Noname manuscript No.
(will be inserted by the editor)

DFIOT: Data Fusion for Internet Of Things

Sahar Boulkaboul · Djamel Djenouri

Received: date / Accepted: date

Abstract In the Internet of Things (IoT) ubiquitous environment, a high volume of heterogeneous data and devices are produced in a quick span of time. In all IoT applications, the quality of information plays an important role in decision making, while providing reliable and accurate information becomes critical for such applications. Making this data precise and highly accurate is challenging an open research trend. Data fusion is one of the current trends that is considered in this paper. We particularly consider typical IoT scenarios where the source's measurements highly conflict, which makes intuitive fusions prone to wrong and misleading results. This paper proposes a taxonomy of decision fusion methods that rely on the theory of belief. It proposes a data fusion method for the Internet of Things (DFIOT) based on Dempster-Shafer (D-S) theory and an adaptive weighted fusion algorithm. It considers the reliability of each device in the network and the conflicts between devices when fusing data. This is while considering the information lifetime, the distance separating sensors and entities, and reducing computation. The proposed method uses a combination of rules based on the Basic Probability Assignment (BPA) to represent uncertain information or to quantify the similarity between two bodies of evidence. To investigate the effectiveness of the proposed method in comparison with D-S, Murphy, Deng and Yuan, a comprehensive analysis is provided using both benchmark data simulation and real dataset from a smart building testbed. Results show that DFIOT outperforms all the above men-

Sahar Boulkaboul
Department of informatics, Faculty of Exact Sciences, University of Bejaia, 06000 Bejaia, Algeria. CERIST Research Center, Algiers, Algeria
Tel.: +213-550-673675
E-mail: boulkaboul.sahar@gmail.com

Djamel Djenouri
Computer Science Research Centre, Department of Computer Science and Creative Technologies, University of the West of England, Bristol, UK
E-mail: djamel.djenouri@uwe.ac.uk

tioned methods in terms of reliability, accuracy and conflict management. The accuracy of the system reached up to 99.18% on benchmark artificial datasets and 98.87% on real datasets with a conflict of 0.58%. We also examine the impact of this improvement from the application perspective (energy saving), and the results show a gain of up to 90% when using DFIOT.

Keywords Wireless sensor networks · Dempster-Shafer theory · Conflict management · Energy consumption · Basic Probability Assignment · Similarity distance · Weighted evidences · Communication networks for data.

1 Introduction

IoT is considered as a global infrastructure for the information, where it is expected that billions of devices or things that are able of sensing, communicating, computing, and potentially of actuation will be connected to the Internet [1]. This includes sensors, Radio Frequency IDentification (RFID), cell phones, smarter watches, smart glasses, etc. The vision of IoT is to allow these things to be connected anytime, anywhere, with anything or anyone, ideally using any path, any network and any service. This will generate a huge amount of data coming from different sources, which arises the need for effective methods for processing such data [2]. Data fusion is defined as the theory, techniques and tools which are used for combining sensor data, or data derived from sensory data, into a common representational format [3]. It is also considered as the combination of information from different heterogeneous sources of measurement [4–6]. The goal in data fusion is to improve the performance of a given system by combining complementary or redundant information. The combination of redundant information makes it possible to reduce the uncertainty of the measurements, whereas the combination of complementary information makes it possible to obtain information that cannot be perceived by a single sensor. Data fusion is commonly used for detection and classification in different application domains [7–9], such as military, robotics, medical, earth sciences, and industrial applications.

Data fusion in distributed and heterogeneous environments such as IoT is challenging. IoT consists of heterogeneous systems where data are generally represented in several feature spaces. This makes it difficult to analyze relationships among different data, even when the datasets are related to each other semantically [10]. Scalability is another challenge in IoT, due to the frequent changes in the shape and size of the networks. Several formal fusion methods may be found in the literature. We consider in this paper the decision fusion, in which the fusion allows to infer decisional information. Decision fusion methods provide a formalism for combining evidence according to the probability theory rules, where uncertainty is represented using the conditional probability terms that describe beliefs.

The Dempster-Shafer (D-S) theory [11] is largely used for uncertainty reasoning, which allows processing uncertain or imprecise information without prior knowledge. This can be well used in IOT environment. It supports the representation of both imprecision and uncertainty, and it allows deriving the probabilities of a collection of hypothesis while dealing with missed information. This can be helpful to process the heterogeneous IoT data. However, under situations where the evidence highly conflicts, it may obtain counterintuitive results. This problem is tackled in this paper in which we first present the literature on decision fusion methods, and then propose a taxonomy of D-S approaches. The most common methods are reviewed by exploring benefits and challenges. A new efficient method is then developed, which emphasizes the importance of reducing the uncertainty of the measurements and conflicts in data fusion. It takes into consideration the contextual IoT parameters while reducing the degree of uncertainty. It also enhance the entropy of Deng, by selecting a sensor report which has a big information volume and well supported by the others sensors. This by exploring similarity between evidence and enhancing credibility. Results show that the proposed solution outperforms all the above mentioned methods in terms of reliability, accuracy, and conflict management.

The remainder of this paper is organized as follows. The related work is given in Section 2. The proposed method is presented in Section 3. In Section 4, simulation based comparison is presented, where the proposed solution is compared to state-of-the-art data fusion methods using benchmark datasets. A comparison using testbed datasets is presented in Section 5. Finally, Section 6 draws the conclusion.

2 Related Work and Background

The basic concept of the D-S evidence theory was originally developed in (1967) [12], and then evolved towards belief functions to model uncertain knowledge on the basis of mathematical formulations [13]. In D-S reasoning system, possible assumptions consist of all elements of indivisible hypothesis that are mutually exclusive and comprehensive. This includes a frame of discernment, denoted T, the space of inference system, θ, which includes all possible subsets of T elements. The number of possible combination (including the empty set) is 2^n, where n is the number of elements in T. There are three main functions in D-S theory, i) the basic belief mass function that specifies the belief mass distribution (m-values) over all possible sub-sets of a frame of discernment, ii) the Belief function, and iii) the Plausibility function.

The Mass function, which is also called a basic probability assignment, associates for every element, $E \in \theta$, $m(E)$ that is the proportion to all available evidence. The value of $m(E)$ therefore concerns only the state E and brings no credit to the subsets of E, which by definition have their own mass. The mass of the empty set is null ($m(\emptyset) = 0$), and the masses of θ's subsets sum up to 1, i.e., $\sum_{E \in \theta} m = 1$. The belief of a set, A, to a node, i, say $Belief_i(A)$, is defined as the sum of the masses of all A's subsets (Eq. 1).

$$Belief_i(A) = \sum_{E_k \subseteq A} m_i(E_k). \tag{1}$$

The plausibility measures the intensity with which the element, A, is found (with no doubt). It can also be interpreted as the maximum belief in A, or the sum of the evidence that is not against A. $Plausibility(A)$ is defined as the sum of the masses of all sets that intersects with A (Eq. 2).

$$Plausibility_i(A) = \sum_{E_k \cap A \neq \emptyset} m_i(E_k). \tag{2}$$

The procedure for fusing multiple evidence using the previously defined mass functions is an important issue in the D-S theory, which provides a method to compute the orthogonal sum ($m = m_1 \oplus m_2$) of two bodies of evidence according to Dempster's combination rule [14].

