modeling and Analysis, Monterey, CA, October 3-6, vol.2, pp.762-774, 2002

[4] Pěchouček, M., Mařík, V., Bárta, J.: *CPlanT: An Acquaintance Model Based Coalition Formation Multi-Agent System*. in Proceedings of CEEMAS' 2001, Krakow, p. 209-216, 2001.

[5] Mařík, V., Pěchouček, M., V., Bárta, J.: *Reduction of Task Complexity in Coalition Planning*. in Proceedings of the 16th European Meeting on Cybernetics and Systems Research (EMCSR 2002) –

Third International Symposium "From Agent Theory to Agent Implementation", Vienna, April 3 - 5, 2002

[6] Barwise, K. J., Seligman, J.: *Information Flow: The Logic of Distributed Systems*. Cambridge University Tracts in Theoretical Computer Science 44, 1997.

[7] Austin, J. L.: *How to do Things with Words*. Claredon Press, Oxford, 1962

Zdenka Winklerova, senior lecturer at Communication Systems Control Department of Command and Staff Faculty of the Military Academy in Brno since 1999, received a university degree in Informatics at Technical University in Brno (1982). Research activities in the field of special data transmissions (1984-1992), database information systems development and implementation (1992-1999). Main recent activities relate to the following areas: Command and Control Systems (C4I), Systems Security, Operation Planning and Scheduling, Knowledge Representation.

Analysis of Free-Form Battlefield Reports with Shallow Parsing Techniques

Dr. Matthias Hecking
FGAN/FKIE
Neuenahrer Straße 20
53343 Wachtberg-Werthoven
GERMANY
Phone: +49 228 9435 576
Fax: +49 228 9435 685
hecking@fgan.de

ABSTRACT

A natural way to communicate with C2 systems would be to use natural language. There are already natural language components used in military systems, e.g. CommandTalk is a spoken-language interface to the ModSAF battlefield simulator. In our project SOKRATES we use shallow parsing techniques for written language to analyze German free-form battlefield reports. These reports are processed by transducers. The extraction result is formalized in feature structures, semantically enriched by the semantic analysis and the augmented result is then stored in the ATCCIS database. After storing in the database, triggers initiate a change in the position of a tactical symbol on the tactical map. Shallow parsing techniques are the basis for information extraction. In this paper, we therefore first introduce to the promising field of information extraction. Then, we describe in detail how shallow parsing techniques are used in our project SOKRATES.

INTRODUCTION

In the NATO technical report *Potentials of Speech and Language Technology Systems for Military Use: an Application and Technology Oriented Survey* (see [Steeneken, 1996]) the *processing of human language* was recognized as a critical capability in many future military applications, among other things in 'command and control'. Though, still a topic of research, there already exist natural language components in military systems, e.g. CommandTalk, a spoken-language interface to the ModSAF battlefield simulator (see [Moore *et al.*, 1997], [Stent *et al.*, 1999], [CommandTalk]) or the Phraselator (see [Phraselator, 2003]) used by the US Army in Afghanistan.

Today, the usability of human language technology is restricted to narrow and well-defined application areas (domains). Another requirement is that the language must be restricted as well. This means, that the vocabulary and the grammatical structures must be limited enough such that processing time becomes acceptable. The military domain and the stereotyped military command language seem to be appropriated for using language technology.

A natural way to input information in a C2 system would be to use written or spoken natural language. In this context, we tried to show in a former approach that the available methods, techniques, and tools of

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

computational linguistics are mature enough to test their applicability to C2 systems. In this former approach we used the ATCCIS database (cf. [NATO, 2000]) as the domain. We have already reported on the progress of our former project using a speech recognizer (cf. [Hecking, 2001]). Furthermore, we showed (cf. [Hecking, 2002]) how to use *deep syntactic analysis* techniques to analyze simple written sentences. This approach however has various deficiencies. Especially, the high demand on processing time and the expectation, that the sentences are all well structured with respect to the grammar, hampers the application of these techniques. Therefore, we were looking for an alternative approach that avoids these deficiencies.

Information extraction (IE) is an engineering approach based on results of computational linguistics to build systems that process huge amount of texts of a specific sort. IE is an approach that avoids the deficiencies mentioned above. Each IE system is tailored to a specific domain and task. IE uses a *shallow syntactic approach*, i.e. that only parts of the sentences (so-called 'chunks') are processed with finite state automaton or transducers. These parts contain the relevant information about the Who, What, When, etc. To realize an IE system the desired output must be specified. This is done through *templates*. Templates represent feature-value structures. During the IE process a domain-specific lexicon and domain-specific rules are used to instantiate the templates.

In this paper, we will first give a short introduction into the promising field of IE. Then, we show how we use the SMES system in our project SOKRATES to realize an IE system for battlefield reports in German. We will describe the various steps during the IE process and we will explain in detail the shallow parsing techniques used.

INFORMATION EXTRACTION

Information extraction (IE) is the task of identifying, collecting and normalizing information from natural language text (see [Appelt, 1999], [Pazienza, 1999]). Relevant information about the Who, What, When, etc. is looked for. The information of interest is described through patterns called *templates*. During the IE task these templates are filled with the collected information. IE therefore can be seen as the process of normalizing from free-form text into a defined structure. The templates are domain and task specific, i.e. for each new task and domain they must be newly created.

To realize an IE system, language resources (lexicon, grammar) and appropriated parsing software are necessary. This software must be language-specific. The IE tools for the English language are not appropriated for analyzing e.g. German text due to the free-order of the language.

In order to achieve robust and efficient IE systems, domain knowledge must be integrated and shallow algorithms must be used. The domain knowledge is tightly integrated with the language knowledge, e.g. the name 'Leopard' in the lexicon has the categorical information 'tank'. This association between words and semantic information is domain-specific and has to be change for other applications.

The current IE technology are used successfully in various application areas, e.g. intelligent information retrieval, linguistically based data mining, automatic term extraction, text classification.

The IE process itself is divided into sub steps. After tokenizing the text, the sentence boundaries must be identified. Then, the morphological component identifies the word stems, the abbreviation and detects the syntactic information (e.g. grammar case and gender). After this, the chunk parsing with transducers selects parts of the natural language text, which are relevant for the specific information extraction task. The chunks are then used to instantiate the templates, which represent the result of the IE process.

Various toolboxes are available to build IE systems. These toolboxes must be language specific. A powerful IE toolbox for German is the SMES system (cf. [Neumann, 2003]). This toolbox offers among other things a morphological analysis component with a huge lexicon, predefined grammars (transducers) for specific phrases (e.g. noun phrases) and the possibility to program arbitrary transducers.

THE PROJECT SOKRATES

The overall objective of the SOKRATES project is to analyze written German battlefield reports. The result of the analysis is stored in the ATCCIS database (see [NATO, 2000]). These stored results can be used for different purposes. One purpose is that location changes of units initiate automatically changes of tactical symbols on the tactical map.

The Architecture

The architecture is shown in Fig. 1. The free-form reports are handed over to the *coordination module*, which is responsible for all the coordination in the system. In a first step, the *syntactic preprocessing* identifies the sentence boundaries. Next, the *information extraction* module uses the lexicon and the grammar transducers to identify and select the relevant parts in the natural language text. These parts are represented as typed feature structures that are coded as an XML document. The result of the information extraction is used by the *semantic analysis* component to deduce more information out of the extracted information with the help of an ontology and the context (see [Schade, 2003a], [Schade, 2003b]). After the semantic analysis the result is pushed into the ATCCIS database and then it is used to alter automatically the position of tactical symbols on the map.

The Information Extraction Module

During the information extraction the structure of a text must be determined. To do this, a grammar with a lexicon and a parser is necessary. There are a lot of different grammar formalisms for natural language processing. To be able to process large amounts of text an efficient approach must be used. Therefore, we use a *shallow parsing approach*, i.e. that only parts of the sentences (so-called ‘chunks’) are processed. These transducers code the necessary grammar.

The result of the syntactic analysis is represented in templates. The necessary templates are formalized in SOKRATES by *typed feature structures* (see [Pollard and Sag, 1994]). These structures consist of name-value-pairs. In the simplest case, the feature values can be strings, numbers or other atomic values. But the values can also be whole feature structures. In the following example

```
(:TIME
    (:SECOND . "???")
    (:MINUTE . 35)
    (:HOUR . 10)
    (:DAY . 9)
    (:MONTH . "september")
    (:YEAR . "???")
    (:TIMEZONE . "???")
    (:TYPE . :TIME)
)
```

the feature with name :TIME of type 'time' ((:TYPE . :TIME)) has a feature structure ((:SECOND . "???") … (:TIMEZONE . "???")) as its value. In contrast, the feature with name :HOUR has an atomic value (10). Unknown values are represented by "???". Feature values can be accessed through paths (e.g. :TIME|:MINUTE gives the '35' value).

The feature structures used form an inheritance hierarchy (see Fig. 2). This hierarchy describes completely all possible structures that the IE module can use and might instantiate. E.g. in Fig. 2 the type 'template' has two subtypes 'report' and 'order'. The report-type feature structure has various name-value-pairs, e.g. ':addressee *partner*'. The value *partner* is itself a feature structure.

The information extraction process by itself is realized with shallow algorithms. These are called *transducers*. Transducers are finite-state automaton that read from an input stream and write to an output stream. Theses automaton can be cascaded, i.e. that the output of one transducer is the input in another one. For example, the names of various locations are the input to a transducer, which constructs a feature structure that formalizes the recognized name of the location. This feature structure is than handed as an input to the transducer responsible for detecting e.g. goal expressions. During the syntactic analysis of the free-form reports with transducers more and more complex feature structures are constructed.

In SOKRATES we use as an IE tool the SMES system (see [Neumann, 2003]). This system contains a huge German lexicon and offers the possibility to create transducers. SMES is implemented in Allegro Common Lisp (see [ACL, 2002]) and therefore offers also the whole functionality of Common Lisp as well. The following example shows a simple transducer for detecting who reports, when and where:

```
(compile-regexp
'(:conc
        (:current-pos start)
        (:seek so-date-time date)
        (:seek so-meldender meldender)
        (:seek so-standort-meldender standort)
        (:morphix-punctuation ":")
        (:current-pos end)
)
:debug *debug*
:register-types '((:register start date meldender standort end))
:output-desc
'(:lisp (multi-acons :speaker meldender :time date :location standort))
:prefix T
:suffix nil
:name 'so-meldung-prolog
:compile *compile*
:write-to-file *trace-file*
)
```

The name of this transducer is 'so-meldung-prolog' (:name 'so-meldung-prolog). It consists of a concatenation (:conc) of calls to other transducers (e.g. (:seek so-date-time date)) and a call to the morphological component ((:morphix-punctuation ":")). If the called transducers recognize successfully the appropriate parts in the report, they construct a feature structure describing their recognition result. This result is than handed over to the calling transducer, e.g. in variable date of the calling statement (:seek so-date-time date). The shown transducer uses the recognition result to construct its own feature structure (:output-desc '(:lisp (multi-acons

:speaker meldender :time date :location standort))) which is than passed up to the calling transducer of the shown transducer.

In the current implementation the SOKRATES system is able to process simple reports about moving objects. One example is the following report: "*09. September 10.35 Uhr von 6./PzMrs332-Zug B in Vinstedt: 18 Fahrzeuge marschieren bei Straßenkreuzung Kr3 nördlich Eppensen nach Ebstorf.*"

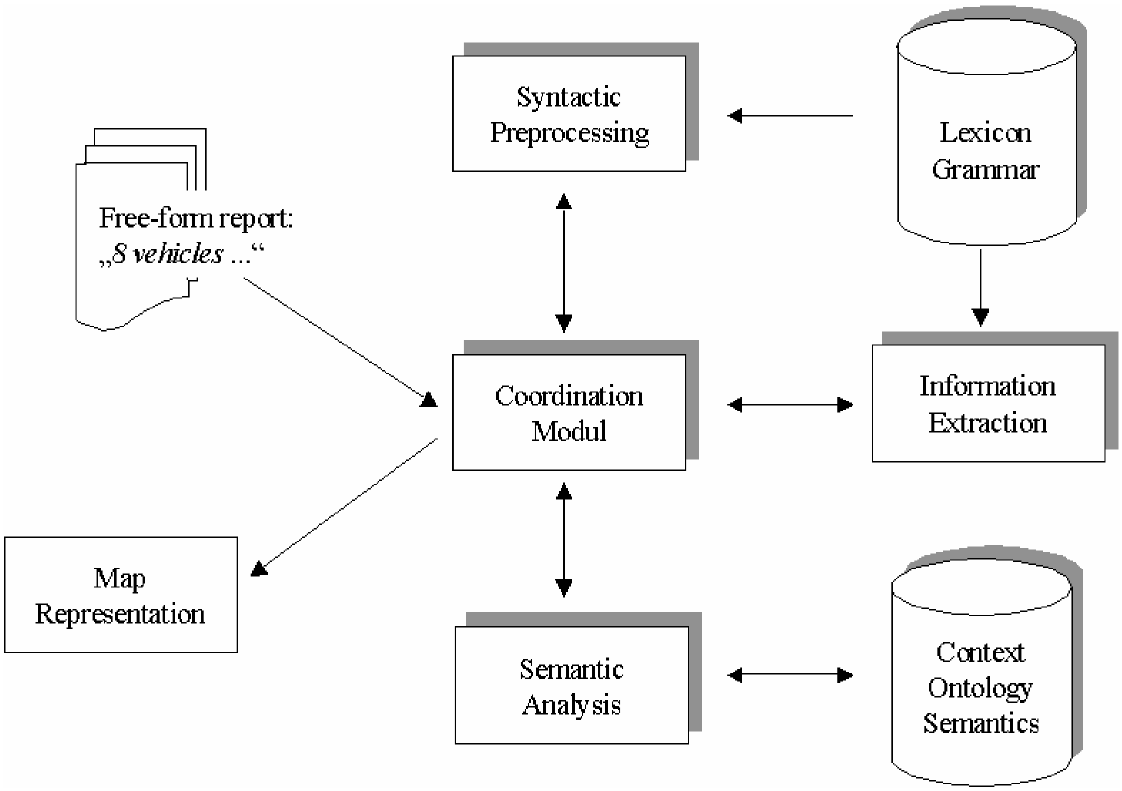


Figure 1: The Architecture of SOKRATES.

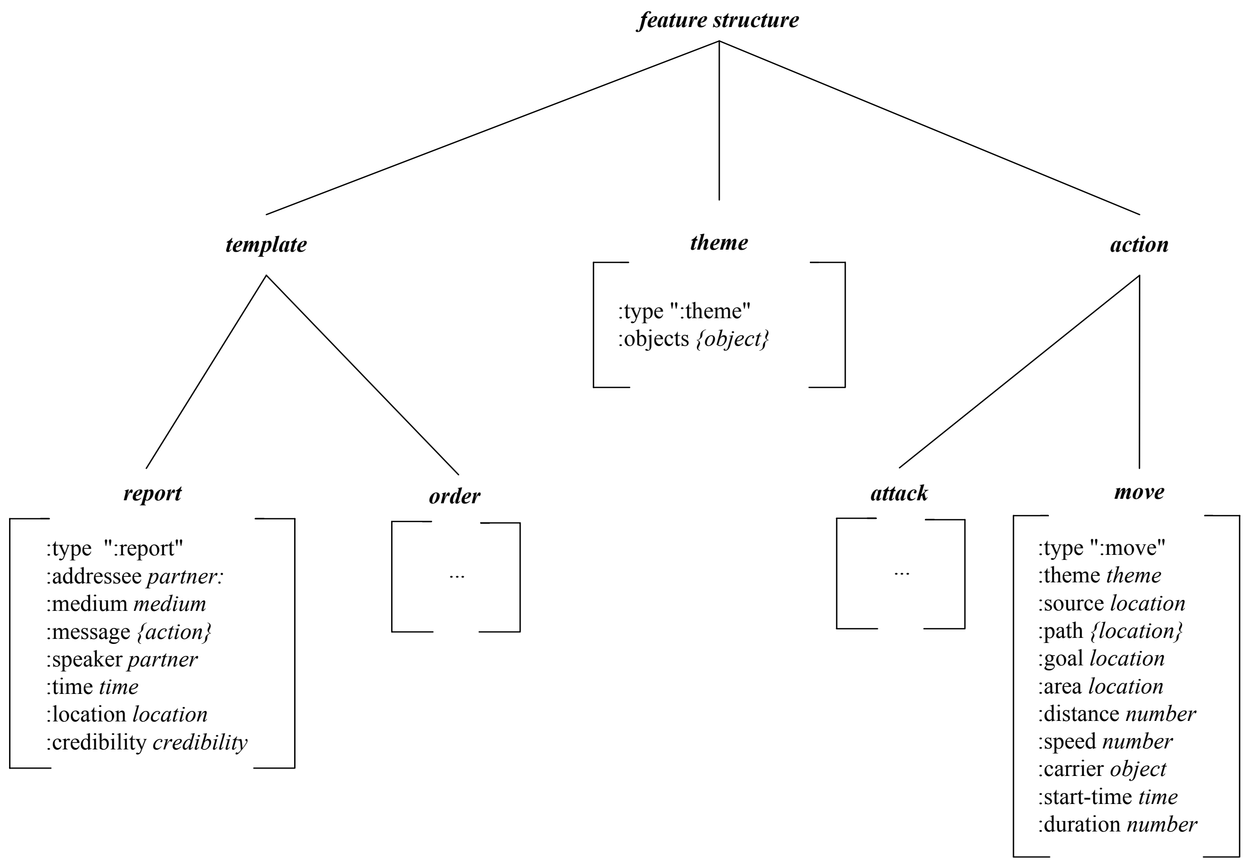


Figure 2: A Part of the Feature Structure Hierarchy.

(September 9th, 10.35 a.m. from 6./PzMrs332-Zug B in Vinstedt: 18 vehicles march at road crossing Kr3 north of Eppensen to Ebstorf.). After processing this report the result is represented as the following typed feature structure:

```
(
        (:CREDIBILITY . "???")
        (:LOCATION
                (:NAME . "vinstedt")
                (:TYPE . :LOCATION)
        )
        (:TIME
                (:SECOND . "???")
                (:MINUTE . 35)
                (:HOUR . 10)
                (:DAY . 9)
                (:MONTH . "september")
                (:YEAR . "???")
                (:TIMEZONE . "???")
                (:TYPE . :TIME)
        )
```

```
(:SPEAKER
        (:NAME . "6/pzmrs/332/zug/b")
        (:TYPE . :UNIT)
)
(:MESSAGE
        (:SET
                (
                        (:DURATION . "???")
                        (:START-TIME . "???")
                        (:CARRIER . "???")
                        (:SPEED . "???")
                        (:DISTANCE . "???")
                        (:AREA . "???")
                        (:GOAL
                                (:QUALIFIER . :TO)
                                (:NAME . "ebstorf")
                                (:TYPE . :LOCATION)
                        )
                        (:PATH . "???")
                        (:SOURCE
                                (:QUALIFIERS
                                        (:SET
                                                (
                                                        (:QUALIFIER . "nördlich")
                                                        (:NAME . "eppensen")
                                                        (:TYPE . :LOCATION)
                                                )
                                        )
                                )
                                (:COORDINATES . "kr3")
                                (:QUALIFIER . :EXACTLY-AT)
                                (:NAME . "Straßenkreuzung")
                                (:TYPE . :LOCATION)
                        )
                        (:THEME
                                (:OBJECTS
                                        (:SET
                                                (
                                                        (:COUNT . 18)
                                                        (:NAME . "Fahrzeuge")
                                                        (:TYPE . :VEHICLE)
                                                )
                                        )
                                )
                                (:TYPE . :THEME)
                        )
                        (:TYPE . :MOVE)
                )
```

```
            )
        )
        (:MEDIUM . :LETTER)
        (:ADDRESSEE . "???")
        (:TYPE . :REPORT)
)
```

The above feature structure is of type 'report'. Each report might contain information about the addressee of the report (:ADDRESSEE), the medium in which it was formulated (in this case :LETTER), the message itself (:MESSAGE), the unit or person who sends the report (:SPEAKER), the time of reporting (:TIME), the location of the unit or person who reports (:LOCATION) and the credibility of the unit or person (:CREDIBILITY). If the IE module doesn't find the appropriate information the string "???" is delivered. The message contains a set (:SET) of action descriptions. In the example it is a move-action (:MOVE). Words like "marschieren", "fahren", "schwimmen", "fliegen" (to march, to drive, to swim, to fly) are mapped to the move-action. The description of the move-action contains various features: the duration (:DURATION), the start-time (:START-TIME), the carrier (:CARRIER), the speed (:SPEED), the distance (:DISTANCE), the area (:AREA), the goal (:GOAL), the path (:PATH), the source (:SOURCE) and the theme (:THEME) of the action. The goal of the move is described with the help of a feature structure of type :LOCATION. It formalizes in our example the city Ebstorf as the goal of the march. The starting point of the move-action is given after the feature name :SOURCE. This is also a feature structure of type :LOCATION. It gives the coordinates ((:COORDINATES . "kr3")) of the crossing ((:NAME . "Straßenkreuzung")) and it gives also a qualifying statement that the crossing is in the north ((:QUALIFIER . "nördlich")) of the city Eppensen ((:NAME . "eppensen")). The theme describes which objects (:OBJECTS) are moving. In our example 18 vehicles ((:COUNT . 18)(:NAME . "Fahrzeuge")(:TYPE . :VEHICLE)) are moving.

The SOKRATES system is implemented and is able to process simple examples as shown above. Up to now, the lexicon was only extended with a few military specific words (but it contains already more then 120,000 German word stems). The next steps will be the extension of each module to enhance the processing capabilities.

CONCLUSION

In this paper, we introduced in the promising field of information extraction and we gave a description of the shallow parsing techniques used in our project SOKRATES. After presenting the overall architecture of our system, we have shown how the syntactic analysis is done with transducers and how feature structures are used to describe and to store potential analysis results. We described how a simple German battlefield report was analyzed and how the analysis result was represented in a feature structure.

REFERENCES

[ACL, 2002] Allegro Common Lisp. Franz Inc., 2002, http://www.franz.com.

[Appelt, 1999] Appelt, D. & Israel, D. *Introduction to Information Extraction Technology*. Stockholm: IJCAI-99 Tutorial, 1999, http://www.ai.sri.com/~appelt/ie-tutorial/.

[CommandTalk] *CommandTalk*. SRI International, http://www.ai.sri.com/natural-language/projects/arpa-sls/commandtalk.html.

[Hecking, 2001] Hecking, M. *Natural Language Access for C2 Systems*. Paper presented at the RTO IST Symposium on "Information Management Challenges in Achieving Coalition Interoperability", held in Quebec, Canada, 28-30 May 2001, and published in RTO MP-064.

[Hecking, 2002] Hecking, M. *Analysis of Spoken Input to C2 Systems*. In: Proceedings of the 7th International C2 Research and Technology Symposium (ICCRTS), Québec City, Canada, 2002.

[Moore et al., 1997] Moore, R. et al. CommandTalk: A Spoken-Language Interface for Battlefield Simulations. In: Proc. of the 5th Conf. on Applied Natural Language Processing, Washington, DC, pp. 1-7, ACL, 1997.

[NATO, 2000] NATO. *The Land C2 Information Exchange Data Model*. AdatP-32 Edition 2.0, 31 March 2000.

[Neumann, 2003] Neumann, G. http://www.dfki.de/~neumann/pd-smes/pd-smes.html, 2003.

[Pazienza, 1999] Pazienza, M. T. (ed.) *Information Extraction*. Berlin, 1999.

[Phraselator, 2003] *The Phraselator*. http://www.phraselator.com/products.htm, 2003.

[Pollard and Sag, 1994] Pollard, C. & Sag, I. A. Head-Driven Phrase Structure Grammar. The University of Chicago Press, Chicago, London, 1994.

[Schade, 2003a] Schade, U. *Ontologieentwicklung für Heeresanwendungen*. Forschungsgesellschaft für Angewandte Naturwissenschaften e.V. (FGAN), FKIE-Bericht, 2003.

[Schade, 2003b] Schade, U. *Towards an Ontology for Army Battle C2 Systems*. In: Proceeding of the 8th ICCRTS, 2003.

[Steeneken, 1996] Steeneken, H. J. M. Potentials of Speech and Language Technology Systems for Military Use: an Application and Technology Oriented Survey. NATO, Technical Report, AC/243(Panel 3)TP/21, 1996.

