Comparative study of evidential reasoning schemes for fusing ESM reports under varying sensor uncertainty and fusion unreliability

Pierre Valin, Pascal Djiknavorian, Eloi Bossé, Dominic Grenier

***Abstract*— We address the problem of fusing Electromagnetic Support Measures reports by two evidential reasoning schemes, namely Dempster-Shafer theory and Dezert-Smarandache theory. These schemes provide results in different frames of discernment, but are able to fuse realistic ESM data. We discuss their advantages and disadvantages under varying conditions of sensor data certainty and fusion reliability, the latter coming from errors in the association process. A thresholded version of Dempster-Shafer theory is fine-tuned for performance across a wide range of values for certainty and reliability, allowing designers who wish to use this method to assess the expected performance. A compromise has to be achieved between stability under occasional miss-associations, and latency under a real change of allegiance. The alternative way of reporting results through Dezert-Smarandache theory is studied under similar conditions, and shown to provide good results which are however more dependent on the unreliability, and less stable.**

***Index Terms*— Evidential reasoning, realistic data, Dempster-Shafer, Dezert-Smarandache**

I. INTRODUCTION

Electronic Support Measures (ESM) consist of passive receivers which can identify emitters coming from a small bearing angle, but cannot determine range. The detected emitters can be related to platforms that belong to 3 classes: either Friend (F=1), Neutral (N=2) or Hostile (H=3), called ESM-allegiance, within that bearing angle. In the case of dense targets, ESM-allegiance can fluctuate wildly due to miss-associations of a bearing-only ESM report to an established Cartesian track. Therefore, all incoming sensor declarations correspond to a frame of discernment of 3 classes, and several theories exist to treat a series of such declarations to obtain a fused result in the same frame of discernment, like Bayesian reasoning and Dempster-Shafer (DS) reasoning [1, 2] (often called evidence theory).

The translation from DS to Bayes can be performed via the pignistic transformation [3], and the result broadcast via tactical data links. In all these implementations, the emitter detected is first correlated to a platform, and then to an allegiance. We will study in the next section the behavior of DS reasoning because it has been adopted on several platforms.

II. DEMPSTER-SHAFER THEORY

Potential users of Dempster-Shafer (DS) theory are often faced at the outset with a list of its pitfalls, which they must somehow solve or at least live with:

1. When the evidence to be fused is too consistent, DS theory will become certain of it after a sufficient number of steps, and will have an extremely hard time to react to a sudden real change in the evidence to be fused. This was solved by Simard et al. [4] by *preventing the ignorance from falling below a certain threshold hereafter called I_{min}. This is done at every fusion step, and the approach is called Thresholded-DS from now on, and is followed here*.
2. When evidence is too conflicting, the normalization step in DS theory can cause wild behaviours from one extreme to another. This is partially a problem in modeling the uncertainty of the data to be fused. We take the approach that the data must correctly modeled by specifying its accuracy and certainty in a reasonable and realistic manner.

At this point, one should make more precise what is meant by data certainty and accuracy:

--*Certainty* is a feature of the sensor that declares that a certain proposition is true with a given mass value m. Without loss of generality, one can assume for simplicity that the sensor declares only one proposition with mass m, and that the rest is assigned to the ignorance. This is likely the case, when the time allowed for decisions is critical, since it provides at each time step only one likely candidate for the declaration. In the example scenario described later, an ESM sensor is likely to provide such a behaviour. In order to stress this point, the article will always mention in the text "*sensor certainty*".

--*Accuracy* refers to how often the data is likely to be wrong. For example the association mechanism that is necessary to select which sensor data is to be associated to which track, can sometimes be erroneous, particularly if it is single scan in nature. Accuracy is therefore a characteristic of the fusion process, not the sensor itself. In the case of the

ESM sensor, miss-associations can occur for the bearing-only reports when the targets are densely placed in that bearing angle. In order to stress this point, the article will always mention in the text "*fusion accuracy*".

One should point out at this time that any sensor will have a value for the uncertainty (or certainty) of its declaration(s), and that, however complex the association mechanism, the association mechanism will occasionally err in its contact-to-track (or track-to-track) correlations, which will provide an inaccuracy in the fusion results. In this sense, the performance characteristics that will be provided later below for thresholded-DS can be applied to a wide range of sensors and positional fusion algorithms, with only very minor modifications.

