23rd International Congress on Modelling and Simulation, Canberra, ACT, Australia, 1 to 6 December 2019
mssanz.org.au/modsim2019

# A protocol for selecting a rule of combination for fusion of imprecise cyber sensor data

**E. El-Mahassni** [a]

[a] *Cyber and Electronic Warfare Division, Defence Science and Technology Group, PO Box 1500, West Avenue, Edinburgh, SA, 5111*

*Email: Edwin.El-Mahassni@dst.defence.gov.au*

**Abstract:** There does not seem to be a day that goes by where there is not news of a new and potentially catastrophic cyber attack infecting computers and networks. In addition, even if it were possible to fully list all possible attacks, cyber sensors might not have the fidelity to correctly identify the type of attack. To further complicate things, it is more likely that these sensors must not only contend with quickly detecting known threats but also attacks which are new and unforeseen. Hence, modelling techniques are needed that will help take into account the possibility that the full possible range of potential threats may not be precisely represented as well model ambiguity and partial information. Dempster-Shafer Theory is one such framework because it not only deals with imprecision but there are also different combination rules which can be used depending on whether it is assumed that the full set of possible outcomes is known or not. Here, a protocol is provided to help the analyst select which combination rule should be used based on the assumptions he or she is making. The importance of these assumptions is illustrated through an example to show how results can vary according to the rule chosen.

***Keywords:*** *Cyber security, cyber sensors, Dempster-Shafer Theory*

## 1 INTRODUCTION

Cyber attacks are difficult to predict and can cause a lot of financial, social and physical damage. One of the ways to increase the chances of detecting and identifying a cyber attack is through sensor fusion. This can be defined as the combination of sensor data obtained from different sources to reduce uncertainty and/or obtain more complete or accurate information than that available from any individual source. In some cases, two or more sensors with complementary capabilities may be used together to obtain a target estimate that is better than when any individual sensor is used alone. The scenario addressed by this paper is that of fusing the data from two or more sensors to support an operator's decision making ability in determining the type of cyber attack.

Dempster-Shafer Theory is a method for combining information from multiple sensors (Dempster, 1968; Shafer, 1976, 1992). Dempster-Shafer Theory may be regarded as a collection of different theories with the same underlying uncertainty calculus and has its origins in the representation of and reasoning under partial probability distributions over a finite set $X$.

In this paper a particular emphasis is placed on when to use different techniques in the fusion of information of sensor data from multiple sources which contains uncertainty within a Dempster-Shafer Theory framework. However, the application of Dempster-Shafer Theory to the cyber domain is relatively new with most applied efforts concentrating in the physical environment.

The objective of sensor fusion in a cyber domain is to correctly detect or identify any potential threats to a computer or network given data from multiple sensors. For a given sensor, it might be difficult to correctly identify an attack among a list of several possible cyber attacks. In the Dempster-Shafer Theory framework, a combination function processes independent evidence and obtains a belief or mass distribution. A Probability Transformation is then performed to determine a discrete probability distribution.

Within the context of cyber, sensor fusion supports decision making by processing uncertain sensor data to arrive at the detection and then identification of cyber attacks. If the identification corresponds to, say a virus, then the decisions of an operator may be quite different than if the cyber threat is that of a Denial of Service (DoS) attack.

However, although here the fusion of sensor information within Dempster-Shafer Theory is discussed, there are several approaches in which information can be fused leading to different results depending on which method is employed. The focus of this paper is to examine these different ways and the assumptions that they hold to through some simple examples so that a better understanding may be gained on their differences. We remark that though the rules described in this paper seem to be the most commonly discussed throughout the literature, there are others which are not considered here. A good review of these rules are given in Florea et al. (2006); Sentz and Ferson (2017). And, whichever combination rule is chosen, for the purposes of decision-making, these masses are then typically converted to probabilities and the most likely outcome is deemed to be the true target. The pignistic transform is the most well-known method for doing this. Finally, through examples, all these concepts will be demonstrated by showing how choosing a particular combination rule can affect the final result after the transformation is performed.