[Stent *et al.*, 1999] Stent, A. et al. *The CommandTalk Spoken Dialogue System*. In: Proc. of the 37th Annual Meeting of the ACL, pp. 183-190, University of Maryland, College Park, MD, ACL, 1999.

The Case for Using Semantic Nets as a Convergence Format for Symbolic Information Fusion

Jean Rohmer
Thales Communications
160, Boulevard de Valmy
BP 82
92704 Colombes Cedex
FRANCE

jean.rohmer@fr.thalesgroup.com

ABSTRACT

We explore how various information formats can be merged into an unique semantic space, using the Semantic Nets formalism. We show that this formalism can then be transformed and reworked to perform classical data analysis computations, which will help in the fusion and discovery process. We advocate for using semantic nets to get sense from heterogeneous informations, in particular texts, as a step towards what could be called "Litteratus Calculus".

Keywords: *Symbolic Information Fusion, Semantic Nets, Data Analysis, Text Mining, Text Understanding, Automatic Classification.*

1.0 INTRODUCING SEMANTIC NETS AS A WAY TO COLLECT HETEROGENEOUS INFORMATIONS INTO AN UNIQUE FORMAT

A key issue in information fusion is to deal with very different natures of data : numerical data, usually in the form of simple tables, more complex structured data like relational databases, semi-structured messages, totally unstructured texts. Moreover, efforts in information standardization recently introduced new formats like XML and its many derivates, TOPIC MAPS, UML models, ontologies models.

We have to deal with an impressive continuum of representations, from fully numeric and structured to totally textual and unstructured.

Solving this situation of heterogeneity is a prerequisite to information fusion processes and algorithms.

In general, Intelligence Information System designers are facing a difficult choice :

- either adopt a structured approach, e.g . choose to unify their data in a large relational database
- or, in the opposite direction, keep all the information under the form of documents

In practice, each model excludes the other one : in the first case, information will be accessed through structured query languages, along with programming of specific applications to interface the user with data.

In the second case, only text search engines are available to retrieve documents containing the desired piece of information. Attempts to put together relational and textual paradigms usually lead to costly and uncomfortable designs.

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

However, in an human brain, this distinction between structured and unstructured data simply does not exist : we, as humans, are able to merge informations coming from a newspaper, an Excel file, a database, an oral conversation …

How can computers mimic our extraordinary capability to make information fusion in our brain ? The solution is to represent information in the machines in a way not too far from the way it may be represented in our heads. This subject has been studied for years in the field of « Artificial Intelligence », and, as early as in the 50's, came the concept of *Semantic Nets* to meet this challenge.

Although in the 80's Artificial Intelligence applications were disappointing to a point that this field of computer science was nearly abandoned in the 90's, we have made recently the proof that Semantic Nets, the representation side of AI –as opposed to its automatic reasoning side- was both an extremely efficient and human friendly way of representing complex informations.

We started developing and using in the early 90's a tool dedicated to the management of Semantic Nets, IDELIANCE. Today we can confirm that Semantic Nets is a practical and efficient way of handling heterogeneous information sources.

Ideliance was originally designed as a personal knowledge management tool. The initial idea is to offer an information representation model which bridges the gap between structured data (like tables in spreadsheets and relational databases) and unstructured data (found in documents written in natural language). Semantic Nets appear to be a nice compromise between data and texts.

They can be viewed as a collection of simple sentences « Subject / Verb / Object » :

Peter / works for / Mary
Mary / lives in / Berlin

A key property in Ideliance is that each sentence is represented in both directions :

Mary / employs / Peter
Berlin / is the place where lives / Mary

Subjects can also be long sentences identifying a complex but precise concept :

The 23rd March 2003 ACM Meeting in Berlin about the ALPHA project

With Semantic Nets, a solution is to forge as many S / V / C sentences as necessary :

Berlin / is the place of / The 23rd March 2003 ACM Meeting in Berlin about the ALPHA project

ACM / is the organiser of / The 23rd March 2003 ACM Meeting in Berlin about the ALPHA project

ALPHA Project / is addressed at / The 23rd March 2003 ACM Meeting in Berlin about the ALPHA project

…

Other pieces of knowledge may be expressed this way :

German chapter of ACM / is located in / Berlin

German chapter of ACM / is member of / ACM

Note that *ACM* , *German Chapter of ACM* and *ALPHA Project* are themselves *Subjects.*

In an Ideliance collection, dizains of thousands of subjects can be found. By constrat, verbs will generally amount only to few to many dizains. They represent the vocabulary which describes the domain of the application.

Interestingly, it becomes easy to know « everything about Berlin » :

Berlin / is the place of / The 23rd March 2003 ACM Meeting in Berlin about the ALPHA project

Berlin / is the place where lives / Mary

Berlin / is the place of / German Chapter of ACM

(Each of these sentences may come from a different source and / or format : text, database, message, …).

When clicking on a given subject, a page is built with all the sentences starting with it. Navigation continues by clicking on one of the subjects at the end of these sentences.

Ideliance can be seen as a general purpose tool for managing such sets of sentences, applicable in many real life situations by non specialists in computer programming. It exists under the form of a personal tool on Windows, and of an HTTP server. Ideliance is in operation for more than three years in various application contexts : military intelligence, knowledge management, competitive intelligence, experience sharing among teams.

Users can edit new sentences through graphical editors, either by reusing the existing vocabulary, or by creating new subjects or verbs. Statements can also be obtained by *automatic translation of structured data* (Excel, SQL, XML …) into sentences. Text mining tools outputs can also be translated into Ideliance sentences.

N.B. The old concept of semantic nets has recently been found in knowledge representation tools developped in the context of Internet. Formalisms for ontologies representation (the coming W3C OWL standard), and more general information representation (the W3C RDF standard notation, based upon XML) are proposed by the Internet community under the general *Semantic Web* umbrella, with the vision that, in the future, information on the Web should be written in such a formalism rather than in textual pages. A tool like Ideliance, dedicated to Semantic Nets management, can be seen as « Semantic Web *avant la lettre* », and also as a practical tool to run dedicated « Semantic Intranets » without waiting for the hypothetical rise of the Global Semantic Net.

2.0 FROM FORMAT FUSION TO INTELLIGENCE FUSION

We can now address the core topic of this paper : once heterogeneous data (from documents, databases, messages, spreadsheets, …) have been gathered in a Semantic Net format, how to process this net to achieve *fusion* ?

Vocabulary remark : for most information technology people, converting various formats and databases into an unique format, is itself called « fusion » : we started from five databases, we end up with just one. We call this kind of fusion « **Format Fusion** ».

For Intelligence people, information fusion (we will call it **Intelligence Fusion**) is a totally different concept : we receive 20 documents or messages describing riots in a town at a given date, we ask ourself the following questions :

- are all these messages about the same unique riot, or do they report about several ones ?
- if there is an unique riot, how to identify and characterise the elements which describe it : number of participants, mode of action, consequences on the population, damages caused … Again, several messages will deal with damages : are they referring to the same damage, or to several ones ?

- if there are several riots, how to distinguish them –i.e. how to refer to them, to name them ? Some papers may be about only one of theses riots, other papers will deal with several ones. And the question of delineating each of the characteristics of each riot raises as well.

Stated like this, **Intelligence Fusion** appears to be much more complex than **Format Fusion**. However, the better the Format Fusion process will be conducted, the better the Intelligence Fusion process will start. In one hand, Format Fusion is a prequisite to deploy automated, computerized procedures for Intelligence Fusion. In another hand, Format Fusion can already help « manual », « human » Intelligence Fusion, simply by offering an unified, seamless way to navigate, browse through the whole set of informations, collected in a unique semantic net.

It is clear that Intelligence Fusion, in general, is extremely complex and difficult.

It can be tempting to use « brute force » to solve it : start from messages texts, do some form of terminology extraction, (text mining), then run statistical tools or « business intelligence » tools. We then face many problems, among which :

- how to differentiate by statistics 20 messages about 2 different riots, or 15 messages about 10 different riots
- if one message mentions 25 casualties, another one 45 casualties, are they figures about two diffrent riots, or about the same one ? In this case, what to do with these two figures : take the minimum, the maximum, take their sum, their average ?
- when a riot is mentioned in a message, does it concern a new incoming event, or is it a reference to a past event ?
- how to take into account the bias followed by the authors of the messages ?

These huge problems may advocate for the need of a fine grain analysis of the natural language used in the documents, including tenses, conditional modes, *nuances,* ...Unfortunately, current state of the art in natural language understanding and interpretation is far behind what is needed here.

Our experience suggest the following steps:

a) Starting from documents, use text minig tools and terminology extraction tools to *prepare the documents for semantic modelling*

b) Translate – mainly « by hand », with the help of tools like Ideliance – the preprocessed documents into an unique semantic net. This realises **Format Fusion**.

c) Perform **Intelligence Fusion** – both manually and automatically – on the unique resulting semantic net

N.B : Automatic translation of structured informations into Semantic Nets is not difficult, since, by definition, the semantics of structured data is known with precision. (Ideliance, for instance offers severals tools to automatically translate spreadsheets and relational databases into semantic nets). It is thus easy to inject structured information (e.g. about geography, weapons) into the semantic net.

3.0 SOME BASIC MECHANISMS FOR INTELLIGENCE FUSION IN A SEMANTIC NET

3.1 How to Formalise the Fusion Problem

We consider now that we start with a set of informations represented in a semantic net. We consider that the Intelligence Fusion process has not yet been processed.

That means for instance that, if we, at the beginning started with :

- message 1, mentioning an event with one person and one car
- message 2, mentioning an event with two persons and two cars

we have created the following subjects :

Event 1, Event 2, Person 1, Person 2, Person 3, Car 1, Car 2

Some sentences are also created, such as :

Person 1 / is mentioned in / Event 1

Car 1 / is mentioned in / Event 1

Person 2 / is mentioned in / Event 2

Person 3 / is mentioned in / Event 2

Car 2 / is mentioned in / Event 2

At this point, it is important to note that :

We do not know if Event 1 and Event 2 are the same or not, (idem for Car 1 and Car 2)

We know that Person 2 is different from Person 3, but each of them may be the same as Person 1

Formally, we can see each subject in our semantic network as a ***variable***, along with constraints (equations, inequations) about theses variables, and we can represent these constraints themselves as sentences in the network :

Person 2 / is different from / Person 3

Event 1 / may be equal to / Event 2

Car 1 / may be equal to / Car 2

Person 1 / may be equal to / Person 2

Person 1 / may be equal to / Person 3

The objective of what we call Intelligence Fusion is to reduce uncertainty, i-e :

- to conclude that some variables are the same
- to conclude that some variables are not the same

As for any system of equations, we need some ***constants*** to ground the system, and to bootstrap the solving process.

Event 1 / takes place / Avenue des Champs Elysées

Event 2 / takes place / Rue de Rivoli

This could lead to the conclusion that :

Event 1 / is different from / Event 2

But certainly not that *Car 1 / is different from / Car 2 !*

3.2 Similarities and Differences, Identities and Distances

We are here in the domain of symbols (a street name, a person name) rather than the domain of numbers (a speed, a pressure, a geometric position).

We have to deal with similarities and differences in a symbolic, discrete world, not in a numerical, continuous world.

Whereas in the latter case, the key point is the –continuous- notion of *distance*, in our case, the notion of ***identity*** prevails.

« Rue de la Paix » is not identical to « Place Charles de Gaulle »

« Rue de la Paix » is identical (only) to … « Rue de la Paix »

In other words, we advocate here, -due to the complexity of the problem- ***not to*** try to transform the symbolic, discrete world, into a continuous world (through fuzzy sets, bayesian networks, possibilities …). There are already enough progresses to do to adress Intelligence Fusion in a discrete symbolic world. (We conjecture that human judgments and decisions –the ultimate goal of Intelligence Fusion output—are more discrete than continuous : « I choose A against B », « I think that James is a nice guy ».

And discrete symbols are capable of describing details and *nuances* :

Given the sentences :

Event 1 / takes place / Rue de Rivoli

Event 2 / takes place / Place Charles de Gaulle

We can add that :

Rue de Rivoli / is located in / Paris 2ème

Avenue des Champs Elysées / is located in / Paris 8ème

Paris 8ème / is member of / Paris Luxury Districts

Paris 1er / is member of / Paris Luxury Districts

We see that ***Event 1 and Event 2 share a common point :***

They both happened in streets belonging to a district among the Paris Luxury Districts

More precisely, we say that two subjects SA and SB have a point in common if there exist sequances of sentences of the form :

(SA / **V0** / SA1) (SA1 / **V1** / SA2) (SA2 / **V2** / SA3) … (SAn **VN** **S**)

(SB / **V0** / SB1) (SB1 / **V1** / SB2) (SB2 / **V2** / SB3) … (SBn **VN** **S**)

(where all SA, SAi are different, and all SB, SBi are different : no loops in the sequence)

We will say that SA and SB have in common the « ***generalised attribute*** »

V0 – V1 – V2 … VN S

In the previous example, Event 1 and Event 2 have in common the *generalised attribute* :

takes place – is located in – is member of Paris Luxury Districts

This attribute is made with three sentences. The simplest attributes are made with one sentence, like :

lives in Berlin

Given a Semantic Net, we will compute the set of all the generalised attributes which are common to at least two subjects. (This set is finite –no loops).

Now, to each subject, we can associate its generalised attributes.

We can build a matrix SA, such that

SA(i,j) = 1 if subject i has attribute j

SA(i,j) = 0 otherwise.

We have finally transformed our complex semantic networked world into a simple binary matrix.

On this binary matrix, we can now -and with more reasons than on the initial texts- apply "brute force" numerical, continuous processes like statistics and data analysis:

- **compute the distance between two subjects** as a function of their shared generalised attributes. There are in the litterature dizains of proposed distances between two objects sharing boolean properties. For instance some of them take into account the frequency of the attributes: two subjects are closer to each other if they share a scarce atttribute rather than a frequent one.
- **build clusters of subjects,** putting together in the same class subjects sharing enough generalised attributes

This latter process gives strong guidance in the fusion decision process:

- subjects found in the same cluster will be candidates to be merged in an unique one
- subjects found in differents clusters will be candidates to be considered as distinct ones

More subtle decisions will be taken by considering point to point distances between two subjects, and, ultimately, by inspecting the very list of shared and non shared generalised attributes between them.

The chain of processes:

Semantic Net → Generalised Attributes Matrix → Distances and Clusters → Fusion Decision

has the advantage of being very systematic, and to combine two modes:

- an *automatic mode* : compute all attributes, distances and clusters
- a *human mode*: visualise the resulting topology and take fusion decisions

In general, the process will be iterative:

a) a step of initial subjects identification and fusion (as being same or different)

b) computation of the new set of generalised attributes after subjects fusion, yielding a new set of shared attributes

c) new evaluation of distances and clusters

d) new fusion decisions

e) iteration on step a)

The process stops at step d) when no new fusion decisions can be taken.

4.0 IDENTIFICATION OF MORE COMPLEX PHENOMENA

In the previous paragraphs, we addressed the problem of identifying individual subjects: an event, a car, a person.

In the real world, more complex entities exist, from concrete ones to abstract ones :

- groups of people (a terrorist group, a sport club)
- groups of groups (a federation of sport clubs)
- ideologies (the british neo-liberalism)
- phenomenas: "the rise of religious confrontations in the South suburbs of Cairo"

How far is Intelligence Fusion concerned by such concepts ?

It is for instance important to discover that Event 1 is a symptom of Phenomenon A (religious confrontations) and that Event 2 is a symptom of Phenomenon B (political rivalry), event if Event 1 and 2 have many attributes in common.

We will illustrate how to discover complex entities with our approach through an example. We consider a semantic net which contains two categories of subjects:

Persons and *Meetings*

The sentences in the semantic net are of the form :

Person P / present at / meeting M

Meeting M / attended by / Person P

(These subjects may have been identified using the fusion steps explained in the previous paragraphs).

We would like now to discover the possible existence of *groups of people*, of *different kinds of meetings, links between people, between meetings* ...

Following our definitions, a group of people is a set of subjects of the Persons category which share a significant set of generalised attributes. Let us look at the possible forms of the generalised attributes:

a) present at Meeting M
b) attended by Person P
c) present at -- attended by Person P
d) attended by -- present at Meeting M
e) present at -- attended by --present at Meeting M
f) attended by -- present at -- attended by Person P
g) present at -- attended by -- present at -- attended by Person P
etc …

Attribute c) means for instance that two persons have in common to attend different meetings, but where the same person P is present.

We call this way of transforming and procession a Semantic Net *"Litteratus Calculus",* to suggest that , in parallel with *scientific calculus* on technical data, a lot of useful computations can be made at a fine grain on data from textual origin.

Imagine now the following situation:

A small political organisation wants to infiltrate large meetings. This organisation is made of cells of a limited number of agents:

Cell A with agents Agent A1, A2, A3
Cell B with agents B1, B2, B3, B4
Each cell responds to a leader (Leader A, Leader B), not member of the cell.
Agents of the same cell are in general present at the same meetings.
There are also Cell Meetings with their Agents and Leaders
A meeting is attended by many Persons.

We have made experiments with Ideliance with simulated data to describe such a situation.

The *only input* of the fusion process is a Semantic Net of sentences like:

Person X / present at / Meeting Y

and indeed no prior knowledge about the structures of groups or roles of persons is available.

The computation of clusters of persons will give the expected results:

Cell A and Cell B will be identified as clusters containing the agents, because, among others and for example :

Agents B1,B2,B3 share the attribute of type c) :

present at -- attended by Agent B4

Agents B1,B2,B4 share the attribute of type c) :

present at -- attended by Agent B3

etc …

We see that a *mesh of relations* links the members of a Cell, and identify it as an interesting result.

Once a cluster is found, the system exhibits the attributes which are shared by most of its members. Thus the result is quite expressive:

Persons B1, …., B4 form a cluster which has in common to be present at meetings with the other members of the cluster.

(theses results are obtained through fixing some threshold to determine how tolerant we want to be on the homogeneity of the groups, and the algorithm works on non perfect situations: not all members of a Cell need to be present at all their meetings …)

Finally, we have discovered several concepts:

- existence of groups of persons, an interesting seed to discover the structure of organisations
- notion of roles of persons
- we can know « which meetings are infiltrated by which cell »

Similarly, we will discover the existence of Cell meetings, and of Cell Leaders: all members of a cell will have in common to be present in meetings with their leader.

Futher steps of analysis could be found:

If we examine ordinary attendees, sympatisers of the infiltrators will be found in clusters which share attributes of type g):

present at -- attended by --present at -- attended by leader A

present at -- attended by --present at -- attended by leader B

In other words: sympatisers of a given organisation often attend meetings where the -masked- infiltrators reporting to the leaders of this organisation are present.

Leaders A and B will themselves be found in the same cluster, characterised by shared attributes of the form

present at -- attended by --present at -- attended by Sympatiser X

This "Leaders cluster" represents, embodies the concept of their organisation.

5.0 CONCLUSION

Our first experiments with Ideliance tend to prove that Semantic Nets can play an important role in Symbolic Intelligence Fusion.

Transforming and merging heterogeneous informations from various formats (databases, tables, messages, texts) into an unique format (what we called *Format Fusion*) is a good basis for *Intelligence Fusion*.

First, it offers an efficient support for "manual" seamless inspection and navigation of the whole set of informations.

Second, it becomes a material upon which powerful data analysis (distances and clusters computation) can be performed, once the *Generalised Attributes* we introduced are computed.

We call this process *"Litteratus Calculus"*.

Our conjecture is that the objects resulting from this analysis form the backbone of the Intelligence Fusion process, which, ultimately is the domain of human decision.

Analysis of Adaptive Data Fusion Approaches within LM Canada's Technology Demonstrator

Elisa SHAHBAZIAN, Louise BARIL, Guy MICHAUD, Eric MENARD, Daniel TURGEON
Lockheed Martin Canada
6111 Royalmount Avenue
Montreal QC H4P 1K6
CANADA
Tel : (514) 340-8343

Email : elisa.shahbazian@lmco.com / louise.baril@lmco.com / guy.michaud@lmco.com / eric.menard@lmco.com / daniel.turgeon@lmco.com

ABSTRACT

Lockheed Martin Canada (LM Canada) has developed a Technology Demonstration environment, which over the last decade has been used to demonstrate initial proof-of-concept and then analyse various approaches for enhancing overall Data Fusion system performance for applications to Canada's defence programs. It has a blackboard-based architecture that permits mix of rule-based and algorithmic approaches, specifically useful in the implementation of higher-level fusion capabilities. Various aspects of these efforts, such as fusion architectures, algorithms and information sharing strategies between multiple collaborating platforms have been presented previously [1,2].

This paper presents currently on-going efforts towards the analyses of concepts for level 4 fusion, i.e. methods for adapting the fusion processes based on the tactical and environmental factors.

1.0 INTRODUCTION

Over the last 13 years, LM Canada's Research and Development (R&D) department in collaboration with Defence Research and Development Canada (DRDC) has been developing data fusion capabilities in support of Canada's defence programs. The initial efforts started with the development of a data fusion capability that fused Above Water Warfare (AWW) onboard sensor data of the Halifax Class frigate [3]. Then Image Fusion capabilities were added to provide data fusion capability for airborne surveillance [4]. To ensure that data fusion developments aimed at different programs can leverage capabilities developed previously, a common Technology Demonstration environment was established where all data fusion techniques and capabilities are developed, evaluated and demonstrated using the same infrastructure. This Technology Demonstration infrastructure is built based on the distributed blackboard-based knowledge based system architecture (Cortex), developed at LM Canada [5], sponsored by DRDC. The blackboard-based architecture was chosen to support the rule-based, concurrent and ad hoc reasoning requirements of higher level fusion. The distributed nature of this architecture was necessary to be able to support various data fusion architectures in different programs as well as to support multi-platform collaboration in a task force and evolution to Network Centric Warfare (NCW). The data fusion methods, techniques and capabilities are being developed and integrated

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

into the Technology Demonstrator (TD) based on the priorities of the various programs that LM Canada becomes involved in, however an overall data fusion model that foresees the evolution into the human-in-the-loop, multi-platform data fusion of levels 1 through 4, supporting NCW has been developed. Currently a Multi-Platform Multi-Source Data Fusion (MP MSDF) – Level 1 Collaborative Data Fusion capabilities have been demonstrated. The TD also contains a subset Situation and Threat Assessment and Resource Management (STA/RM) capability – Level 2, 3 Fusion and RM decision support functionality, which is available on any platform , however which provides decision support for the platform alone (MP STA/RM capabilities are not yet developed). At its current state the TD's MP MSDF includes bearing-only association, track-to-track fusion, backwards data integration, bearing intersect fix management, etc. in an architecture permitting data exchange between collaborating platforms. Currently work is on-going to mature the MP MSDF and add more sophisticated fusion techniques, to incrementally add capabilities and evolve Level 2, 3 Fusion and RM, and to introduce Level 4 Fusion, specifically fusion management, i.e. an adaptive fusion capability.

The objective of this paper is to describe the current developments in fusion management and the envisaged evolution of Level 4 Fusion capabilities. To be able to show the path for the evolution of data fusion capabilities, specifically in the area of Level 4 Fusion, first the LM Canada Data Fusion Model is described below.

2.0 THE DATA FUSION MODEL

The LM Canada's data fusion model that is shown in Figure 1.

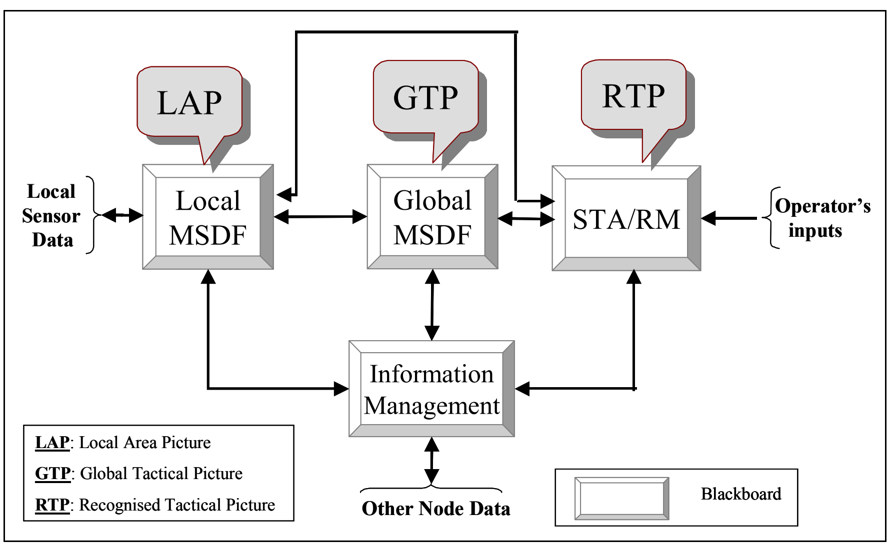


Figure 1: The Data Fusion Model for each Fusion Node on a Network.

