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.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **00 MAR 2004**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE
Analysing an Identity Information Fusion Algorithm Based on Evidence Theory

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Defence Research and Development Canada Valcartier 2459 Pie-XI Blvd North Val-Bélair, Quebec G3J 1X5 CANADA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001673, RTO-MP-IST-040, Military Data and Information Fusion (La fusion des informations et de données militaires)., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**46**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

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 Pl(A)=1-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.

*D*1 **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.

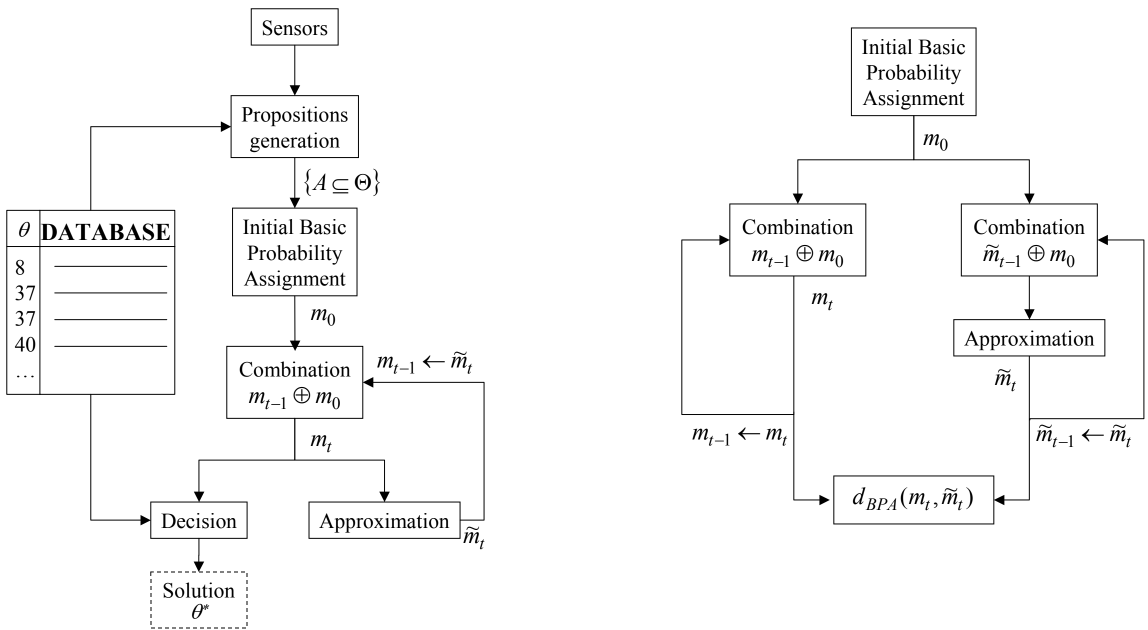
The distribution of the masses of the prunes focal elements lies on successive iterations and accounts for the relations between these focal elements and the remaining ones (objects in common, etc.). After the last iteration, the remaining mass is affected to the frame of discernment, which is then the 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-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]:

$$NS(m) = \sum_{A \subseteq \Theta} m(A) \log_2(|A|) \tag{8}$$

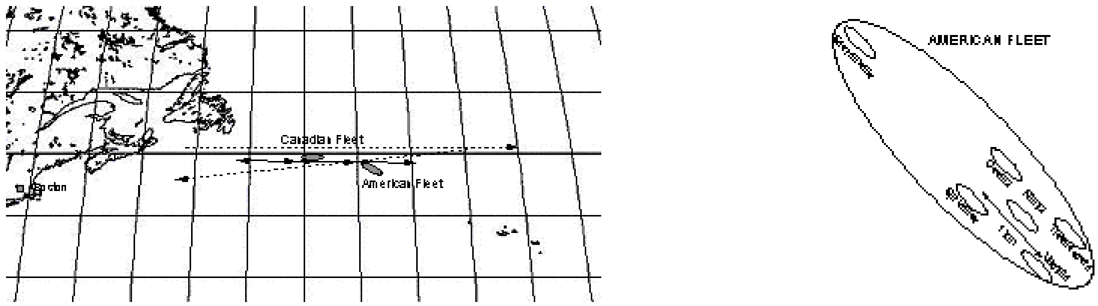
where |.| denotes the cardinality. The expression is widely accepted as the only one measure of non-specificity in evidence theory, because it is the only function satisfying the five properties of (1) additivity (for non-interactive BPAs), (2) subadditivity (for interactive BPAs), (3) normalization, (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.

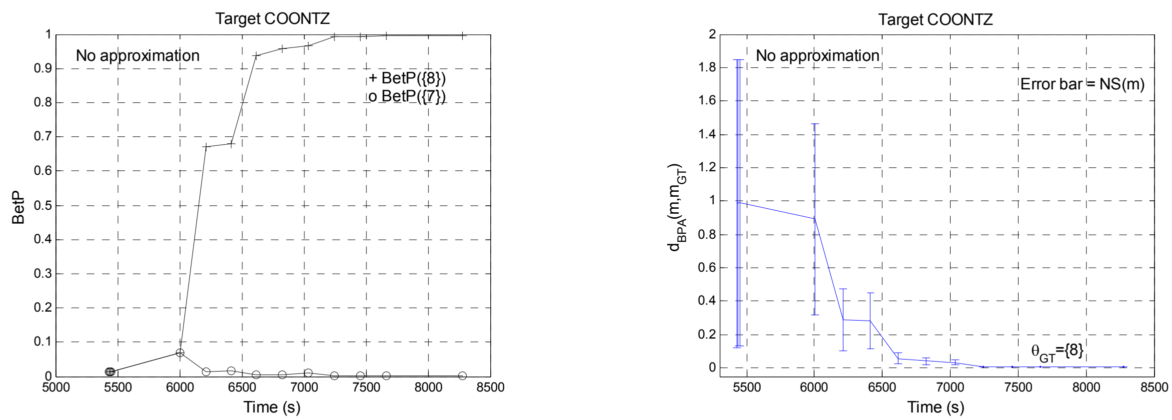
One thus expects a larger amount of false associations of ESM-to-track than in the Canadian fleet case, due to the combined effect of large (classified) ESM bearing accuracy and intrinsic filtered radar track accuracy. In order to follow the ID evolution, Table 1 shows the emitter list for the American 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 $A_i = \{\theta \in \Theta | \ \theta$ 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.



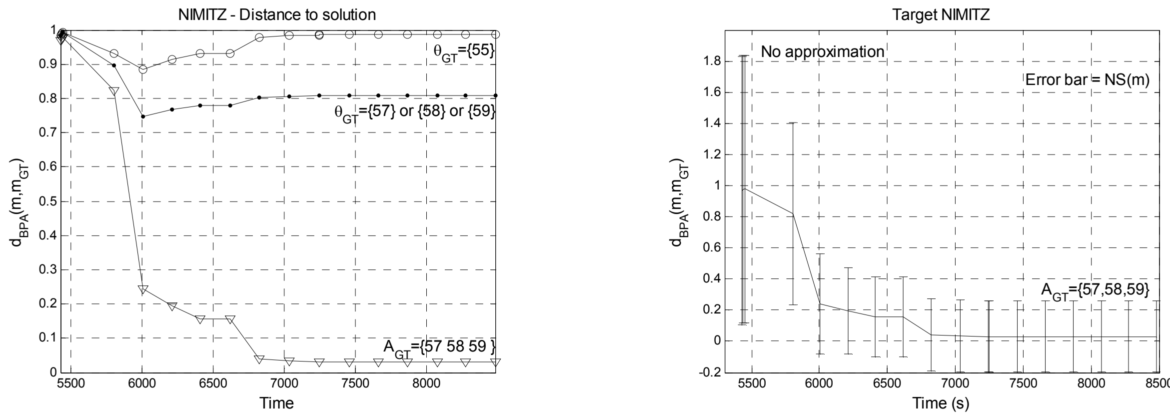
(a) Pignistic probability of singleton {8} without approximation

(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 $\theta_{GT}=\{57\}$, 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.

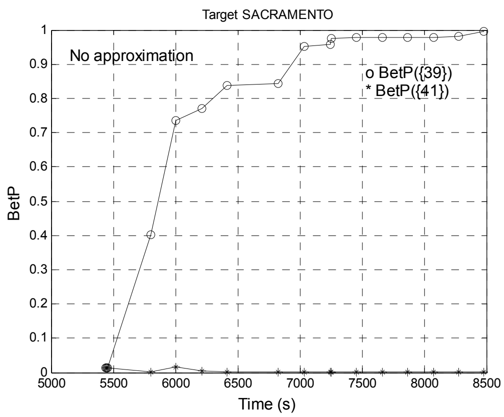
4.4 Convergence to a "Wrong" Solution

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.

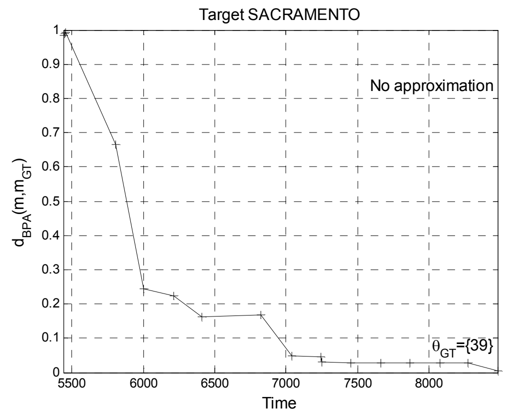
4.5 Impact of the Approximation Rule

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.

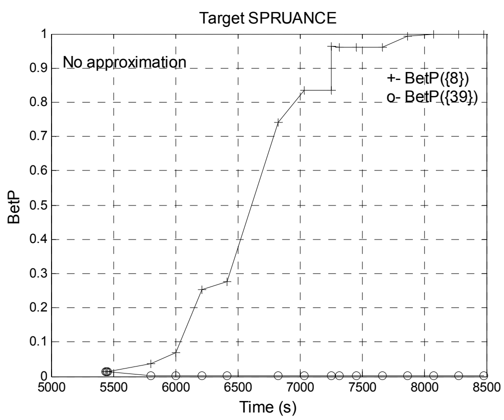
The comparisons of $D1$ and k-l-x algorithms for the six targets of the scenario are shown on figures 7-(a) and 7-(b), with a number of focal elements set to 8. Following these results the two algorithms are equivalent, both leading in finality to very low distances to the original BPA (*i.e.* without approximation), and this although a previous analysis on Monte-Carlo simulations identified $D1$ algorithm as the best one [3, 10].



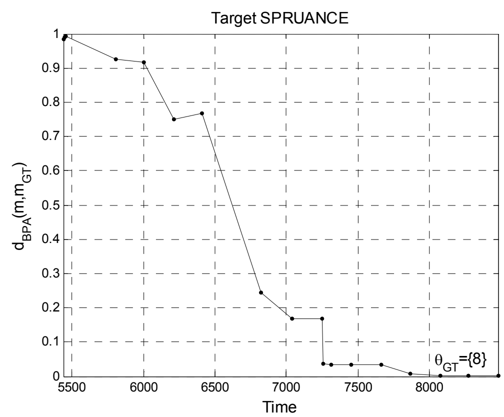
(a) Pignistic probability of singletons {41} and {39} for target SACRAMENTO.



(b) Distance to solution m_{GT} ({39})=1 for target SACRAMENTO.