When there are multiple sources and the observations are assumed to be independent of each other, the combination of evidence in D-S provides away to combine these observations. For every proposition A in, θ, the combination rule between an object with a mass, m_i, and another object with a mass, m_j, is given by Eq. 3,

$$m(A) = \frac{\sum\limits_{B \cap C = A} m_i(B) m_j(C)}{1 - K}, \tag{3}$$

where K is a measure of conflict between the sources, which is also called inconsistency of the merger. It is given by Eq. 4.

$$K = \sum_{B \cap C = \emptyset} m_i(B) m_j(C). \tag{4}$$

The mass function obtained after combination tends to reinforce the belief of decisions for which the sources are consistent. Note that this combination of evidence rule is both associative and commutative. That is, the mass function can be the result of a combination of evidence between two other objects. The evidence of combination process for multiple sources can be chained, and the order in which the sources are combined does not affect the final results. The combination of conflicting evidence has always been challenging in D-S evidence theory [15]. Many methods have been proposed to solve this problem [14, 16, 17], but there is no universal solution thus far.

Several protocols have been proposed for data fusion in IoT that aims at improving decision-making. Some of these protocols rely on the combination rule, while some others are interested in the measurement of distance between nodes (weight) or amount of information. The relevant literature is reviewed in this section based on different categories of D-S approaches. We propose a taxonomy of the state-of-the-art protocols that are based on D-S theory using several criteria as shown in Fig. 1, and then we describe the most common methods. Two main categories may be distinguished, 1) protocols based on modified models in which measure of conflict is modified in combined rules, and 2) improved D-S based on modified methods in which a weight coefficient is used in the calculation of BPA.

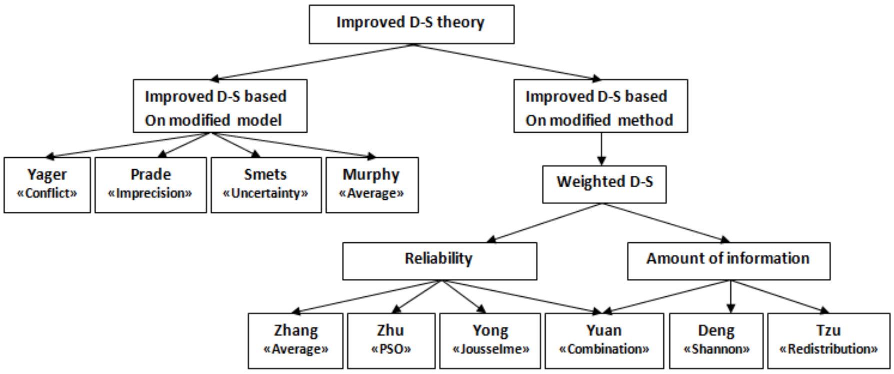


Fig. 1: Taxonomy of data fusion solutions

Starting with the first category, Yager [18,19] suggested that all conflicting evidence is unable to provide effective information, so he assigned all conflicts to unknowns $m(\theta)$. The improved formula can be used in high conflicting evidence combination, but the irrational distribution will lead to unreasonable results for assigning all conflicting evidence to the unknown. Smets [20] considers that the data sources are reliable. Based on this assumption, the conflict can only come from an ill-posed problem, i.e., the non-inclusion of one or several assumptions in the frame of discernment. Therefore, the author recommends redistributing the conflict mass K, but only on the empty set. Dubois and Prade [21] consider that the data sources are the unreliable part and assume that when a conflict exists between two data sources, at least one of the two sources is reliable. Given the impossibility of identifying the reliable source, they opted for redistributing the conflicting mass on the union of the two sources. Dubois and Prades rule doesnt work for dynamic fusion problems when a singleton or a union of singletons becomes empty. This problem is fixed by the sum S2 in the general D-S rule of combination. Murphy's [22] averaging approach suggested that if all the evidence are available at the same time, the average of evidence masses is calculated and then combined N time using D-S theory, where N is the Number total of evidence. However, this approach does not consider the association relationship and difference among the evidence.

The second category is further divided into two classes, reliability methods vs. the amount of information based methods. In reliability methods, the evidence distance of Jousselme [23] is used. The distance between two bodies of evidence, $d_{BOE}(m_1, m_2)$, is defined in Eq. 5.

$$d_{BOE}(m_1, m_2) = \sqrt{\frac{1}{2}(\overrightarrow{m_1} - \overrightarrow{m_2})^T D(\overrightarrow{m_1} - \overrightarrow{m_2})}. \tag{5}$$

where $\overrightarrow{m_1}$ and $\overrightarrow{m_2}$ are the vector forms of the evidence bodies. The size of each body is 2^θ. D is a $2^\theta * 2^\theta$ matrix, whose elements are given by Eq. 6,

$$D(s_1, s_2) = \frac{|s_1 \cap s_2|}{|s_1 \cup s_2|}, \quad \mathrm{s}_1, s_2 \in 2^\theta. \tag{6}$$

Yong [24] applied the evidence distance to obtain a weighted average combination and thus measure the conflict degree among evidence. The higher the distance between two bodies of evidence is, the less these two bodies of evidence support each other. If evidence conflicts highly with others, it will have less effect on the final combination result. The support degree and the credibility degree of each evidence are defined, respectively, with Eq. 7 and Eq. 8.

$$Sup(m_i) = \sum_{j=1, j \neq i}^{N} (1 - d(m_i, m_j)). \tag{7}$$

$$Crd_i = \frac{Sup(m_i)}{\sum_{j=1}^{k} Sup(m_j)}. \tag{8}$$

The credibility degree represents how reliable evidence is. The higher the credibility degree is, the more effective the evidence will have on the final combination result. Zhang [25] proposed a new method of combining conflicting evidence based on average. This method considered the association relationship among the evidence collected from multi-sources, and it weighs the evidence based on the distance of evidence. Zhu [26] proposed a new method for weighting evidence using PSO (Particle Swarm Optimization) algorithm to optimize the calculation of sources weight.

Context-aware data fusion [27–29] is employed in the basic concepts of IOT. The amount of information based methods use efficient tools to quantify information. These approaches can be applied in evidence theory where the uncertain information is represented by BPA. Deng entropy [30] is one of these methods and a generalization of Shannon entropy [31]. It is an efficient way to measure uncertainty, not only under the situation where the uncertainty is represented by a probability distribution, but also in the situation where the uncertainty is represented by BPA. This enabled its wide application in D-S evidence theory. When the uncertainty is expressed in the form of a probability distribution, Deng entropy degenerates to Shannon entropy.

The related concepts are given in the following. Let A_i be a proposition of BPA m; the cardinality of the set A_i is denoted by $|A_i|$. Deng entropy E_d of the set A_i is defined by Eq. 9

$$E_d = -\sum_i m(A_i) log \frac{m(A_i)}{2^{|A_i|} - 1}. \tag{9}$$

Tzu [32] proposes another method to calculate information volume, which reapportions the mass in the null set among the other subsets that are originally assigned to the null set. This is using Eq. 10.

$$I_{ev} = \sum_{i=1, A_i \neq \emptyset}^{n(A_i)} \frac{m(A_i)}{|A_i|}. \tag{10}$$

Yuan [33] combines the weighted credibility method in [24] to reduce conflict between evidence, and Deng entropy when calculating information volume.