This figure shows a set of processes and communication paths between these processes that can exist within a fusion node on a network. Depending on the Data Fusion requirements of a specific defence platform, e.g. naval, airborne, army units, land-based Command and Control Centres, etc., certain processes, and certain communications paths may not be required, or more that one such fusion nodes may be required, however we believe that when building a Data Fusion application it is necessary to examine the requirements for each process and communication path in this figure. A high level description of the processes and communication links is given below, with specific emphasis on those which play a role in the Level 4 Fusion processing.

2.1 Processes

There are 4 processes:

1) Local MSDF – Level 1 Fusion engine which fuses primary (only I/O to that node) data/information - and generates the Local Area Picture (LAP)

2) Global MSDF – Level 1 Fusion engine which fuses data/information from external (non primary, could be other fusion centres) sources with the LAP and generates the Global Tactical Picture (GTP)

3) STA/RM – Level 2,3, 4, and RM processes that provide decision support for:

   a) The interpretation of LAP and GTP – Level 2 and 3 DF processing,

   b) The MSDF process performance and refinement – Level 4 DF processing

   c) Sensor Management – Level 4 DF processing

   d) Weapons Management – RM processing

   The modifications of GTP based on STA/RM reasoning as well as operator refinements of the GTP are maintained in the Recognised Tactical Picture (RTP).

4) Information Management (IM) – A process that manages the information flow between the Fusion Node and the network

These processes are implemented as blackboards (or set of blackboards) in our TD. Although the Cortex architecture has been mainly developed to satisfy the requirements of STA/RM, all processes in the TD have been developed in Cortex to benefit from its modularity and the facility to perform concurrent development of different components in the system. Furthermore, Cortex permits easy breaking of an application into multiple parallel blackboards, or to combine all processes into one backboard, depending on the specific needs of the application, e.g. processor resource needs or information exchange needs between the processes Overall the analysis of information exchange needs is a very important aspect of the system design. The specific information interchange needs for Level 4 fusion have formed and will form a significant portion of analyses for the establishment of level 4 DF capabilities for the Canadian defence programs, and Cortex will facilitate experimentation with and validation of these capabilities.

2.2 Communication Links

The high level description of information interchange in the various communication links supporting the Level 4 DF processing are:

1) LAP estimates are made available to STA/RM for evaluation of Local MSDF performance

2) STA/RM recommendations are sent to Local MSDF to perform a number of actions to enhance Local MSDF performance including:

   a) Select an alternate association mechanism for a subset of observed targets

   b) Select an alternate filtering approach for a subset of observed targets

c) Modify MSDF parameters for a subset target processing

d) Select different association, filtering, or parametric modifications for data/information coming form a specific source

e) Recommend a Sensor Management action, e.g. provide sensor with target information to support its processing or request information of specific type, location, etc.

3) Local MSDF may also receive requests from other fusion nodes to perform a Sensor Management action through IM

4) GTP estimates are made available to STA/RM for evaluation of Global MSDF Performance

5) STA/RM recommendations are sent to Global MSDF to also enhance the Global MSDF performance. The type of recommendations regarding the Global MSDF algorithms and processing of input data would be the same as for Local MSDF algorithms (bullets a, b, c, d above), while recommendations for processes and information management in external Fusion nodes STA/RM could be send out either via the Global MSDF or directly to IM.

6) IM may also provide Global MSDF with recommendation from other fusion nodes about fusion processing performance (e.g. track quality issues, track number or ID conflicts can lead to changes in MSDF algorithm or parameter modifications.

The description of what already exists within LM Canada's technology demonstrator, what is currently being developed and what are the near term plans for developing, analysing and maturing a Level 4 Fusion capability in the context of NCW are described in the sections below.

3.0 THE TECHNOLOGY DEMONSTRATOR CURRENT DESIGN

Figure 2 shows LM Canada's technology demonstrator's current design.

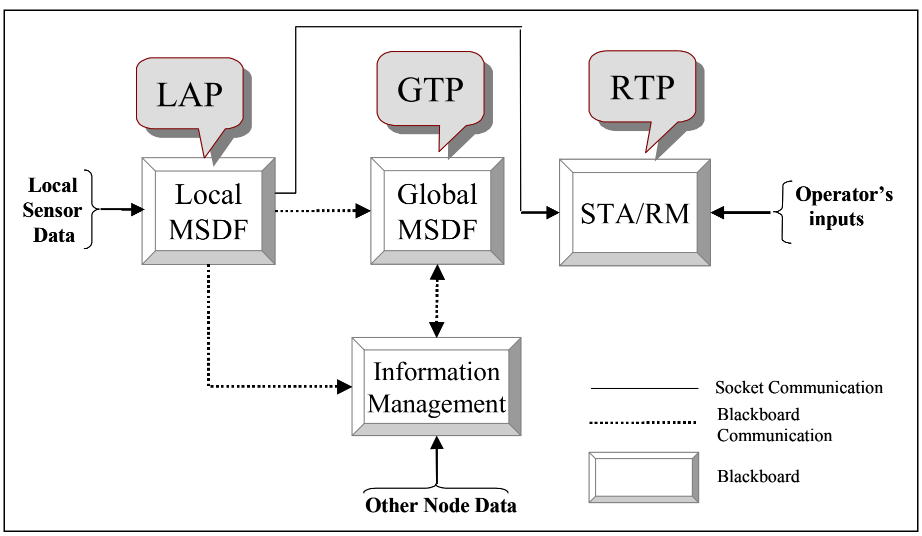


Figure 2: Technology Demonstrator's Current Design.

This design is consistent with the data fusion model described in Figure 1. In Figure 2 communication links are implemented in 2 different ways, through blackboard's intrinsic communications mechanism and through sockets. This design choice is application dependent, determined based on growth and other considerations. For Example in the case of STA/RM a socket communication was used because it is a process that is designed to run as an external process, to be able to integrate also with the company's Naval engineering prototype. The current Level 1 Fusion capabilities (Local and Global MSDF) are relatively mature. STA/RM is currently designed to only perform reasoning regarding the LAP, hence STA/RM has no need to communicate its recommendations on the network. LAP tracks are broadcast on the network for fusion in the Global MSDF of other nodes and GTP tracks can be sent to other nodes using either broadcast or point-to-point link, but the fusion of such data has not yet been fully analysed, in terms of data incest handling.

The current network architecture provides a consistent tactical picture. Each platform maintains parallel track databases. LAP contains the local track estimates produced by the fusion of the measurements created from the local sensor tracks. GTP contains the fused information of the local track databases of the host and remote platforms. RTP contains the reasoning results of the STA decision support tools and the operator inputs regarding the target identification and position refinements. Each one of the high level processes (Local MSDF, Global MSDF, STA/RM and IM) may need to maintain additional layers of information refinements and the blackboard architecture permits to add additional blackboards where the results of different layers of reasoning capabilities regarding the tactical situation are maintained and made available for further analysis if required. The philosophy has been to maintain the reasoning at different levels separate initially primarily to facilitate the handling of data incest issues when fusing data provided by collaborating platforms and in the long run to facilitate the analysis of information interchange requirements between the different reasoning levels. On the other hand the current architecture of the technology demonstrator does not yet support all communication links (information exchange interfaces shown in Figure 1 as well as communication between any additional layers of reasoning) that would be required to analyze this. Based on DND program priorities, these communication links as well as the capabilities in the processes to handle the data that is provided by these links are being implemented incrementally.

The on-going enhancements to the processes and communication links to be able to demonstrate some initial Level 4 Fusion capabilities as well as the on-going experimentation with Level 4 Fusion concepts are described below.

3.1 The On-Going Enhancements

As mentioned above the Level 4 Fusion is responsible for 2 areas:

1) Data Fusion process refinement
2) Sensor/Information source Management

The Level 4 fusion decision support for performing various actions for DF refinement and sensor management are part of (or additional reasoning layer on the) STA/RM process or operator actions.

The current developments are in the area of Data Fusion process refinement, namely which should be the criteria to initiate modification in the data fusion processes, and how to modify MSDF to be able handle processing of different sub-sets of targets using different MSDF methods.

The aspects of the fusion processes that are being considered for refinement currently include:

1) Choice of the association mechanism,

2) Choice of filtering techniques,
3) Modification of parameters within the algorithms for association and filtering.

One can observe from comparison of Figure 1 and Figure 2 that new communication links and processing are required to be implemented to be able to demonstrate and analyse Level 4 Fusion capabilities.

3.1.1 Process Modifications

<u>Addition of new approaches for association and fusion within the fusion centres</u>.

Currently there are 2 methods of association, nearest neighbour (NN) and Jonker Vongenant Castanon (JVC), which are single scan association methods. One multi-scan method, a multi-hypothesis (MH) association has also been added.

Currently there are also 2 methods for filtering, an adaptive Kalman filter and an Interactive Multiple Model (IMM) filter. Both of these filters do not address the data incest issues. Initially, when - mixing of local and global data were not allowed in the position fusion, i.e. local data overwrote global data, if available, these methods were sufficient, however an approach that deals with data incest is necessary, especially when information sharing between fusion centres is recommended by a level 4 fusion decision. Currently 2 types of tracklet fusion approaches have been added for this purpose, however this implies that the covariances are passed between platforms, which is not the case currently in datalink and other communication protocols.

<u>Addition of new STA capability.</u>

As mentioned above STA agents will be incrementally added to detect certain context and criteria that could be considered as basis for decision to modify fusion processes. As a starting point enhancements to the Human Computer Interface (HCI) have been made to be able to interactively input decisions about fusion or sensor management. A capability to select certain targets or sector where the operator can chose modifications to the fusion processes have been demonstrated to be very useful in analyzing Level 4 fusion impact on the overall level 1 fusion performance.

3.1.2 Communication Links Modifications

<u>Addition of communication link between Global MSDF and STA.</u>

A communication link to provide GTP to STA/RM is required and an analysis of information interchange requirements between LAP, GTP and STA to understand the two-way information flow between STA and both Local and Global MSDF is required.

Currently the STA reasoning has been developed mainly to understand the behaviour of some example STA tools and to analyse the human interface requirements for identification refinements for LAP. The impact of STA analyses on the local or global fusion results are only beginning to be a topic of analyses. The results of STA capabilities will provide context and criteria for level 4 fusion. These capabilities applied against either the LAP or GTP would support decision to modify the association and fusion processes of each fusion centre. Similarly STA capabilities can provide decision regarding sensor management for the local database and they can provide decision regarding information interchange between the collaborating platforms. We view this as a longer term research.

Addition of communication link between STA/RM and IM.

This link will permit analysis of information that could be shared between collaborating platforms and modification of interfaces between the platform's databases and the Information Management capability within the platform. As a longer term research, STA functionality will analyse the Global MSDF performance dependence on data from other nodes and recommend fusion modifications in other nodes, or will receive recommendation about the Global MSDF performance from other nodes. This can be classified as Force STA/RM functionality.

4.0 FUTURE PLANS

As in all LM Canada developments work priorities are dictated by the Canadian program priorities. The implementations done and on-going so far are specific bottom-up developments to demonstrate specific capabilities necessary in the current projects. In parallel top-down design activities are on going to analyse:

1) Data exchange requirements in each communication link
2) STA algorithms and heuristics to be used in decision support for adapting the fusion processes. The various context and criteria that could be considered as basis for decision to modify fusion processes include the mission, source reporting the target, the target density, identification, behaviour, geolocation, operator selection, etc. STA capabilities will be developed which will analyse both LAP and GTP based on the various criteria and recommend modifications in the Local and Global MSDF processes.
3) Metrics to quantify MSDF performance at run-time.
4) MSDF architecture and new algorithms in both Local and Global MSDF centres to deal with multiple fusion approaches being used to fuse different subsets of input information/data.

It is clear that the design to be developed here will continuously evolve, therefore the currently chosen system architecture based on Cortex and well documented and enforced message interchange standards. These will ensure that the evolution occurs with minimal re-work.

In the near term rule-based and algorithmic approaches are foreseen to select an alternate association or fusion mechanisms. Then an adaptive fusion approach under development at DRDC Valcartier, which proposes to modify parameters in MSDF based on some criteria, will be analyzed for integration.

Sensor/information source management capabilities can also be incrementally added. For example one of the most obvious cases of such an action is the detection of an ID conflict between the LAP and remote report. A number of sensor/Information source analyses can be initiated and the decision support can range from performing an IFF interrogation, activating an imaging sensor (if available), re-examining the local MSDF ID estimation and instructing a Participating Unit (PU) to do such actions.

5.0 CONCLUSIONS

LM Canada's Technology Demonstrator environment for the incremental development, analysis and demonstration of level 1 through 4 data fusion capabilities for Canada's defence programs was described. The path for specifically building Level 4 Fusion capability and currently on-going analyses and implementations was also shown. However it is understood that the problem of developing the appropriate level 1,2,3 and 4 fusion capabilities for a system, and especially a network of collaborating systems is extremely large. LM Canada's technology demonstrator has been used to incrementally add, analyse and demonstrate capabilities towards this goal and once some capability has been sufficiently investigated, it is prototyped and demonstrated on-board the Halifax Class frigates.

REFERENCES

[1] Jouan, A., Baril, L., Michaud, G., Shahbazian, E.,Distributed Data Fusion Architecture for the Evaluation of Information Sharing Strategies, SYSTEMS CONCEPTS AND INTEGRATION PANEL, the SCI-116 Symposium, Norfolk, Virginia, 21 to 23 October 2002

[2] Shahbazian, E., Building Fusion Systems, IDC'2002 - Information, Decision & Control (Cosponsored by ISIF, IEEE), February 11-13, 2002, Adelaide, Australia, Proceedings on CDROM.

[3] Bégin F., Boily E., Mignacca T., Shahbazian E. and Valin P., Architecture and Implementation of a Multi-Sensor Data Fusion Demonstration Model within the Real-time Combat System of the Canadian Patrol Frigate, AGARD symposium on Guidance and Control for Future Air-Defence Systems, Copenhagen, 17-20 May 1994, AGARD-CP-555, pp. 28.1-28.8.

[4] Shahbazian, E., Gagnon, L., Duquet, J.-R., Macieszczak, M. and Valin, P., Fusion of Imaging and Non-Imaging Data for Surveillance Aircraft, in Sensor Fusion: Architecture, Algorithms, and Applications, SPIE Aerosense 97, Orlando, 20-25 April 1997, Proc. Conf. 3067, pp. 179- 189.

[5] Bergeron, P., Couture J., Duquet, J.-R., Macieszczak, M., and Mayrand, M., A New Knowledge-Based System for the Study of Situation and Threat Assessment in the Context of Naval Warfare in FUSION 98, Las Vegas, 6-9 July 1998, Vol II, pp.926-933.

A Lessons Learned Knowledge Warehouse to Support the Army Knowledge Management Command-Centric

Dr. Pierrette Champoux, Martin Trudel M.Sc.
DMR Consulting
Place Iberville Trois 2960, boulevard Laurier, bureau 400
Sainte-Foy, QC G1V 4S1
CANADA
Phone: 1 (418) 653-6881
Fax: 1 (418) 653-4428
Email: {pierrette_champoux, martin_trudel}@dmr.ca

Gaétan Thibault M.Sc.
Defence R&D Canada - Valcartier
2459 Pie-XI North
Val-Belair, QC G3J 1X5
CANADA
Phone: 1 (418) 844-4000
Fax: 1 (418) 844 4538
Email: gaetan.thibault@drdc-rddc.gc.ca

ABSTRACT

The Canadian Army Lessons Learned Knowledge Warehouse (ALLKW) is at the heart of the environmental Knowledge Management (KM) strategy to support the Canadian Land Force Operational Model centered on Command. This paper presents the Army Lessons Learned System and Process (ALLP) that is part of the approach for gathering observations and comments from operations and exercises in order to support lesson elicitation, action identification as well as proper follow-up. Along with the description of the LL Processes, a focus will be made on how the current system could be improved in order to facilitate its integration within the Intelligence, Surveillance, Target Acquisition and Reconnaissance (ISTAR) Information-Centric Workspace Processes, in particular the interaction and use of such a system with the Data Fusion Processes and System.

1.0 INTRODUCTION

Over the years, the Army has faced numerous changes with regards to personnel and resources availability, aging and knowledgeable workforce reductions, re-engineering, and culture upheaval. The shift to more frequent, complex and demanding peace support missions and the perspective of a possible engagement in a war reinforces the needs for the Army to migrate to a '*…knowledge-based and command-centric institution capable of continuous adaptation and task tailoring across the spectrum of conflict…*'[1]

The Command function is foreseen as the centre of activities linking all other operational functions, Sense, Act, Sustain and Shield within a single multi-level operation concept designed to simultaneously achieve tactical, operational and strategic goals. [2] '*The command-centric view of knowledge management is integral to the command support capability. It focuses on the usefulness of implementing knowledge management processes, techniques and tools to support Army commanders during decision-making in an operational context.*' [3]

The Army Knowledge Management Strategy overall intent is the '*…optimal development of common/shared intent and improvement of the synchronization between commanders and staff…*' [3]. One way of improving this synchronization is by the transformation of key tacit knowledge into explicit knowledge, which will

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

increase the quality and availability of information in context (knowledge) to the decision maker in a timely manner.

One of the first Knowledge Management (KM) initiatives that was foreseen and undertaken by the Army is the development of the Lessons Learned Knowledge Warehouse (LLKW). The aim of the LLKW System is to support the Army throughout the entire process of gathering, organising, analysing, tracking action, and disseminating knowledge (e.g. lessons learned) related to field operations and exercises and, thus, improving effectiveness of the Army's combat capabilities. In the Army KM Command-Centric-View, the LLKW System is a supporting element to the Commander's decision (figure 1) [3].

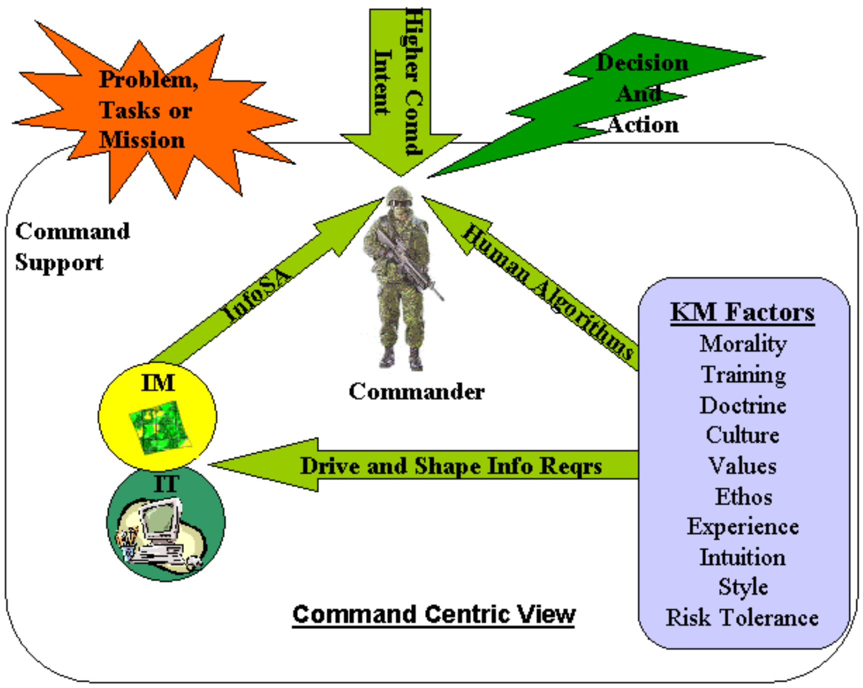


Figure 1: Army KM Command-Centric View [3].

Also, the LLKW is one of the systems integrated into the Intelligence, Surveillance, Target Acquisition and Reconnaissance (ISTAR) System of systems Information-Centric Workspace [4], whose aim is to provide a more coherent information management approach to better support the Commander. The Systems integration features numerous challenges and is highly influenced by the available technical skills and the ability to adapt processes, doctrine, organisation and culture. [5]

Data Fusion is one of the main processes in the ISTAR Information-Centric Workspace. From the initial Data Fusion Lexicon, produced by the Joint Directors of Laboratories (JDL) in 1987, Data Fusion is defined as:

> '*A process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats, and their significance. The process is*

> *characterized by continuous refinements of its estimates and assessments, and the evaluation of the need for additional sources, or modification of the process itself, to achieve improved results.'* [6]

This paper provides a description of the ALLKW development context to identify the scope of the projects that lead to its realisation. An introduction of the Data Fusion Process Model is made along with a presentation of possible interactions and similarities between LL and Data Fusion. Then, each phase of the LL process is described whereas potential improvements to the current system are identified with regards to the Data Fusion context.

2.0 ALLKW DEVELOPMENT CONTEXT

The creation, sharing and re-use of Lessons Learned lead the way among the best practices in any organisation [7], [8] and [9]. In particular, Lessons Learned (LL) have always been at the core of the fundamental knowledge assets in Command and Control environments. As a part of their traditional culture, the military personnel has reported observations or lessons after operations or exercises. Past operations lessons convey an important experiential knowledge (both about successes and failures) that can be learned, re-used or avoided in future similar situations in a way to make the best decisions and undertake the best actions (e.g. preparation of future operations, selecting best course of actions, etc.). The sound re-use of human and intellectual capital is the main constituent for the creation of a safe and efficient culture within the Army.

In 2001, the Canadian Forces had a requirement to provide the Army Lessons Learned Center with a new approach for the management and publication of the Army Lessons Learned. The past Lessons Learned Systems were based on technologies dating from the early 1980's which at the time relied little on IT applications while requiring a lot of human resources. This approach did not easily support the evolution of the Army requirements, nor the re-use of the knowledge conveyed in the Lessons Learned.

The main requirements for the ALLKW were as follow: [9]:

(1) Accessing the System with a Browser similar to Internet Explorer

(2) Facilitating or enabling the capture of observations and comments from the Chain of Command

(3) Accelerating updates

(4) Providing tools to assist the Lessons Learned Analysts in the search and validation of specific issues as well as the identification of lessons

(5) Facilitating action tracking

(6) Enabling fast publishing

For the ALLKW System, these requirements aim at supporting the Army throughout the entire process of gathering, organising, analysing, tracking action, and disseminating knowledge (e.g. lessons learned) related to field operations and exercises and, thus, improving effectiveness of the Army's combat capabilities.

An Interim Operational Capability (IOC) of an Interactive Lessons Learned Knowledge Warehouse (ILLKW) was developed in 2001 in collaboration with the Defence R&D Canada Valcartier. This prototype [10] was instrumental in presenting the system to the Army personnel and was utilized to clarify the functional requirements for the Final Operational Capability (FOC) in 2002. The main objective in the development of the FOC was to ensure that the LLKW would provide functionalities covering the entire LL Process with respect to budget, allowing the Army personnel to acknowledge the organisational changes to be pursued and capture future system requirements. The methodology used to develop and deploy the ALLKW follows the Canadian Forces requirement '*to develop a single methodology linking requirements, research, and acquisition through experimentation and fielding'*. [11]

The development approach was crucial in ensuring that the system would meet the Army requirements, in balance with the design of the entire LL Process, with the close participation of all stakeholders and end-users. The KM approach undertaken to develop the LLKW was a combination of both methods and techniques retrieved from the following three main concepts [12]:

- The User-Centric Approach which puts the emphasis on the tasks and the user early in the system development while measuring the reactions using prototypes, interfaces or any other means of simulation. The system development is realised using multiple iteration.
- The System-Centric Approach is mainly preoccupied by the flow of information and objects that are taken into account by the system while considering the user's perspective of validating and accepting the work or missions to be performed.
- The Process-Centric Approach considers the process from a centralised perspective, at the very core of the development. This process consists of a chain of activities that have an orderly priority in both space and time, with a beginning and an end for each activity as well as a triggering event.

The definition of the development context allowed for the identification of the scope and the extent of the current LLKW application. In order to facilitate the elicitation of new and future requirements, we must position the system in the context where it will interact and integrate with the other systems and their related processes, in this case the Data Fusion Process.