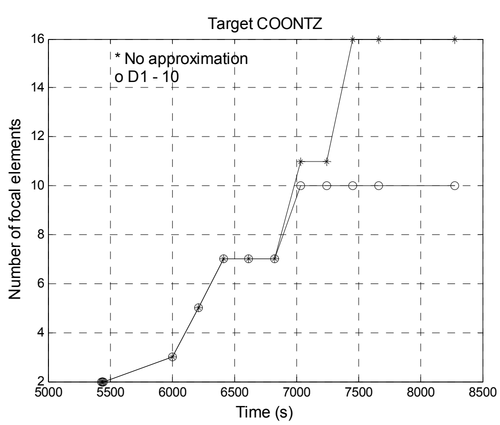


(c) Pignistic probability of singletons {39} and {8} for target SPRUANCE.



(d) Distance to solution m_{GT} ({8})=1 for target SPRUANCE.

Figure 5: Convergence to wrong solutions due to errors of association.



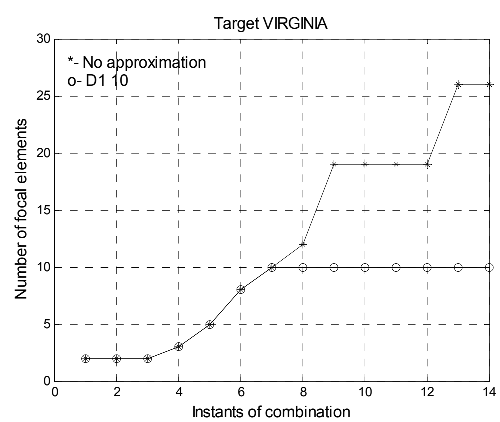


Figure 6: Impact of the approximation rule on the number of focal elements – Targets COONTZ and VIRGINIA - Algorithm $D1$ with k =10.

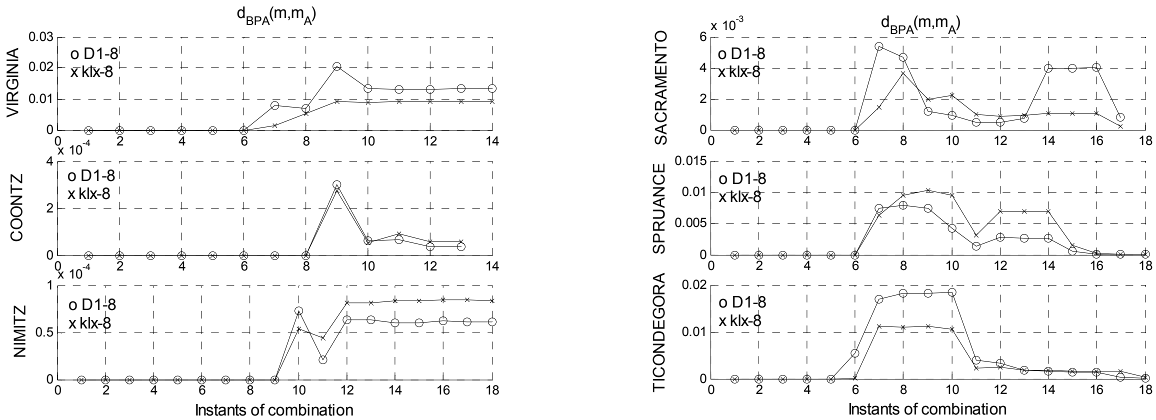


Figure 7: Comparison of approximation algorithms *klx* -8 and *D1*-8 based on the distance to the original BPA.

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) $d_{BPA}(m_t,m_{\mathrm{GT}})$ 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 $d_{BPA}(m_t;m_{\mathrm{GT}})$ 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.

DEFENCE DÉFENSE

Analysing an Identity Information Fusion Algorithm Based on Evidence Theory

Anne-Laure Jousselme, Éloi Bossé, Alexandre Jouan

'Military Data and Information Fusion', NATO IST-040 RSY-012
Prague, Czech Republic, 20-22 October 2003

Defence R&D Canada R et D pour la défense Canada

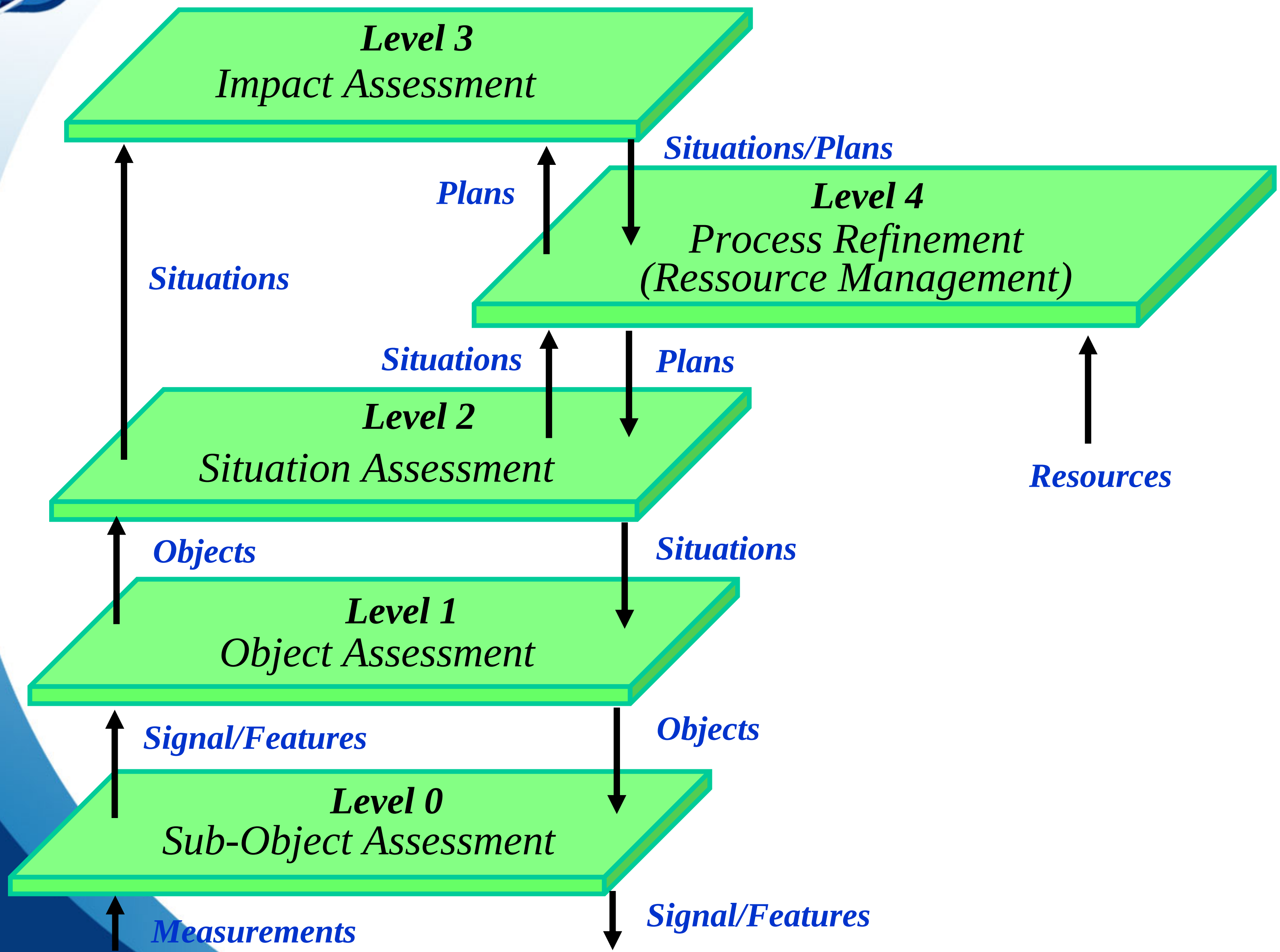
Canada

Outline

- Data fusion
- Identity information fusion process
- Dempster-Shafer approach
- Analysis of an algorithm
- Conclusion

The JDL Data Fusion Model

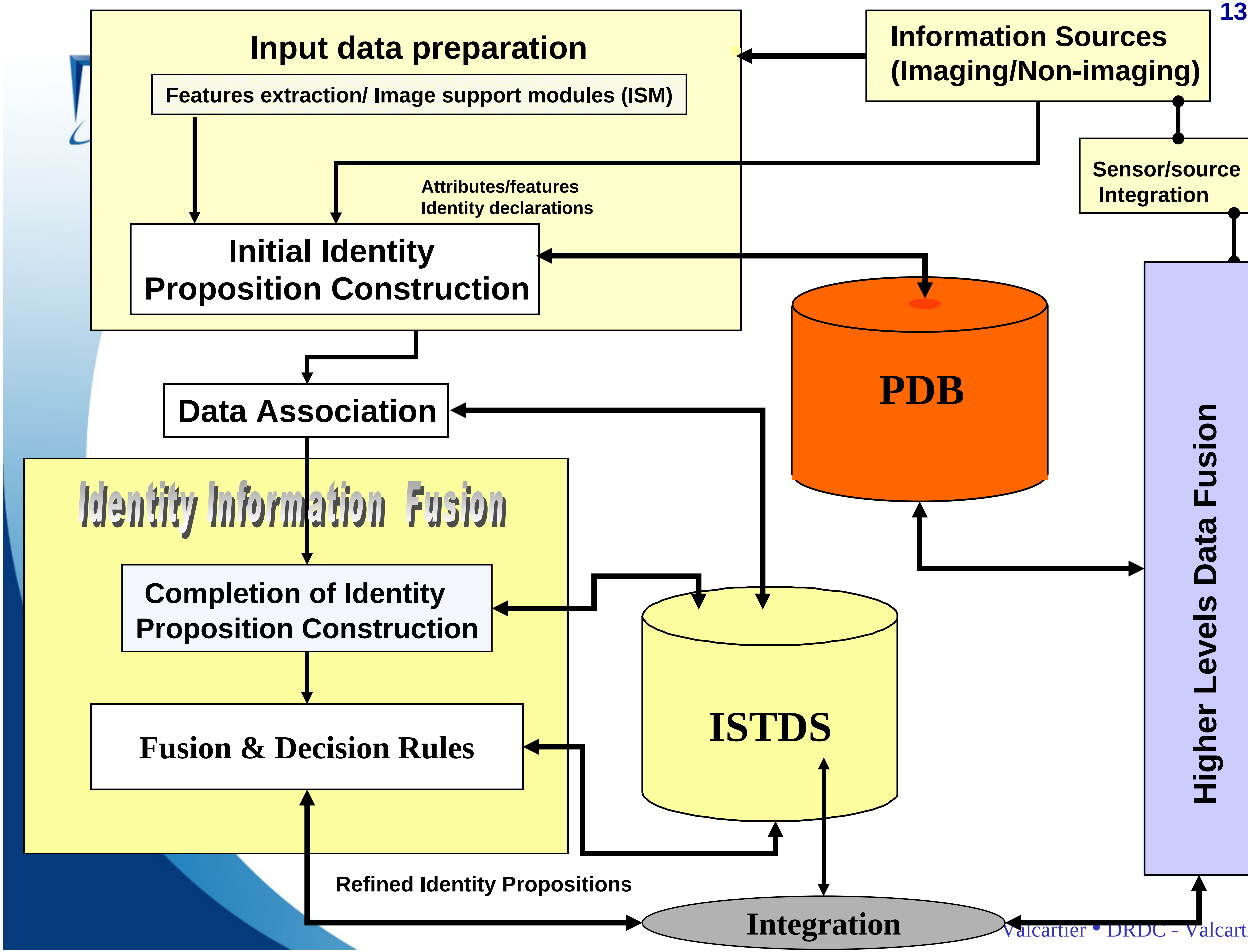
(Revised JDL Model: A. Steinberg/C. Bowman/F. White, 1998)