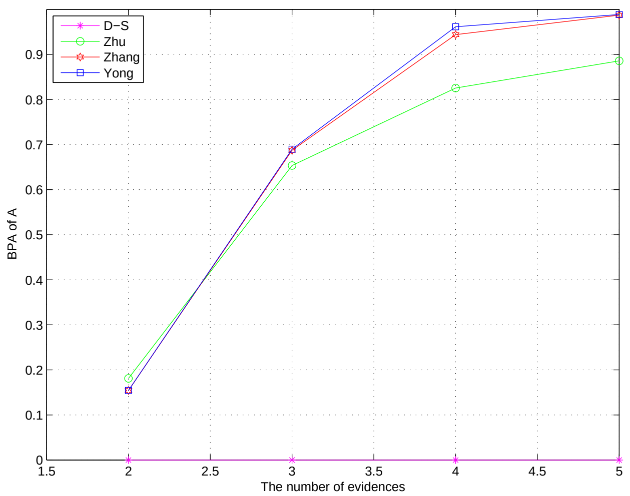
Contrary to most of the previous methods that are pointed to only one parameter to calculate the weighted evidence, the one proposed in this article improves the credibility degree by considering the evidence relationships, and enhancing the uncertainty degree using inner properties of evidence. We are particularly interested in the reliability of a sensor which is determined by the distance between the sensor and the entity in question, as well as the time validity of the information (lifetime).

3 Solution Description

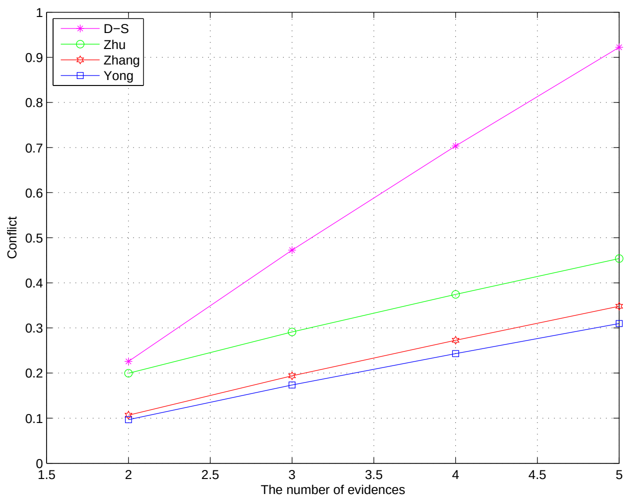
In this paper, a weighted evidence combination method is used. It is based on weight which represents the degree of confidence that is given to a data source. The method is largely used for uncertainty measure [33] to handle conflicting evidence combination and to take into consideration heterogeneous data in IoT [25,26]. Our solution addresses the features of IoT data fusion which are: 1) the uncertainty, as data provided by sensors, is always subjected to some level of uncertainty and inconsistency. Data fusion algorithms reduce uncertainty by combining data from several sources. 2) Conflicts to present quality data to users, which is critical to resolve conflicts and discover parameters' values that reflect the real world. 3) Energy consumption has always been challenging in wireless sensors networks, and consequently in IoT. To justify the choice of the weighted method upon which we rely for developing the proposed solution, a comparison between several state-of-the-art methods in each subcategory has been carried out. The results are presented in the following.

3.1 Comparison Between Weighted Methods

We implemented and compared the methods based on reliability, which are all based on jousselme distance. A benchmark numerical example has been used [25]. The results are given in Fig. 2a and Fig. 2b. They show that both Yong and Zhang methods give approximately the same BPA of hypothesis A (up to 98%). Yong method, however, has less conflict and increases its performance as the number of evidence goes up.



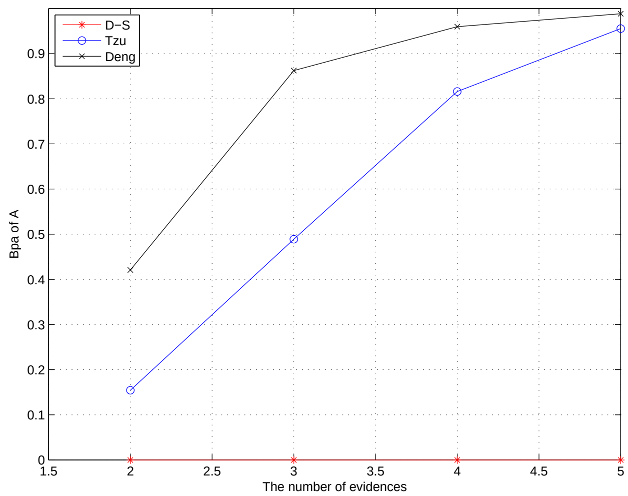
(a) BPA



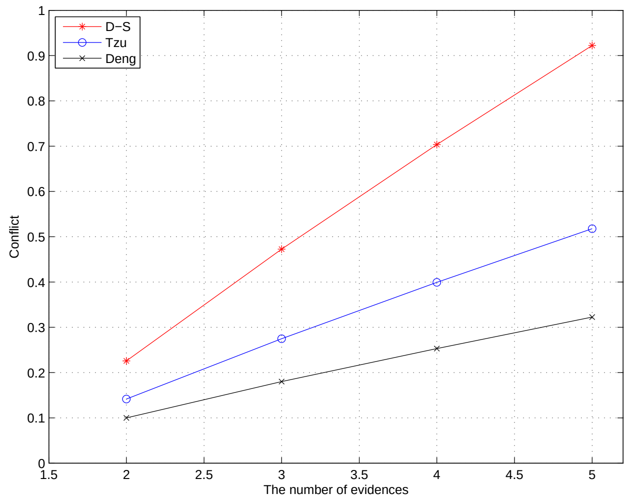
(b) Conflict

Fig. 2: Comparison between weighted data fusion methods based on reliability

We implement also the second category of data fusion methods based on the amount of information, and compare them using a benchmark numerical example [30]. The results are depicted in Fig. 3a and Fig. 3b. Deng entropy has clearly the best results in terms of accuracy (BPA) and reaches up to 98%, while D-S's BPA is equal to 0. Deng also has the best performance in terms, e.g., it remains at 0.32 for 5 bodies of evidence while the other solutions exceed 0.50. Based on these results, we use Yong method to calculate the credibility of evidence and Deng entropy to capture the information volume when combining evidence.



(a) BPA



(b) Conflict

Fig. 3: Comparison between weighted data fusion methods based on amount of information

The aim is to select the best improved method and use it to calculate the credibility degree of evidence in the first category and the uncertainty degree in the second one. The simulation results justify our choice of Yong and Deng entropy.

3.2 DFIOT Steps

DFIOT is based on D-S theory and Deng entropy. It proposes three methods IDeng, WDST and WDSD, to improve the information processing. IDeng is an improved Deng method that gives more importance to a sensor which has low uncertainty, i.e., more volume of information, and less evidence distance vs. other sensors. That is, if a sensor report has a big information volume and a less evidence distance, it will be well supported by other sensors and thus will have a higher weight proportion. Otherwise, a small weight proportion will be assigned to sensors with conflicting readings (having high evidence distance) or with a low volume of information. The contextual parameters used in DFIOT are, 1) the lifetime of sensed data, which is represented by Weighted Dempster Shafer method based on Time (WDST), and 2) the distance between sensors and the entity which is represented by Weighted Dempster Shafer method based on Distance (WDSD). DFIOT includes four steps as illustrated in Fig. 4, and detailed in the following:

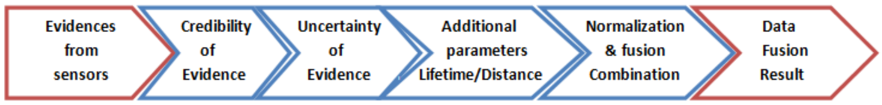


Fig. 4: The flowchart of DFIOT method

3.2.1 Calculate the Credibility Degree of Evidence:

For every piece of data collected by sensors, Eq. 5 is used to calculate the distance between every two bodies of evidence, and Eq. 7 to obtain the support degree of each evidence. Algorithm. 1 describes the calculation of evidence credibility.