3.0 THE JDL DATA FUSION MODEL AND THE ALLKW

The Joint Directors of Laboratories (JDL) Data Fusion Group, established in 1976, has created and ever since maintained the Data Fusion Model used for categorising Data Fusion-related functions. [13]. This model was developed with the intent to facilitate the communications between military researchers and system developers. [14]

The top-level JDL Data Fusion Model gives a functional-oriented model, which was designed to be general but also suitable for different application domains. The top-level model (see figure 2) is composed of Sources of Information, Human Computer Interaction, Level 1 through 4 Processing and Database Management System.

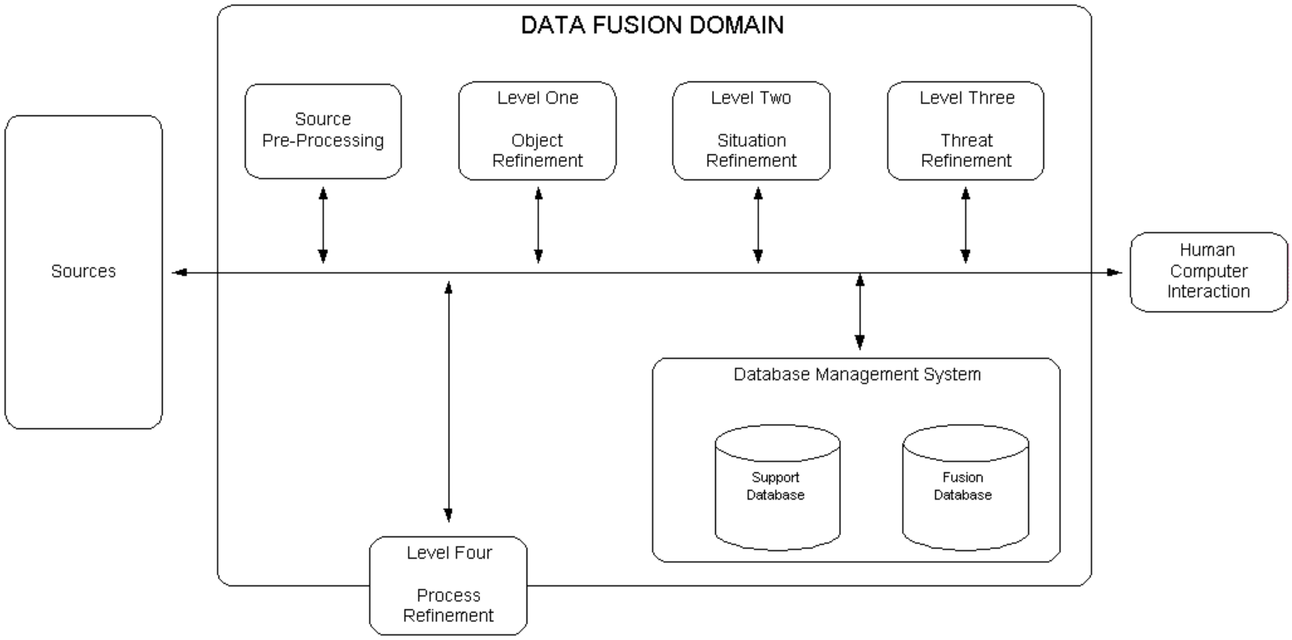


Figure 2: JDL top-level Data Fusion Model [13].

Each level is defined as follows [13]:

Level 1 – Object Assessment: estimation and prediction of entity states on the basis of observation-to-track association, continuous state estimation (e.g. kinematics) and discrete state estimation (e.g. target type and ID);

Level 2 – Situation Assessment: estimation and prediction of relations among entities, to include force structure and cross force relations, communications and perceptual influences, physical context, etc.

Level 3 – Impact Assessment: estimation and prediction of effects on situations of planned or estimated/predicted actions by the participants; to include interactions between action plans of multiple players (e.g. assessing susceptibilities and vulnerabilities to estimated/predicted threat actions given one's own planned actions);

Level 4 – Process Refinement (an element of Resource Management): adaptive data acquisition and processing to support mission objectives.

And, the Sources of Information and Human Computer Interaction concepts are defined as follows [14]:

- *Sources of Information*: indicate that a number of sources of information may be available as input including local and distributed sensors, reference information, geographical information, knowledge base, Intelligence data (HUMINT, SIGINT, etc.); and
- *Human Computer Interaction (HCI)*: allows human input such as commands, information requests, human assessments of inferences, reports from human operators, etc. It includes methods to assist humans in direction of attention, and overcoming human cognitive limitations (e.g. difficulty in processing negative information).

In our effort to map the LLKW System, Process and information with the JDL Data Fusion Model, we asked ourselves the following questions:

- How can the LLKW system be incorporated within the Data Fusion Process from the HCI point of view?
- Can the LL Knowledge Base provide any valuable information to achieve better estimates of the objects state? If this were the case, what should be the steps or actions taken?

From these interrogations, we were able to identify two main areas where these processes may interact. First, the current LLKW System may be foreseen as a Decision-Based System (DBS) assisting in the HCI (figure 3, item 1) that would be standardized into the Info-Centric Workspace of ISTAR [4]. Second, the LL Knowledge Base may be a source of information (figure 3, item 2) for the Data Fusion Process itself, thus provide contextual knowledge used in the different levels of fusion.

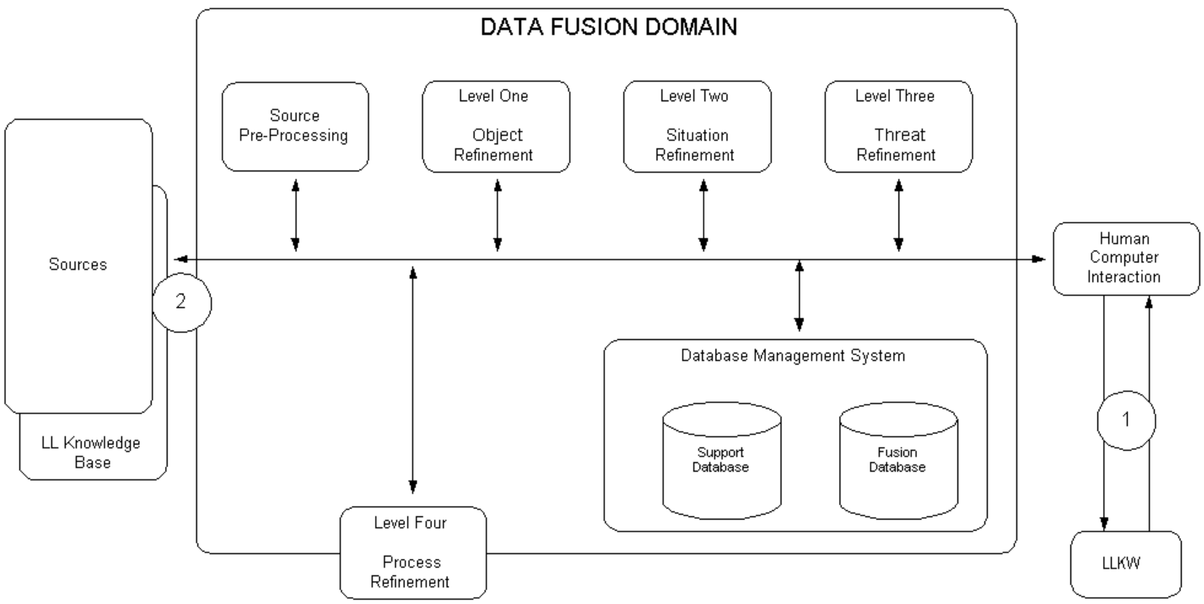


Figure 3: LLKW and Data Fusion interactions.

The relation between Data Fusion and the Command-Centric Approach envisioned for the Army of the future leads into considering the HCI as important input in the realisation of the overall process. From [15], Artificial Intelligence (AI) turned out to be an important tool for the application but lacked sufficient functional insight therefore putting emphasis on the feedback obtained through the HCI.

With this in mind, it is clear that task-support and decision-based systems will help in providing doctrinal, operational and contextual information essential for the realisation of the fusion operator’s tasks. A LLKW application, comprised of fusion-related information, can be positioned as a two-way feedback system (figure 3, item 1) in the JDL model.

First, the application can provide a shared decision-based tool. This means that information, given through past experiences, raised issues and lessons, will be contextually available to the “fusion operators” while they are performing their tasks and thus providing:

- knowledge valuable for taking command level decisions;
- contextual knowledge while executing the required task; and
- re-use of past experiences (doctrine, best practices, etc.).

In the current LL System, this implies the creation of a Management Board, similar to a digital dashboard driven by its contextual inputs. It also means that the underlying knowledge base will have to hold information on the circumstances under which the knowledge was produced. This will eventually lead into the production of a more efficient search engine and, thus, better information would be pushed to the fusion operators.

Second, by defining a questionnaire based on the Data Fusion, the fusion operator will become a contributor or a Knowledge Analyst for the LL System and thus participate to the LL Processes as described in the next section. This enables the LL System to re-use the knowledge acquired by the operator during the execution of the Data Fusion Process.

The other main area of interaction between LLKW and Data Fusion is when the LLKW is used as a Source of Information automatically feeding the Data Fusion Process levels (figure 3, item 2).

The next section will present the Lessons Learned Process, with regards to what has been deployed for the Army and what should be modified in the future to support the Data Fusion Processes and make the integration possible in the context of the Canadian Forces Command-Centric view.

4.0 ALLKW CONTENT AND PROCESS

The actual LLKW is mainly composed of textual observations, from over 45 Operation/Rotation since 1996, that have been converted from the previous system into the newly Knowledge Model as shown on figure 4. The Army Lessons Learned Center (ALLC) is currently managing all other documents (e.g. Bulletin, Dispatch and Report Analysis) where one can extract past lessons through the Army On-line System. Further work by the ALLC is under development to link these historical documents to appropriate Knowledge Objects of the ALLKW, thus, giving direct access to relevant knowledge without any duplication of documents.

The Knowledge Model of the ALLKW presented in figure 4 consists of the followings knowledge object facets:

- The Knowledge structure facet regroups all objects related to the questionnaire structure: Phase, Theme, Subject, Question and Questionnaire.
- The Response facet is composed of the observation and comments objects. An observation or a comment is linked to specific questions of a questionnaire version and type (i.e. 2001, International Operation). Within the LLKW an observation is a response to a specific question and a comment is a response or a clarification on an observation.
- The Lesson Facet is composed of the Issue, the Lessons and the Staff action objects.
- The Organisational Structure facet is composed of the Organisation, the User, the Role and the Reporting Level, defined as the level of authority of an organisation when it answers a questionnaire.
- The Reference facet consists of the Workproduct object, which is a document produced by ALLC and the Reference Document object produced by other organisations.

- The Mission facet is composed of the Exercise, Operation and Rotation objects. Each Exercise or Operation/Rotation is linked to the current questionnaire at the time of its activation.
- The ETSS facet is composed of the Process, Task, Procedure and Step objects.

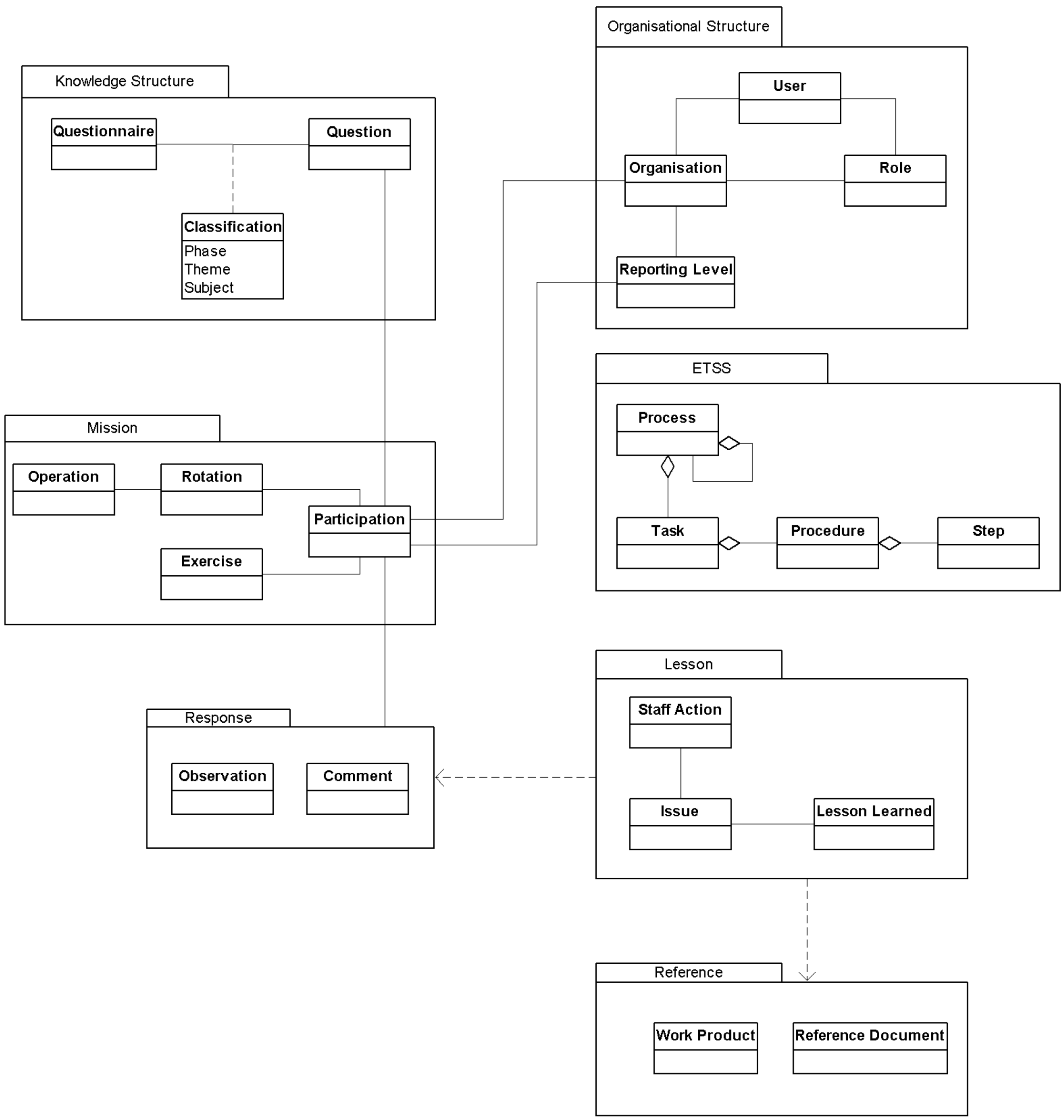


Figure 4: ALLKW Knowledge Object Model.

From the figure above, we observe that the Army is not only sharing Lessons but also sharing observations and comments from operations and exercises, which becomes the raw data that the Analyst will use to extract issues that eventually could become lessons or actions to be implemented.

The ALLKW is currently a standalone system that may be less effective because it forces the user to master a separate process [16]. In order to overcome this limitation, an Electronic Task Support System was embedded into the ALLKW, adding Knowledge about the Processes within the ETSS facet (figure 4). An Electronic Task Support System improves how military personnel are trained. It guides the users on how to use the LLKW and gives them access to the right amount of knowledge, at the right time, in relation with their role and, thus, enhances their tasks performance. The ETSS structures the knowledge in such a way that help is provided in context and dependent on what the user does, what role he has to play and what should be the minimum information about processes, tasks or procedures.

The ETSS brings forward all the explicit knowledge necessary to support the newly design LL Process, fulfilling, by such, the need for the '*Army to develop effective concept-to-fielding Cycle that include the integration of technology with the appropriate doctrine, organisation and training*' [11]. The ETSS is also a sub-system that is part of the LFC2IS and integrated into the ISTAR Canadian Info-Centric Workspace [4]. The same functionalities can be envisioned for the Data Fusion Processes where a user may have access simultaneously to knowledge about how to perform a task and what are the lessons learned associated to it, bringing forward the notion of a dynamic process of learning and sharing best practices [5].

It is mostly recognised in literature that the LL Process is a strategy to elicit, retrieve and re-use Lessons obtained from experiential knowledge [16], [8] and [9]. Processes of the Lessons Learned Process (LLP) are regrouped, supporting that strategy, under four main processes (figure 5) such as Knowledge Organisation, Knowledge Gathering, Knowledge Analysis and Knowledge in Action. Each phase is further described in the context of the LLKW currently deployed for the Army. The Canadian Forces agreed upon the composition of the LLP and are currently pursuing the development of the LLS to leverage the work overtaken by the Army and make it a Canadian Lesson Learned System where they will share knowledge and learn from it (figure 6).

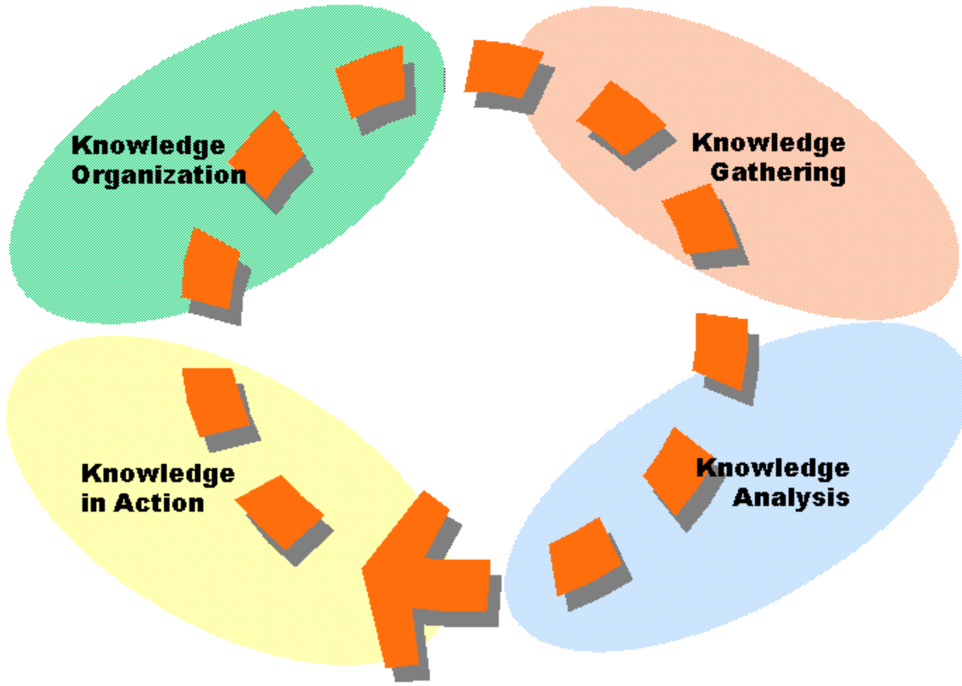


Figure 5: Phases of the Canadian Lessons Learned Process.

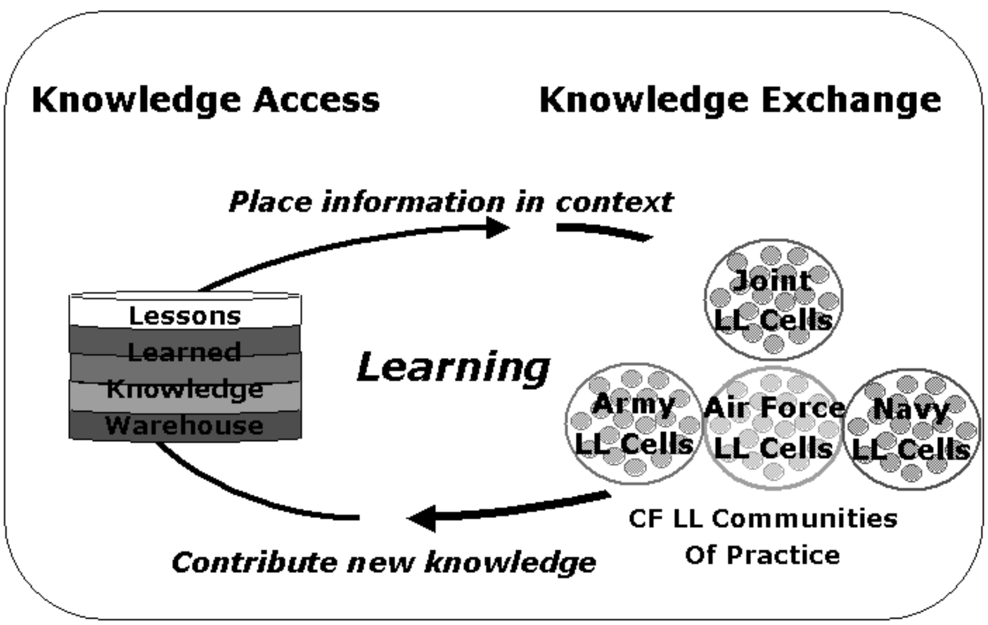


Figure 6: LLKW as part of the Canadian Knowledge & Learning Strategy.

4.1 Knowledge Organisation

At this particular phase, the aim is to organise or structure knowledge in such a way that it can easily be visualised, searched and managed. From the Army Lessons Learned Center perspective, and specifically for their analyst, the Knowledge Organisation phase is the entry point of the LL Process.

The sub-process of the Knowledge Organisation phase relates on how to manage and modify the knowledge structure such as the evolution or the creation of questionnaires, on how to manage the creation and edition of operation or exercise, on how to manage the Army organisational structure, and finally on how to profile the user's roles within current operations and/or exercises (figure 7).

When creating a mission or an exercise, the ALLC Analyst has to identify which organisation is involved at which Reporting Level within the Chain of Command and who will be the representative acting as contributor or Commanding Officer.

The use of questionnaires at the tactical level for operations and exercises is the Knowledge Gathering strategy for the Army. These questionnaires are known as Post Operation Report (POR) and Post Exercise Report (PXR). The structure to be managed for operations is based on questionnaires that follow a Phase, Theme and Subject hierarchy and a Theme and Subject hierarchy for exercise.

The LLKW manages all of the relationships between the different versions of questionnaires (i.e. Domestic Operation questionnaire). Some questions may be updated and kept within the same meaning (minor modification, clarifying a question) while some may be created, replaced or deleted in the future structure. The management of the historic of questionnaire allows the system to extract all observations from a current question and it's previous version. There is currently three official versions of the International Operational questionnaire.

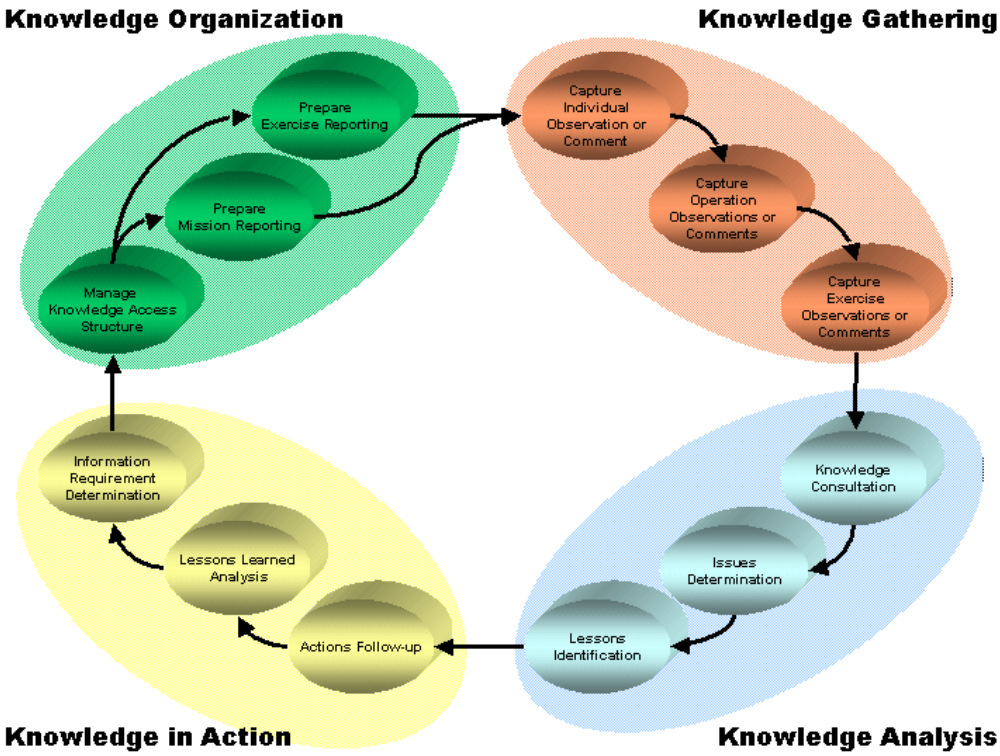


Figure 7: Main Sub-processes all Phases of the LL Process.