A. *Statement of the problem and scenario*

The list of the pre-requisites that any scenario must address are:

- should have a clearly defined ground truth, which is sufficiently complex to test stability and latency in the response time.
- should contain sufficient miss-associations, leading to values of average fusion accuracy that are in a realistic range.
- should only provide partial knowledge about the ESM sensor declaration and to varying degrees, which therefore leads to sensor uncertainty (or sensor certainty) values that are in a realistic range.

The following scenario parameters have therefore been chosen accordingly:

1. the known ground truth is Friend (1) for the first 50 time steps of the scenario, and Hostile (3) for the last 50 time steps. Note that the change at time step 50 is uniquely there to test if the algorithms could change their decisions were there a real change in allegiance. It is a convenient way of combining several Monte-Carlos into just one.
2. the percentage of correct associations is approximately Acc%. If the accurate allegiance is Friend (as is the case for the first 50 time steps), then the declarations which correspond to miss-associations are equally distributed between Neutral and Hostile. Similarly, for the last 50 time steps when Hostile is the correct allegiance, the miss-associations are distributed evenly between Friend and Neutral.
3. the ESM declaration has a mass of m, with the rest (1- m) being assigned to the ignorance, reflecting a certainty percentage Cer% = m in the declaration.

Thresholded-DS should be able to adequately represent the main features of the known ground truth, namely

1. show stability under occasional miss-associations, namely show stability when fused data are generally consistent, namely for the first 50 time steps (after a short ramp-up time) and the last 50 time steps (after the ramp-up time, or latency, due to the allegiance change),
2. switch allegiance when the ground truth does so, namely have a reasonable measured latency in the response time (or delay) when an abrupt change occurs in the data to be fused.

A typical scenario with *Acc%* = 80% is shown in Figure 1 below.

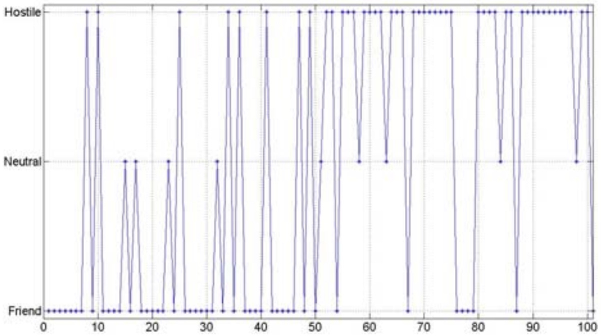

Fig. 1. Chosen scenario.

With I_{min} arbitrarily chosen to be 0.02, the results of Thresholded-DS for the fused mass as a function of time are shown in Figure 2 below.

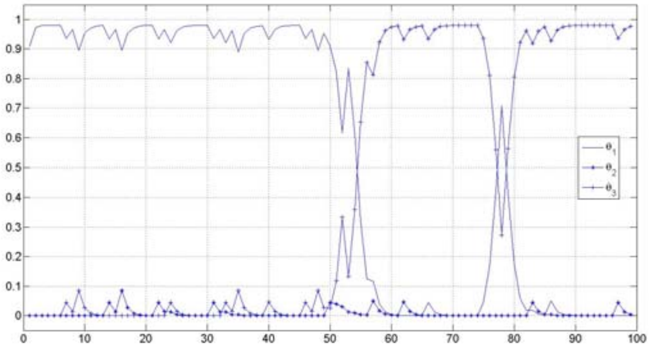

Fig. 2. Thresholded-DS result for the chosen scenario.

Thresholded-DS never becomes confused, reaches the ESM-allegiance quickly and maintains it until time step 50. It then reacts reasonably rapidly and takes about 6 reports before switching allegiance as it should. Furthermore after being confused for a time step around the sequence of 4 Friend reports starting at time step 76, it quickly reverts to the correct Hostile status. This fairly quick reaction is due I_{min} = 0.02, which translates to Thresolded-DS never being more than 98% sure of an ESM-allegiance, as can be seen by the curve topping out at 0.98.

We then performed Monte-Carlo runs of similar scenarios to assess the general performance of Thresholded-DS. Figure 3 shows a sample of a good decision rate of the target identification for thresholded-DS using an input case such as the one from figure 1 generated randomly 100 times, with an ESM sensor having *Acc%*= *Cer%* = 80%, and with I_{min} = 0.05

at every fusion step.