3.2.2 Calculation of the Uncertainty Degree of Evidence:

We propose an Improved Deng method (IDeng) to calculate the information volume (degree of uncertainty) associated with each evidence. The Deng entropy presented in Section 2 is used. In fact, each evidence suffers from a degree of uncertainty that influences on the degree of its involvement in the final combination.

Three cases can be identified: 1) Evidence has a high degree of uncertainty (low information volume); its weight must be set to a small value (case of uncertainty). 2) Evidence has a low degree of uncertainty and significant evidence distance with the other evidence; its weight must be set to a small value (case of negative certainty). 3) Evidence has a low degree of uncertainty and low evidence distance; its weight must be set to a high value (case of positive certainty). The information volume of sensed data $I(SD)$ will be calculated for each evidence in two steps as follows:

Algorithm 1: Credibility computation of each evidence

Input : <m1,m2,m3,.> [vector of evidences]
N: number of evidence
$T = h1, h2, h3, ...$ frame of discernment.
Output: CRD vector

$CRD = zeros(N) : Initialization$
Sum=0;
SumCRD=0
Begin
z **for** $i = 1$ *to* N **do**
    for $j = 1$ *to* N **do**
        if $i \neq j$ **then**
            $Sum = Sum + 1 - d_{BPA}(mi, mj)$
            end if
        endfor
    CRD[i]=Sum;
    SumCrd=SumCrd+CRD[i];
    Sum=0;
    endfor
for $i = 1; i <= N; i++$ **do**
    $CRD[i] = CRD[i]/SumCrd$
    endfor
Return CRD vector
End

First, the entropy of each evidence is calculated with Eq. 9. Second, the distance of evidence is verified in terms of decision with other bodies of evidence. To determine if the certainty is positive or negative, the sum of the distance is calculated and the farthest evidence is ignored. If the sum reaches a certain threshold then the certainty is considered negative, otherwise it is considered positive. This is explained in Algorithm. 2. The final weight for each evidence is calculated using Eq. 11.

$$ED = 1 - \frac{ED}{sum(ED)}, \tag{11}$$

where $sum(ED)$ is the sum of all evidence.

3.2.3 Addition of Contextual Parameters:

In IoT applications, a large collection of sensors, devices, and users provide a large amount of information in different contexts. This information is usually prone to errors and lacks reliability and credibility. The information to be used must be analyzed, substantiated and motivated. Therefore, merging contexts while increasing the confidence to bring new information, and giving a complete view of the environment is important when fusing data in IoT.

To enhance the quality of data fusion in large heterogeneous and distributed wireless sensors networks used in IoT applications, we consider contextual parameters that are essential for any type of IoT application. These

Algorithm 2: Information volume computation of each evidence

```
Input  : <m1,m2,m3,.> [vector of evidences]
         N: evidence number
         T = h1, h2, h3, ··· frame of discernment.
Output: ED vector; information volume of sensors
ED= zeros(N); NC= zeros(N); booleen , NC:Negative Certainty
SumDisatance = 0; MaxDistance = 0;
Begin for i = 1 to N do
    SumDistance=<Sum of evidences>
    MaxDistance=<the largest distance between evidence i and the others>
    ED[i]=DengEntropy(mi);
    if SumDistance − MaxDistance) > 0.5(N − 2) then
        NC[i] = true;
    endif
    done
for each evidence i do
    ED = ED ∗ N/Sum of ED
    if NC[i] = true then
        RetunED = ED/2;
    endif
    endfor
Return the vector ED
```

parameters measure the degree of conformity of the IoT environment as perceived by the measuring device. The first parameter is sensor confidence distance, which is a quality parameter that measures the accuracy of the information in a context object. The quality of the decision made by a sensor is strongly affected by the resolution of the space, namely, the distance between the sensor and the entity in question. The second parameter that has been taken into account is the time validity or lifetime of the information. In fact, the information taken at time t is more important than information taken at $t-1$. Furthermore, the information becomes invalid after a certain time, which depends on the context of the application.

Lifetime of sensed data: A new parameter is added into WDST that measures the validity of contextual information. The lifetime of sensed data(SD) is normalized by Eq. 12

$$Lifetime(SD) = 1 - \frac{Age(SD)}{\Delta T_{max}}. \tag{12}$$

where ΔT_{max}, the maximum lifetime or total duration of the information when we recover all sensed data, Age(SD) represents the lifetime of the observation, which is given by Eq. 13 :

$$Age(SD) = T - T_{mes}(SD) \tag{13}$$

where T is the current time and $T_{mes}(SD)$ the time when SD was collected.

Distance between sensors and the entity: In the WDSD, the quality of the decision−making is strongly affected by the resolution of the space, i.e., the distance between the sensor and the monitored entity. The normalized function is given in Eq. 14 is used. The following function:

$$Dis(S) = 1 - \frac{D(S,E)}{D_{max}} * \delta. \tag{14}$$

where $D(S,E)$ denotes the distance between the sensor S and the entity E, $Dmax$ the maximum distance upon which the sensors observation can be trusted, and δ the accuracy of a sensor measured on the basis of statistical estimation.

3.2.4 Normalization of the weights and data fusion combination:

For each evidence, i, the weight W_i is defined with Eq. 15,

$$W_i = CRD(i) * ED(i) * Lifetime(i) * Dis(i). \tag{15}$$

Assume there are k bodies of evidence; the normalized weight is given by Eq. 16.

$$\widetilde{W_i} = \frac{W_i}{\sum_{j=1}^{k} W_j} (i = 1, 2, ..., k). \tag{16}$$

We use the weights obtained in Eq. 16 to calculate the BPA. The weighted bodies of evidence are then combined using orthogonal (Eq. 17). Notice that the Dempster's rule of combination has the attractive property of commutativity and associativity [34].

$$M = m_1 \oplus m_2 \oplus ... \oplus m_k. \tag{17}$$

4 Simulation and Comparative Analysis

A numerical analysis with MATLAB [35] is presented in this section. The proposed solution is compared with five state-of-the-art methods, including D-S [16], Murphy [22], Yong [24], Deng [30], and Yuan [33]. These methods are the most efficient from the literature for conflict management, and thus the most relevant for comparison. A benchmark numerical example [25] has been used. In a system of automatic target recognition $\{A, B, C\}$ based on different types of sensors (CCD, sound, infrared, radar and ESM), assume that the current target is A and the number of sensors is five. The system has collected five bodies of evidence, the results obtained are shown in Table. 1.