The effectiveness of a structured reporting context, i.e. the questionnaire, has been proven by its usage. When comparing LL applications where in one case a more rigid and task-oriented structure for capturing knowledge is defined and whereby, in the other case, a mission-oriented and free form is used, we observed that the essence of the observations was more meaningful with regards to the subject and related question, upon which it was supposed to be reported.

The task of capturing knowledge involves a cognitive process establishing the information that will be given, i.e. "knowledge mapping", and how it will be formulated. By structuring and classifying the required inputs, it reduces the cognitive load and helps in the gathering of higher quality observations.

The questions were formulated in order to gather specific and important information on a particular aspect of the operation/exercise. The Warning, Mounting and Deployment phases gather the information on mission planning aspects, while the Employment and Re-deployment phases cover the mission execution aspects.

The following are questions taken from the POR International Questionnaire version 2002:

Phase: 3 - Deployment

Theme: Movement

Problems/Successes during the deployment Phase

a. What were the problems/successes encountered during the deployment phase with respect to:

(1) equipment availability?

(2) allocation of AORs?

(3) establishment of accommodation sites and infrastructure?

(4) communications?

(5) security?

(6) standing up the Contingent HQ?

Phase: 4 – Employment

Theme: Command and Control (C2)

Unit Deployment Detail

Briefly describe the unit deployment in terms of:

a. Tour dates

b. Sub-units

(1) Number and type of sub-units deployed (to include sub-unit function, major eqpt and parent unit title).

(2) Indicate if there was a change in unit organisation at sub-unit level for this mission.

(3) If re-roling was required, identify for applicable sub-units.

c. Total number of people deployed (Reg/Res)

d. Description of the location of the deployment (country/province/ major cities)

e. Were there momentous incidents or changes to the situation in the deployment area that had a major impact on the operation?

Theme: Information Operations

Climate Effects on Operations

a. How and to what extent were the operations affected by the climate?

b. How were the problems resulting from the climate overcome?

Theme: Operations

Manoeuvre Characteristics

a. Did the unit possess the mobility, firepower and protection necessary to conduct its mission?

b. If not what would be a better ORBAT for this mission?

Mine Threats and Counter-mine Measures

a. Was there a mine threat?

b. Were engineer resources adequate to deal with this threat?

c. Were there any mine incidents during the tour?

Requirements for Doctrinal Review

a. Are there tactics, organisations or procedures that should be reviewed as a result of this mission?

b. Were there any TTPs developed specific to this operation that could be of future use? Is so, provide a brief description of the subject matter and a Point of Contact (POC).

The actual LL system mainly incorporates knowledge specifics to planning and execution of a mission in light of the way the questionnaires were defined. Also, the LLKW mostly contained unstructured knowledge from which it was not easy to perform any analysis or identify the reporting context. In order to structure the underlying knowledge so it can be accessed in the Data Fusion Processes, the following actions were identified:

a) Review the LLKW to capture the context in which the information was gathered, to categorize and structure the underlying knowledge, to control the way knowledge is captured and thus associating concepts with users inputs;

b) Define a unifying terminology that would be used in both the LL and the Data Fusion; and

c) Define a questionnaire to gather the domain-specific knowledge and identify at which Data Fusion Process level it is related;

a) Knowledge Base Improvements

Since the LLKW content is mostly textual, further work has to be performed to add more functions for capturing context, finding similarities, structuring users inputs, etc. The following describes some of the aspects that should be considered.

First, context information can be obtained by answering the following questions:

- **when** did the event happened, the conditions under which it happened, the actions that triggered the event;
- **where** did it take place, the elements that defined the surroundings, the weather, the location; and
- **what** was the task performed at the time of the event, the lesson to be learned, the topics covered, the content, etc.

Second, to store such information, we must categorize the knowledge by adding appropriate attributes to the existing knowledge structure. Keeping in mind that the environment is associated to a lesson by mapping concepts, a lesson learned with contextual information could have the following representation [17]:

- Originating Action
- Result
- Lesson Contribution
- Applicable Task(s)
- Conditions
- Suggestions

Categorization can be accomplished by associating a unifying terminology to the already existing knowledge. This facilitates knowledge elicitation while providing a mechanism to extract and analyse information in order to feed it back into the Data Fusion Process.

b) Unifying Terminology

To help provide information integration between the LL System and top-level Data Fusion Processes, the use of a unifying terminology is proposed.

'*Ontologies range from controlled vocabularies to highly expressive domain models. They are integrated data dictionary designed for human understanding, structured data models suitable for data management, and computational ontologies* [18]. '

Different categories are used to classify ontologies, but the one we are most concerned with are Domain ontologies. These structures represent specific concepts from a **domain** or field of expertise. In the case of knowledge retrieval from the LLKW, the domains targeted are top-level processes of Data Fusion level 2 and 3, i.e. Situation and Threat Assessment.

'*With ontologies, information integration from heterogeneous sources can be addressed at the structural, syntactic or semantic levels. These are used to make the content more explicit* [18].'

As a result from the Joint Directors of Laboratories (JDL) Data Fusion Working Group to codify the terminology related to Data Fusion, an effort was made to create a Data Fusion Lexicon [6].

c) Questionnaire Approach

Using the questionnaire approach, domain experts and Knowledge Analysts have to link the knowledge structure with the different concepts used in the construction of the Domain Ontology. By doing so, information can be easily re-used for further analysis or as a feed for the Data Fusion with additional inputs.

For example, if we refer to the questionnaire sample given in scenario 1, the question categorized under
Phase: 4 – Employment,
Theme: Command and Control (C2)
Subject – Unit Deployment Detail
can give information regarding level 2 fusion for the Force Deployment concept.

Also, for level 3, the question associated with
Phase 4 – Employment
Theme: Operations
Subject – Mine Threats and Counter-mine Measures
can give some insights on the Enemy Force Capability, activities, etc.

4.2 Knowledge Gathering

The main objective of this phase is to manage observations and comments gathered in relation to an operation or exercise. This phase is the entry point of military personnel involved in an operation or an exercise.

The sub-processes for the Knowledge Gathering phase relates on how to capture observations, how a Commanding Officer approves observations and comments within his organisation, how to add additional observation that doesn't fall under an official POR or a PXR and finally how to integrate related reference documents (figure 7).

Observations and comments are captured on-line by contributors and approved by the Commanding Officer for each organisation associated to a specific Reporting Level. The distinction between observations and comments allows users to easily visualise and follow a discussion on a particular subject within the context of an operation or an exercise. The sequence of observations and comments follow the Chain of Command from the unit level up to the higher level. One of the benefits of the on-line capture of knowledge is the possibility to access observations and comments as soon as the Commanding Officer has approved his knowledge acquisition.

At this phase, one of the potential outcomes from applying a unifying terminology for categorisation of the observations and comments is that the users inputs could be formed and controlled by series of concepts taken from that same terminology. This controlled vocabulary ensures consistency between LL knowledge and Data Fusion information.

4.3 Knowledge Analysis

The main objective of the Knowledge Analysis phase is to seek knowledge on different topics, validate issues and stimulate learning while doing operations and exercises. This phase is the entry point of a wide range of users that are interested in learning about previous experiences and increase their state of preparedness for operations or training.

The sub-processes for the Knowledge Analysis phase relates on how to consult, analyse and extract knowledge, how to manage issues and lessons, how to capture comments on issues and lessons from an Office of Primary Interest (OPI) and finally, how to link knowledge objects such as observations to support Issues or Lessons (figure 7).

The actual LLKW offers two ways of accessing knowledge; either by browsing through the site or by searching with specific criteria related to knowledge objects or knowledge structures.

When browsing through the LLKW, the Analysts will always have access to knowledge objects through Views presenting knowledge by specific attributes (i.e. by name, by date, by OPI). A filter function is also available to assist the Analyst when he browses through a questionnaire structure. This function limits access to specified knowledge objects by filtering the pages content. This enables to discard any irrelevant information that would, otherwise, be treated.

The knowledge extraction can be accomplished by using the full text search capabilities in combination with criteria, such as the type of operation or exercise, the phase, theme, or subject of a questionnaire and/or from the organisation (figure 8).

Several research projects at the Defence R&D Canada Valcartier are under development to find ways of assisting the Analyst during the Knowledge Analysis phase. Among those, some projects examine how to find similar observation or lesson [9] or how to better classify and find knowledge with the use of multiple ontologies and intelligent agents [19].

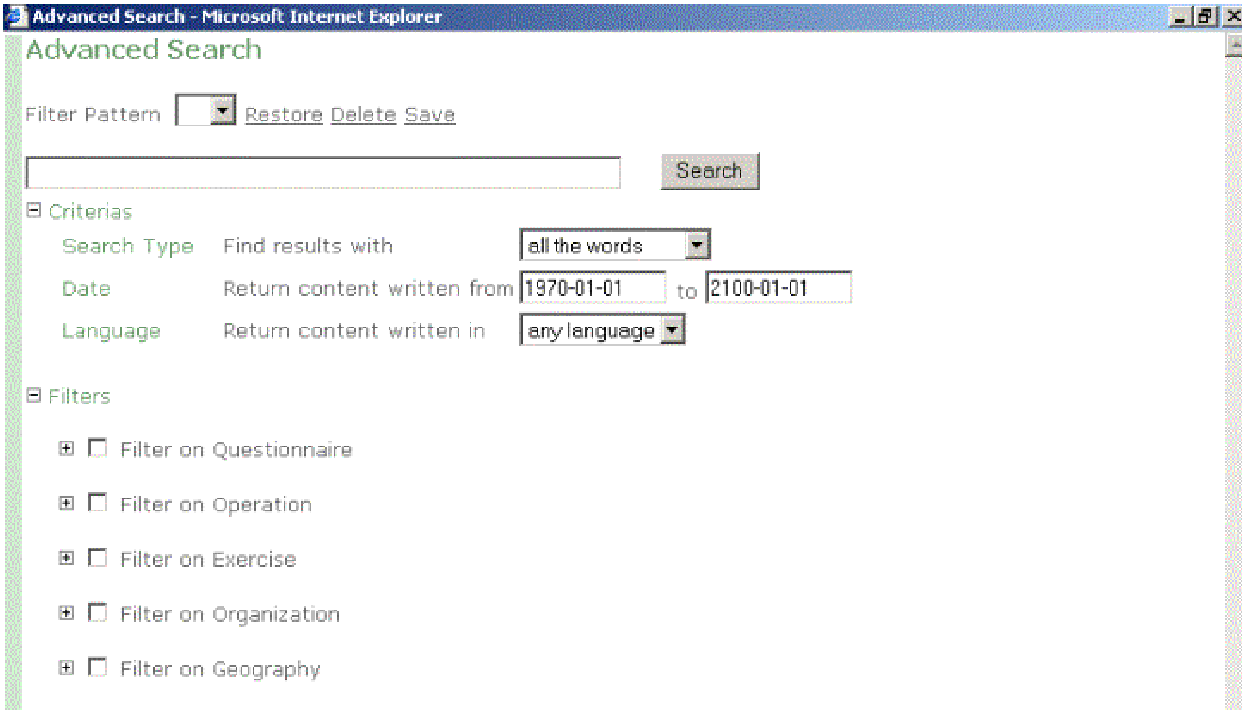


Figure 8: Advanced Search Criteria of the ALLKW.

4.4 Knowledge in Action

The main objectives of the Knowledge in Action phase is to gather recommendations and decisions from staff authorities, insuring that proper actions are taken and lessons are learned. This phase is the entry point of the stakeholder, OPI, Staff, Commander, ALLC Staff with the main concern to follow up on actions but also to contribute to the LLKW by adding their comments, recommendations or decisions on issues that falls under their command.

The sub-processes for the Knowledge in Action phase relates on how to manage Staff Action, how to capture comments, recommendations and decisions on Issues and how to confirm whether or not a Lesson has been learned (Figure 7).

This phase was the most difficult one to design during the development phase of the ALLKW. Part of this is due to the fact that the ALLC, at the time of development, weren't empowered to make things happen without direct leadership and responsibility. Their limit of responsibility was to propose, suggest, and try to gather comments and decisions from Staff. The direct involvement of Staff would occur only when it fell under their roles and responsibilities.

The ALLKW offers a way of accessing and managing Staff Action but these sub-processes will need to be improved as soon as the Staff becomes aware of what the system can offer. Unfortunately, Staff personnel weren't involve during the ALLKW design phase, thus the proper balance between the users, processes and system views, regarding the Knowledge in Action phase, will have to be obtained through future developments in order to maximize the benefits of the new technologies.

5.0 CONCLUSION

The current Army Lessons Learned System is deployed as a standalone system but will soon be integrated into the Land Force Command and Control Information System (LFC2IS) through the proposed Canadian ISTAR Information-Centric Workspace System of Systems Vision [4]. The System will provide knowledge to Commanders within their operational context. For doing so in a timely manner and without to much processing effort, we have already identified that knowledge must be properly organised, structured and classified in such way that it can be easily exploited to automatically provide (i.e. push and pull) relevant lessons for specific situation. The right balance between text and structured text within the database will have to be addressed in light of all the possibility the technology offers to extract relevant knowledge.

The Data Fusion System and Process is also an important component of the ISTAR Information-Centric Workspace and, as such, it will benefit from the integration of the LLKW in gathering information and knowledge linked to the process itself or by providing an easy access to any relevant information already captured in similar situation.

The Army Lessons Learned Process can easily be mapped with the top-level Data Fusion Process. The main difference is the notion of level of abstraction for knowledge objects or the context from which we look at the model.

The top-level Data Fusion Process Model was intented to be a generic model that can be applied in different situations. For this reason, we compared the Data Fusion process levels to the ones described in the LLP.

The following table shows the association that was made between the different processes and components. We can observe that only the nature of the objects differ.

Data Fusion		Lessons Learned	
Processes	**Components**	**Components**	**Processes**
Level 1 – Object Refinement	Data alignment Data object/correlation Position/kinematic and attribute estimation Object identity estimation	Observations Comments	Knowledge Gathering
Level 2 – Situation Refinement	Object aggregation Event/activity interpretation Contextual interpretation	Issues Lessons	Knowledge Analysis
Level 3 – Threat Refinement	Aggregate force estimation Intent prediction Multi-perspective assessment	Actions Lessons Learned Recommendations	Knowledge in Action
Level 4 – Process Refinement	Performance evaluation Process control Source requirement determination Mission management	Questionnaires management Organisational Structure modification Reporting Context definition Users and Roles determination	Knowledge Organisation

Moving towards a Command-Centric vision implies that knowledge must be shared between organisations.

For the last two years, the Lessons Learned Joint Staff has hosted several workshops in an effort to bring all environmental parties (Army, Navy, Air Force, Joint) towards building a common understanding of the Lessons Learned Process and a shared vision of the Canadian Forces Lessons Learned Knowledge Warehouse. Each of them agreed that the Army LLKW System could be used as a starting point to provide a better understanding of technical and organisational requirements. From these emerged the need to manage multiple knowledge structure or context, to modify their actual LL Process or to identify specialised personnel. Each environment also agreed to the fact that they need to share information as soon as it's validated by the Commanding Officer and transmitted forward into the Chain of Command.

Sharing Knowledge will be one of the most important principles of the Canadian Forces LLKW with relevance to simplicity, sustainability and credibility. The CF LLKW System shall provide an essential KM tool that becomes a significant factor in the development and the learning curve within the Canadian Forces while enabling a faster and more accurate decision making/sharing process. This will reinforce the Canadian Force Command-Centric Vision, providing a unified way of sharing knowledge between environments while merging the tactical, operational and strategic level into one workspace.

We can only hope that the work that we are currently doing will eventually benefit Canadian Forces as well as the ones from our Allies as time and precision in the decision making process is becoming an overwhelming preoccupation for everyone.

ACKNOWLEDGEMENT

This project was supported by the Directorate Land Command Information, National Defence Headquarters Canada and by the Defence R&D Valcartier Canada.

REFERENCES

[1] Department of National Defence Canada (DND), ARMY STRATEGY-Advancing with Purpose 2002. 33p.

[2] Department of National Defence Canada (DND), The Future Security Environment, Report Number 99-2, 1999.

[3] Department of National Defence Canada (DND), Army Knowledge Management Strategy Outline, Revision 1.4, September 2002, 10p.

[4] Gaétan Thibault, François Le May. Introducing the Canadian Information-Centric Workspace Concept, *Research & Technology Organisation Symposium on Military Data and Information Fusion*, 20-22 October 2003.

[5] David S. Alberts Information Age Transformation Getting a 21st Century military. Information Age Transformation series.

[6] Data Fusion Lexicon, Data Fusion Subpanel of the Joint Directors of Laboratories Technical Panel for C^3, F.E. White, Code 4202, NOSC, San Diego, CA, 1991.

[7] Thomas H. Davenport and L. Prusak. Working knowledge: How organizations manage what they know. Boston, MA: Harvard Business School Press, 1998.

[8] Rosina Weber, David W. Aha, Irma Becerra-Fernandez. Intelligent Lessons Learned Systems. International Journal of Expert Systems Research & Applications, Vol. 20, No.1., pp. 17-34.

[9] Anne-Claire Boury-Brisset, Marlène Gauvin, Pierrette Champoux, A Knowledge Management Approach to the Creation and Sharing of Canadian Forces Lessons Learned, *ICCRTS* September 2002, 17p.

[10] Pierrette Champoux, Marlène Gauvin, Anne-Claire Boury-Brisset, Applying Knowledge Management to Lessons Learned, *DND Knowledge Management Blue Print*, (*Chapter 10 Appendix)*, Department National Defense Canada, 2002, 10p.

[11] Department of National Defence Canada (DND), Future Army Capabilities, January 2001 48p.

[12] Champoux Pierrette. *Study on Task Analysis and Usefullness of a Case Base Tool during the Analysis Phase of Geomatic Development Project*. Ph.D. Thesis, Geomatic Science Department, Forestry and Geomatic Faculty, Université Laval, Québec, 289 p. 1997.

[13] Alan N. Steinberg, Christopher .L. Bowman, and Franklin .E. White *Revision to the JDL Data Fusion Model*, Joint NATO/IRIS Conference, Quebec City, October 1998.

[14] David L. Hall and James Llinas An introduction to Mutlisensor Data Fusion, Proceedings of the IEEE, Vol. 85, no1. January 1997, pp. 6-22.

[15] Tore Smestad, *Data Fusion- for Humans, Computers of Both*, Translation of an article in the periodical 'Mikroskopet' of the Norvwegian Defence Research Establishment published in February 2001, Norway, 6p.

[16] David W. Aha, Rosina Weber, Héctor Munoz-Avila, Leanard A. Breslow, Kalyan Moy Gupta, Bridging the Lesson Distribution Gap. *IJCAI'0,* Seatle, WA August 2001, Morgan Kaufmann Publishers, Inc. pp. 987-992.

[17] Weber, R., Aha, D.W., & Becerra-Fernandez, I. (2000). *Categorizing Intelligent Lessons Learned Systems*. In D.W.Aha & R. Weber (Eds.) (2000). *Intelligent Lessons Learned Systems:* Papers from the AAAI 2000 Workshop, 63-67. (Technical Report WS-00-008). Menlo Park, CA:AAAI Press.

[18] Anne-Claire Boury-Brisset. *Ontology-based Approach for Information Fusion*. Sixth International Conference on Information Fusion, Cairns, Queensland, Australia, July 2003.

[19] Anne-Claire Boury-Brisset, Ontological Approach to Military Modeling and Knowledge Management, *Research & Technology Organisation Symposium on Military Data and Information Fusion*, 0-22 October 2003.

The Visualisation of Diverse Intelligence

Dr. Hans-Joachim Kolb / Dr. Michael Towsey /
Mr. Stefan Maetschke / Dr. Ulla Uebler
MEDAV GmbH
Gräfenberger Straße 32-34
D-91080 Uttenreuth
GERMANY

Email: Hans.Kolb@medav.de / Michael.Towsey@medav.de /
Stefan.Maetschke@medav.de / Ulla.Uebler@medav.de

1.0 INTRODUCTION

At last years' NATO workshop "Massive Military Data Fusion" in Norway 2002, among others, it was agreed that there is an increasing need for visualisation of data. This need is a increased by obtaining more data coming from more sensors and different sources.

In the intelligence context, the fusion of information is a multi-layer process. First, raw data are collected and filtered and brought to archives. The analysis of the data is the most time consuming process, since it can be automated only in a limited range. Thus, there is a need for a "human-in-the-loop". For humans, the best way to analyse amounts of data is their pattern recognition abilities, thus using powerful algorithms of visualisation. In this paper, we will present our approaches to visualisation in different levels.

First, we will describe data modelling in the intelligence context, presenting our approach of the MEDAV archive as data model and the stages at which visualisation is a powerful means. In the following section, we define approaches to visualisation, and further on, show our approaches to visualisation for different information retrieval tasks.

2.0 DATA MODELLING

One of the most important parts of the information fusion task is the data model. The purpose of data modelling is to develop an accurate model, or graphical representation, of the client's information needs and business processes. The data model acts as a framework for the development of the new or enhanced application. Since the amount of data to be analysed can be enormous, the data model as a core of the system must be efficient and well-organised for the domain of diverse intelligence.

Data modelling can be defined as the design of the logical and physical structure of one or more databases to accommodate the information needs of the users in an organization for a defined set of applications. In the intelligence context the basic need is

- to offer possibilities to store huge amounts of data,
- allow access to all data without regard to the type of original data, e.g. recording source, text or speech signal.

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

The amount of data is rising, since data come from more and more input channels. Channels in a typical application in the intelligence context may be speech from telephones, mobiles, HF and many other sources. Signals are found that might incorporate speech in some parts. Furthermore, information is gathered from texts from different sources, e.g. internet, mails, newspapers. Another source of information is images, such as from satellites. All these sorts of data have to be analysed according to a special task or question. For each of the channels, the way of analysing data may be different, but none of the channels may be neglected.

A good data model finds an efficient way in which input from different channels is stored in a somewhat similar way in order to access the data for further analysis in a similar manner. Usually, data would be stored according to different criteria, e.g. depending on the source of the data. Another criterion may be the type of pre-processing that classifies the data into different categories, and attaches the gathered meta information to each datum.

For the analysis of data, there is a need for a human-in-the-loop, i.e. at some point, a human operator must check and evaluate the data or the automatically gained analysis. The work of the human-in-the-loop can be assisted in different ways, mainly in two different types: 1) pre-selection of data that are of more relevance than other sets in the data. This can be performed automatically up to a certain level.. 2) analysis of data, among others by means of visualisation. The main advantage of the using visualisation at this step of processing is that humans tend to quickly understand complex issues when presented visually. Thus, the process of analysis is speeded up when using visualisation.

In the next section we will describe our approach to data modelling with the MEDAV archive in more detail.

3.0 DATA MODELLING IN THE MEDAV ARCHIVE

For the past 20 years, MEDAV GmbH has been a developer of intelligence processing hardware and software. We have been engaged in studies and commercial projects for the German Federal Armed Forces and for other German and international government agencies. Our challenge has been to develop hardware and software systems that incorporate the different levels of information fusion within the one architecture.

The need, especially in the intelligence context, is to archive huge amounts of data of different sources and to find all kinds of information regardless of the format the data are stored. The goals are:

- Allow storage of amounts of data coming from a variety of input sources.
- Retrieve information within all data types in the same manner, regardless of the type of source.

The task to be solved by the archive is to search information or documents in the archive and for example to print out the relevant documents. This task can be solved with our data modeling together with the MEDAV archive.

The main principle is, that all data are stored together with attributes. The attributes consist of standard attributes like the date, size of source data as well as with an unlimited number of additional attributes that are defined by the application and by the users. Possible attributes are the author of a documents, the transcription (of sound files for example), keyword or other comments on the content of a data set. New attributes can be added at any time when desired by the user of the archive. The data are compressed and can be encrypted. The access can be limited to special data and a special user group. The access to the archive can be logged.

The search can easily be performed within the attributes. When a datum is stored, it is also entered into a searchable index file. Furthermore, the data can be removed from the archive, but the index can still be searched further on.

Most importantly, the front-end towards the user remains the same for different tools inside the archive. All types of data can be searched in the same way regardless of their origin. No knowledge about data modelling detail or the data architecture is necessary for the user.

An overview of the data modelling in the MEDAV archive is given in Figure 1.