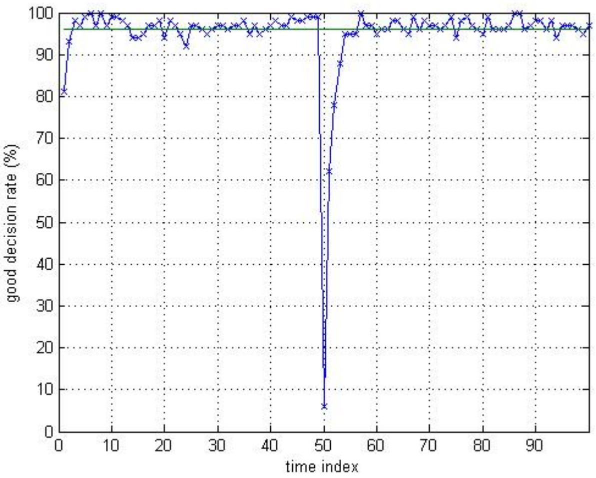

Fig. 3. Good decision rate for the scenario with *Acc%* = *Cer%* = 80%, and 100 runs

In order to evaluate the latency in the reaction time around time step 50, we first determine the empirical mean averaged over time index 15 to 45 and 65 to 95, and then we subtract three times the value of the empirical standard deviation averaged over the same interval. This interval has been chosen arbitrarily to exclude most of the instability that is due to the initialization instability and the change of allegiance instability. The measure of latency then starts at time index 50, and ends at the time index at which the good decision rate reaches the threshold for reaction time performance shown as a green horizontal line in Figure 3.

Standard deviations, which indicate the stability in the above mentioned time periods are shown in Figure 4a for 100 and in Figure 4b for 1000 time steps. Note that they tend to the same value of 0.16 (%, in this case), but show less noise. This underscores that this a dynamical feature of the process, rather than being dependent on the number of Monte-Carlo runs (which do not change the value, but just remove noise).

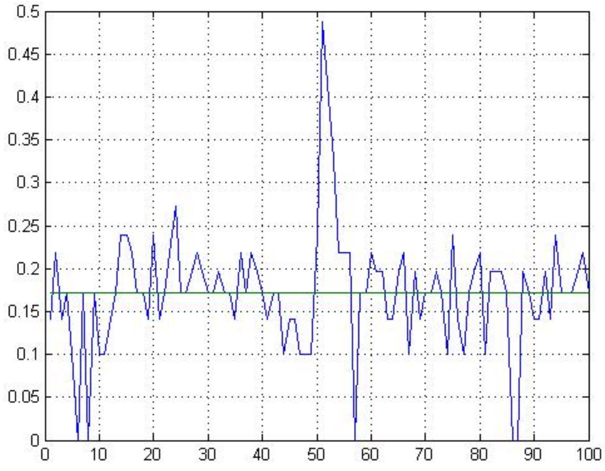

Fig. 4a. Standard deviations for stability after 100 runs

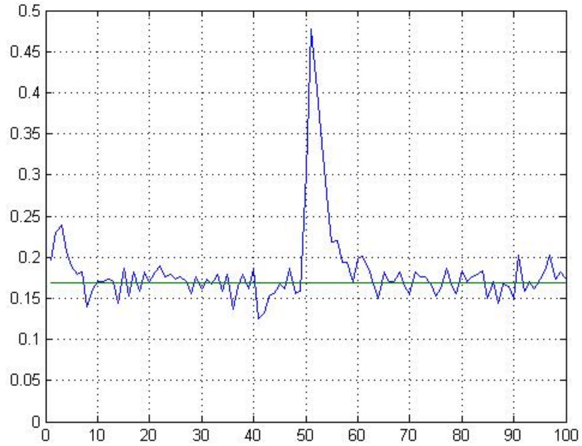

Fig. 4b. Standard deviations for stability after 1000 runs

B. *Monte-Carlo numerical results*

This section shows typical graphs for the standard deviation indicating stability (if small) or instability (if large), and the reaction time latency (or delay) for 1000 Monte-Carlo runs, for I_{min} = 0.03 in Thresholded-DS, as a function of certainty and accuracy.