Notice for the BPAs given above that the detection of $S2$ is abnormal. This may lead to a contradictory result after fusion. As the information concerning the lifetime of sensed data, and the distance between sensor and entity are required in the benchmark example, we used our solution without considering

Table 1: BPAs of the example

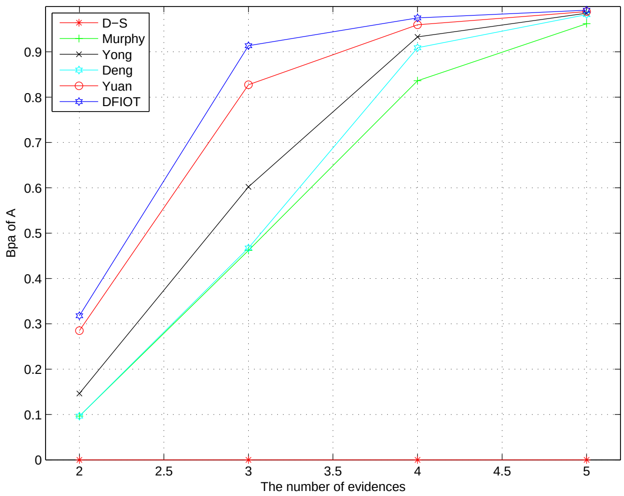
	A	B	C	A,C
S1:m1(.)	0.41	0.29	0.3	0
S2:m2(.)	0	0.9	0.1	0
S3:m3(.)	0.58	0.07	0	0.35
S4:m4(.)	0.55	0.1	0	0.35
S5:m5(.)	0.6	0.1	0	0.3

contextual parameters. Two metrics are used in the simulation, 1) certainty in the decision (BPA), and 2) the conflict between bodies evidence. Table. 2 depicts the fusion results when varying the combination rules and the number of bodies of evidence. The calculation process about the last column of the proposed method is given.

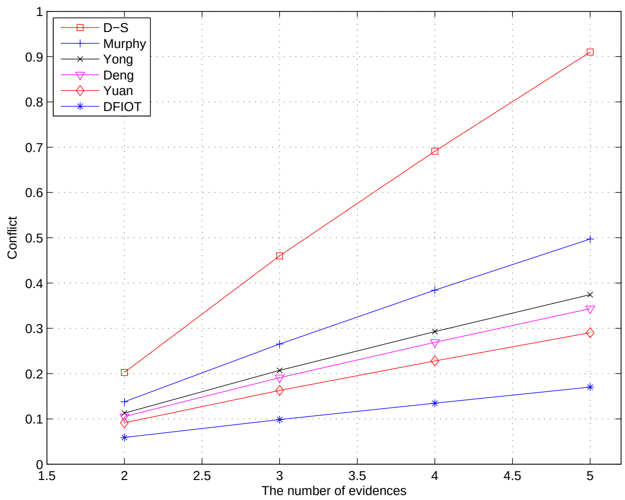
Table 2: The results of different combination rules

Method	Fusion Results			
	m1,m2	m1,m2,m3	m1,...,m4	m1,...,m5
D-S	m(A)=0 m(B)=0.8969 m(C)=0.1031	m(A)=0 m(B)=0.6575 m(C)=0.3425	m(A)=0 m(B)=0.3323 m(C)=0.6679	m(A)=0 m(B)=0.1422 m(C)=0.8578
Murphy	m(A)=0.0964 m(B)=0.8119 m(C)=0.0917 m(AC)=0	m(A)=0.4619 m(B)=0.4498 m(C)=0.0792 m(AC)=0.0090	m(A)=0.8362 m(B)=0.1147 m(C)=0.0410 m(AC)=0.0081	m(A)=0.9620 m(B)=0.0210 m(C)=0.0138 m(AC)=0.0032
Yong	m(A)=0.1463 m(B)=0.7620 m(C)=0.0917 m(AC)=0	m(A)=0.6021 m(B)=0.2907 m(C)=0.0990 m(AC)=0.0082	m(A)=0.9330 m(B)=0.0225 m(C)=0.0353 m(AC)=0.0092	m(A)=0.9851 m(B)=0.0017 m(C)=0.0096 m(AC)=0.0035
Deng	m(A)=0.0964 m(B)=0.8119 m(C)=0.0917 m(AC)=0	m(A)=0.4674 m(B)=0.4054 m(C)=0.0888 m(AC)=0.0084	m(A)=0.9089 m(B)=0.0444 m(C)=0.0379 m(AC)=0.0089	m(A)=0.9820 m(B)=0.0008 m(C)=0.0089 m(AC)=0.0036
Yuan	m(A)=0.2849 m(B)=0.5306 m(C)=0.1845 m(AC)=0	m(A)=0.8274 m(B)=0.0609 m(C)=0.0986 m(AC)=0.0131	m(A)=0.9596 m(B)=0.0032 m(C)=0.0267 m(AC)=0.0106	m(A)=0.9886 m(B)=0.0002 m(C)=0.0072 m(AC)=0.0039
DFIOT	m(A)=0.3178 m(B)=0.5233 m(C)=0.1589 m(AC)=0	m(A)=0.9134 m(B)=0.0039 m(C)=0.0395 m(AC)=0.0333	m(A)=0.9748 m(B)=0.0002 m(C)=0.0066 m(AC)=0.0183	m(A)=0.9918 m(B)=0.0001 m(C)=0.0061 m(AC)=0.0020

Fig. 5a shows the evolution of belief's value assigned to the target A (the right decision) after each combination of five bodies of evidence for the compared methods. Fig 5a shows that the accuracy (BPA) of the proposed solution is always superior to all the other solutions and reaches up to 99.18%. More importantly, while all solutions converge to BPA values beyond 95%, DFIOT grows very fast. It exceeds 90% with only three bodies of evidence, while Yuan's method reaches about 82% and the others remain below 60%. DFIOT also provides the best performance for conflict that was smoothly increasing between 0.058 and 0.170 as shown in Fig 5b, while all the other solutions exceed 0.29 for five bodies of evidence (some go up to 0.9). These results con-



(a) The belief value (BPA) of the target A



(b) The value of conflict between ~~evidences~~evidence

Fig. 5: The fusion results comparison between DFIOT and different rules

firm the efficiency of the proposed method and the effectiveness of considering conflict, information volume.

5 Experimental Performance Evaluation

The model is evaluated through an extensive set of experiments realized in (CERIST-ALGERIA) research center lab in the context of IoT and smart building project. In addition to the simulation study presented in the previous section. A real dataset collected from a testbed has been used for further investigation on the performance of DFIOT in comparison with state-of-the-art solutions. An IoT based building automation application has been considered in the office scenario, where presence and ambient light are permanently monitored with IoT enabled wireless sensors to optimize lighting and energy control of appliances and HVAC systems. 4 sensors have been placed on the ceiling of an office, including 3 PIR sensors and a light sensor. The reason behind using three sensors is the need to detect even small movements when the office is occupied (e.g., typesetting on the keyboard, head movements, etc.), which usually occur at the desk space. No placement of a single sensor can guarantee total coverage of this space. Optimal positions of the PIR sensors have been calculated using an integer linear programming (ILP) model and CPLEX solver, where the space has been uniformly split into equal distance grids. The outcome for a $30cm$ granularity unit provides optimal number of sensors that cover the whole office space (centers of cells in the space) to be four, with the positions depicted in Fig. 6.