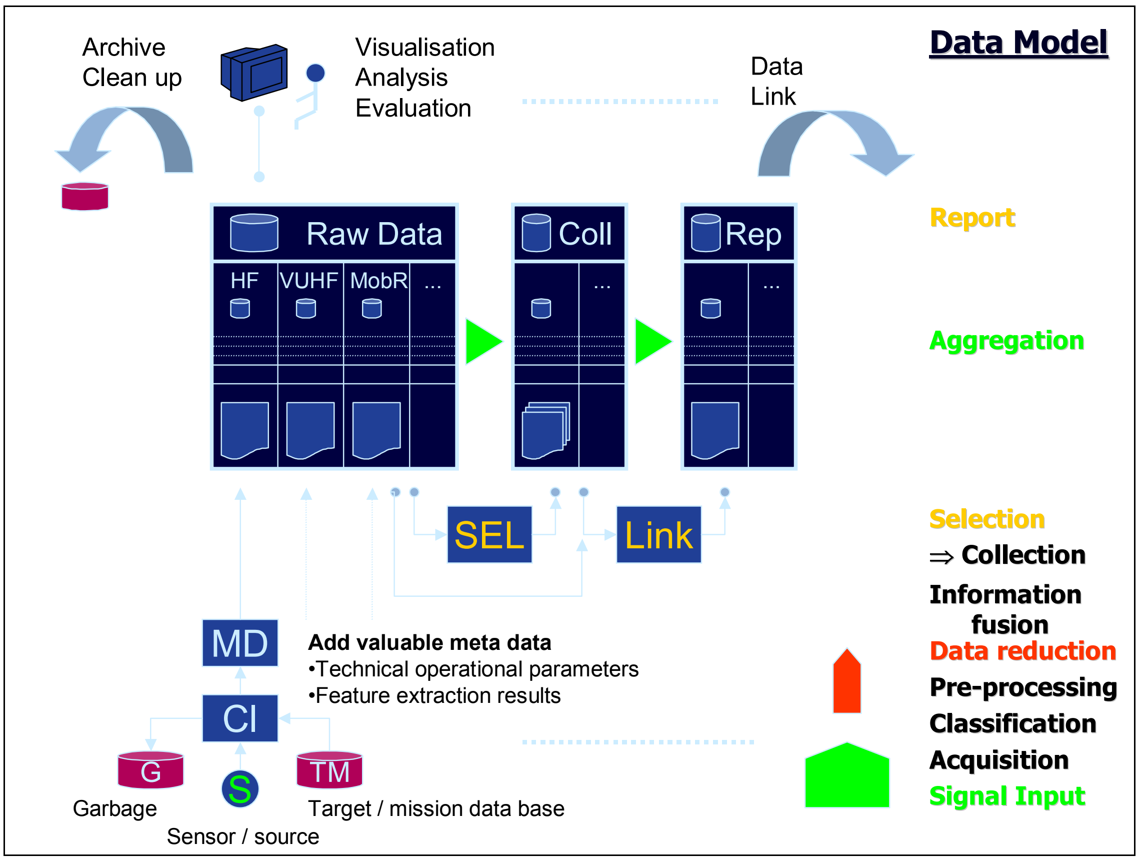


Figure 1: Data modeling in the MEDAV archive.

At the bottom on the left the sources *S* are shown. Input is gathered from different sensors. These data pass a pre-processing step, in which they are classified (*CL*) using a target/mission data base *TM*. Data that are not of interest are filtered to the garbage. The kept data are passed to a feature extraction step obtaining meta data information that may be of interest for further analysis.

The data are labeled with additional information like recording time and spoken language. For each data file that is obtained that way, information is attached as meta data (*MD*) that may the used for a faster search than when looking up in the source data. These raw data are stored in an archive.

Each information source is kept separately, as HF, VUHF, MobR in the example in Figure 1. The raw data part of the archive contains the original data together with the additional meta data. For further analysis, a reference to the original data is always added in order to be able to trace the analysis back to the original data, for example for verification.

The next step in the archive is a selection of the data (*SEL*) with respect to the information retrieval. Selected data according to a choice of criteria are put to the *Coll* part of the data base. Selection can be performed automatically for some task as well as manually by a human operator. In this step, information fusion can be performed both by selection of a number of data files as well as by reducing each data file itself, e.g. cutting speech signals to the relevant parts or by summarizing text. Visualisation of data files is a feature that is very helpful at this step. A request to the selection of raw data may be “show all data containing KEYWORD_1 and KEYWORD_2 and make statistics about the frequency of occurrence in a certain time frame”. Another application could be to make text summaries of text files containing certain keywords or by a certain author.

The result of the information fusion is a collection of relevant data and/or summaries and statistics of the raw data base. Now, the evaluation is completed by writing a report (*Rep*) to the organisation of the operator. The report contains the evaluation results from the collection part of the data base. Furthermore, the report contains links to the original data in order to be able to check the completeness and correctness of the data and to provide quality assurance.

During the analysis of the raw data base with respect to a certain task, an aggregation takes place from raw data to collection and towards report. This aggregation on the one hand is a selection of the important data as well as a reduction of the data towards to important features for a chosen task.

Evaluation of the data happens in two places: first, in the selection step and further on, in the report. The evaluation can be performed automatically or semi-automatically or manually by the human operator. An interface for automatic evaluation is provided to make the evaluation easier for the operator.

Visualisation in this application can be used at two stages: in the process of selection and in the report. The task of the visualisation is to support the operator to easier find important data and to make the obtained results more visible, i.e. easier to understand at a glance.

4.0 VISUALISATION

As we have seen in the previous section, visualisation is an important feature in the analysis of diverse intelligence. In this section, we will present our approach to visualisation. First, we should define our understanding of visualisation and how it can be classified. Therefore, we present a visualisation model. The human operator and the reader of a report are the users of the output of visualisation. An important topic is the human interaction between the human and the visualisation interface as well as the best way to show contents visually.

The goal of visualisation is, generally speaking, to make large data sets more accessible and easier to understand in short time. Different means can be used for visualisation: graphics, that show statistics for example in a histogram, but also compressed texts that are shorter than the original texts and therefore save time to read. Additionally, important parts may be highlighted.

Classifying visualisation by means of output we can find three categories:

- Traditional static predefined displays

- Augmented reality displays (where interactive iconic and textual information are embedded in realistic life scenarios)
- Virtual reality displays (where the user is located within the image and becomes part of it)

The emphasis of this paper is set on the first item, additionally we will show some examples from the other types.

4.1 Visualisation Model

At last years' workshop, a visualisation model was developed that describes the visualisation process, see Figure 2. The user of the visualisation interface accesses the multimedia displays which optimally cover the senses of humans, i.e. visually, aurally and even haptically as appropriate for the application in question. The user interacts with the multimedia displays in several modalities, traditionally with keyboard and mouse, but also with voice and gestures according to the requirements of the application.

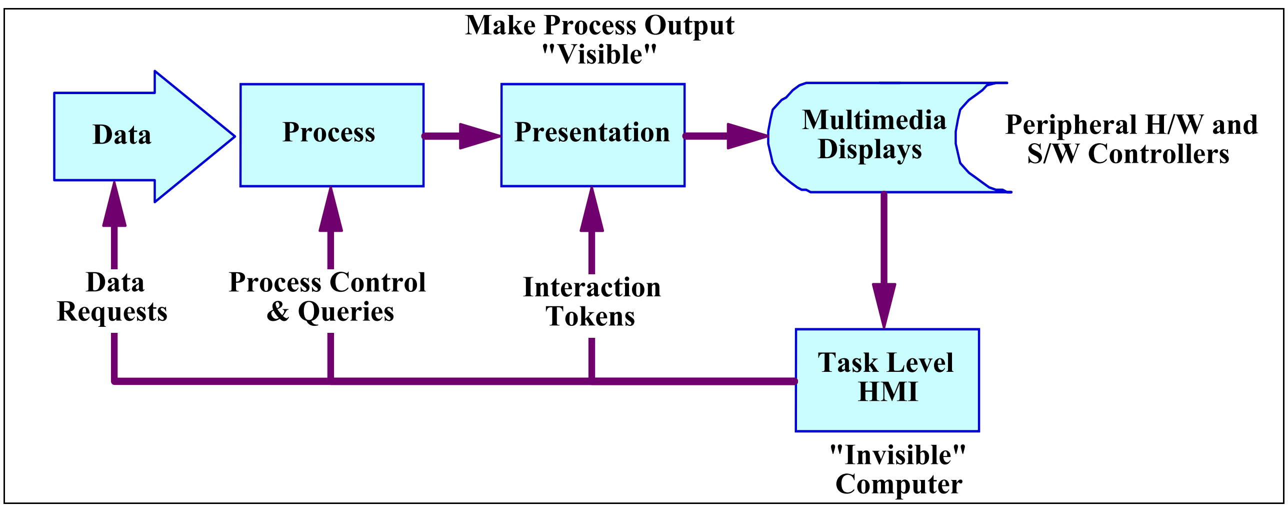


Figure 2: The Visualisation Model.

From the multimedia displays, the user interacts with the task level HMI (human machine interface) in the most natural way, since at this level, humans should not be burdened with extraneous cognitive tasks required to operate a computer system — a well-designed HMI should make the computer "invisible" to its users, in the words of Donald Norman [Norman, 1998 #1286]. Using the task level HMI, the user should be able to obtain the desired information in the wanted type of display, the HMI controls the modalities of presentation by interaction tokens.

4.2 Human Interaction

The goal of visualisation is to provide easy and fast understanding of complex issues. Therefore, the HMI should be designed corresponding to natural human communication. The first item is to study the results of physiology and psychology. The *grok box* project (http://vader.mindtel.com/concepts.html), among others, deals with these aspects.

A visualisation interface should be understood and handled intuitively. It should be represented by well-known symbols. Ideally, the symbols in the used iconography have a relation to what they stand for, such that understanding the visual representation becomes easier. Furthermore, the human understands best, if several sensory modalities are employed at the same time such as visual, aural, tactile stimuli. This effect can be consolidated by using different colors, different sounds etc.

For optical representation different shapes and colors should be used. Large or frequently occurring aspects should be represented by large symbols, important topics should be placed in the middle and/or be highlighted. Still, we must pay attention not to overload the image with too much information and too complex iconography to keep the image intuitively understandable.

The visual representation depends much on the content that is to be displayed. There is no best set of symbols or way to display information. Some complex data may need more interaction with the user, such that he may want to change the view on the presented data depending on his interests.

4.3 Our Approach to Visualisation

The overall principle of our approach to visualisation is to provide an intuitive means for fast understanding of complex issues and facts. The important parts of the presented information should directly come into mind and view.

During the design of a visualisation tool, it is important to specify the preconditions of both the domain of analysis as well as the knowledge and preferences of the user. Thus, an estimation of effort must be done regarding, computing power that might limit the range of possible visualisations. The preferences of the user regarding the type of output, and the display of information must be determined. It must also be estimated if any type of knowledge is needed as a prerequisite for visualisation like a language's grammar or an organisation's structure.

One need that we have found in our studies is to provide a generic tool that can be specified to a certain amount by the user towards his means. On the other hand, the tool should be specific enough to meet the needs of the user providing fast access. We will provide different levels of visualisation ranging from text output to virtual reality graphics. We will use colors and highlighting for making the optical access easier. Drill-down, i.e. interactive focusing on a special part of the image will be provided if desired. Depending on the application and prerequisites we provide different modules for visualisation. In the following section, we will provide examples from our work.

5.0 VISUALISATION EXAMPLES

In studies with the German Federal Armed Forces, MEDAV has evaluated a variety of information visualisation tools. While it is true that " a picture tells a thousand words", we find that users need to have different views over the same data. For example a graphical display of a military command structure can be usefully complemented by a simple tabular display of other information. However, neither display by itself is adequate. As another example, a graphical content summary of a text document may be useful if the document is large but for small documents (or parts of documents), the user may prefer textual summaries.

In this section, we will demonstrate visualisation tools at different levels of specialization, representation and display from studies we have performed. We can roughly classify the visualisation examples as shown in Table 1. The first two examples show the visualisation of signals.

Visualisation at the signal level is shown in Figure 3. A basic approach is shown in Figure 4, showing a two-dimensional graph of communication structures. Figure 5 and Figure 6 show further solutions for the visualisation of flows. Figure 6 and Figure 7 present interactive visualisation HMIs, the first one with a complex choice of visualisation options. Figure 8 gives an example of a virtual reality application. Figure 9 and Figure 10 show applications that work based on the analysis of the contents of texts and tables using artificial intelligence techniques.

Table 1: Visualisation solutions

Figure 3: Sonagram in 3D, Eye Diagram in 3D	Signal, 3D
Figure 4: Command Hierarchy	Classical, clustering, static
Figure 5: Bandwidth of a communication network	Flow, static
Figure 6: Transmitter traffic in real sites	Flow, photo, interactive
Figure 7: Communication cube	Cube, interactive
Figure 8: Virtual Reality Application	Virtual reality
Figure 9: Communication of employees	Text mining, interactive
Figure 10: Visual Summary	Visual summary, interactive

5.1 Visualisation of Signals

At the level of signals, there is already a use for visualisation. Both example in Figure 3 show that the use of colors, and three-dimensional representation enhance the easy understanding of signals. The sonagram can be visualised in 3D representation on the left side of Figure 3. The frequency is shown from left to right, the time axis is from front to back, and the intensity is shown in the height and colors. Another example is the eye diagram on the right side of Figure 3. The quality of the signal can be seen better using colors and 3D representation.

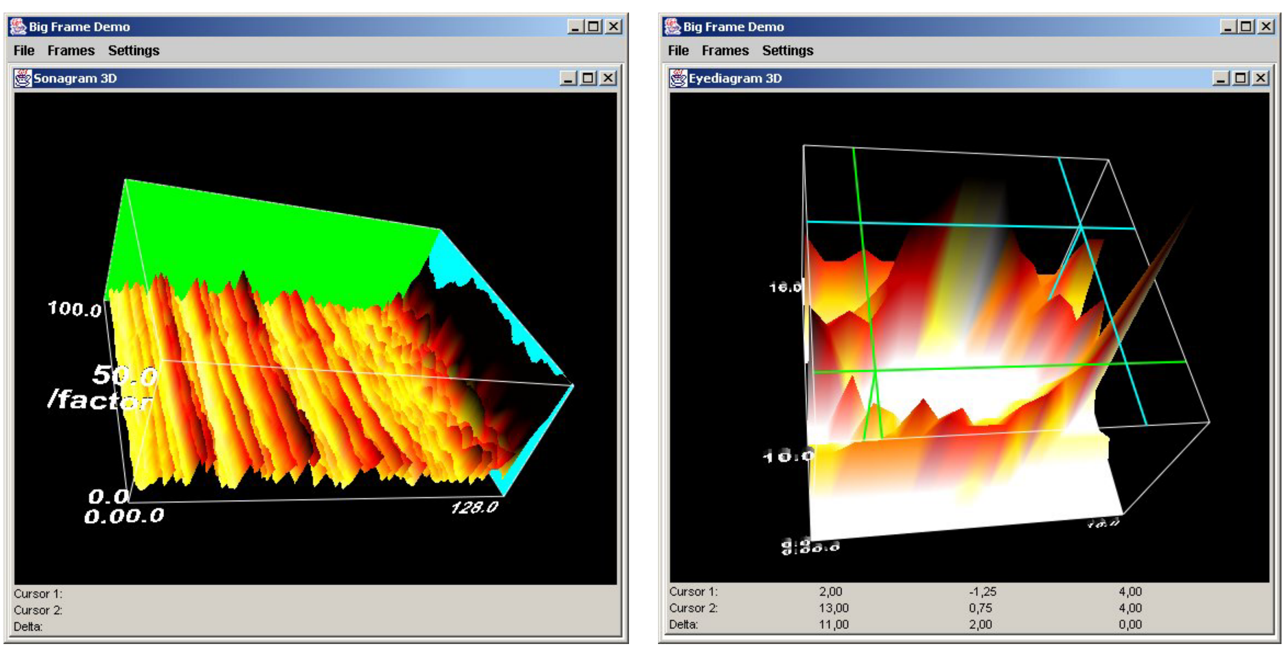


Figure 3: Sonagram in 3D, Eye Diagram in 3D.

5.2 Classical Visualisation

A very simple and easy-to-realise visualisation is to represent data in a diagram or histogram in 2D or 3D. Colors help to show correlation, additional graphics may emphasize results. Figure 4 shows a representation of a command structure obtained by cluster analysis of communications frequency. The command hierarchy is illustrated in the vertical dimension while the colours and shapes are easily interpreted by reference to the index at top-right. (Führung Ebene 1 = leadership level 1, Bearbeiter = administrative assistant, Admin = administration). The y-axis shows written e-mails, the x-axis shows confirmations (or very short replies) on these e-mails. From the type of e-mails and their answers, conclusions of the hierarchy of a company can be drawn.

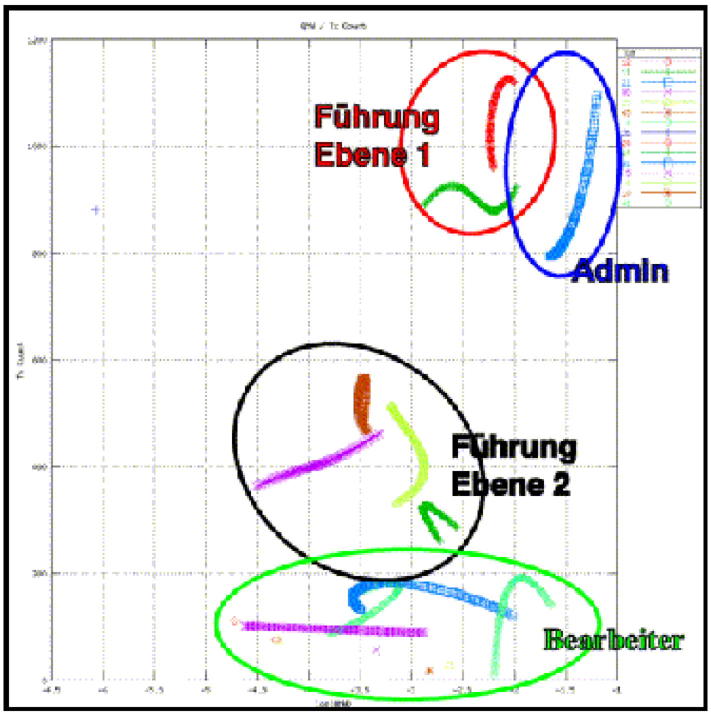


Figure 4: Command Hierarchy.

5.3 Communication Flows

Communication flows can be visualised by lines with different color or thickness depending on their value. In Figure 5 the bandwidth is shown by the thickness of the lines from different places. For intuitive understanding, the network is projected on to a map of Europe. At the first glance, the main network streams can be identified.

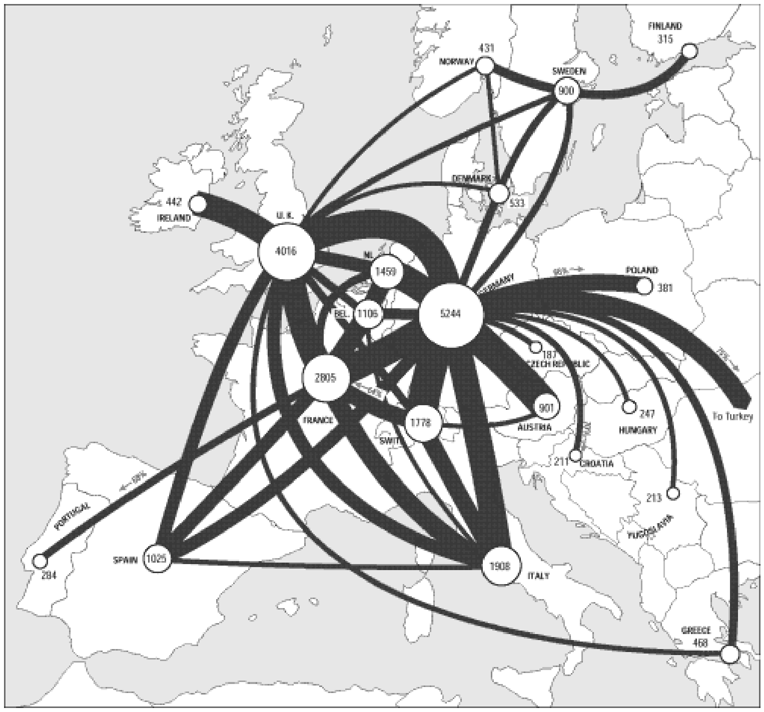


Figure 5: Bandwidth of a communication network.

Another example is the interactive view of communication streams in Figure 6. Instead of the map, a high resolution 3D landscape is shown, and the position of the transmitters is projected onto the site.

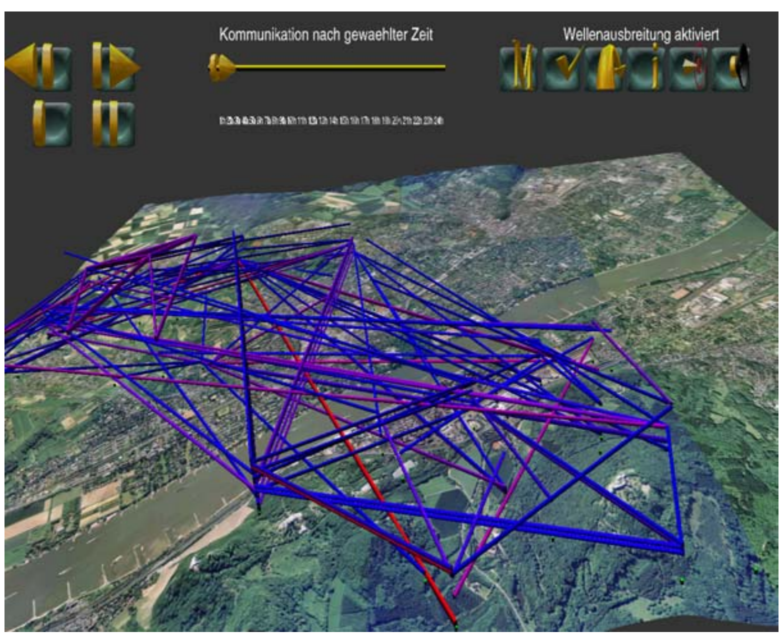


Figure 6: Transmitter traffic in real sites.

The user may scroll the icons at top of the graphic and choose a time and place of interest as well as a choice of options for the display of the transmitters. The color of the communication lines distinguish between data and speech communication. The user has several options for interaction with this visualisation interface (top right of Figure 6).

- Text drill-down: the user can select a transmitter and get text information about this transmitter.
- Audio drill-down: the user can listen to the selected channel.
- Wave propagation: shows the wave propagation around a selected transmitter
- Labeling of a transmitter: e.g. for establishing a hierarchy of transmitters.
- Inspection of transmitters: further properties of the transmitter can be studied.
- Message: observation can be recorded as text.

Another example of displaying a communication flow with many data is a communication cube as in Figure 7. For example, the communication flow over time can be visualised in this way. Two axes in this cube represent the sender and receiver, the third axis represents the time window. The intensity is shown by the size of the dots in the cube. Periodically occurring communications can be seen at a glance. The cube can be rotated or moved by the user in order to obtain a better view to his specific field of interest.