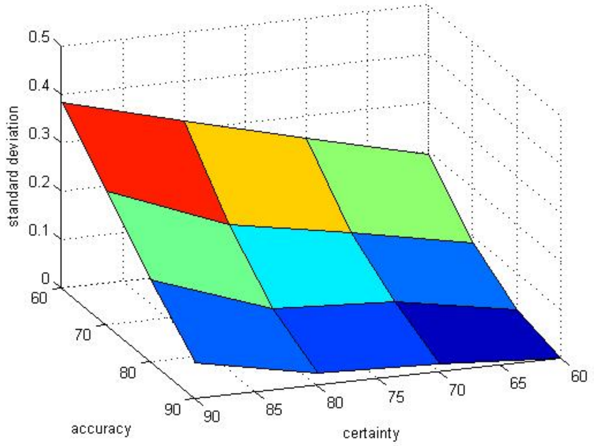

Fig. 5. Standard deviation for I_{min} = 0.03.

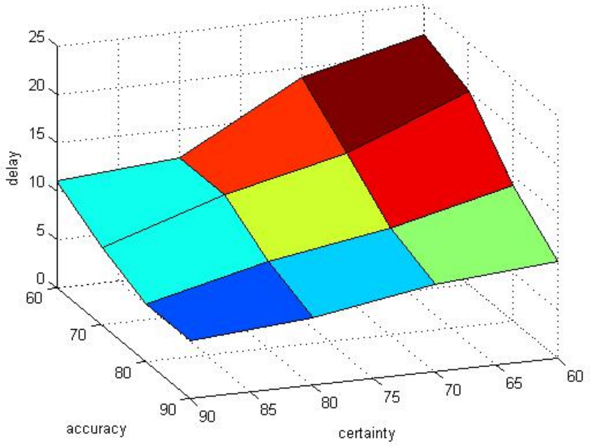

Fig. 6. Reaction time latency (or delay) for I_{min} = 0.03.

We have analyzed the evolution of the standard deviation of the stability, as well as the reaction time latency, as a function of certainty and accuracy, revealing the following trends for the evolution of these complex non-linear figures as I_{min} *changes in steps of 0.01 from 0.01 to 0.05*. This can be easily seen by an animation which will be shown during the lecture.

For the standard deviation of the stability, we have the following trends:

1. For a fixed value of certainty, the value of instability increases when the accuracy decreases (already visible in Fig. 5, and confirmed by the animation).
2. For a fixed value of accuracy, the value of instability increases when the certainty increases (already visible in Fig. 5, and confirmed by the animation).
3. For fixed values of certainty and accuracy, the value of the instability *increases* when the value of the total ignorance threshold I_{min} is *increased* (visible only through the animation).
4. A change in accuracy affects more the instability than the certainty does (already visible in Fig. 5, and confirmed by the animation).
5. Lower values of instability (good) are achieved with higher accuracy and lower certainty, and vice versa (already visible in Fig. 5, and confirmed by the animation).

For the reaction time latency (or delay) we have the following trends:

1. For a fixed value of certainty, the value of the delay increases when the accuracy decreases (already visible in Fig. 6, and confirmed by the animation).
2. For a fixed value of accuracy, the value of the delay increases when the certainty decreases. (already visible in Fig. 6, and confirmed by the animation).
3. For fixed values of certainty and accuracy, the value of the delay *increases* when the value of the total ignorance threshold filter I_{min} is *decreased* (visible only through the animation).
4. A change in accuracy affects more the delay than the certainty does (already visible in Fig. 6, and confirmed by the animation).
5. Lower values of delay (good) are achieved with higher accuracy and higher certainty, and vice versa (already visible in Fig. 6, and confirmed by the animation).

Points 3 in the above two lists clearly show that a compromise must be achieved when using thresholded-DS between being responsive to any real change in the data, yet not being too responsive to fluctuations in the data, due to either poor sensor data certainty or fusion accuracy. In general, a high value for I_{min} will tend to respond to a stream of false reports rather quickly (bad) but will be very responsive to a real change in the data (good). A low value for I_{min} will provide excellent stability (good), but will react slowly to a real change in the data (bad).

III. DEZERT-SMARANDACHE THEORY

Decision makers would like the target platforms to be identified on a more refined basis than the 3 studied above, belonging to 5 classes, namely the 3 studied above (Hostile or foe, Neutral and Friend, but adding Suspect (S), and Assumed Friend (AF). This will henceforth be referred to as STANAG 1241 allegiance, or just STANAG allegiance for short [5]. Unlike Dempster-Shafer theory, Dezert-Smarandache (DSm) theory can coherently, with well-defined fusion rules, lead to an output amongst those 5 classes, even though the input classes number only 3, because the theory allows for intersections. For example,

- "Suspect" might be the result obtained after fusing "Hostile" with "Neutral", and
- "Assumed Friend" might be the result obtained after fusing "Friend' with "Neutral".