## 2 BASIC CONCEPTS OF DEMPSTER-SHAFER THEORY

In this section, we describe some of the basic concepts of Dempster-Shafer Theory. For instance, whether the frame of discernment is assumed to be closed (the answer is in the solution space defined) or open (it might lie outside of that solution space).

### 2.1 Frame of Discernment

At the heart of Dempster-Shafer Theory is the frame of discernment which is typically denoted by $\Theta_D$ and is a model that describes the set of possible hypotheses. Suppose that the frame of discernment is $\Theta_D = \{\theta_1, \theta_2, \ldots, \theta_k\}$, where the $\theta_i, 1 \leq i \leq k$ are mutually exclusive.

### 2.2 Closed World Assumption (CWA)

A Closed World Assumption (CWA) implies that $\Theta_D$ is exhaustive, so that at least one of the $\theta_i$ must be true. For example, if it is known that $\theta_1$ is not true, then the CWA implies that the conclusion must correspond to one of the other $k-1$ possibilities. In other words, there is zero probability or chance that the true solution might lie outside of $\Theta_D$. Consider a situation where a CWA might be made. If the frame of discernment may

be known to be complete so that all possible solutions are listed. For cyber applications, such an assumption means that the sensors all have fully encoded knowledge of the possible types and kinds of attacks and are able to produce a belief mass distribution based on a library of cyber attacks for subsequent identification fusion. But, if the frame of discernment is regarded as being complete but is actually incomplete so that the actual solution may lie outside it, then the misidentification of an attack may result from evidential reasoning applied to a frame of discernment that is not exhaustive.

### 2.3 Open World Assumption (OWA)

Dealing with a lack of complete knowledge about the complete or entire solution space is known as making an Open World Assumption (OWA). While the CWA may lead to inference errors if the frame of discernment is actually incomplete, the OWA places restrictions on the deductions that can be made from the available information. Assuming a frame of discernment $\Theta_D$, the OWA assumption implies that if an inference is made about $\theta_1, \ldots, \theta_{k-1}$ then we cannot conclude anything about $\theta_k$. The true frame of discernment is actually $\Theta_F = \Theta_D \cup \Theta_E$, where $\Theta_E$ is another frame of unknown size such that $\Theta_D \cap \Theta_E = \emptyset$. That is, both frames have distinct elements and there is no "doubling-up".

### 2.4 Belief Function Assignment

The most common means of encoding is via the *basic belief assignment* (bba) which is a function $m$ defined on the power set of the frame of discernment $\Theta_D$ as follows: $m : \wp(\Theta_D) \to [0, 1], X \mapsto m(X)$ subject to the constraint that $\sum_{X \subseteq \Theta_D} m(X) = 1$. The power set of a set of elements is simply the set of all possible subsets of elements from that set. For instance, for a frame of discernment $\Theta_1 = \{\theta_1, \theta_2, \theta_3\}$ consisting of $|\Theta_1| = 3$ elements, the power set is simply

$$\wp(\Theta_1) = 2^{\Theta_1} = \{\{\emptyset\}, \{\theta_1\}, \{\theta_2\}, \{\theta_3\}, \{\theta_1, \theta_2\}, \{\theta_1, \theta_3\}, \{\theta_2, \theta_3\}, \{\theta_1, \theta_2, \theta_3\}\}\,.$$

For a given set $X \subseteq \Theta_D$, the belief mass $m(X)$ represents the proportion of all relevant and available evidence that supports the claim that a particular element of $\Theta_D$ belongs to the set $X$, but to no particular subset of $X$. We remark that mathematically, $m$ is akin to a probability distribution when the only non-zero belief masses are all assigned to subsets containing a single element, also known as a *singleton*. However, as already noted, no interpretation of the belief masses as probabilities is made in some interpretations of the Dempster-Shafer Theory. To combine evidence from two independent belief functions represented as bbas $m_1$ and $m_2$, Dempster's rule of combination is typically used and the resulting belief function, denoted by $m_{1,2} = m_1 \oplus m_2$.