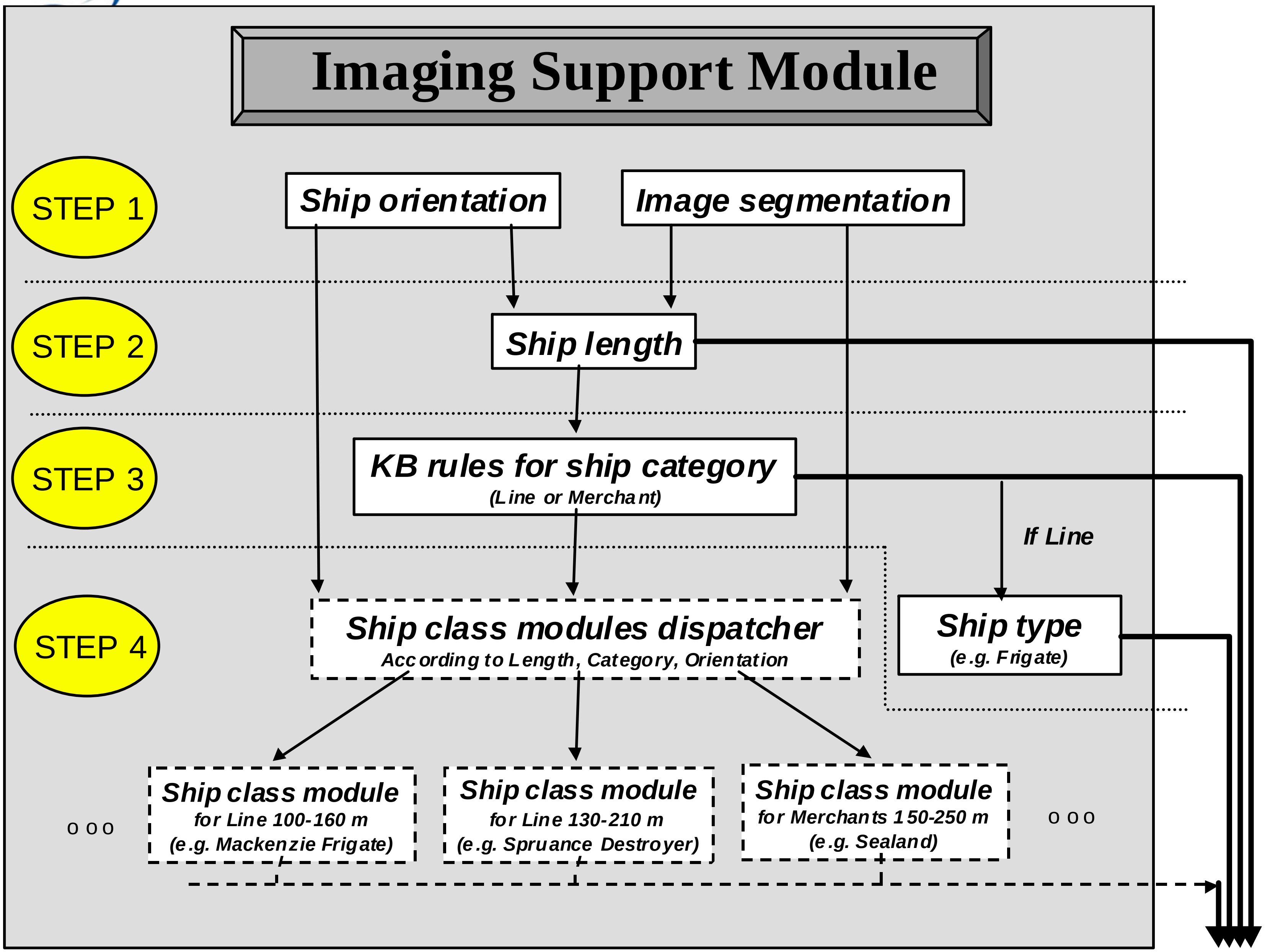
Integration of All DF Levels

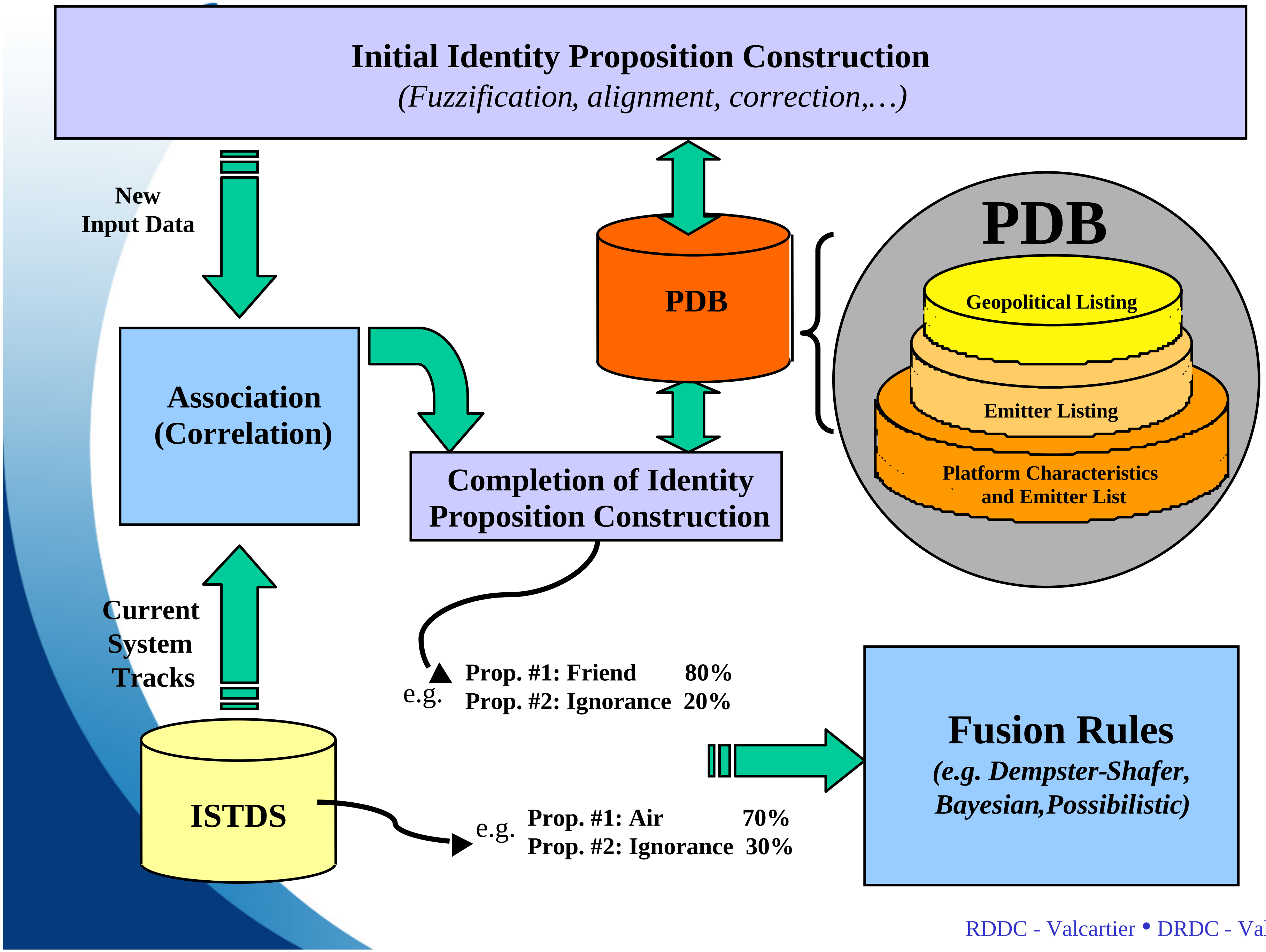
- The data fusion model implicitly supposes that all levels are implemented within a closed loop framework (level 4: process refinement).
- The levels of fusion are linked together and their tight integration is essential to gain the maximum benefits from this process.
- Functional decomposition of the data fusion process is different whether the process is implemented as an open or a closed loop

Identity Information Fusion Process

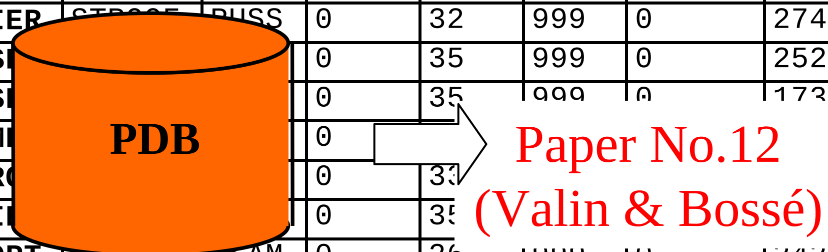

Input data preparation
Features extraction/ Image support modules (ISM)
Information Sources
(Imaging/Non-imaging)
Sensor/source
Integration
Attributes/features
Identity declarations
Initial Identity
Proposition Construction
PDB
Data Association
Higher Levels Data Fusion
Identity Information Fusion
Completion of Identity
Proposition Construction
ISTDS
Fusion & Decision Rules
Refined Identity Propositions
Integration