We will consider, for this paper, that the extreme result of an association between Friend and a Hostile is impossible, so that the set intersection $1 \cap 3$ yields the null set (having therefore null mass), which is a constraint in DSm, leading to the use of its *hybrid rule*. This leads to Figure 7, and would correspond to a peace-keeping mission where such as association is quite unlikely.

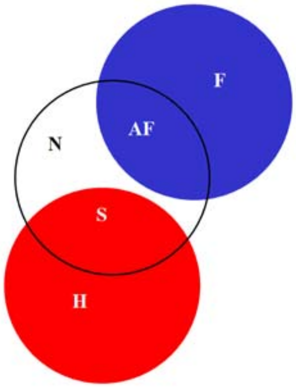

Fig. 7. Venn diagram for the STANAG allegiances

DSm theory encompasses DS theory as a special case, namely when all intersections are null. Both use the language of masses assigned to each declaration from a sensor (in our case, the ESM sensor). The reader is referred to a series of books [6, 7, 8] on DSm theory for lengthy descriptions of the hybrid rule.

Since the incoming sensor reports are in DS-space: Friend (F=1), Neutral (N=2) or Hostile (H=3), then Figure 7 has the interpretation in DSm output space (allowing intersections during the fusion step) of:

Friend = $\{\theta_1 - \theta_1 \cap \theta_2\}$
Hostile = $\{\theta_3 - \theta_3 \cap \theta_2\}$
Assumed Friend = $\{\theta_1 \cap \theta_2\}$
Suspect = $\{\theta_2 \cap \theta_3\}$
Neutral = $\{\theta_2 - \theta_1 \cap \theta_2 - \theta_3 \cap \theta_2\}$

For the chosen scenario of Figure 1, DSm theory outputs the results of Figure 8 below.

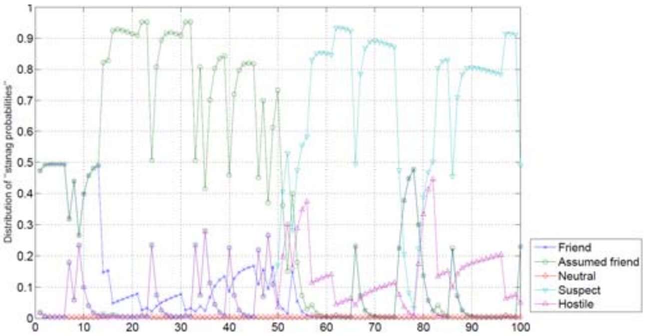

Fig. 8. DSm result for the chosen scenario

This should be compared with the Thresholded-DS result of Figure 2. The decision maker would clearly be informed that miss-associations have occurred, since Assumed Friend dominates for the first 50 time steps and Suspect for the latter 50. DSm is more susceptible to miss-associations than thresholded-DS (the dips are more pronounced), but it has the advantage of giving extra information to the decision maker, namely that the fusion algorithm is having difficulty with associating ESM reports to established tracks. Just as for thresholded-DS, the Friend declarations starting at time step 76 cause confusion, as it should. The change in allegiance at time step 50 is detected nearly as fast as DST. What is even more important is that F and AF are clearly preferred for the first 50 time steps and S and H for the last 50, as they should. With this more refined STANAG-allegiance, a decision maker would probably take no aggressive action against either a friend or an assumed fiend (although he/she would monitor an assumed friend more closely). Similarly a decision maker would probably take aggressive action against a foe and send a reconnaissance force (or a warning salvo) towards a suspect.

Note that one can remove the restriction that the set intersection $1 \cap 3$ yields the null set, leading to a different interpretation of what situations can generate a Suspect declaration, as shown in Figure 9, which would correspond to a war-fighting mission, where it would be prudent to associate Suspect with any track that has been associated with a Hostile contact. This is presently being studied, and is more complex since all intersections have to be kept in intermediate results.