## 3 Rules of Combination

Dempster's Rule of Combination is the most common approach in Dempster-Shafer Theory for fusing information from multiple sources. In this section, this is described along with Smets' unnormalized rule of combination from the Transferable Belief Model (TBM), Yager's Rule and Dubois and Prade's Rule.

### 3.1 Dempster-Shafer's Rule of Combination

To combine evidence from two independent belief functions represented as bbas $m_1$ and $m_2$, Dempster's rule of combination is typically used and the resulting belief function, denoted by $m_{1,2} = m_1 \oplus m_2$, is given by: $(m_1 \oplus m_2)(X) = \frac{1}{(1-\kappa)} \sum_{Y \cap Z = X} m_1(Y) m_2(Z)$, where the conflict mass $\kappa$ is given by $\kappa = \sum_{Y \cap Z = \emptyset} m_1(Y) m_2(Z)$. In Dempster's rule of combination, the empty set mass is zero, i.e. $m(\emptyset) = 0$ Sentz and Ferson (2017). It should be noted that Dempster's rule is commutative, i.e. $m_1 \oplus m_2 = m_2 \oplus m_1$ and associative, i.e. $(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$, when the basic belief masses are compatible, i.e. are not in complete conflict.

### 3.2 Smets Unnormalized Rule of Combination

The most well-known rule for explicitly accounting for conflict is the Transferable Belief Model (TBM) Smets (2000). In this case, the combination rule is given by $m^s(X) = (m_1 \oplus_s m_2)(X) = \sum_{Y \cap Z = X} m_1(Y) m_2(Z)$, where the mass for the empty set $m(\emptyset) \geq 0$ in contrast to Dempster's rule of combination and many other rules of combination where $m(\emptyset) = 0$. The effect of normalisation with the conflict mass $\kappa$ in Dempster's rule of combination is that a closed world is assumed. However, by allowing the possibility of the empty set mass to be non-zero in Smets TBM, this implies that an open world is assumed. The TBM is commutative and associative, just like Dempster's rule of combination.

### 3.3 Yager's Rule

Yager's rule of combination is similar to Shafer and Smets' TBM but it differs in the treatment of mass that is assigned to the empty set (Yager, 1986). Yager defines the *ground probability assignment* associated with $X$ as $r(X) = \sum_{Y \cap Z = X} m_1(Y) m_2(Z)$. The bba arising from Yager's rule of combination is given by $m^y(X) = r(X), m^y(\Theta_D) = r(\Theta_D) + r(\emptyset)$ so that any conflict $\kappa = r(\emptyset)$ is assigned to the frame of discernment $\Theta_D$. That is, Yager's rule implies a CWA because the possible solutions are contained in the frame of discernment $\Theta_D$.

### 3.4 Dubois and Prade's Rule

Dubois and Prade (1986) took a set-theoretic approach in arriving at their rule $(m_1 \oplus_{dp} m_2)^{dp}(X) = \sum_{Y \cup Z = X} m_1(Y) m_2(Z)$. The use of the set union in Dubois and Prade's rule avoids consideration of the conflict mass and so there is no requirement for normalisation. This rule assumes that the true target is being reported from at least one of the masses' sources. The Dubois and Prade rule of combination is commutative and associative.

### 3.5 Other Rules of Combination

There are other rules of combination which have been developed within the Dempster-Shafer Theory framework such as Zhang's Center combination rule (Smets, 1994), Mixing or Averaging (Ferson and Kreinovich, 2001), the Adaptive Combination Rule (ACR) (Florea et al., 2006), Convolutive X-Averaging (Ferson and Kreinovich, 2001), the cumulatative rule (sang, 2010), the cautious rule (Denoeux, 2006), and the collection of proportion conflict resolution rules (PCR) (Florea et al., 2006; Dezert and Smarandache, 2009).

Finally, Yager (2007) proposed another rule which combines both their original disjunctive rule with Smet's TBM. These are not described further in this paper but thorough reviews of various Dempster-Shafer Theory rules of combination have been provided elsewhere such as in Florea et al. (2006) and especially Sentz and Ferson (2017). Lastly, most of these rules are also more thoroughly reviewed in (Nguyen and Docking, 2015).