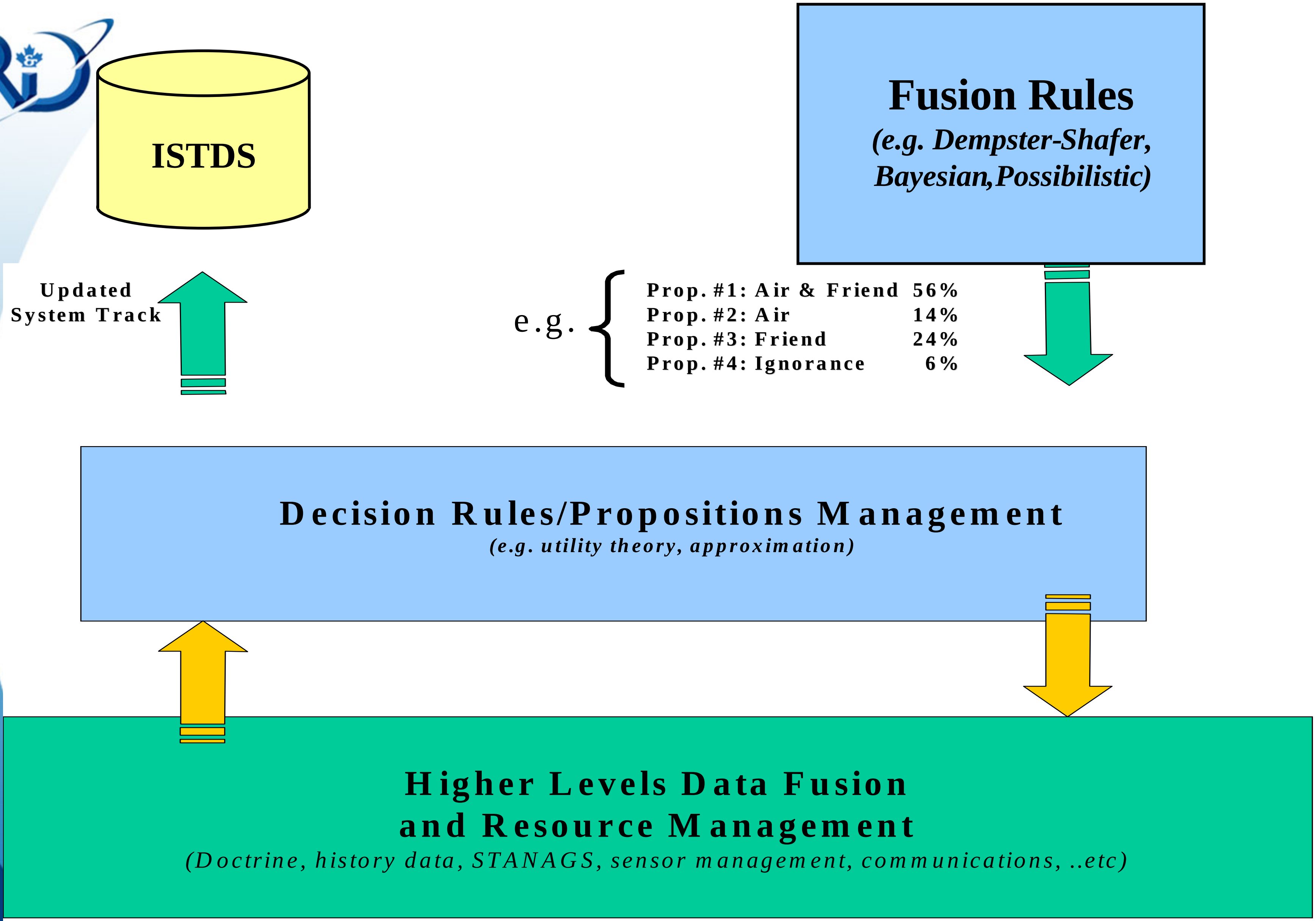
Input Data Preparation



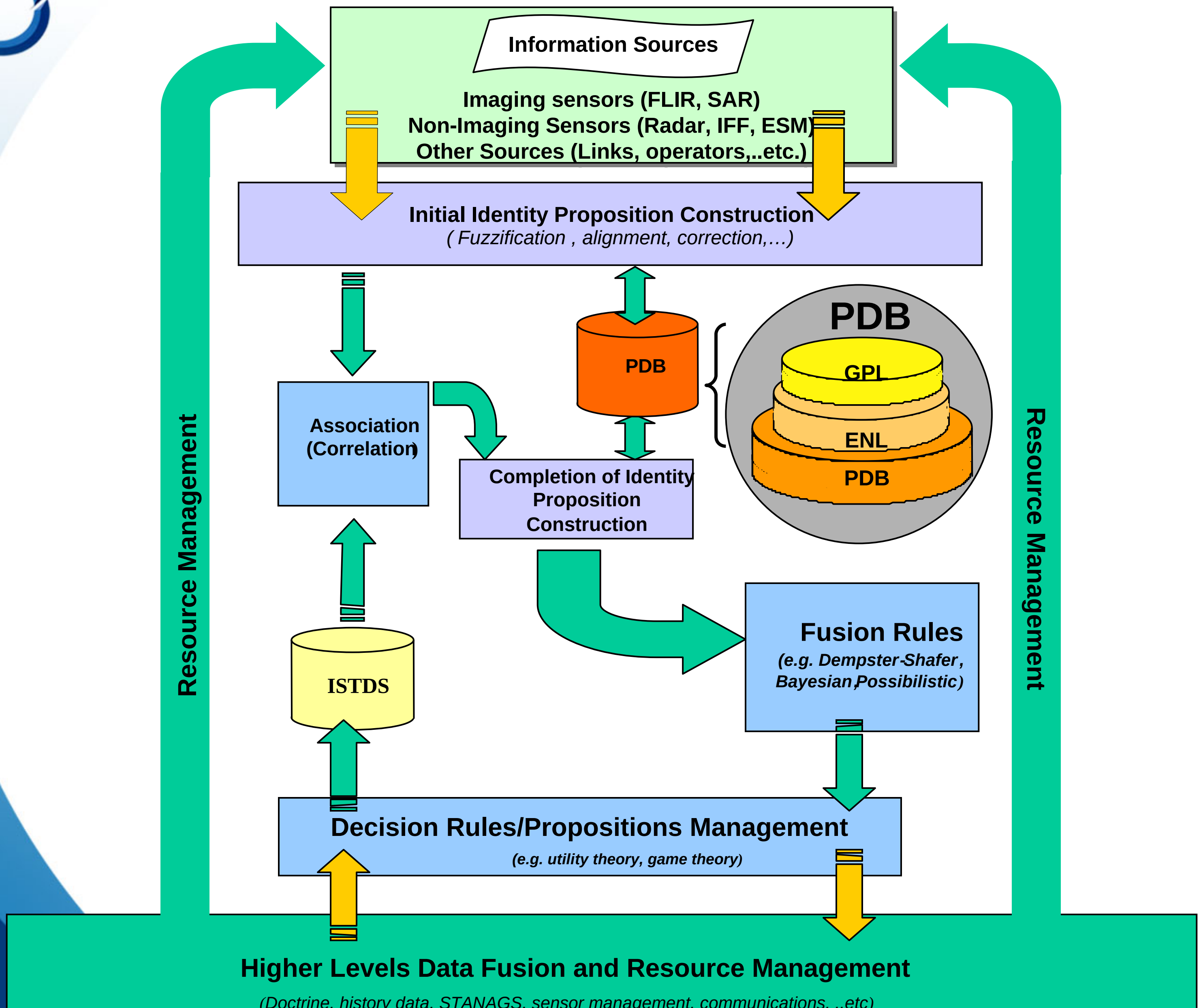

Initial Identity Proposition Construction
(Fuzzification, alignment, correction,…)
New
Input Data
Association
(Correlation)
PDB
Completion of Identity
Proposition Construction
PDB
Geopolitical Listing
Emitter Listing
Platform Characteristics
and Emitter List
Current
System
Tracks
e.g.
Prop. #1: Friend 80%
Prop. #2: Ignorance 20%
ISTDS
e.g.
Prop. #1: Air 70%
Prop. #2: Ignorance 30%
Fusion Rules
(e.g. Dempster-Shafer,
Bayesian,Possibilistic)

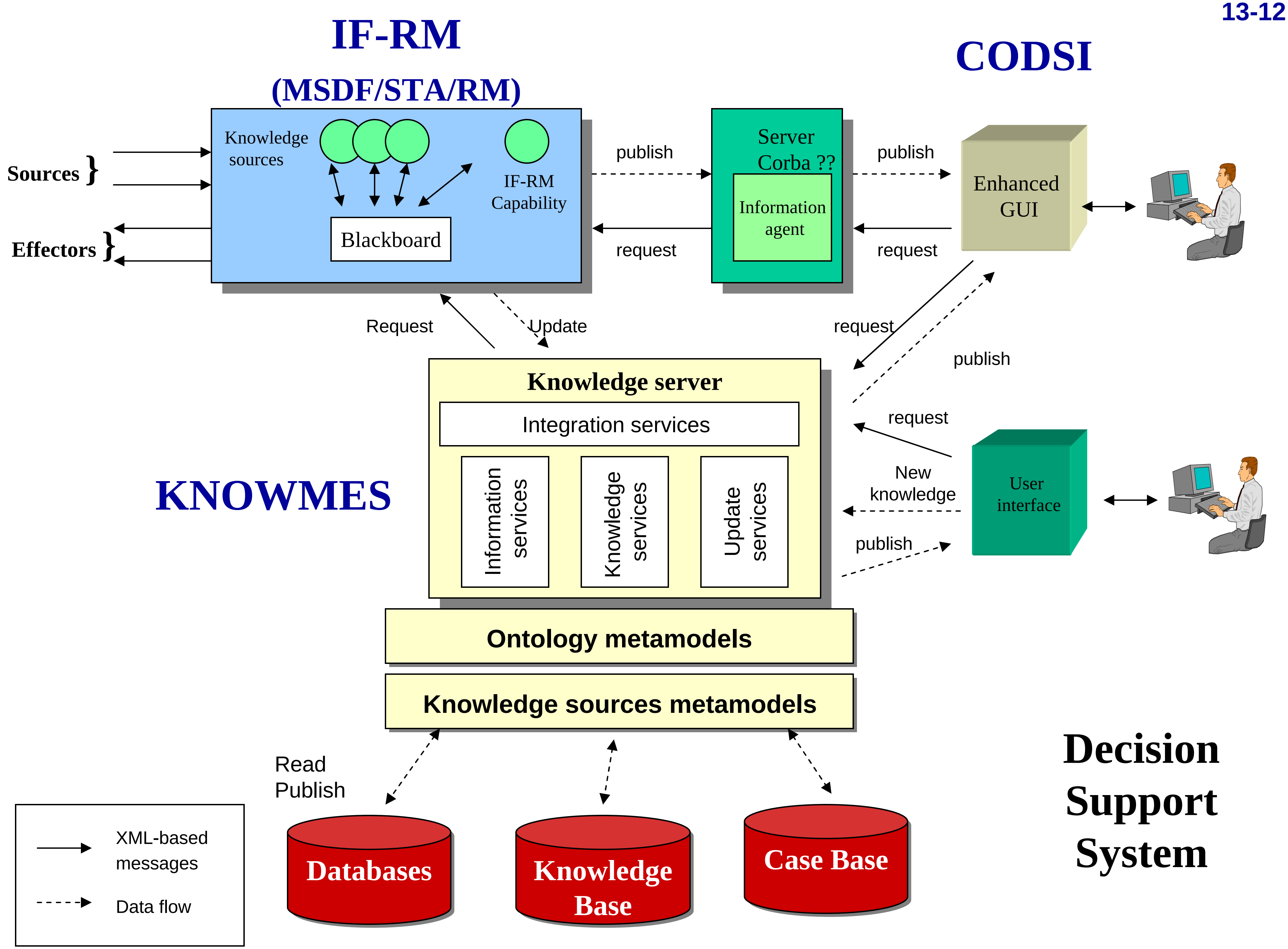
ID #	NAME----------------	PLATYPE	SUBTYPE	OFFENS	CONT	V_ MINI	V_ MAXI	ACC	ALT_ MAXIM	LEN	HEI	WID
000001	**JAHRE-VICKING-------**	SURFCOM	**TANKERV**	HARMLE	DANM	0	35	999	0	460	33	51
000002	**HALIFAX-CPF---------**	SURMILI	**FRIGATE**	MEDIOF	CANA	0	35	999	0	130	5	16
000007	**CALIFORNIA----------**	SURMILI	**CRUISER**	STROOF	USAM	0	35	999	0	182	10	19
000008	**COONTZ--------------**	SURMILI	**DESTROY**	STROOF	USAM	0	35	999	0	156	5	16
000017	**VIRGINIA------------**	SURMILI	**CRUISER**	STROOF	USAM	0	35	999	0	178	10	19
000018	**MIRKA-I-------------**	SURMILI	**FRIGATE**	WEAKOF	RUSS	0	32	999	0	82	3	9
000019	**MIRKA-II------------**	SURMILI	**FRIGATE**	WEAKOF	RUSS	0	32	999	0	82	3	9
000025	**IROQUOIS------------**	SURMILI	**DESTROY**	WEAKOF	CANA	0	30	999	0	130	5	15
000026	**ADELAIDE------------**	SURMILI	**FRIGATE**	MEDIOF	AUST	0	30	999	0	138	5	14
000027	**IMPROVED-PROVIDER---**	SURMILI	**SUPPORT**	VEWEOF	CANA	0	21	999	0	172	9	23
000028	**QUEST---------------**	SURMILI	**MISCELL**	HARMLE	CANA	0	11	999	0	72	5	13
000029	**KNOX----------------**	SURMILI	**FRIGATE**	MEDIOF	EGYP	0	27	999	0	134	5	14
000030	**IVAN-ROGOV----------**	SURMILI	**ASSAMPH**	WEAKOF	RUSS	0	25	999	0	158	8	25
000031	**KARA-KERCH----------**	SURMILI	**CRUISER**	STROOF	RUSS	0	35	999	0	173	7	19
000032	**MODIFIED-KIEV-------**	SURMILI	**CARRIER**	STROOF	RUSS	0	32	999	0	274	10	51
000033	**KIROV-ADM-USHAKOV---**	SURMILI	**CRUIS**[illegible]	[illegible]	[illegible]	0	35	999	0	252	9	29
000037	**TICONDEGORA---------**	SURMILI	**CRUIS**[illegible]	[illegible]	[illegible]	0	35	999	0	173	10	17
000038	**TARAWA--------------**	SURMILI	**ASSAM**[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]	[illegible]	[illegible]	8	40
000039	**SPRUANCE------------**	SURMILI	**DESTR**[illegible]	[illegible]	[illegible]	0	33	[illegible]	[illegible]	[illegible]	6	17
000040	**NIMITZ--------------**	SURMILI	**CARRI**[illegible]	[illegible]	[illegible]	0	35	[illegible]	[illegible]	[illegible]	11	41
000041	**SACRAMENTO----------**	SURMILI	**SUPPORT**	WEAKOF	USAM	0	26	999	0	242	12	33
000042	**KIROV-ADM-NAKHIMOV--**	SURMILI	**BATTLES**	VESTOF	RUSS	0	35	999	0	252	9	29
000043	**KIROV-ADM-LAZAREV---**	SURMILI	**BATTLES**	VESTOF	RUSS	0	35	999	0	252	9	29
000044	**KIROV-PYOTR-VELIKIY-**	SURMILI	**BATTLES**	VESTOF	RUSS	0	35	999	0	252	9	29
000045	**KARA-AZOV-----------**	SURMILI	**CRUISER**	STROOF	RUSS	0	35	999	0	173	7	19
000046	**KARA-PETROPAVLOVSK--**	SURMILI	**CRUISER**	STROOF	RUSS	0	35	999	0	173	7	19
000047	**KARA-VLADIVOSTOK----**	SURMILI	**CRUISER**	STROOF	RUSS	0	35	999	0	173	7	19
000064	**UDALOY-II-----------**	SURMILI	**DESTROY**	STROOF	RUSS	0	30	999	0	164	8	19
000065	**UDALOY-AND-KULAKOV--**	SURMILI	**DESTROY**	MEDIOF	RUSS	0	30	999	0	164	8	19
000066	**SOVREMENNY-II-------**	SURMILI	**DESTROY**	STROOF	RUSS	0	32	999	0	156	7	17
000067	**SOVREMENNY-OSMOTRITE**	SURMILI	**DESTROY**	STROOF	RUSS	0	32	999	0	156	7	17
000068	**UDALOY-SPIRIDONOV---**	SURMILI	**DESTROY**	MEDIOF	RUSS	0	30	999	0	164	8	19
000074	**TYPHOON-------------**	SUBSURF	**NUCPSTR**	STROOF	RUSS	0	26	999	900000300	165	13	25