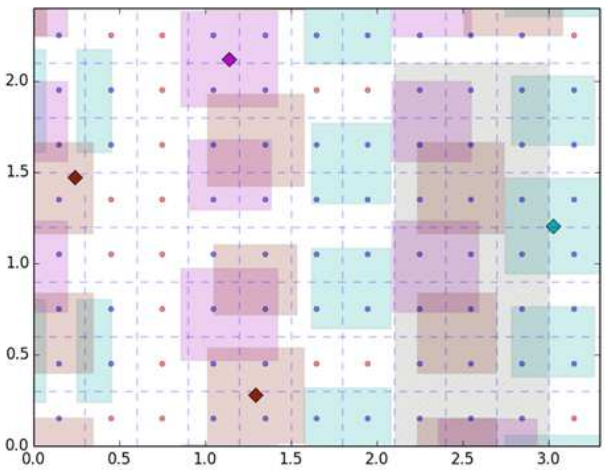


Fig. 6: Deployment of sensors in office

A handy device at the entrance has been added as a ground truth sensor for the presence, with buttons the occupants are asked to push on for every entrance/exit during the experiment. Every node acts as a source and sends data periodically to the central station that stores the data in a dataset. We define 4 hypothesis $(H1, H2, H3, H4)$ as follows. $H1$: office occupied and light value is more than $580lux$, $H2$: office empty and light value is more than $580lux$, $H3$: office occupied and light value is not exceeding $580lux$. $H4$: office empty and light value is not exceeding $580lux$. The frame of discernment is $o = H1, H2, H3, H4$. Without loss of generality, we use a simple scenario;

light control, where the data fusion method is applied to make a decision of switching on/off the light. The BPA has been calculated using the mean values in each state and the environment has been simulated with a standard error and confidence interval of 98%. In the following, we consider the situations when hypothesis $H1$ and $H2$ are verified. The ground truth has been used to filter out entries in the dataset where each hypothesis is verified. In Table. 3, the system has collected bodies of evidence when hypothesis $H1$ is verified. This includes PIR sensors, light sensor, and a ground truth sensor:

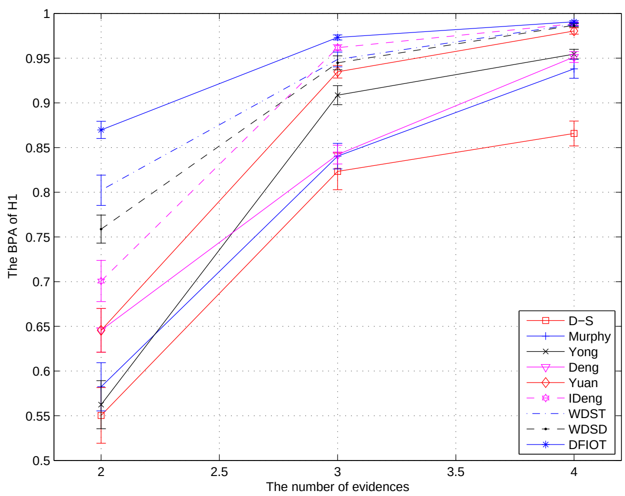
Table 3: BPAs of the solution

	H1	H2	H3	H4
S1:m1(.)	0.72	0.17	0.10	0
S2:m2(.)	0.69	0.08	0.22	0.01
S3:m3(.)	0.81	0.06	0.11	0.02
S4:m4(.)	0.83	0.07	0.09	0

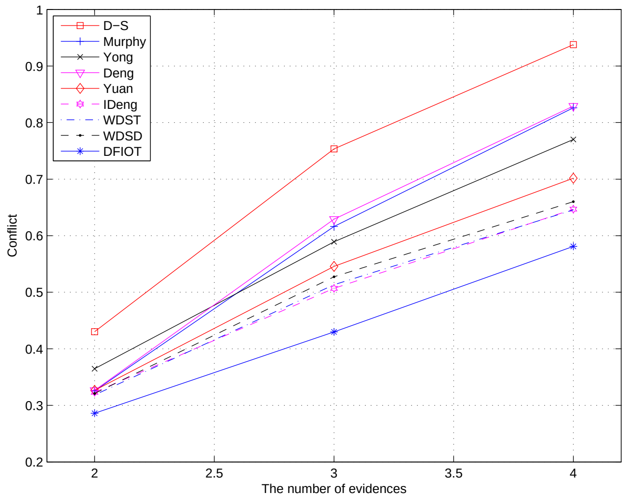
~~The BPA of Hypothesis~~ To evaluate the performance of our solution, we present the BPA and the conflict degree ~~are presented~~ of both Hypothesis H1 and H2 vs. the variation of evidence numbers.

5.1 BPA/Conflict of Hypothesis H1

We illustrate the three proposed methods (IDeng, WDST and WDSD) separately to investigate the efficiency of each method of DFIOT. We compare the results with five state-of-the-art methods as shown in Fig 7a and Fig 7b.



(a) BPA of Hypothesis H1



(b) Conflict of Hypothesis H1

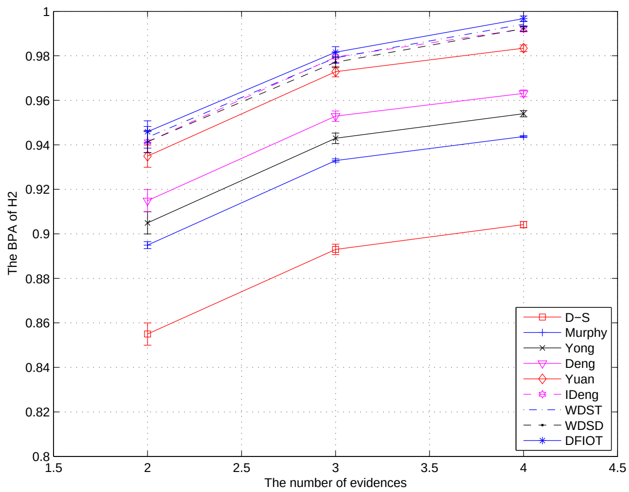
Fig. 7: The fusion results for Hypothesis H1

In D-S evidence theory, the BBA of H1 is 87%, and it decreases gradually with the increase of evidence. The conflict hugely increases and reaches 0.92. In Yuan method, the belief degree of H1 is 98%, while the new method has a higher belief degree of 99.18%. The main reason is that the proposed method takes into consideration the dynamic reliability measured by evidence distance and entropy while adding new parameters that enhances the Deng entropy. This decreases the conflict to less than 0.6.

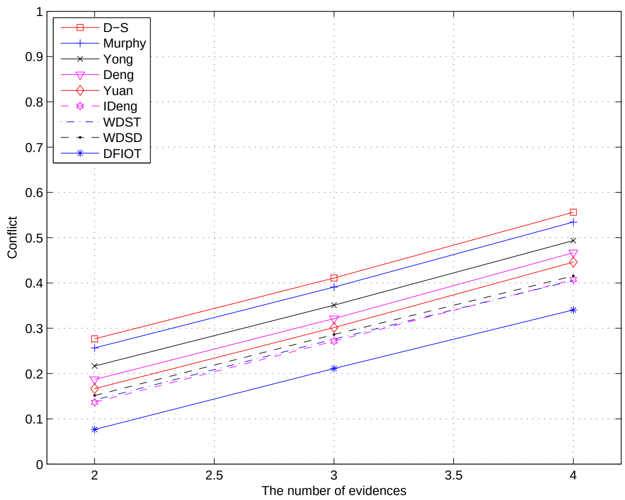
These results improve accuracy and confirm the efficiency of DFIOT. Each parameter added considerably increases the BPA. The results presented in Fig 7a and Fig 7b confirm those presented in the previous section and the superiority of the proposed solutions. Similarly, BPA of DFIOT and all its versions grows much faster than the other solutions, and have lower conflict values.