## 4 THE PIGNISTIC TRANSFORMATION

Once all the evidence has been aggregated, the basic belief masses from the resulting belief function may be used to construct the *pignistic probability distribution*. As mentioned, in the decision-making (or pignistic) level, it is typical to take the Dempster-Shafer masses and transform them to probabilities. The most well-known method is the the pignistic transform. Originally conceived in Smets (2000), it moves the belief mass from the union elements of the power set and distributes it equally amongst the singleton members. For an element $x \in \Theta_D$, the transform gives $P(x) = \sum_{\emptyset \neq A \subseteq \wp(\Theta_D): x \in A} m(A)/|A|(1 - m(\emptyset))$ for each $x \in \Theta_D$. A hard decision may then be made by declaring the target most highly supported by the amassed evidence to be $x_k$, where $x_k = \arg\max_{i=1,\ldots,k} P(x_i)$.

Note that although that there are other probability transformations the Pignistic transformation is perhaps the most popular and widely used of all transformations.

## 5 DESCRIPTION OF PROTOCOL

Above, four different combination rules have been described. The task now is to decide which one to use at any given time. Through a series of questions, it can be shown that a particular rule should be preferred over another under certain conditions. Hence, these rules can be categorized through a series of questions, which would help clarify some assumptions under which sensor fusion is being carried out.

1. Is the frame of discernment open? That is, are there possibly types of attacks other than those which are assumed to be possible. If the answer is yes, then that means that the cyber attack database is incomplete and so the TBM rule should be used. Note that because conflict is not resolved using the rule, this leaves the possibility that the correct type of attack might lie outside the defined frame of discernment. If the answer is negative, then the next question is posed.

2. Is at least one of the sensors' reports unreliable? If the answer is yes, then Dubois and Prade's rule, which assigns the product value to the union of the focal elements should be used. If the answer is no, then the next question is posed.

3. Is part of a sensor's evidence in complete conflict with part of another sensor's evidence? If the answer is positive then Yager's rule should be used. If the answer is in the negative so that $\kappa$ is 0, then all three of the appropiate choices for the combination rule reduce to the same rule, the classical Dempster's rule of combination, and so this should be used in this case.

For Figure 1, the protocol is summarised in the following flowchart.

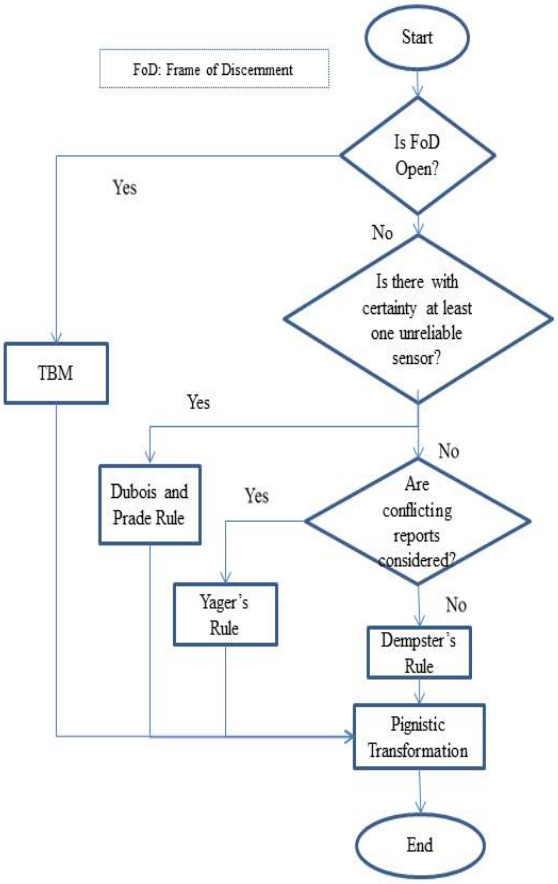


**Figure 1.** Flowchart of Protocol

In the next section, an example is demonstrated to show the implications of choosing a particular combination rule and how that might affect decision-making.