ISTDS
Fusion Rules
(e.g. Dempster-Shafer, Bayesian,Possibilistic)
Updated System Track
e.g.
Prop. #1: Air & Friend 56%
Prop. #2: Air 14%
Prop. #3: Friend 24%
Prop. #4: Ignorance 6%
Decision Rules/Propositions Management
(e.g. utility theory, approximation)
Higher Levels Data Fusion and Resource Management
(Doctrine, history data, STANAGS, sensor management, communications, ..etc)

Identity Information Fusion Process




IF-RM
(MSDF/STA/RM)
CODSI
Sources
Effectors
Knowledge sources
Blackboard
IF-RM Capability
publish
request
Server Corba ??
Information agent
publish
request
Enhanced GUI
Request
Update
request
publish
Knowledge server
Integration services
Information services
Knowledge services
Update services
KNOWMES
request
New knowledge
publish
User interface
Ontology metamodels
Knowledge sources metamodels
Read
Publish
Databases
Knowledge Base
Case Base
XML-based messages
Data flow
Decision Support System

Dempster-Shafer Approach

Evidential Reasoning

Frame of discernment: $\Theta = \{1,2,\ldots,N\}$ (elements of PDB) Ignorance, singletons

Power set: $P(\Theta) = 2^{\Theta}$ (subsets of PDB)

Basic Probability Assignment: $\sum_{A \in P(\Theta)} m(A) = 1$ and $m(\phi) = 0$ (A, subset of PDB)
(Probability that the object is in A)

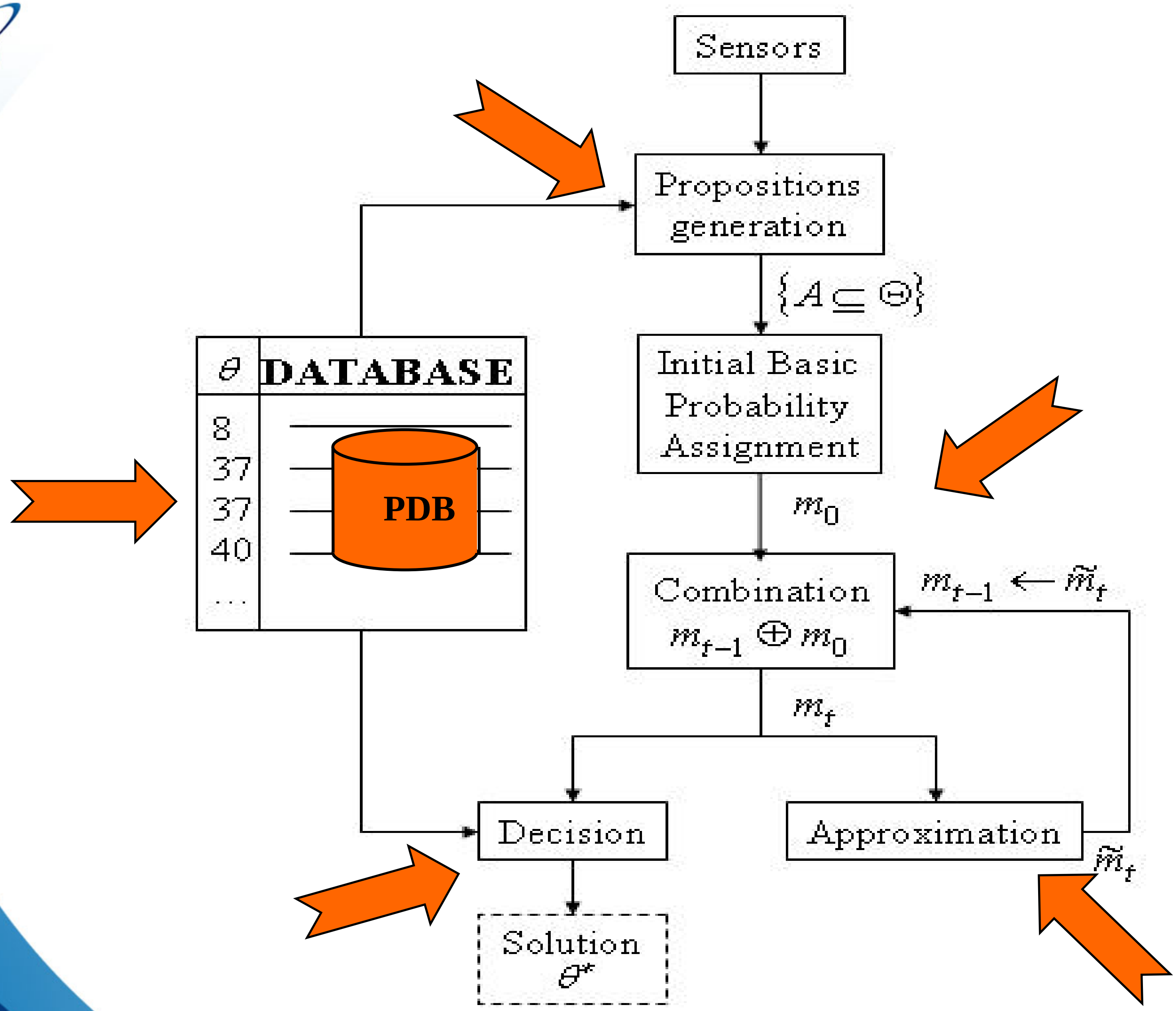
Belief: $bel(A) = \sum_{B \subseteq A} m(B)$ *Plausibility:* $pl(A) = \sum_{A \cap B \neq \phi} m(B)$

Body of evidence: $\{B, m\} = \{[A_1, m(A_1)], [A_2, m(A_2)], \ldots, [A_n, m(A_n)]\}$

Set of focal elements A ($m(A) \neq 0$) associated to their masses.

Fusion rule (Dempster 's combination):

$$m(A) = \frac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - K} \quad \text{with} \quad K = \sum_{A_i \cap B_j = \phi} m_1(A_i) m_2(B_j)$$


Sensors
Propositions generation
$\{A \subseteq \Theta\}$
Initial Basic Probability Assignment
m_0
Combination $m_{t-1} \oplus m_0$
$m_{t-1} \leftarrow \tilde{m}_t$
m_t
θ DATABASE
8
37
37
40
...
PDB
Decision
Approximation
$\tilde{m}_t$
Solution θ^*

Analysis

- Pignistic probability of the singletons of the database as a decision rule
- Distance between a BPA and the ground truth
- Distance between an approximated BPA and an original
- Non-specificity of the BPAs

Decision

$$\theta^{*} = Arg\left[\max_{\theta\in\Theta} BetP(\theta, m)\right]$$

and

$$BetP(\theta, m) = \sum_{\theta\in A\subseteq\Theta} \frac{m(A)}{|A|}$$

Distance between two BPAs

$$d_{BPA}(m_1, m_2) = \sqrt{\frac{1}{2}\left[\left\|\vec{m}_1\right\|^2 + \left\|\vec{m}_2\right\|^2 - 2\left\langle \vec{m}_1, \vec{m}_2 \right\rangle\right]}$$

$$\left\langle \vec{m}_1, \vec{m}_2 \right\rangle = \sum_{i=1}^{2^N} \sum_{j=1}^{2^N} m_1(A_i) m_2(A_j) \frac{\left|A_i \cap A_j\right|}{\left|A_i \cup A_j\right|}$$

Distance to the solution (GT)

For any BPA solution focused on one set A^*:

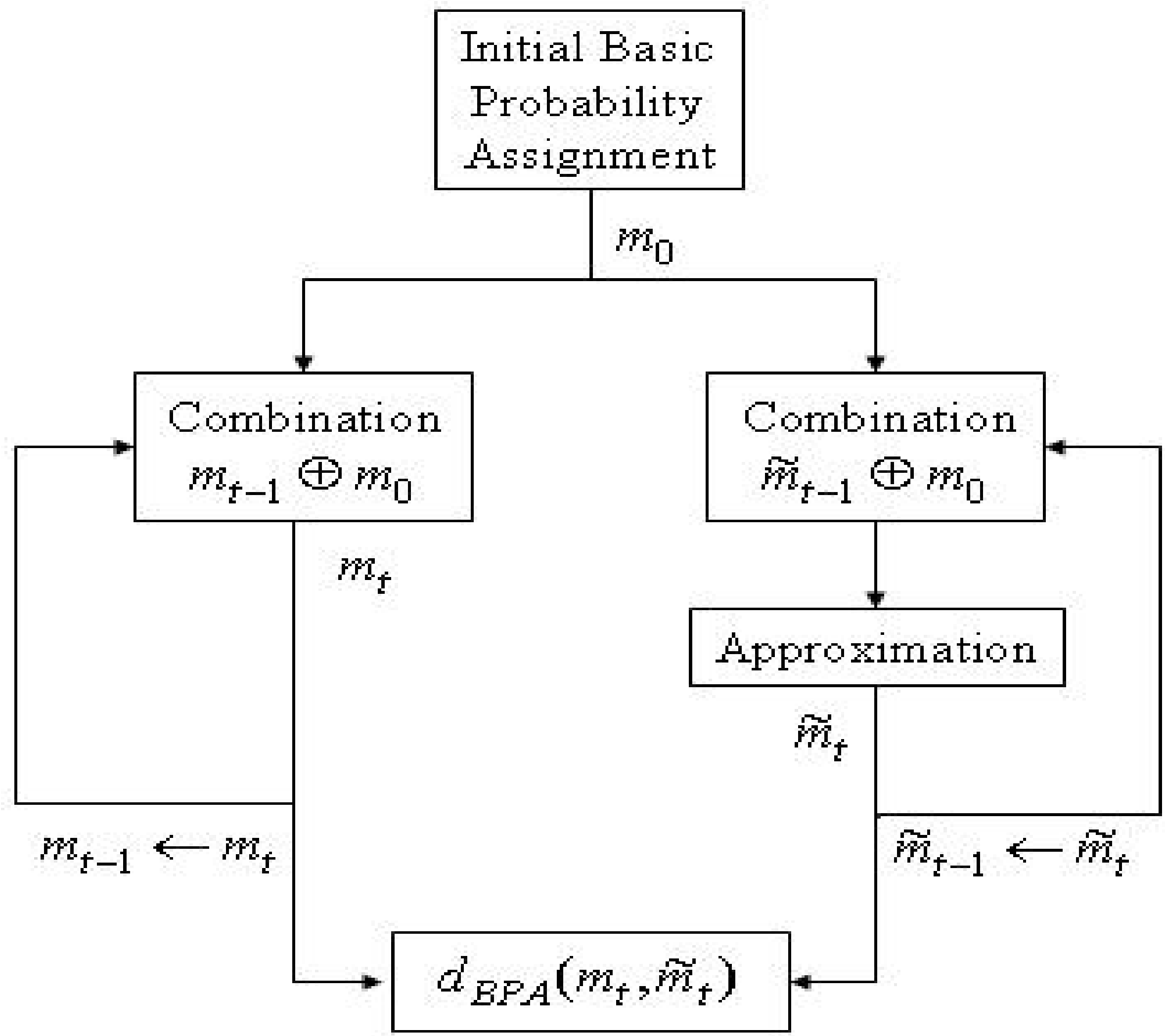
$$d_{BPA}(m^*, m_t) = \sqrt{\frac{1}{2}\left[1 + \left\|\vec{m}_t\right\|^2 - 2\sum_{B\subseteq\Theta} m_t(B)\frac{\left|A^* \cap B\right|}{\left|A^* \cup B\right|}\right]}$$

When A^* is a singleton θ^*:

$$d_{BPA}(m^*, m_t) = \sqrt{\frac{1}{2}\left[1 + \left\|\vec{m}_t\right\|^2 - 2BetP(\theta^*)\right]}$$

where *BetP* is the pignistic probability.

Approximated BPA vs Original

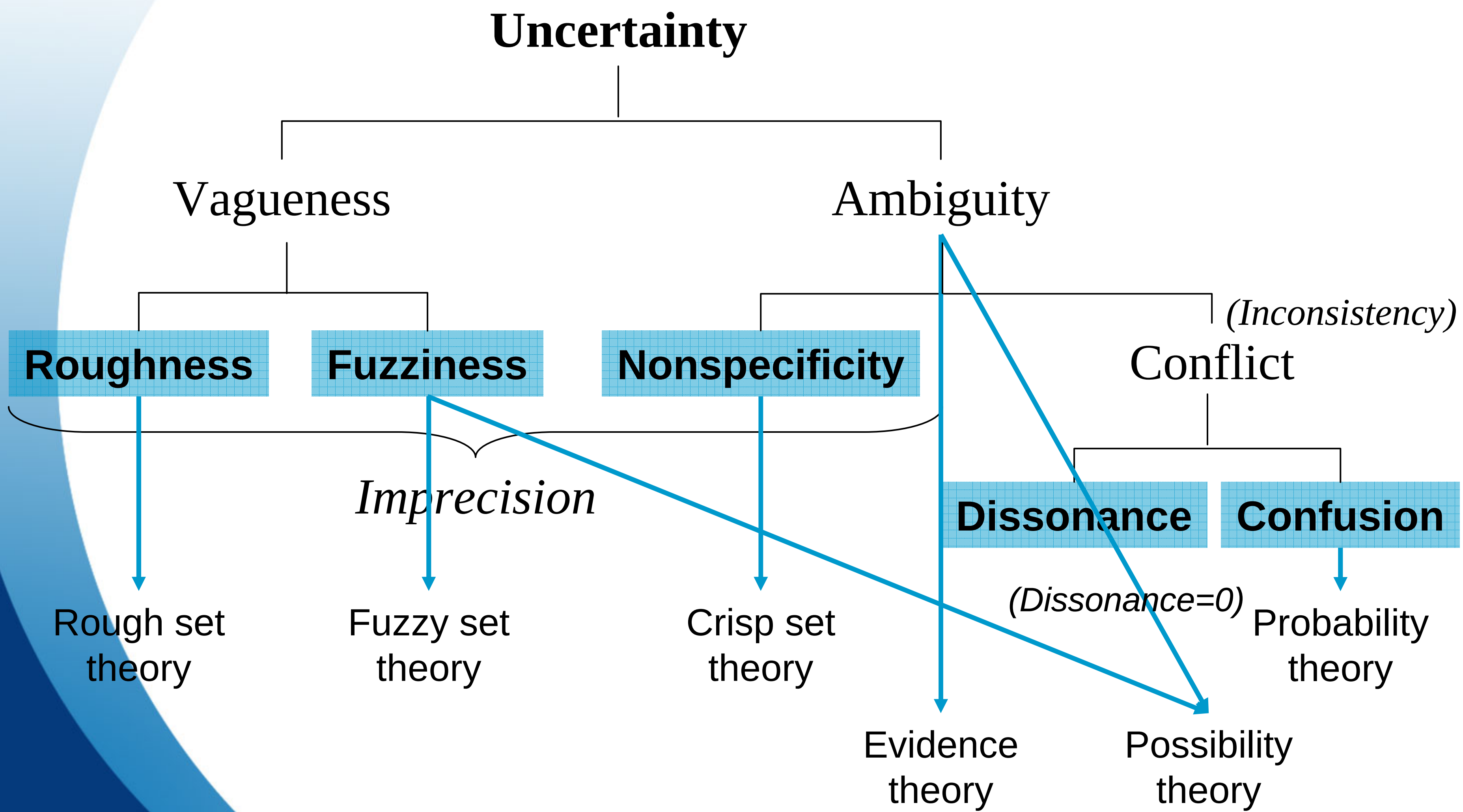


Non-specificity

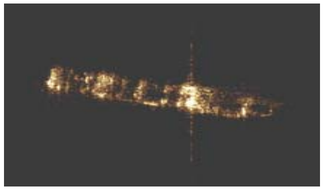
A BPA represents two types of uncertainty: non-specificity and conflict [5]. In this paper, we consider only the non-specificity, as a complementary measure of the quality of the identification algorithm. The non-specificity of a BPA *m* on Θ is defined as [12]:

$$NS(m) = \sum_{A \subseteq \Theta} m(A) \log_2(|A|)$$

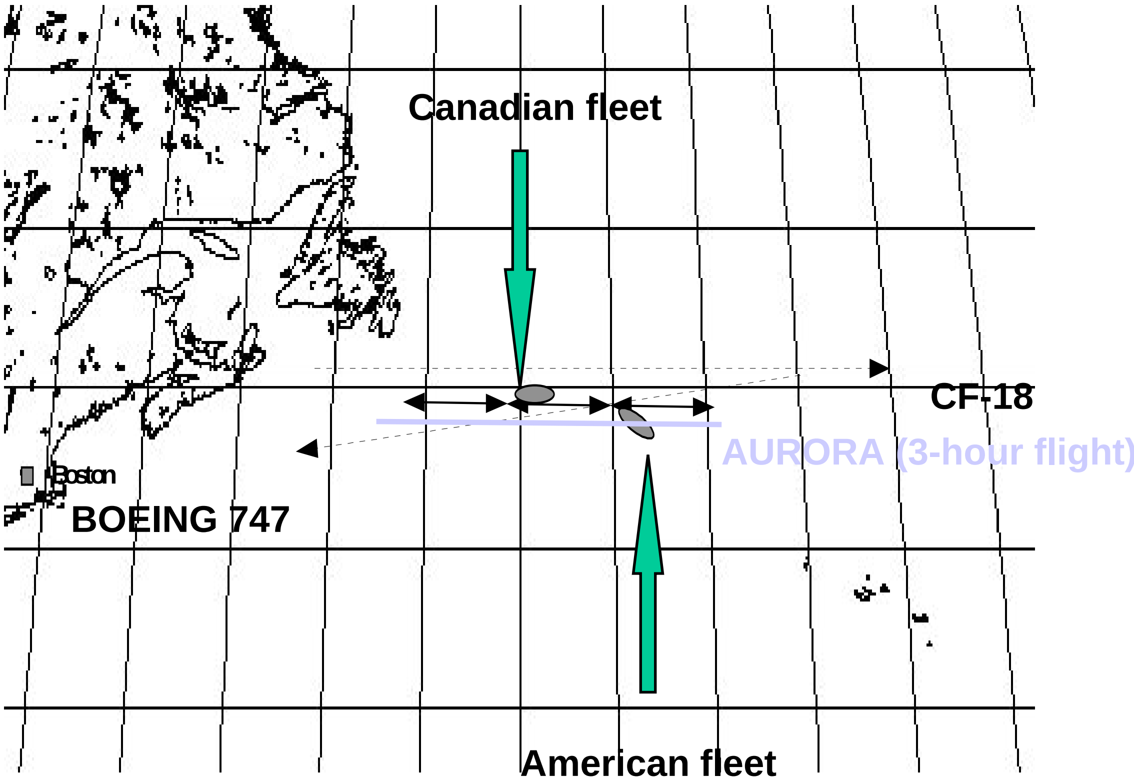
3. Types of uncertainty and uncertainty-based information