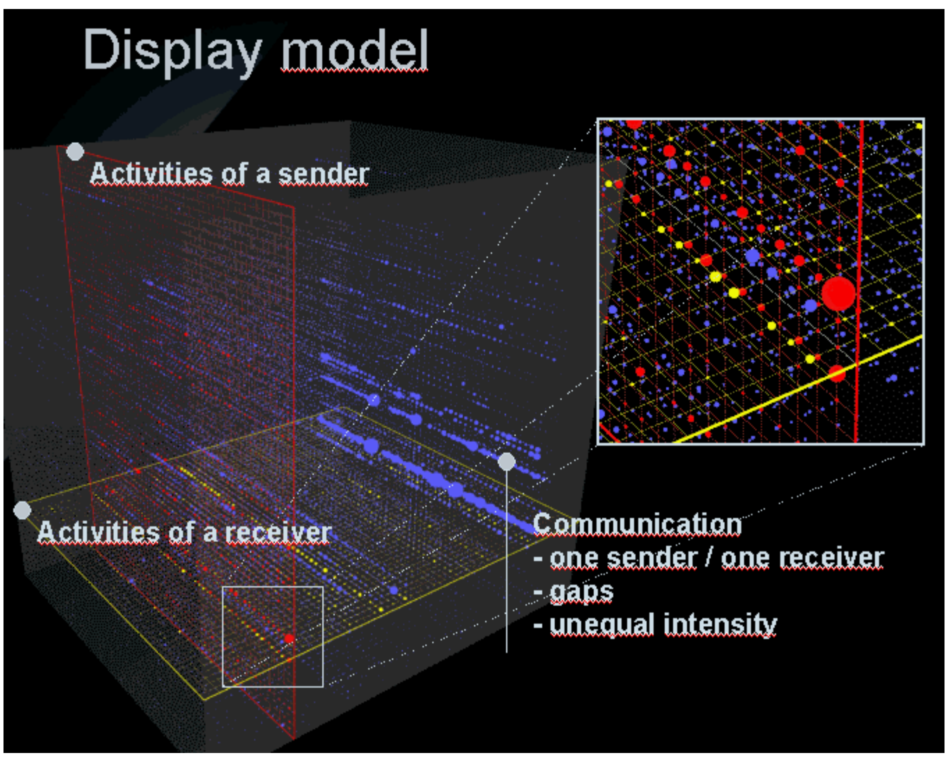


Figure 7: Communication cube.

5.4 Augmented and Virtual Reality

In augmented reality, a live vision is combined with additional information gathered from some type of analysis. For example, a photograph of a city is enriched by data classifying the buildings in the photograph. Interactive systems in augmented reality have special requirements concerning real-time processing in order to obtain an alignment between the calculation and the ongoing reality.

In virtual reality, the real image is substituted by a virtual scenery. There is an interaction between the human user and the virtual reality. An example is shown in Figure 8. A special requirement is the fast and exact estimation of new positions, in this case of the position of the hands of the person as well as an estimation of the view and the eye position of the user. In addition, the technique for displaying the virtual reality may be complex for the calculation of the current position and the update after a movement of the user.

Figure 8: Virtual Reality Application.

5.5 Data/Text Mining

In the examples presented until now, the input to visualisation are numbers, signals. The visualisation process presents the data as they were in a format that can easily and intuitively be understood. Text mining as a specialisation of data mining takes texts as input and visualises the *contents* after analysing them according to algorithms deriving from artificial intelligence research.

The data mining technique is employed when large amounts of data must be analysed. After processing of the data and a classification process, correlations, trends, and dependencies can be found.

Automatic clustering can lead to a categorisation of the input data into different classes as the circles in Figure 4, that show to which group of employees a person belongs to with the highest probability. Another strategy is visualisation of dependencies by means of a classification tree. A classification tree is also trained automatically and can draw conclusions after analysing a data base. For example, let us have a data base with information about car drivers including their age, the size of their car and their frequency of accidents. Using this data base, a classification tree can be trained that estimates the probability of a person being involved in an accident.

Association techniques can be used to find correlations between different events. As an example, shopping lists in the supermarket often show strong correlation between different products. As a result, the products in a supermarket can be sorted according to this correlation. This process can be automated.

Another feature is fuzzy search. Looking for a person whose name or orthographic spelling is not known exactly becomes easier when allowing fuzzy search. The suspected name is entered into the category first name or last name and the persons in the data base are returned who most resemble the data entry.

Text mining takes text or tables as input. Given a data base containing data of the internal communication of employees, different results can be found using text mining: Figure 9 shows the habits of communication for two different users. The user on the left side prefers to write the majority of communication in the morning, the user on the right side has a quite stable quota during the day, with a peak towards the evening. This way, people can be identified. Using other views, the hierarchy of the company can be found, explained by the fact, that each employee writes preferably to his boss.

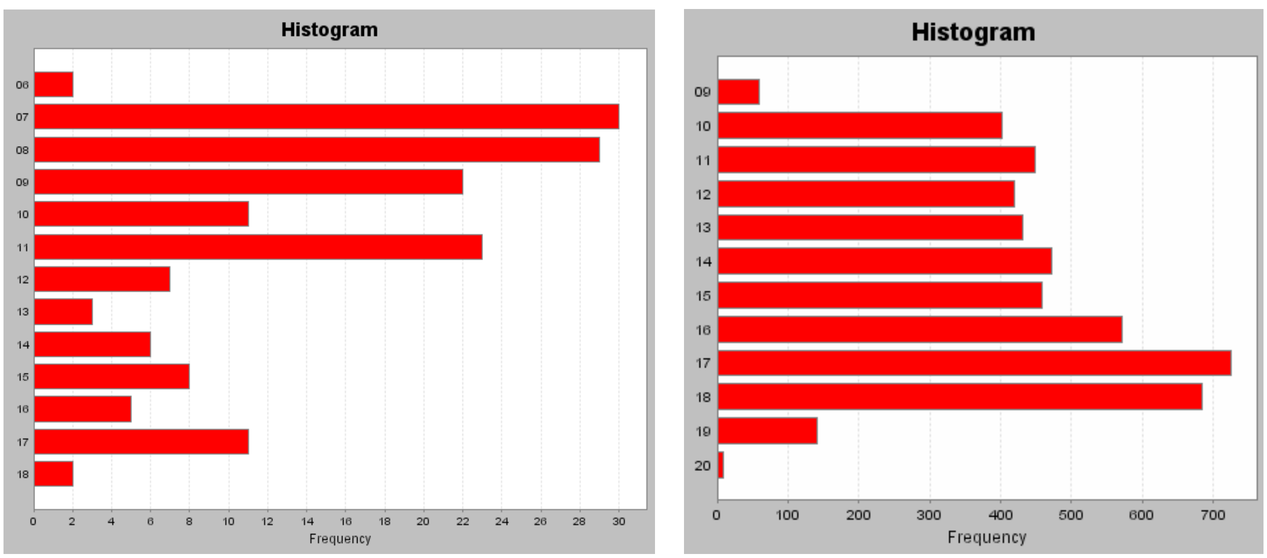


Figure 9: Communication of employees.

5.6 Visual Summary

Another application for visualisation is to generate a visual summary from text files of any kind. There are two possibilities:

1) text tagging: from each text, frequent words and keywords are extracted and brought into relation. The output consists of the original text with highlighted keywords. In addition, the highlighted words contain cross-references to other keywords. Statistics of the text are produced.

2) visual summary: the most frequent words of a text or the associations between words are presented as interactive graphic.

For the visual summary, keywords are extracted from texts: these keywords can be determined automatically as the most frequent words in a text. From the list of keywords, function words (the words that are frequent in any text) are removed resulting to a list of words that are important to this special content.

Now, associations between the keywords are estimated, i.e. the frequency of words occurring in each others' neighbourhood. Drawing an image of the keywords with lines between the words as an indication of their association, interesting relations between persons and actions become visible as in Figure 10.

The image shows the frequency of keywords in a source text of 130,000 words. The large dots describe the frequency of the keywords. The thick lines show the associated words. With this type of graphic, the dependencies in complex texts can be found automatically. On the right side more detailed information on each word is given, e.g. the frequency of each word, the most associated words, links to other occurrences of the same word etc.

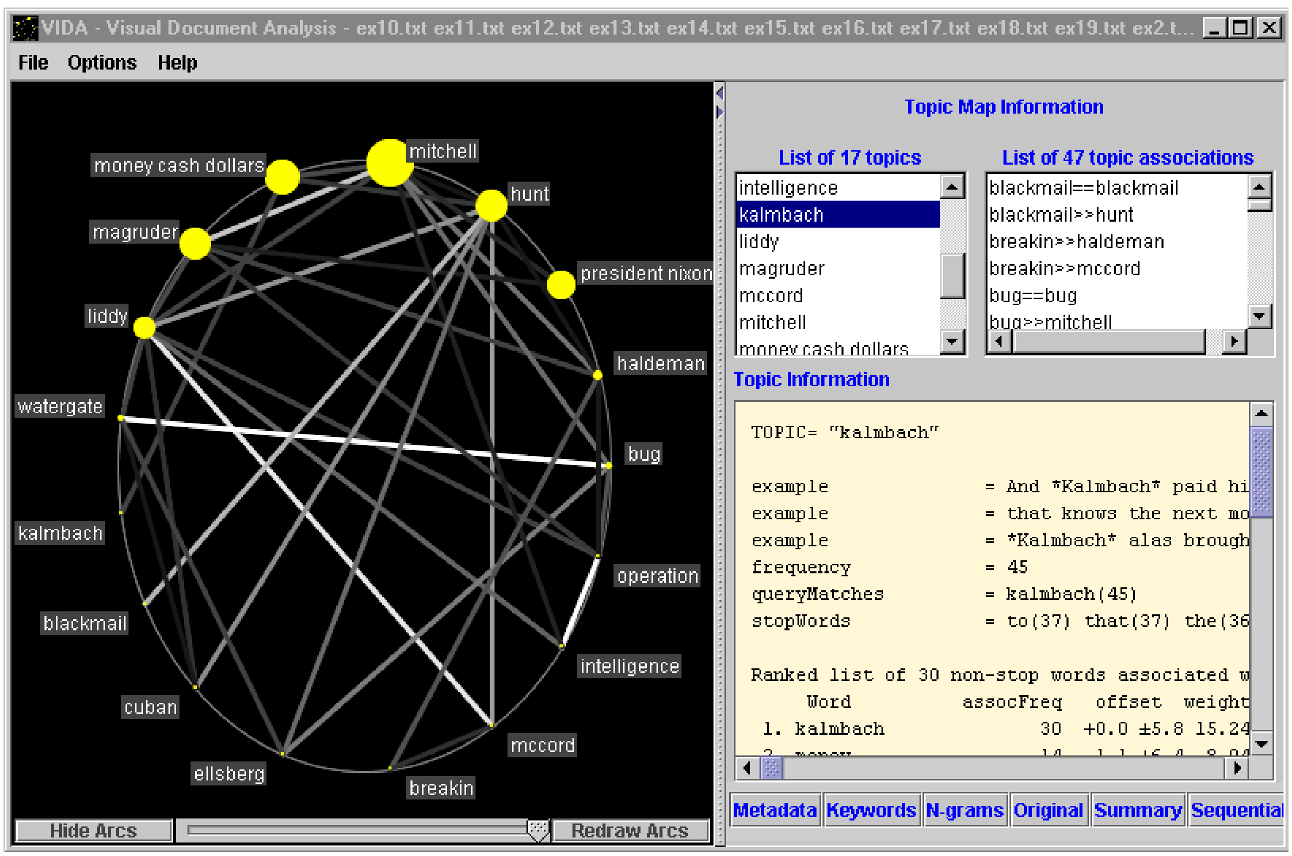


Figure 10: Visual Summary.

6.0 SUMMARY

A picture tells a thousand words. This is true, if the picture is well developed and designed. For information fusion and the visualisation of large amounts of data, it is very important – and becoming more and more important – that there is a good means to reduce the data in a sensible way to quickly obtain the desired information. Some of the ways of designing visual interfaces are intuitive and therefore easy to use. In order to present more complex facts, more of an iconic language may have to be introduced. Still, it seems quite difficult to learn a new iconography that is not obviously intuitive.

This paper showed the results of our work on visualisation that we carried out in order to get the most intuitive visualisation tools for each of the many applications in our studies. As shown, there is a huge potential for the analysis of data bases. Another goal that will be realised in the future is to automatically detect the similarity of data. Using this technique, one can easily find for example all documents belonging to the same topic. For the future, it remains a challenging task for us to continue finding more and intuitive visualisation interfaces for emerging applications.

Web-Based Countrywide Resources Planning System with the Centralized Support of National Mapping Center

Mr. Jaroslav Pejcoch / Mr. Jan Duda
T-SOFT s.r.o / Administration of State Material Reserves
Novodvorska 1010/14, 142 01 Praha / Olbrachtova 3/1677, 140 00 Praha
www.tsoft.cz / www.sshr.cz
CZECH REPUBLIC

Email: pejcoch@tsoft.cz / jduda@sshr.cz

ABSTRACT

The paper discuses a specific area of the information gathering and processing – at the boundary between military and civilian usage. Crisis management is important for both areas, increasingly in the current situation of asymmetric threat and terrorist activities. The availability of resources is one of a key needs for the crisis management. The system ARGIS, described here, represents a part of the information support of the countrywide crisis management system, providing the function of an "information pump" to the responsible staff at various command level in the country. The system is based on a loose reporting structure, based mainly on the rules defined by the legislation and on the activities of the bodies acting in the crisis management. The ARGIS users have a detailed country maps available as a service of a "State mapping centre". All the services of the system are supplied as web-based application, with no specific requirements to the users (except a security ones). The base of the system has been in operation since the year 2001.

FOREWORD

There is a specific area of activities at the military-civilian boundary - Crisis management. In the current period of asymmetric threats, where, besides the "classical" risks of natural and industrial catastrophes, we must count with hostile acts of terrorists, the importance of a co-operation between military and civilian areas is becoming crucial.

Crisis management is an area, where not all obvious tightly managed information input streams are available for use, where the need to know and act is mixed and interfered with the standard human rights, citizen's freedom and business, where for the most of the time does not exist centralized command as in military environment and where many subjects, governed by their own, group, territorial and other aims live and act. One of the most important things to assure during the crisis situations is to have enough resources. Resources of various kinds – material, financial and also some special services.

The military approach to the logistics is in some way easier to accomplish – there is the centralized command, there are well-established messaging structures, there is an obligation and responsibility of each element in the command structure to follow the orders. There is also a system of planned resources, maintained by the military forces or government and dedicated fully just for their usage. In the civil crisis management there is not such a structure or obligations, in spite of this the command, which is taking over the management during the crisis, has to have resources available to save lives, properties and environment.

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

In the “peace-time”, i.e. the time when the region or country is not in the crisis situation, the government and civil organizations have obviously very loose tools for assuring some structured activities. There are only some rules defined by the legislation, which might determine the proper behaviour of various organizations, businesses and citizens in order to be prepared for crisis situations. There is also a factor of the minimization of cost for such an “insurance” of the society. The solution for that has been found in some countries – to minimize the volume of resources directly owned or financed by government and replace this system by the more effective one.

The new approach to crisis management resources planning takes into account the fact, that most of the resources needed to cope with the crisis situation are permanently present in the territory, used for the daily business of both public and private organizations. The base for this new approach is very simple – we could plan for the utilization of various resources that are in the territory during the crises, if we have information on them. This approach minimizes the need to stock a vast amount of material just for the purpose of crisis. The planning is accomplished in a sort of “virtual” and “fuzzy” environment, that is why a substantial information support is a must.

This is the purpose of the system **ARGIS**, which collects information about the various resources in the countrywide scale, having thousands of users in the public and private organizations and which gives the planners and crisis managers tools to fulfil their responsibilities. This system is designed to support planning and use of all available resources in the manufacturing environment in favor of fulfilling the government tasks. For ARGIS system development the modern information technologies have been used. The a goal was to simplify as much as possible the user access to the information, minimize the operation costs and open it for integration with other systems - local, regional and international. As some of the system features (especially the data gathering) are opened to the public, both information security and interoperability have been treated as important factors.

This approach requires, though, a corresponding legislation, which has been valid in the Czech Republic since the year 2000.

Processes Supported by ARGIS

There are several basic processes needed for crisis management.

- Risk analysis
- Preparation of operation plans
- Resources planning
- Actual management of the specific crisis situation.

ARGIS supports the last three of them. The needs for resources to solve any specific crisis situation (flood, earthquake, epidemic, extensive black-out, extensive shortage of oil, monetary crisis, terrorist attack, …) result from operations plans. ARGIS allows summarizing the needs and balancing the resources found in the territory against the needs and report the deficits found to the higher level, up to the national bodies, responsible for the coverage of the whole country.

In the specific crisis situation, the information stored in ARGIS serves as a first point for the resources lookup, getting a fast response for the crisis management teams at various levels of country management.

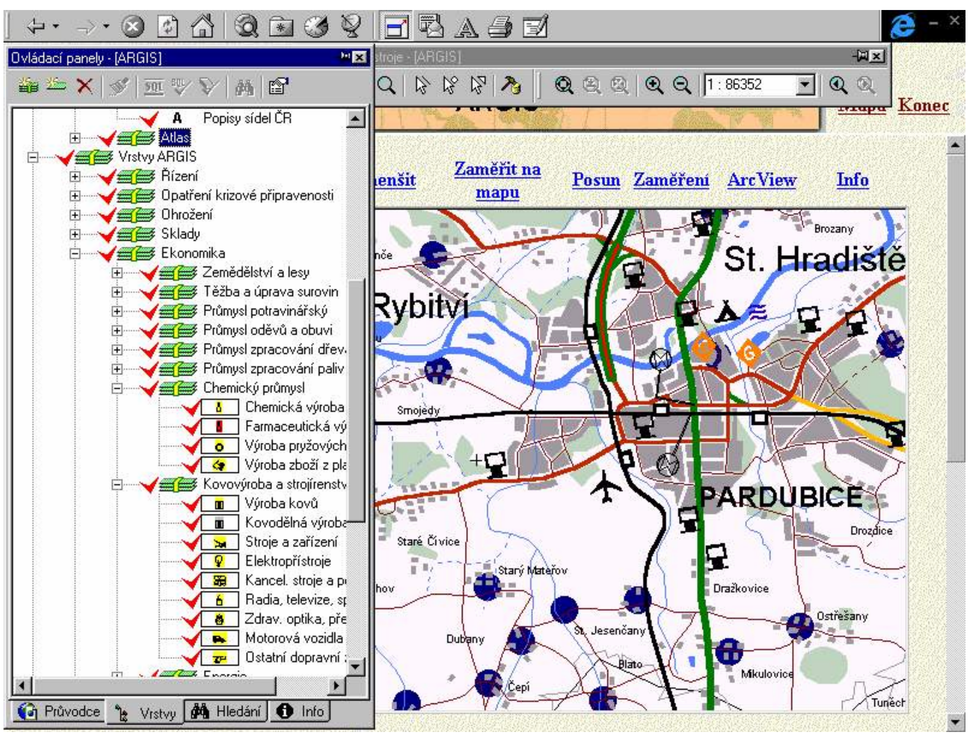
Besides organizational arrangements and standard operation procedures, material means and specific services there are needed during the crisis situation. We are talking about “Necessary Resources” which the government should have prepared in this case. These resources might be assured by the special arrangement at the specific supplier, or by addressing of supplier, which normally produces the requested commodity.

The government deals with the emergency states at the regional level, in case of large disasters or war situation at the central level. Then it needs to fulfill the role of coordinator of the activities. The philosophy of treating of reserves has changed from the simple stocking to the preserving and generating the flexible production of resources.

The activities mentioned require strong information support. Decision-making during the crisis is rather complicated process, carried under stress, with the extreme demand for quality and quantity of verified information. Uncertainty may have large negative consequences. For the qualified decision the crisis management authorities need to process large quantities of information, to sort, group and analyze the information by the current situation needs. This is a purpose of the countrywide information system ARGIS.

GOALS OF THE SYSTEM AND IT BASIC FEATURES

The basic purpose of the system is a permanent collection and maintenance of the information on the resources, with all the linked information. This applies both to the resources maintained by the government for the war situations and any other resources placed on the territory, which belong for example to companies. ARGIS collects all the information describing the production capabilities and also the agreements or contracts with the state. The system filled by all this information eases the planning and utilization of resources. Regions are able to see both their own resources and the resources available in the neighboring regions.



Basic features of ARGIS system are:

- Gathering and permanent updating of information about resources in the countrywide scale.
- Searching for resources based on various criteria (types, properties, ownership, justification of use, geographic location, ...)
- Analytical functions for planning and optimization of the resources.

The system has been build based on the following criteria:

- Availability of the relevant information in the understandable and usable form.
  - Use of cumulative and vivid data (maps, schematics, plans, graphs, ...)
  - Permanent operation of information resources, system backups, information security
  - Ease of user operation
- Efficient system of collecting and updating of information
  - System approach - description of the overall strategy of information acquisition and usage of specialized tools for the selected user actions.
  - Standardization - use of standard indexes and formats as for example
    - Index of economic subjects
    - Standard Products Classification
    - Material Catalog of the Czech Army
    - Territorial ID Index
    - Countrywide index of streets and public places
    - Military and civilian maps background
    - etc.
  - Automatic replication of information with the central management. Automatic replication of information among the system nodes.
  - Information Security
- Support of interoperability of various organizations by cultivating of a common information culture and also by using interoperability standards as for example NATO STANAG5500 (ADatP-3) for the formatted messaging. This and other features allow using ARGIS even in the multinational environment for the cooperation in the area of resources management.

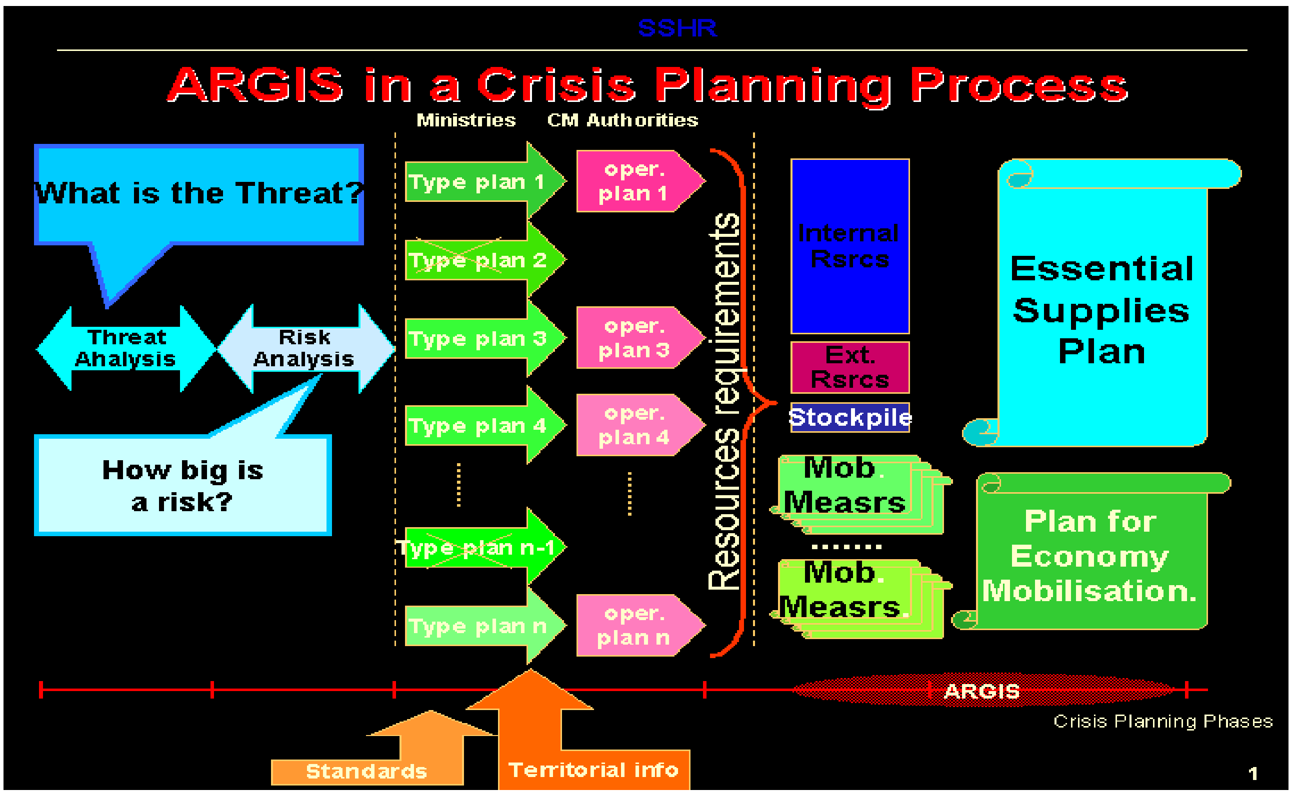
ARGIS ARCHITECTURE

The system is centralized, but distributed for the operation continuity reasons (replicated servers for the case of network interruptions). ARGIS is available to all the government and municipal subjects, dealing with the crisis situations, to the operation centers and to the Central Crisis Staff, which is being activated in time of large disasters or similar crisis situations. It represents a specific resources monitoring system with the instant visualization of the geographical placement of resources of various kinds. It also supports the planning activities – the resources needs and balance evaluation etc.