5.2 BPA/Conflict of Hypothesis H2

The hypothesis, in this case, is H2 that is when we have no presence detection (empty office) and light exceeding $580lux$. The results are depicted in Fig 8a and Fig 8b and confirm the superiority of our method. The results are very similar to those obtained for hypothesis H1, except that Yuan provides a bit better performance in terms of BPA (compared to its performance in hypothesis H1), but still clearly lower than DFIOT.



(a) BPA of H2



(b) Conflict of Hypothesis H2

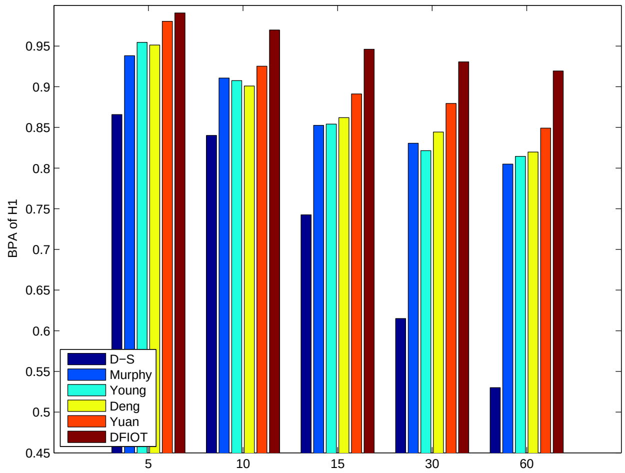
Fig. 8: The fusion results comparison with Hypothesis H2

A sensor may be misled by many factors and provide wrong evidence, and such abnormal measurement can generate a conflicting mass during the evidence combination which leads to the conflict. As shown in the simulation section, the classical Dempster's rule cannot support the correct hypothesis H1 and cannot eliminate the conflict between bodies of evidence. and a wrong final result. With incremental evidence, our solution provides better results.

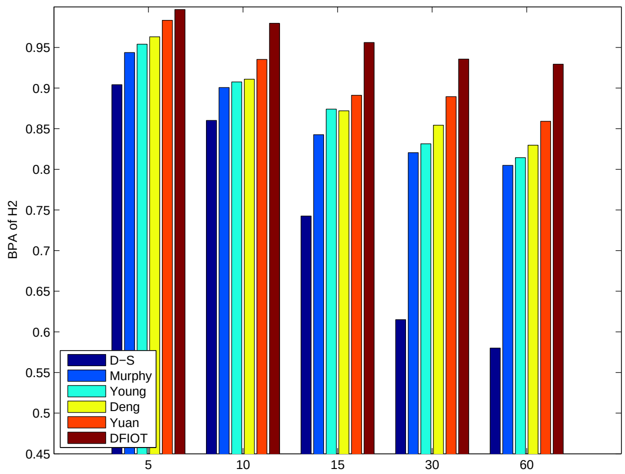
5.3 Impact of Data Fusion Period

5.3.1 Impact on BPA

We investigated the impact of the data fusion period on the BPA. Results in Fig. 9a shows that the proposed solution provides the best values and is less affected by increasing this period compared to the other solutions. It remains beyond 90% even for as a high period as 60min.



(a) Hypothesis H1



(b) Hypothesis H2

Fig. 9: Data fusion period

Similarly in Fig. 9b, when Hypothesis $H2$ has applied the results show the effectiveness of our method when reducing the period of data fusion. While the other solutions are influenced considerably by the increase of the data fusion period, especially in D-S method where the BPA is less than 50% in both cases of H1 and H2, which leads to a wrong decision (switching on/off the light).

~~Hypothesis H1~~ ~~Hypothesis H2~~ ~~Data fusion period~~

5.3.2 Energy Gain

We investigated the impact of this data fusion on the application performance, in our case energy gain is used as the application related metric. It is defined in Eq. 18.

$$Gain(A) = \frac{\sum_{i=1}^{n} \alpha_i}{\sum_{j=1}^{k} \gamma_j}, \tag{18}$$

where α_i represents the total period when the office is unoccupied (hypothesis $H2$ and hypothesis $H4$ are verified.), and γ_j represents the total period of data fusion in a day. As shown in Fig. 10, the gain increases with the increase of the fusion frequency (decreasing the period). The increase is fast for values bellow $30min$ and reaches up to 90% for a data fusion period of $5min$, which is a reasonable period.

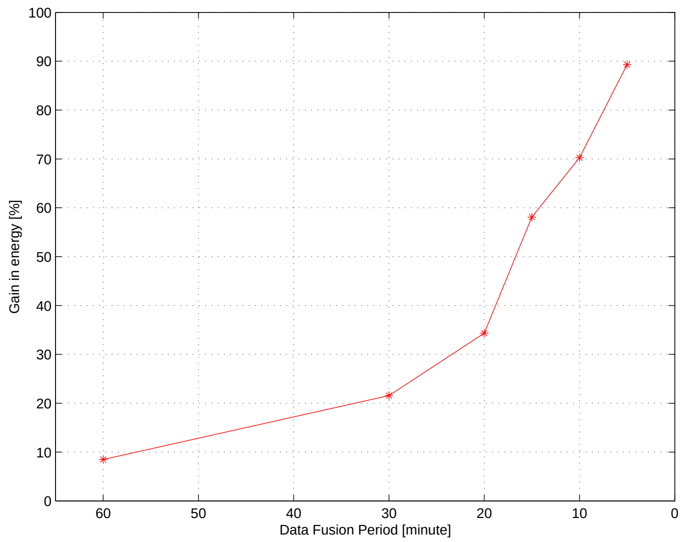


Fig. 10: Gain in energy consumption

Contrary to the previous weighted methods that either calculate the weight of each evidence based on similarity distance to enhance credibility degree, or take into consideration the information volume to upgrade the uncertainty degree of evidence. DFIOT takes into consideration several parameters when calculating the mass function. DFIOT has several advantages: First, It improves the Deng entropy in uncertainty degree of evidence by giving more importance to sensors which have low uncertainty, i.e., more volume of information and less evidence distance vs. other sensors and by eliminating the evidence with the farthest distances when making decisions. It applies evidence distance in measuring conflict degree and credibility. The second advantage that has been considered is the contextual parameters, which assures accuracy of the information affected by the resolution of the space. It focuses on the most recent information and avoids invalid information using the lifetime of sensed data. Another critical advantage of our method is that it helps to take the right decision when fusing the critical information. Examples of this include critical medical areas that need the lifetime of measurements to make sensitive decisions, in military applications that take the distance between entity and sensors as essential parameters, and in smart buildings that help to combine measurements and take optimal actions.

6 Conclusion

A novel data fusion method for the Internet of Things, we called DFIOT, has been proposed in this paper. This method uses an adaptive weighted fusion algorithm based on D-S theory. The reliability of devices in the network and the conflicts between devices are considered in DFIOT. This by considering the information lifetime, the distance separating sensors and entities, reducing the computation and by using combination rules based on the Basic Probability Assignment, which allows to represent uncertain information or to quantify the similarity between two bodies of evidence. We compared the proposed solution with some state-of-the-art data fusion methods, including D-S, Murphy, Deng and Yuan, and using both benchmark data simulation and real dataset from a smart building testbed. Results show that DFIOT outperforms all the above mentioned methods in terms of reliability, accuracy and conflict management. The impact of this improvement from the application performance perspective has also been investigated, and the results show a gain of up to 90% in energy saving when using DFIOT.