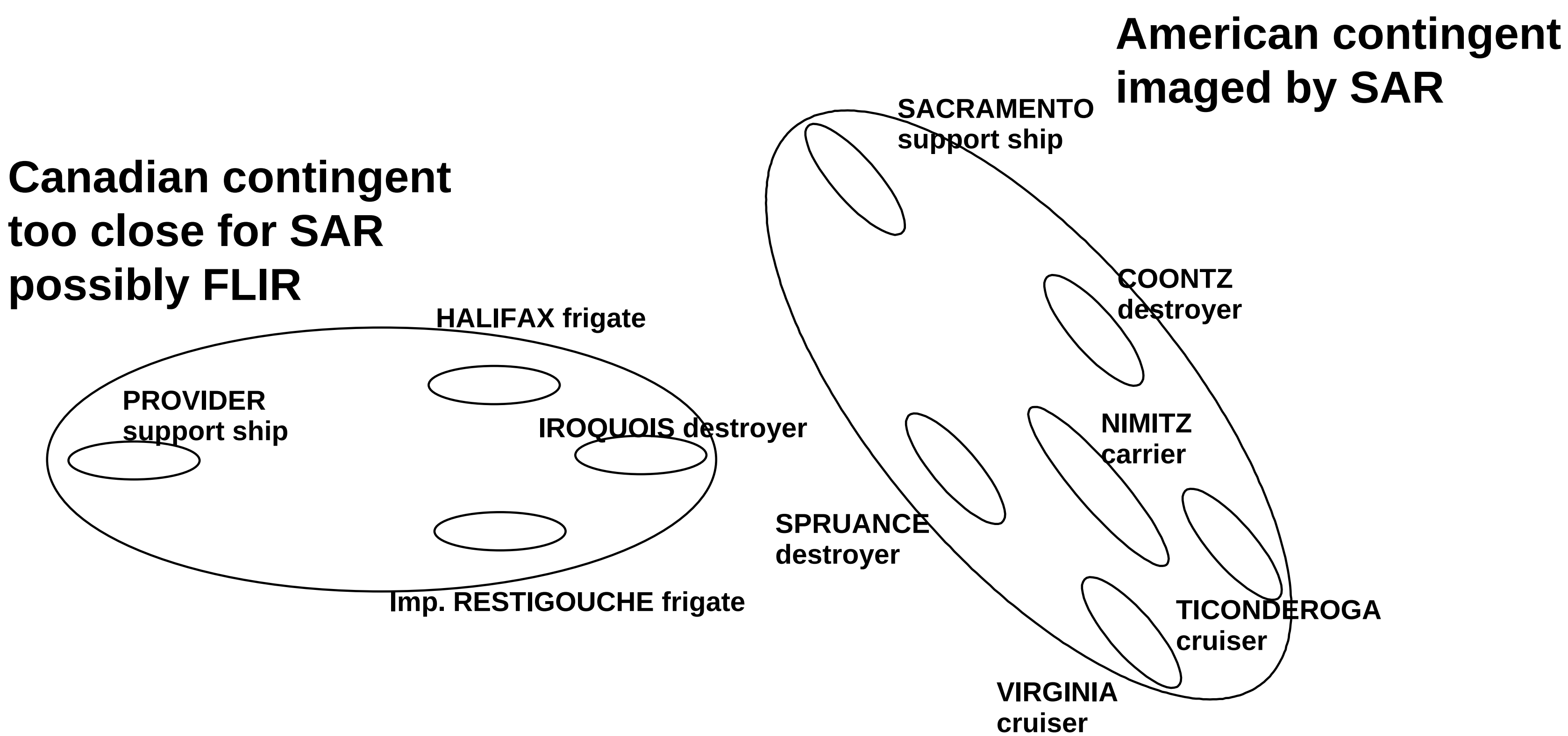
Airborne maritime surveillance



Scenario 2 - Direct Fleet Support

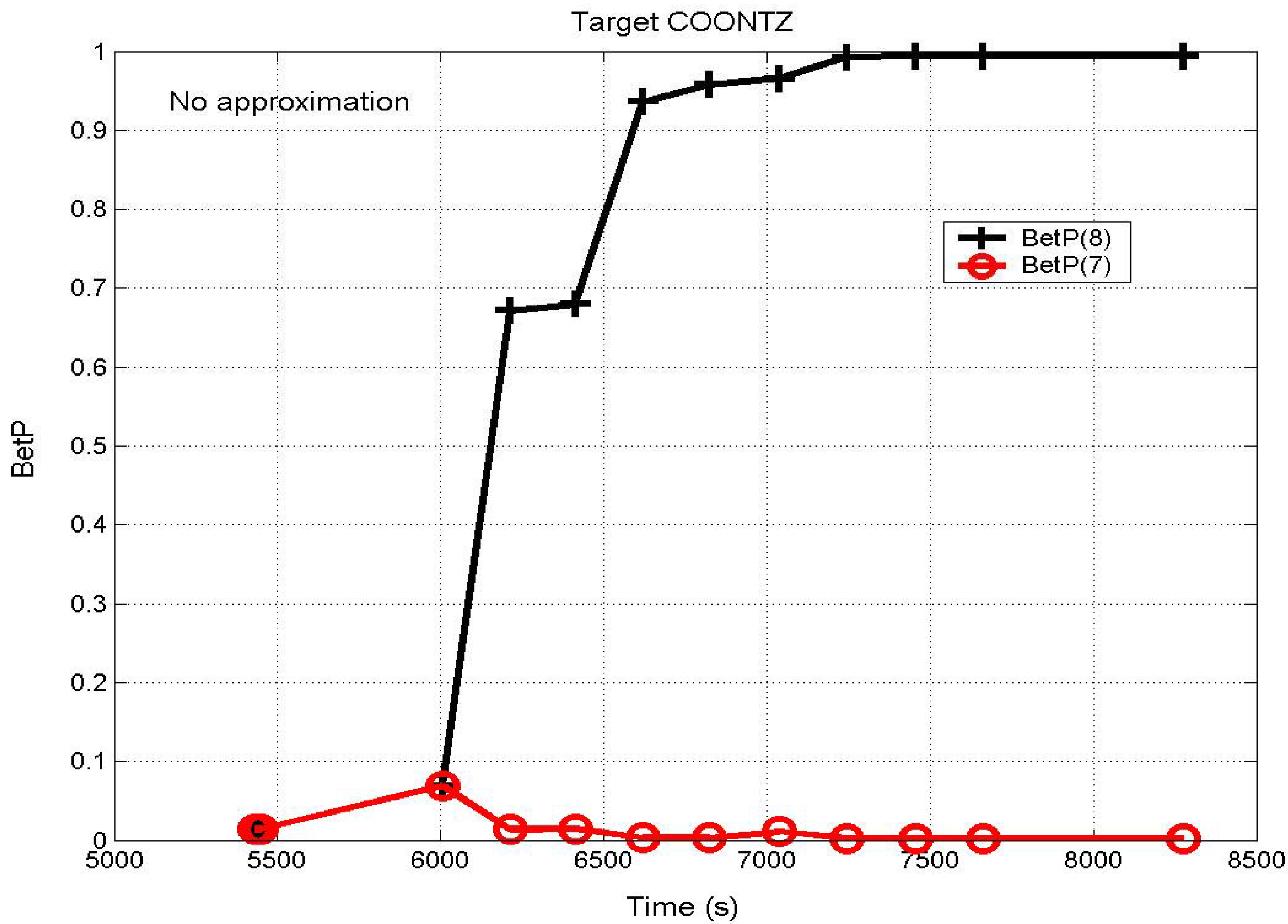


Direct Fleet Support

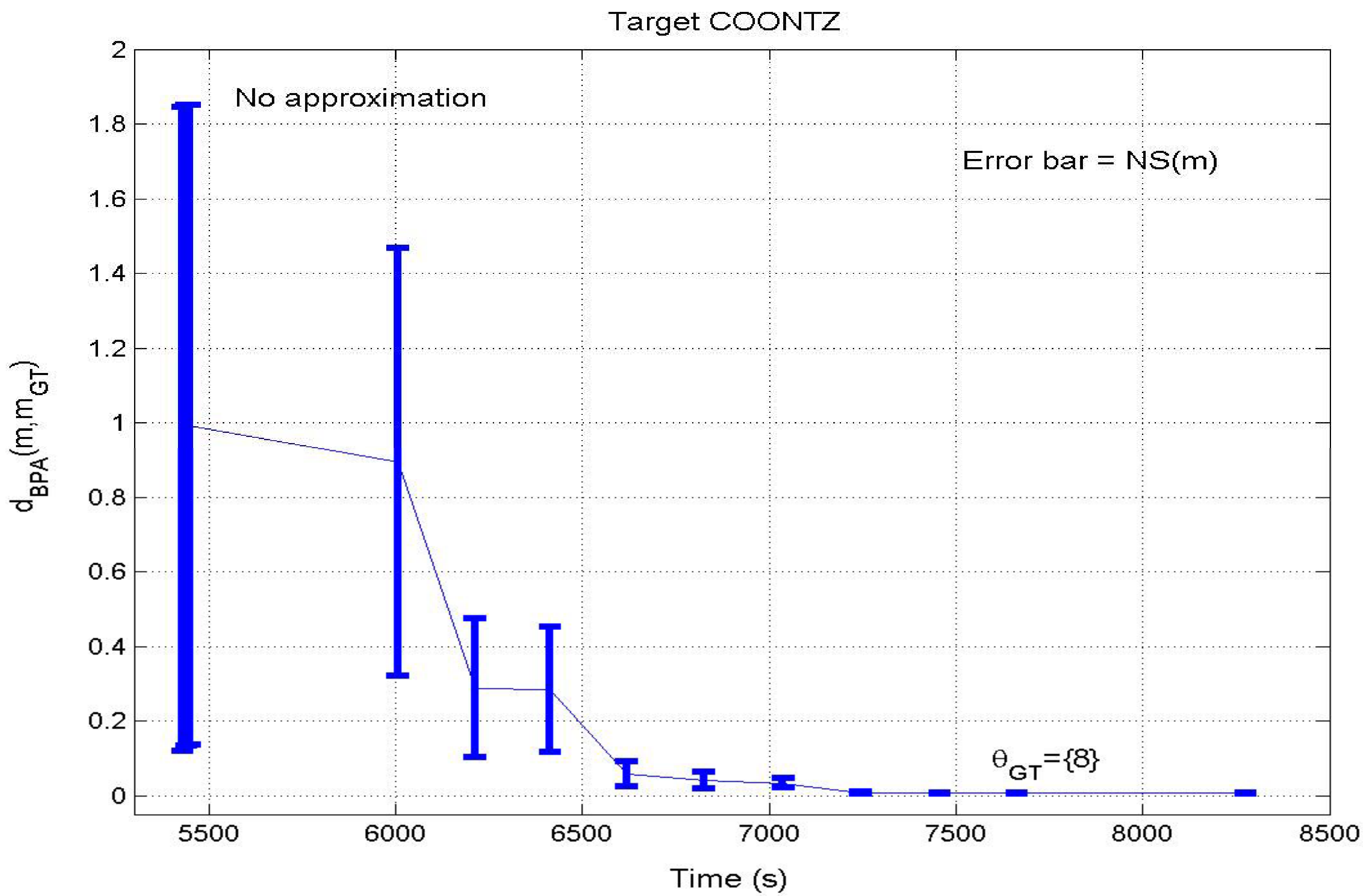


Possibility of many other line ship choices

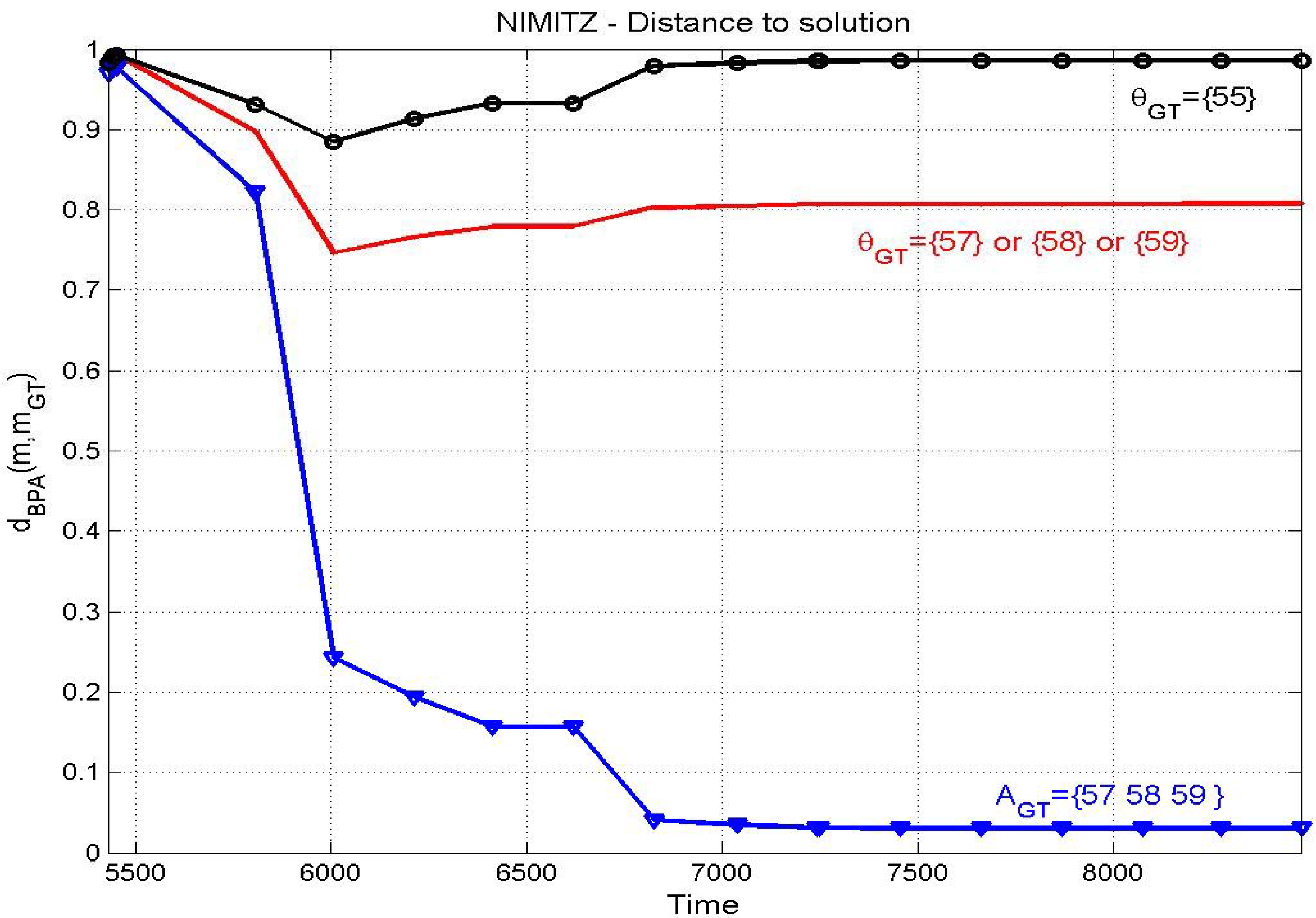
Pignistic probability (decision rule)



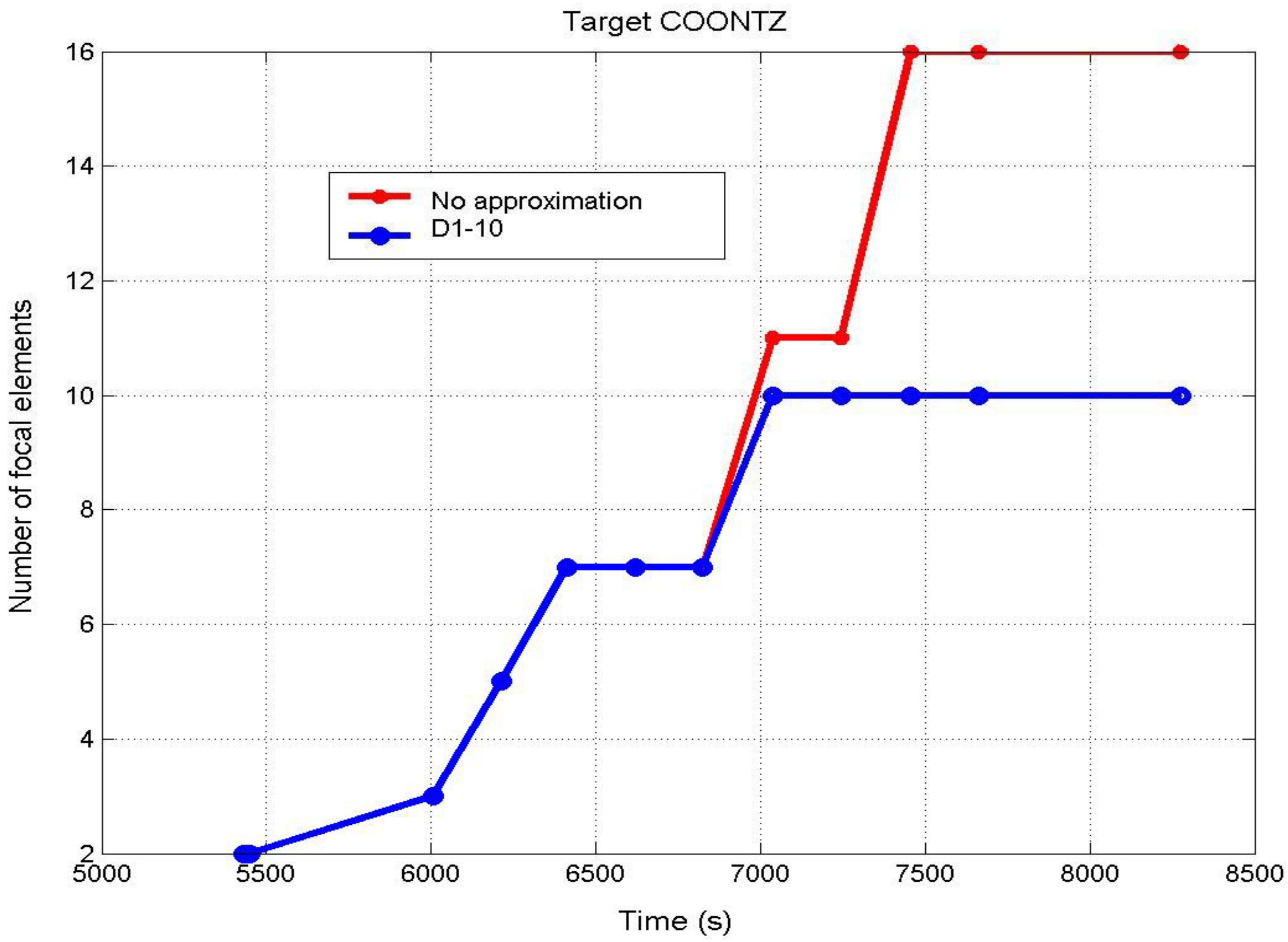
Non-specificity



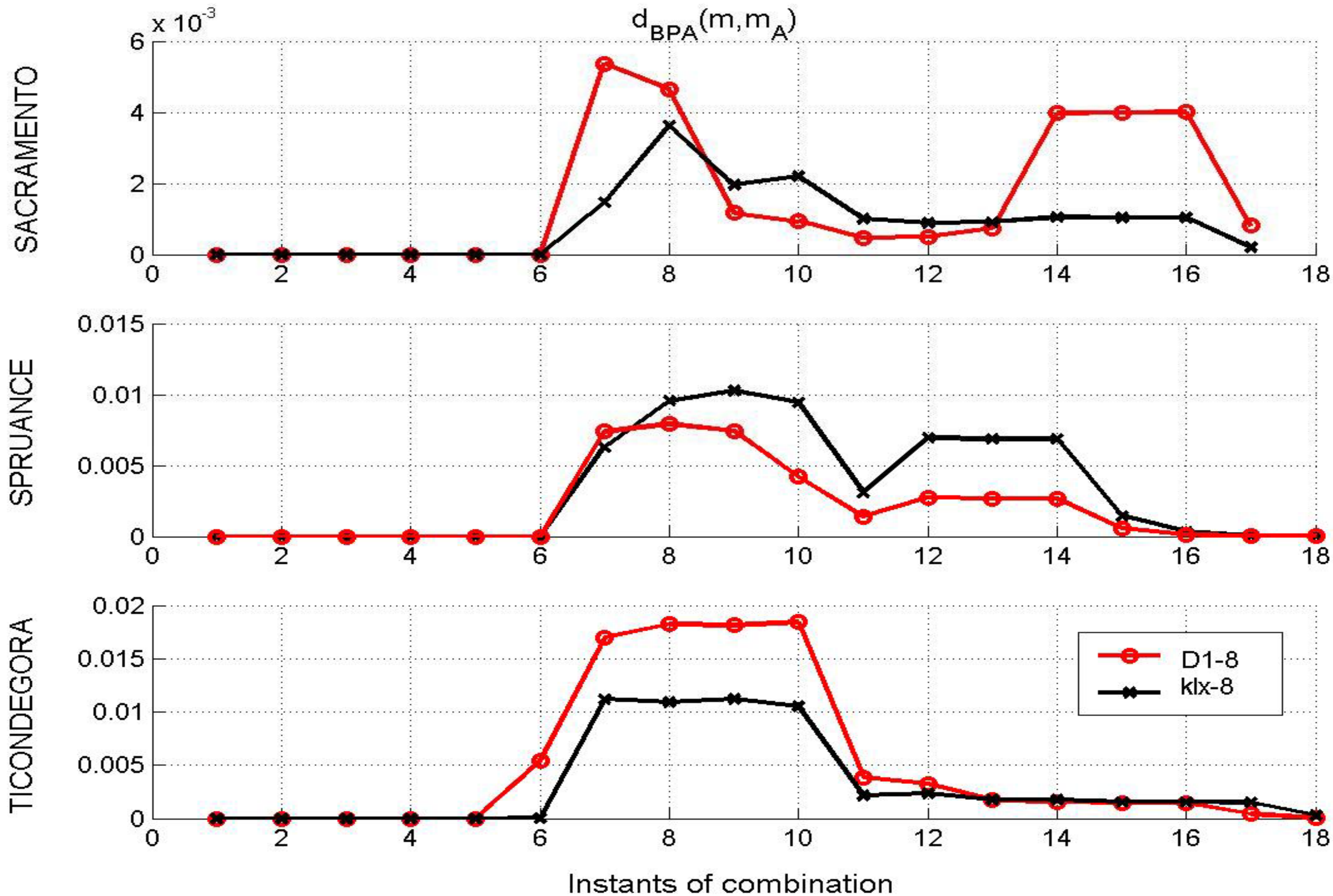
Distance to the solution



Impact of approximation rule



Comparison of approximation algorithms



Conclusions

- **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.**

- **The distance between a BPA and a "solution", to analyse the convergence of the identification algorithm. *mGT* can represent the ground truth (expected solution) but also any other subset of the database towards which the algorithm converges.**

- **The distance between two Basic Probability Assignments to quantify the performance of different approximation algorithms.**

- ***NS*(*mt*) 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 *dBPA*(*mt;mGT*) tends to 0 means a convergence towards a singleton.**

DEFENCE R&D DÉFENSE