## 6 EXAMPLE

Below, we have two sensors with the following sensor reading outputs which are being fused. It is assumed that there is only 1 attack present, but there are 3 distinct possibilities as to what type of attack it might be. Note that for non-singleton masses, such as $m(\alpha_i \cup \alpha_j) = x$ this is represented by $(\alpha_i, \alpha_j, x)$. This is followed by Table 1 which lists the fused masses after applying each combination rule.

$$S_1 : \quad \{(\alpha_1, 0.2), (\alpha_2, 0.2), (\alpha_1, \alpha_2, 0.2), (\alpha_1, \alpha_3, 0.1)\}$$
$$S_2 : \quad \{(\alpha_1, 0.1), (\alpha_2, 0.1), (\alpha_3, 0.1)(\alpha_1, \alpha_2, 0.3), (\alpha_1, \alpha_3, 0.1), (\alpha_2, \alpha_3, 0.2), (\alpha_1, \alpha_2, \alpha_3, 0.1)\}.$$

Finally, Table 2 lists the results from using the pignistic transformation on the above masses to probabilities.

**Table 1.** Fused Masses from Example Using Different Combination Rules

| Combination Rule | | | |
|---|---|---|---|
| Dempster | TBM | Dubois & Prade | Yager |
| $0.28^{(\alpha_1)}$ | $0.23^{(\alpha_1)}$ | $0.02^{(\alpha_1)}$ | $0.23^{(\alpha_1)}$ |
| $0.28^{(\alpha_2)}$ | $0.23^{(\alpha_2)}$ | $0.02^{(\alpha_2)}$ | $0.23^{(\alpha_2)}$ |
| $0.07^{(\alpha_3)}$ | $0.06^{(\alpha_3)}$ | $0.26^{(\alpha_1,\alpha_2)}$ | $0.06^{(\alpha_3)}$ |
| $0.2^{(\alpha_1,\alpha_2)}$ | $0.17^{(\alpha_1,\alpha_2)}$ | $0.07^{(\alpha_1,\alpha_3)}$ | $0.17^{(\alpha_1,\alpha_2)}$ |
| $0.06^{(\alpha_1,\alpha_3)}$ | $0.05^{(\alpha_1,\alpha_3)}$ | $0.06^{(\alpha_2,\alpha_3)}$ | $0.05^{(\alpha_1,\alpha_3)}$ |
| $0.07^{(\alpha_2,\alpha_3)}$ | $0.06^{(\alpha_2,\alpha_3)}$ | $0.57^{(\alpha_1,\alpha_2,\alpha_3)}$ | $0.06^{(\alpha_2,\alpha_3)}$ |
| $0.04^{(\alpha_1,\alpha_2,\alpha_3)}$ | $0.17^{(\emptyset)}$ | | $0.2^{(\alpha_1,\alpha_2,\alpha_3)}$ |
| | $0.03^{(\alpha_1,\alpha_2,\alpha_3)}$ | | |

**Table 2.** Probability Distributions After Applying the Pignistic Transformation on the Masses from Table 1.

| ID | Probability Transformation for Different Combination Rules | | | |
|---|---|---|---|---|
| | Dempster | TBM | Dubois & Prade | Yager |
| $\alpha_1$ | 0.42 | 0.42 | 0.38 | 0.39 |
| $\alpha_2$ | 0.43 | 0.43 | 0.37 | 0.39 |
| $\alpha_3$ | 0.15 | 0.15 | 0.26 | 0.22 |

## 7 Discussion

Clearly for at least some cases, knowing whether the frame of discernment is open or closed can have a real impact on the most likely outcome after the probability transformation is applied to the masses. Furthermore, also being aware if at any given time the analyst can be sure if at least one sensor is accurate could impact his or her choice of combination rule and thus fused results. However, it is worth noting that there might be instances where the analyst is not sure whether the frame of discernment is open or closed. In that case, it is unclear whether the analyst should take a cautious approach (open frame of discernment) or not (closed frame of discernment). Perhaps, a trade-off and comparison between both approaches should be carried to determine the sensitivity of using one method over another. This may be the subject of further study.