This solution opens perspectives for the application of our method in the emerging IoT applications. Integrating a virtual assistant to facilitate interaction with the system is a possible perspective. As the second perspective, the realized method can be combined with approaches implementing the data analysis techniques to exploit the saved data. The third perspective is the integration of machine learning to allow the system to learn habits of users and plan actions. The last perspective in our agenda is the integration of artificial intelligence algorithms to enable the system making autonomous decisions when necessary. A typical example of this (on which we are currently working) is the use of augmented and virtual reality based to develop mobile interfaces for IoT application system. The data fusion framework proposed in this paper facilitates the development of these solutions.

References

1. L. Atzori, A. Iera, G. M. The internet of things: A survey. *Comput. Netw. 54*, 15 (Oct. 2010), 2787–2805.
2. Gubbi, J., Buyya, R., Marusic, S., and Palaniswami, M. Internet of things (iot): A vision, architectural elements, and future directions. *Future Generation Computer Systems 29*, 7 (2013), 1645 – 1660.
3. Mitchell, H. Multi-sensor data fusion: An introduction. https://www.amazon.com/Multi-Sensor-Data-Fusion-H-B-Mitchell-ebook/dp/B000SHOM3G *Springer Verlag*,(2007).
4. Niu, W., Lei, J., Tong, E., Li, G., Chang, L., Shi, Z., and Ci, S. A survey of fault management in wireless sensor networks. *Journal of Network and Systems Management 22*, 1 (Jan 2007), 50–74.
5. Data fusion. https://algo-data.quora.com/Data-Fusion-an-overview-of-some-relevant-works.
6. Orchestrator, J. E. E. O. I. O. T. https://docs.oracle.com/en/middleware/index.html Jd edwards enterprise one internet of things orchestrator, 2015.
7. Abu-Elkheir, M., Hayajneh, M., and Ali, N. A. Data management for the internet of things: Design primitives and solution. *Sensors 13*, 11 (2013), 15582–15612.
8. R. Ranjan, M. Wang, C. P. P. J. M. Z. P. S., and Shyamsundar, P. City data fusion: Sensor data fusion in the internet of things. *Int. J. Distrib. Syst. Technol. 7*, 1 (Jan. 2016), 15–36.
9. Shen, G., and Liu, B. Information resources management association. In *The Internet of Things: Breakthroughs in Research and Practice* (2017), p. 530.
10. Santos, J., Wauters, T., Volckaert, B., and De Turck, F. Resource provisioning for iot application services in smart cities. In *2017 13th International Conference on Network and Service Management (CNSM)* (Nov 2017), pp. 1–9.
11. Guan, and Bell. Evidence theory and its applications. In *Studies in Computer Science and Artificial Intelligence 7, Elsevier* (1991), vol. 1.
12. Shafer, D. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.
13. Dempster, A. Upper and lower probabilities induced by a multivalued mapping. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*. 2008, pp. 57–72.
14. A. Tazid, P. D., and Boruah, H. A new combination rule for conflict problem of dempster-shafer evidence theory. *International Journal of Energy, Information and Communications 3*, 1 (2012), 35.
15. S. Le Hegarat-Mascle, I. B., and Vidal-Madjar, D. Application of dempster shafer evidence theory to unsupervised classification iin multisource remote sensing. *IEEE Transactions on Geoscience and Remote Sensing 35*, 4 (jun 1997).
16. Boston, J. A signal detection system based on Dempster-Shafer theory and comparison to fuzzy detection. *IEEE Trans. Systems, Man, and Cybernetics, Part C 30*, 1 (2000), 45–51.
17. Y. Li, J. C., and Lin, Y. An efficient combination method of conflict ~~evidences~~evidence. *International Journal of Hybrid Information Technology 8*, 12 (2015), 299–306.
18. Yager, R. Decision making using minimization of regret. *International Journal of Approximate Reasoning 36*, 2 (2004), 109 – 128.
19. Yager, R., and Filev, D. Including probabilistic uncertainty in fuzzy logic controller modeling using dempster-shafer theory. *IEEE Trans. Systems, Man, and Cybernetics 25* (1995), 1221–1230.
20. Smets, P. The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence 12* (1990), 447 – 458.
21. Dubois, D., and Prade, H. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence 4*, 3 (1988).

22. Murphy, C. Combining belief functions when evidence conflicts. *Decis. Support Syst. 29*, 1 (July 2000), 1–9.
23. A.L. Jousselme, D. Grenier., and É. Bossé. A new distance between two bodies of evidence. *Information Fusion 2*, 2 (2001), 91–101.
24. D. Yong, S. WenKang, Z. Z., and Qi, L. Combining belief functions based on distance of evidence. *Decis. Support Syst. 38*, 3 (Dec. 2004), 489–493.
25. Z. Zhang, T. Liu, D. C., and Zhang, W. Novel algorithm for identifying and fusing conflicting data in wireless sensor networks. *Sensors (Basel, Switzerland) 14*, 6 (2014), 95629581.
26. P. zhu, B. X., and Xu, B. An Improved Particle Swarm Optimization for Uncertain Information Fusion. *Springer Berlin Heidelberg, Berlin, Heidelberg*, 2011, pp. 494–501.
27. Gite, S., and Agrawal, H. On context awareness for multisensor data fusion in iot. In *Proceedings of the Second International Conference on Computer and Communication Technologies* (2016), Springer, pp. 85–93.
28. Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. Context aware computing for the Internet of things: A survey. *IEEE Communications Surveys Tutorials 16*, 1 (First 2014), 414–454.
29. Baloch, Z., Shaikh, F. K., and Unar, M. A. A context-aware data fusion approach for health-iot. *International Journal of Information Technology.Springer 10*, 3 (Sep 2018), 241–245.
30. Deng, Y. Deng entropy: a generalized Shannon entropy to measure uncertainty, Jan. 2015.
31. Shannon, C. E. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev. 5*, 1 (Jan. 2001), 3–55.
32. Lin, T. Improving D-S evidence theory for data fusion system. 2015.
33. K. Yuan, F. Xiao, L. F. B. K., and Yong, D. Conflict management based on belief function entropy in sensor fusion. *SpringerPlus 5*, 1 (May 2016), 638.
34. Judea, P. Reasoning with belief functions: An analysis of compatibility. *Int. J. Approx. Reasoning 6*, 3 (1992), 425–443.
35. Moore, H. *MATLAB for Engineers. Prentice Hall Press.* 2014.

Author Biographies

Sahar Boulkaboul Obtained her master degree from the University of Bejaia. Now she is a Ph.D student and also a permanent Full-time Researcher at the Center of Research (CERIST) in Algiers. She is a member of WSN group, where she worked mainly on data processing in IOT and smart cities.

Djamel Djenouri Obtained the PhD in Computer Science from USTHB, Algeria, in 2007. He was a postdoctoral researcher at NTNU, Norway, from 2008 to 2009. We worked for CERIST research center in Algeria, where he was a director of research and deputy director. He is currently with the university of the West of England (UWE), UK. He is working on IoT, WSN, ML and application for smart environments. He published more than 100 papers in journals and conferences and gave several keynotes and invited talks. He organised many conferences and participated in the TPC of numerous international conferences. He is a senior member of the ACM and AGYA.