ARGIS architecture is based on several premises

- Powerful, centralized and distributed web-based countrywide system.
- Automated collection system for easy gathering of information from the subjects, with the minimization of occurrence of errors.
- User interface based on maps with a simple geographical identification of resources, with the possibility of geographical analysis and internet-like user interface
- System interface, enabling working with multiple heterogeneous information resources (various maps, databases, sketches, tables, etc.), enabling also the automatic upgrading and maintenance of software.

The basic ARGIS placement in the crisis-management planning environment is on the following picture:



Base of ARGIS is so called Collector System, used for the permanent collection and update information. At the companies level the Collector is implemented as a special service (ladle), which is assigned by any regional office to a company or other resources provider. Each user is assigned a login and password and is able to begin the data input. The information is collected centrally at the main server of the State Administration of Material Resources and is immediately available to users at all levels.

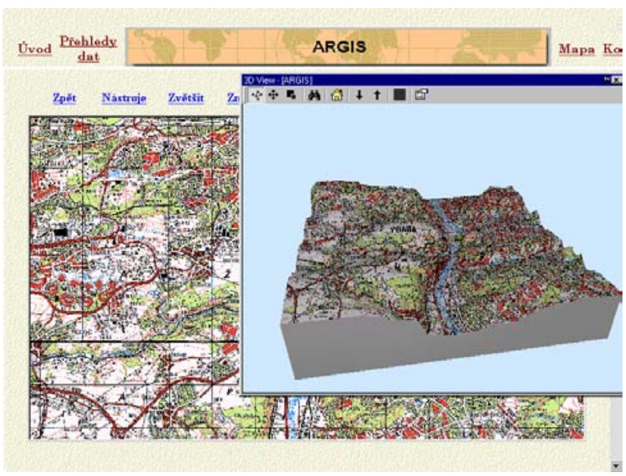
To the objects on the map any type of information including multimedia can be attached. Thus the picture of a building, plans, video sequences, specific charts, spreadsheets, text files and so on may be used. ARGIS is built to use as much as possible of COTS products and technologies.

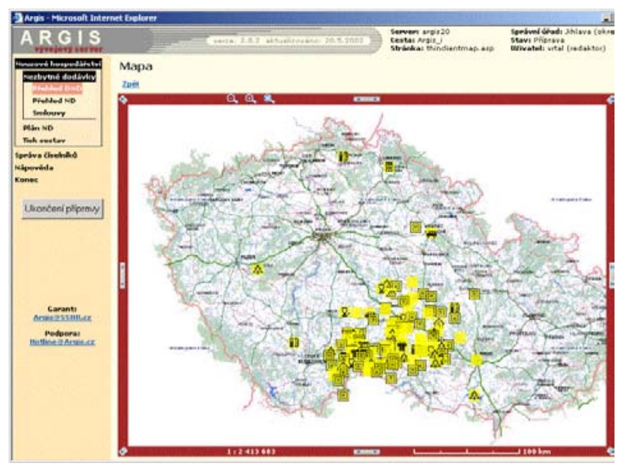
Information is assigned to various layers. These layers are possible to switch on and off separately. To the standard set of layers describing for example roads, rivers, forests, factories, etc. are associated special ARGIS layers describing the economy, based on the standard indexes.

Placing the query about for example cranes of a specific type, ARGIS brings the answer in the list of resources available and display their locations at the map. Similar request may be raised towards the sources of water, specialists and other resources.

Maps and National Mapping Centre

One of the substantial features of the ARGIS system is the possibility to work with all digital maps available in the country, which are maintained by both government and military agencies. This is being accomplished by usage of the "National Mapping Center", which holds all the maps data in various layers and supplies them as a service over the net. The special set of symbols, maintained by the ARGIS management can easily depict the various kinds of resources at the territory as food, machinery, shelters, etc.





COMMUNICATIONS AND SECURITY

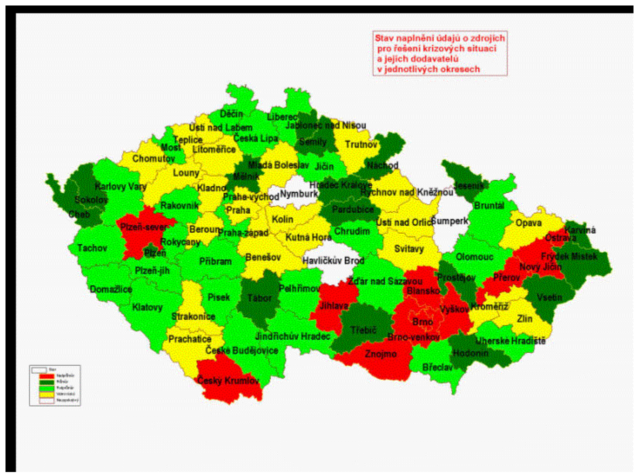
ARGIS operates over the encrypted Virtual Private Network with the information protection at both network and application levels. The usage of encrypted Virtual private network, enables to operate a data collection part over the Internet. This feature allows to handle the data input individually by each subject (companies, entrepreneurs, …), to keep the data actual and to decrease the operation costs and error level in the whole system.

There are various roles assigned to the users. On one side the basic set of functions assigned to the individual company and on the other side the full set of analysis functions over the whole country and the system management.

The Information Inflow

As said, there is not the tight organizational structure behind crisis management during the "peace time" and also the information gathered for the purposes of planning and control is not exact and reliable. Even though the system represents a solid base for the planning and operational decision support.

At the beginning of the ARGIS usage the picture of the system fill up was as on the following picture. One very simple conclusion can be derived out of that – the quality of information is significantly better in the regions, which were hit by the floods in the years 1997 (eastern part of the country – Moravia).



CONCLUSION

System ARGIS represents a modern solution of a countrywide system for the management of information related to the emergency and crisis management. It covers automated data collection from the companies level up to the national center. ARGIS users are regional and national crisis planners and managers in various organizations. The new legislation rules were used to build the information source based on fuzzy information inflow and loose organizational structures in favor the support of crisis planning at the minimal costs.

Usage of Internet as a communication agent together with the encrypted Virtual Private Network technology bring down the operation costs, maintaining sufficient data security.

ARGIS assures the effective access to the relevant information not only in the crisis states, but also during all other emergencies.

Data Collection and Global Data Fusion in C4ISR in the Czech Armed Forces

mjr. Ing. Radek Orsák
VTUL a PVO Praha
Mladoboleslavská ul.
197 06 Praha-9, Kbely
CZECH REPUBLIC
Phone: +420 – 541 182 450
Fax: +420 – 545 246 180
E-mail: radek.orsak@vtul.cz

INTRODUCTION

The beginning of the research of control and information systems started in the deep past. At first the systems were based on simple mechanic and electro mechanic utilities, able to control single element, but during the development of new technologies the systems became much more difficult. These systems allowed to control more elements and their command posts, support operators and their decide processes and also included needed communication infrastructure.

From original purpose to control single weapons systems research gets to building of whole complexes, which provides control of units command, fire coordination, data gathering and so on.

During complex system realizations there is one serious problem, which is not as resoluble by technical tool, as system's intelligence. This problem is in processing a big number of data, fusing its content to explicit context, interpretable by automated systems. This is the main problem of systems for fire control.

HISTORY

Original fire control systems operated with very simple data and knowledge model of battlefield situation. Simple example: target was engaged by element, if the expected path of the target leads directly to the element. The very intelligence of fire elements use, the moment and the method of its use remain on commander's decision.

In near history, about 1989, the Czech Air Forces were using legacy systems K1M and SENĚŽ ME with implemented algorithms of fire control. The algorithms were quite as simple as it has been described in the previous part. The Czech republic stopped using of these systems in 1998 because they were not able to be adapted to work in new environment, in cooperation with new systems and they were not be able to implement new algorithms for control of ground based air defense systems. Despite these facts, when we started development of the new systems (SHARC, LADIC), we were using simple algorithms based on legacy systems.

The gathering information about target position was also taken in very primitive way and the interpretation depended on operator's skills and intelligence. By growing abilities of data gathering systems and data automatic processing the information data capacity get past the system operators abilities. For example for fire control system there is a data gathering process from various sources such as RADAR

Paper presented at the RTO IST Symposium on "Military Data and Information Fusion", held in Prague, Czech Republic, 20-22 October 2003, and published in RTO-MP-IST-040.

rangefinders, altimerts, 3D systems, passive sensors, intelligence systems, etc… Interpretation and processing of this data to superior information is relatively difficult and multilevel operation.

If the operator has to gauge characteristics of only a few objects, his activity can be effective. To gauge the whole battlefield situation when there is a lot of targets is impossible. The characteristics are much more difficult, if the gauge of object in history is needed for gauge its current behavior.

Because in these days the operator is still only a human, we have had to find out a solution how to make the decision easier. It is needed to get important information from the big packet of data – in and create very simple air picture. After that, the operator is able to work with the air picture.

Another problem of fire control is to get together this information with information about fire elements, which are also in various forms and include beside exact information also inexplicit information such as operators skills, and so on. After input of all this information to common picture of the battlefield it is possible to make the optimal decision.

The information about fire units is in very different form – statement of maximum range of missile, maximum range of reconnaissance but also level of operator’s skill or influence of weather to fire result. This information isn’t exact but it is indefinite and mostly it is expressed with measure of persuasion. Common system is not able to work with measure of persuasion by the common math apparatus but it is needed to use for example expert system, fuzzy logic system or neural network.

In scope of development of air defense command and control systems for Czech armed forces, systems SHARC and LADIC, there were improved a technology of data fusion to extract relevant information for fire control operators decide process. Used technologies issues from known and published algorithms and procedures and also from its new implementations and from new novel mechanisms.

HOW TO ACHIEVE THE RIGHT EFFECT?

For the optimal process of command and control we have to see the battlefield like in a single piece. It has its own history, its own actual status and its own foreseeable future. Therefore we can’t control the fire by the way, how the battlefield is looking now, but we have to take its history and requisite future into our consideration.

The whole process starts with collection of information about air picture. The information contains number of targets in the air, their position, identification (if it is available) and other information that is detectable by the primary and secondary detectors. This information includes false alerts too, that it is needed to get away.

In this moment processing of air picture is starting. Next parameters, velocity, alpha code, unified trajectory, are added here. At the same time the false alerts are filtered. At last we have “Recognized air picture (RAP)” but it isn’t the last operation. At the total end we have to join our picture to the information delivered from intelligence service. This is our perfect and exact air picture.

Next task is to collect information about our fire units that we are able to receive from their reports about their status and abilities to fight that they transmit to us. There is something like tertiary processing, where we are able to capture information about the time, when the unit will be able to fight, about the possibility to move and we are able to influence the ability to kill target in the air. This is very important relation. There aren’t only our air defence fire units in the battlefield but there are other friend units and foe units too. We have to get all this information together to create total picture of the ground situation.

If we have “Recognized air picture” and “picture of the ground situation”, we will create “Common operation picture (COP)” at the end. COP is complete description of the battlefield and it is data reality of the real situation. This description is more or less exact; it depends on the precision of data-in and precision of their processing.

It seems to be all in the area of data fusion in C4ISR systems but the opposite is right. If we have complete COP, next task will be making a final decision for own fire unit. By this complicated process we take the COP and our experience from previous battles into our consideration. The operator usually makes the decision by heart but we have to put his experience on software and hardware. So we should create sophisticated knowledge base to save experience and it is needed to create some algorithms to get information out. In this case we can use previous experience to make decision automatically. This process is just the most difficult data fusion process in command and control systems. Nowadays we do our best to find the best solution of this problem for systems SHARC and LADIC. If we solve out this process, our systems will learn to make decision based on previous experience. By this way we have come to the last problem data fusion in C4ISR systems.

As soon as we create decision, we will watch how the battlefield behaves. We still collect data about battlefield behavior and we have to extract the most important things from everything. The fusion with previous experience is following now and it is very difficult and complicated process, of course. Full version of the process has never been included in our systems.

CONCLUSION

Data fusion in C4ISR systems is very difficult problem, which is running at several levels. We could say that data fusion is getting more difficult at every layer. Data fusion is at the lower layers simple math problem but at higher layers we can’t solve needed algorithms by direct math apparatus. Here is a place for neural and experts systems.

Battlefield situation data gathering methods and data fusion methods that are presented in contribution are used in C4ISR systems at the Czech armed forces, especially at Czech air forces. Presented advances and methods issues from work and publications on systems SHARC and LADIC in the Research and Development Institute of Air Force and Air Defense Prague and the Military academy in Brno between 1998 - 2003.

Prepared Remarks and Open Discussion: Critical Issues and Needs in Information Fusion

Wednesday, October 23rd, 2003,
Prague, Czech Republic

PROTOCOL

Discussion Panel Members:

- LtCol. Baron Louis de CHANTAL
- Dr. James LLINAS
- Dr. Per B. SVENSSON
- Dipl. Math. Joachim BIERMANN
- Dr. Jürgen GROSCHE

1.0 DISCUSSION PANEL MEMBERS REMARKS SUMMARY

LtCol. Baron De **CHANTAL**: Prioritisation on operational aspects, especially (1) evaluating of information, (2) relevance and importance of information, (3) visualization / display with respect to specific needs, and (4) main focus on the more complex ground operations, in opposition to an already quite satisfying basis for air and maritime operations.

Dr. **LLINAS**: Comments on Adaptive Information Fusion: Regarding Level 1, the methodological approach and performance evaluation can be seen as satisfying. Beside still existing semantic misinterpretations for level 2 and higher tasks, the role of adaptation (in the Steinberg sense) needs to be defined. Further level 4 refinements are necessary and ongoing research is on processing control itself and multi-algorithm management. For the so called quest for "the golden algorithm", the key elements are named as (1) dynamic thresholding, (2) robustness performance, (3) input control, (4) sensor management and (5) awareness of other input resources. For level 2 and higher level tasks, space/time and condition control and the role of human interaction are as important as computational steering. This should be kept in mind on behalf of technical realizations. One should care more about methods of formal control theory (i.e. provable convergence).

Dr. **SVENSSON**: The fusion process should be seen as real time simulation challenge. Better models have to be developed for closed world representation. Lessons can be learned from simulators and modellers (i.e. Dr. Andreas Tolk). One should look at relations between the Simulation and real world including the C4I community and apply cognitive modelling and theory to the simulations. New developments should include (1) agent simulations, (2) tool simulation control frameworks with ontology's as building blocks for simulators, (3) evaluation of closed world models with representation of uncertainty, and (4) human-in-the-loop training for improved trust and confidence.

Mr. **BIERMANN**: The information overload problem has to be tackled both in a semantic / syntactic and ontological approach. Decision support has currently not left the stage of intelligent information management. There are interesting studies in incorporating high level knowledge into mature level 1 methods for level 1 problems like physical object classification and identification, but the applicability of level 1 mathematical methods to level 2 and 3 problems like abstract domain object aggregation is still an open question. For an appropriate approach to level 2 and 3 fusion and to gain the still lacking knowledge for a deep understanding of the problem domain, priority should be given to more cooperative experiments together with military personnel. Results from such studies and already available information and knowledge about high level fusion should be used for thorough scientific analysis and modelling of the battle-space domain and its information fusion methods.

Dr. **GROSCHE**: In summary it can be said, that a common understanding of the state of research in the area of data and information fusion related to JDL level 2 and higher was reached. But one thing is the theoretic models another is the applicability. Further outcomes are (1) scepticism for higher level data fusion automation, so far only weapon automation systems are operational, (2) improvement in a feedback-cycle process with military personnel is necessary, (3) intelligent data "compression" due to bandwidth problems and the graceful degradation topic have to be developed, (4) more system flexibility with regard on network-centric-warfare tasks is necessary. (5) No ontology is possible without customers common language, (6) Ontology is a way of communication between developer and domain experts, (7) for level 1 scenarios (e.g. tracking) challenges as test and competitions have to be promoted as (8) overall information fusion test scenarios have to be created for method evaluation on level 2 fusion and higher levels.

2.0 OPEN DISCUSSION

- Dr. Shahbazian: There is a problem with communication between people working with fusion. They are not always aware of others work, partly because they come from different domains with different ontology. Why is it so and how can we improve the communication between people performing fusion or fusion-like work?
  a) Dr. Svensson: The international conference on Information Fusion taking place Summer 2004 is an opportunity.
  b) Dr. Llinas: A survey in US education system on fusion showed that no organised education on fusion aspects existed. Many of the individual topics are covered but they are synthesised into fusion. Fusion is a matter of thinking and remains a 'colleague industry'.
  c) Dr. Grosche: An event like this symposium is a way to increase awareness on Information Fusion.

- Dr. Snajder: Proposes a competition. US offers text containing information with terrorist activities to the R&D industry which then build systems to handle this information. An open competition will increase the capability of the systems.

- Dr. Rohmer: See similarities between the military and civilian use, especially with respect to process control: the mixing of knowledge based systems with information fusion systems.
  a) Mr. Biermann: The Research Task Group in Information Fusion Demonstration is organising a specialist meeting broadening the topic outside the military use and would very much appreciate hints on people from civilian domains who can contribute to this event.

- Mr. Porta: Interested in fusion on web-based data and found Dr. Hecking`s paper very interesting and important for use outside the military. I would extent the scope to web security and intrusion-detection. But why is it a problem to go outside level 1 fusion for the information technology field?
  a) Mr. Biermann: Network Centred applications is a new way of seeing the problem domain: It is our task to deal with the increasing amount of information. We can abstract from the military use and go back to general research on information fusion. Important is the research for situation awareness independent from information load and independent from system architecture.
  b) Dr. Llinas: There are many problems with higher level fusion: Problems with association and estimation, assessing the quality of information, the organisational aspects, who are the

participants? A number of dimensions which influence the development of the domain have to be taken into account. Therefore new strategies have to be set up.

- Dr. Gentleman: It has been mentioned that 'human in the loop' is important but of equal importance is visualisation which has to take a cognitive model of the problem area into account. A complex graph does not help in itself.

  a) LtCol. Baron de Chantal: Visualisation is more than just display issues. We need a stronger attachment to requirements. It should be coupled to the instant needs of the users and provide with just the needed information and nothing more.

  b) Dr. Llinas: Visualisation has to be matched with the mental understanding and is especially important in Time Critical Missions. We should focus on actionable information and impedance matching in information management. This means to set up new paradigms for display/ visualization problems.

  c) Dr. Svensson: What the user wants to see is a model of the opponent forces, including aspects of uncertainty. Further we should keep in mind, that not only on the tactical level a strong hierarchical dependency of information exists. Not everybody needs all kind of information, and the same information has to be displayed differently e.g. for assessment on different military levels.

  d) Mr. Biermann: A main disadvantage of visualisation community was that they tried to model the user without having a sufficient understanding of the problem domain. We need to have a better understanding of how the user processes the information, e.g. a staged development of intelligence assistance and support of operators. Approaches starting basically with "user profiles" do not concentrate on still existing fundamental problems. Complete understanding of the active process of information fusion is most important and therefore preferable.

- Dr. Uebler: Comments to visualisation aspects: We have time constraints and need to optimise the systems. But a VR-simulation helps the solider to orientate in urban areas. We have to distinguish between display/visualization – confusion. Means of visualization and human cognitive capabilities have to be brought into agreement.

Mr. Biermann gave an announcement for a planned follow-up meeting in 2005 and expresses the strong interest of the members of the RTO Task Group on Information Fusion Demonstration (RTG on IFD) to enlarge the involved community to the e.g. homeland-security and criminal investigation field.

Final remarks and conclusions were given by Dr. Grosche.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO-MP-IST-040 AC/323(IST-040)TP/23	ISBN 92-837-0037-6	UNCLASSIFIED/ UNLIMITED

5. Originator

Research and Technology Organisation
North Atlantic Treaty Organisation
BP 25, F-92201 Neuilly-sur-Seine Cedex, France

6. Title

Military Data and Information Fusion

7. Presented at/Sponsored by

The Symposium organised by the RTO Information Systems Technology (IST) Panel in Prague, Czech Republic, from 20 to 22 October 2003.

8. Author(s)/Editor(s)	9. Date
Multiple	March 2004

10. Author's/Editor's Address	11. Pages
Multiple	328 (text) 541 (slides)

12. Distribution Statement

There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.

13. Keywords/Descriptors

C4ISR
Correlation techniques
COTS (Commercial Off The Shelf)
Data fusion
Decision making
Geo-spatial intelligence
Identification systems
Information fusion
Information systems
Integrated systems
Intelligence fusion
Interoperability
Lessons learned
Models
Multisensors
Natural language processing
Ontologies
Operations research
Requirements
Scenarios
Situational awareness
Systems engineering
Tactical intelligence
Target classification
Threat evaluation
Visualization techniques

14. Abstract

This volume contains two Keynote Addresses, 24 papers, and a summary of the final panel discussion presented at the Information Systems Technology Panel (IST) symposium held in Prague, Czech Republic, from 20th to 22nd October 2003.

The papers presented covered the following headings:

- Military Requirements and Experiences
- Fusion System Concepts I
- Fusion System Concepts II
- Fusion Methods
- Fusion Methods for Classification and Identification
- Semantic Approach to Information Fusion
- Applications and Lessons Learned
- Applications

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DIFFUSION DES PUBLICATIONS
RTO NON CLASSIFIEES

Les publications de l'AGARD et de la RTO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre est la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la RTO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (www.rta.nato.int) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Friedrich-Ebert-Allee 34, D-53113 Bonn

BELGIQUE
Etat-Major de la Défense
Département d'Etat-Major Stratégie
ACOS-STRAT – Coord. RTO
Quartier Reine Elisabeth
Rue d'Evère, B-1140 Bruxelles

CANADA
DSIGRD2
Bibliothécaire des ressources du savoir
R et D pour la défense Canada
Ministère de la Défense nationale
305, rue Rideau, 9e étage
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1, P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

ETATS-UNIS
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GRECE (Correspondant)
Defence Industry & Research General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Centro di Documentazione Tecnico-Scientifica della Difesa
Via XX Settembre 123
00187 Roma

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25, NO-2007 Kjeller

PAYS-BAS
Royal Netherlands Military Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE
Armament Policy Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE
DIC Czech Republic-NATO RTO
VTÚL a PVO Praha
Mladoboleslavská ul.
Praha 9, 197 06
Česká republika

ROYAUME-UNI
Dstl Knowledge Services
Information Centre, Building 247
Dstl Porton Down
Salisbury
Wiltshire SP4 0JQ

TURQUIE
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320
ETATS-UNIS

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Acquisitions, Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants :

Scientific and Technical Aerospace Reports (STAR)
STAR peut être consulté en ligne au localisateur de ressources uniformes (URL) suivant:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR est édité par CASI dans le cadre du programme NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
ETATS-UNIS

Government Reports Announcements & Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 2216
ETATS-UNIS
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

NORTH ATLANTIC TREATY ORGANISATION

RESEARCH AND TECHNOLOGY ORGANISATION

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DISTRIBUTION OF UNCLASSIFIED RTO PUBLICATIONS

AGARD & RTO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO reports, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your Organisation) in their distribution.

RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of RTO reports as they are published, please visit our website (www.rta.nato.int) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Etat-Major de la Défense
Département d'Etat-Major Stratégie
ACOS-STRAT – Coord. RTO
Quartier Reine Elisabeth
Rue d'Evère
B-1140 Bruxelles

CANADA
DRDKIM2
Knowledge Resources Librarian
Defence R&D Canada
Department of National Defence
305 Rideau Street
9th Floor
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC
DIC Czech Republic-NATO RTO
VTÚL a PVO Praha
Mladoboleslavská ul.
Praha 9, 197 06
Česká republika

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72
92322 Châtillon Cedex

GERMANY
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Friedrich-Ebert-Allee 34
D-53113 Bonn

GREECE (Point of Contact)
Defence Industry & Research General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HUNGARY
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ICELAND
Director of Aviation
c/o Flugrad, Reykjavik

ITALY
Centro di Documentazione Tecnico-Scientifica della Difesa
Via XX Settembre 123
00187 Roma

LUXEMBOURG
See Belgium

NETHERLANDS
Royal Netherlands Military Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25, NO-2007 Kjeller

POLAND
Armament Policy Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide, P-2720 Amadora

SPAIN
SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

TURKEY
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

UNITED KINGDOM
Dstl Knowledge Services
Information Centre, Building 247
Dstl Porton Down
Salisbury, Wiltshire SP4 0JQ

UNITED STATES
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076-1320

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320
UNITED STATES

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
STAR is available on-line at the following uniform resource locator:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
UNITED STATES

Government Reports Announcements & Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 2216
UNITED STATES
(also available online in the NTIS Bibliographic Database or on CD-ROM)

ISBN 92-837-0037-6