In our example, both the TBM and Dempster's methods gave $\alpha_2$, while Dubois and Prade gave $\alpha_1$ as the most likely outcome. Yager could not decide between either of them. This is perhaps not so significant given that in all cases they were all very close to one another but it would be the case if decision making is based on the highest probability value. This underlines the importance of knowing the assumptions behind the fusion method being used.

Within the context of cyber, it is not believed that one combination rule is optimal. Given (i) the environment, (ii) what is sought to be protected, (iii) with the information available and (iv) confidence of sensor information, a specific combination is chosen. Whatever the combination rule chosen, all of them allow for ignorance and imprecision in a way which other techniques, like Bayesian approaches are not able to do. The introduction of time decay of information is another approach which can also be implemented and studied.

## 8 Concluding Remarks

This paper has given an overview of several of the main combination rules for Dempster-Shafer Theory, as well as some applications to sensor fusion in a cyber environment. More importantly, a protocol was given for when to use which rule given the assumptions under which the fusion is taking place. Through the use of a simple example, it was shown how an erroneous assumption in this respect can have profound implications at the decision-making level. It is hoped that the work presented here gives greater insight on the subtle differences

found in the different combination rules. Future potential work could concentrate on which combination rules would be applied given a specific cyber context.

## REFERENCES

Dempster, A. P. (1968). A generalisation of bayesian inference. *J. Royal Stat. Soc., Series B 30*, 205–247.

Denoeux, T. (2006). The cautious rule of combination for belief functions and some extensions. In *Proc. 9th Intl. Conf. on Inf. Fusion*, Florence, Italy.

Dezert, J. and F. Smarandache (2009). *Advances and Applications of DSmT for Information Fusion*, Volume 3. Perfect Paperback.

Dubois, D. and H. Prade (1986). A set theoretic view on belief functions: Logical operations and approximation by fuzzy sets. *Intl. J. Gen. Sys 12*, 193–226.

Ferson, S. and V. Kreinovich (2001). Representation, propagation and aggregation of uncertainty in risk analysis - from traditional probabilistic techniques to more general, more realistic approaches: A survey. University of Texas at El Paso Report UTEP-CS-01-33, University f Texas at El Paso.

Florea, M. C., J. Dezert, P. Valin, F. Smarandache, and A.-L. Jousselme (2006). Adaptive combination rule and proportional conflict redistribution rule for information fusion. In *Proc. Cogis'06*, Paris, France.

Josang, A. (2010). Cumulative and averaging fusion of beliefs. *Inf. Fusion 11*(2), 192–200.

Nguyen, V. and M. Docking (2015). Influx: A tool and framework for reasoning under uncertainty. DST Group Technical Report DST Group-TR-3142, Defence and Technology Group, Department of Defence, Australian Government.

Sentz, K. and S. Ferson (2017). Combination of evidence in dempster-shafer theory. SANDIA Report SAND2002-0835, Sandia National Laboratories.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.

Shafer, G. (1992). Dempster-shafer theory. In S. Shapiro (Ed.), *Encyclopedia of Artificial Intelligence, 2nd Ed.*, pp. 330–331. Wiley.

Smets, P. (1994). Representations, independence, and combination of evidence in the dempster-shafer theory. In R. Yager, M. Fedrizzi, and J. Kacprzyk (Eds.), *Advances in the Dempster-Shafer Theory of Evidence*, pp. 51–69. Wiley.

Smets, P. (2000). Data fusion in the transferable belief model. In *Proc. 3rd Intl. Conf. on Belief Functions*, Paris, France, pp. 21–33.

Yager, R. (1986). A set theoretic view on belief functions: Logical operations and approximation by fuzzy sets. *Intl. J. Gen. Sys 12*, 193–226.

Yager, R. (2007). Representation and combination of uncertainty with belief functions and possibility measures. *Comp. Intel. 4*(3), 244–264.