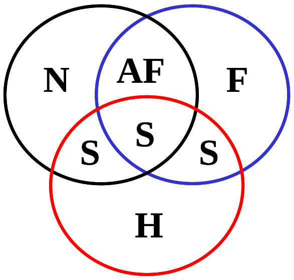

Fig. 9. Another Venn diagram for STANAG allegiances

In order to sample the parameter space in a different way, the simulations below correspond to $Acc\%$ = 90%, $Cer\%$ = 60%, and an I_{min} = 0.02 as before. The number of Monte-Carlo runs was set to 100. For these values, the result for DS is shown in Figure 10. As expected, since DS reasons over the 3 input classes, Suspect and Assumed Friend are not involved.

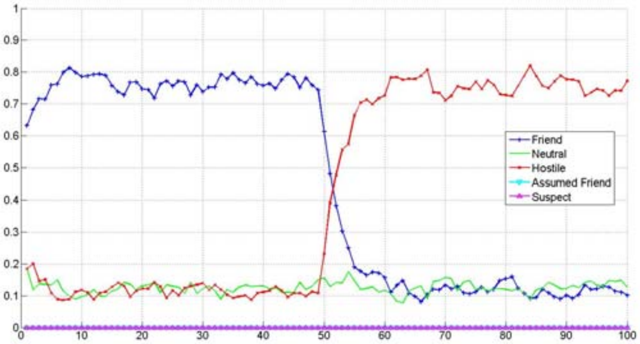

Fig. 10. DS result after 100 Monte-Carlo runs

The equivalent result for DSm is shown in Figure 11 below. In this case, AF dominates for the first 50 time steps, on average (over 100 runs) and S for the last 50, confirming that the chosen scenario was representative of the behaviour of DSm. The response times are similar on average also. DSm is slightly less sure (plateau at 70%) than DS (plateau at 80%).

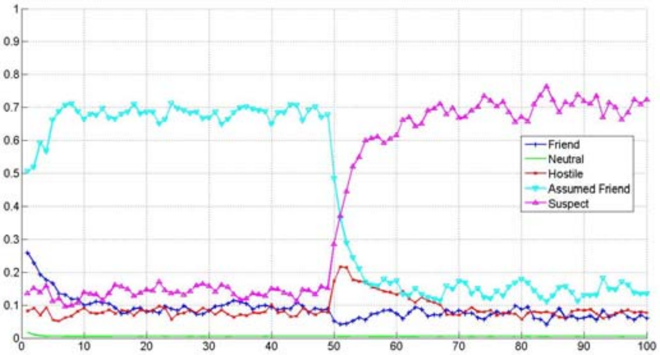

Fig. 11. DSm result after 100 Monte-Carlo runs

IV. 4 CONCLUSIONS

We addressed the problem of fusing Electromagnetic Support Measures reports by two evidential reasoning schemes, namely Dempster-Shafer theory and Dezert-Smarandache theory. For DS theory, a compromise has to be achieved between stability under occasional miss-associations, and latency under a real change of allegiance. The alternative way of reporting results through Dezert-Smarandache theory was studied under similar conditions, and shown to provide good results, which are however more dependent on the unreliability, and less stable.

REFERENCES

[1] A.P. Dempster, “Upper and lower probabilities induced by a multivalued mapping”, *Ann. Math. Statist*. 38, 325--339 (1967).

[2] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, NJ (1976).

[3] Ph. Smets, “Data Fusion in the Transferable Belief Model”, *Proceedings of the 3rd International Conference on Information Fusion*, *FUSION 2000*, Paris, pp. PS21--PS33 (2000).

[4] M.A. Simard, P. Valin, and E. Shahbazian, “Fusion of ESM, Radar, IFF and other Attribute Information for Target Identity Estimation and a Potential Application to the Canadian Patrol Frigate”, *AGARD 66th Symposium on Challenge of Future EW System Design*, Ankara, AGARD-CP-546, pp. 14.1--14.18 (1993).

[5] STANAG 1241: *NATO Standard Identity Description Structure for Tactical Use,* North Atlantic Treaty Organization, April 2005.

[6] F. Smarandache, and J. Dezert, Editors, *Advances and Applications of DSmT for Information Fusion, vol. 1*, American Research Press (2004).

[7] F. Smarandache, and J. Dezert, Editors, *Advances and Applications of DSmT for Information Fusion, vol. 2*, American Research Press (2006).

[8] F. Smarandache, and J. Dezert, Editors, *Advances and Applications of DSmT for Information Fusion, vol. 3*, American Research Press (2009).