

DRO

Deakin University's Research Repository

This is the published version:

Izadi, Davood, Abawajy, Jemal H., Ghanavati, Sara and Herawan, Tutut 2015, A data fusion method in wireless sensor networks, *Sensors*, vol. 15, no. 2, pp. 2964-2979.

Available from Deakin Research Online:

<http://hdl.handle.net/10536/DRO/DU:30070772>

Reproduced with the kind permission of the copyright owner.

Copyright : 2015, MDPI

Article

A Data Fusion Method in Wireless Sensor Networks

Davood Izadi ¹, Jemal H. Abawajy ¹, Sara Ghanavati ¹ and Tutut Herawan ^{2,*}

¹ School of Information Technology, Deakin University, 3220 Waurn Ponds, Geelong, VIC 3216, Australia; E-Mails: dizadi@deakin.edu.au (D.I.); jemal.abawajy@deakin.edu.au (J.H.A.); sghanava@deakin.edu.au (S.G.)

² Department of Information System, University of Malaya, 50603 Pantai Valley, Kuala Lumpur, 50603, Malaysia

* Author to whom correspondence should be addressed; E-Mail: tutut@um.edu.my; Tel.: +60-3-7967-2509; Fax: +60-3-7957-9249.

Academic Editor: Leonhard M. Reindl

Received: 11 September 2014 / Accepted: 5 December 2014 / Published: 28 January 2015

Abstract: The success of a Wireless Sensor Network (WSN) deployment strongly depends on the quality of service (QoS) it provides regarding issues such as data accuracy, data aggregation delays and network lifetime maximisation. This is especially challenging in data fusion mechanisms, where a small fraction of low quality data in the fusion input may negatively impact the overall fusion result. In this paper, we present a fuzzy-based data fusion approach for WSN with the aim of increasing the QoS whilst reducing the energy consumption of the sensor network. The proposed approach is able to distinguish and aggregate only true values of the collected data as such, thus reducing the burden of processing the entire data at the base station (BS). It is also able to eliminate redundant data and consequently reduce energy consumption thus increasing the network lifetime. We studied the effectiveness of the proposed data fusion approach experimentally and compared it with two baseline approaches in terms of data collection, number of transferred data packets and energy consumption. The results of the experiments show that the proposed approach achieves better results than the baseline approaches.

Keywords: wireless sensor network; data fusion; fuzzy logic controller; data redundancy

1. Introduction

Wireless sensor networks (WSNs) have been used in various domains such as military applications (e.g., military surveillance) and civil applications (e.g., industrial control and wildlife monitoring). Generally, a large number of sensor nodes, capable of collecting data, processing and communicating among themselves as well as with a base station (BS) are deployed in the sensing field to collect data according to a specific application. The sensor measurements are referred to as a sensor reading or sensor value. The corresponding correct value of an event in the environment is referred to as the true value. If a sensor reading and the true value disagree, the sensor reading is said to be incorrect.

A fundamental issue in WSNs is how various applications such as event detection, target tracking, and decision making can use the sensor measurements with increased confidence and as minimum energy consumption as possible in the presence of imprecise sensor readings. The incorrect sensor readings can be attributed to a number of factors. For example, strong variations of pressure, temperature, radiation, and electromagnetic noise in the monitored area might interfere with the sensor node readings and subsequently could lead to imprecise sensor readings. Furthermore, a sensor node itself might in some cases collect incorrect data due to failure, spatial and temporal coverage problems. Moreover, the neighboring sensors within the sensing field often generate duplicate and highly correlated data, which might also decrease the QoS. In order to overcome these problems, a data fusion mechanism can be used to remove the incorrect and duplicated data from the sensor measurements.

Data fusion mechanisms process the data from multiple sensors and thereby create meaningful new information that cannot be obtained from any single sensor. The main purpose of data fusion mechanisms in WSNs is to provide a greater QoS for the purpose of arriving at reliable and accurate decisions about the events of interest. The QoS here can mean reliable delivery of accurate, complete, and dependable information. In fact, fusing data ensures that not only the data quality of the WSN is enhanced, but also energy consumption can be lowered as it removes redundant information as well.

There are many data fusion mechanisms with the purpose of reducing the energy consumption for WSNs [1–3]. These mechanisms use different techniques such as probability theory [4], fuzzy set theory [5], a combination of fuzzy sets and neural network [6] and Dempster-Shafer evidence theory (DSET) [7]. Most of these approaches are able to eliminate duplicate data in the fusion process. However, these approaches do not consider specific limitations of the sensor devices. For example, they assume that the sensor nodes are always functioning properly and generating accurate data. These are unrealistic assumptions as the environment may change. For instance, the temperature can influence the accuracy of the sensor node functionalities. Moreover, the existing approaches transfer both necessary and unnecessary sensed data to the processing center, which results in excessive energy consumption.

In this paper, a fuzzy-based data fusion approach for WSNs that increases the QoS whilst maximizing the network lifetime by minimizing the energy consumption is presented. By virtue of distinguishing and aggregating only the true values of the sensed data, the proposed approach is able to reduce the transmission as well as the processing of the entire sensed data. It is also able to eliminate redundant data and consequently reduce energy consumption thus increasing the network lifetime. In summary, we make the following contributions:

- (a) We propose a data fusion approach to improve the performance of a WSN with respect to the level of QoS generated about the events of interest.

- (b) We minimize energy consumption by transferring only the calculated result of the events instead of the entire fused data.
- (c) We show that the proposed approach is robust in terms of the events of interest with respect to the sensor node failures as it combines all received data from the sensor nodes.
- (d) We studied the effectiveness of the proposed data fusion approach experimentally and compared it with two baseline approaches in terms of data collection, number of transferred data packets and energy consumption. The results of the experiment show that the proposed approach achieves better results than the baseline approaches.

The rest of this paper is organized as follows: Section 2 presents an in-depth analysis of the existing approaches. Section 3 describes the system model and problem overview. In Section 4 the proposed data fusion approach is explained in details. In Section 5 the performance of the proposed approach is analyzed. Finally, the conclusions are presented in Section 6.

2. Related Works

In this section, a review of some common protocols that have been proposed to aggregate data is presented. An overview of type-2 fuzzy logic is also discussed. Wireless sensor network applications such as surveillance and reconnaissance generate a large amount of redundant data. There have been many approaches to filter redundant sensing data using data aggregation techniques. These approaches can be classified as cluster-based, tree-based, grid-based and structure-free approaches. A hybrid clustering-based data aggregation mechanism that concurrently combines static and dynamic clustering methods is discussed in [8]. The algorithm chooses a suitable clustering technique based on the status of the network. However, the authors assume that the entire collected data from each sensor node is the true value. That assumption obviously reduces the ratio of packet transmission in the WSN. Another cluster-based data aggregation scheme is discussed in [9]. As the approach is a geographic location-based multicast protocol, data is aggregated based on their geographic locations in their clusters. However, the energy consumption efficiency is not fully addressed as a GPS system is used in each sensor node. Moreover, the sensor nodes simply transfer the data to the base station as the true value.

Tree-based methods arrange the entire sensor nodes into a tree [10]. Xin and Fei-Qi [11] discuss a tree-based protocol where the parent node receives data packets from leaf nodes to aggregate them with the data that is coming from the surrounding node. Then, the parent nodes pass on the new series of data packets to their parent nodes until it reaches the BS. GBDAS [12] is a grid-based data aggregation scheme that partitions the sensor field into a 2-D logical grid of cells. In each cell, the node with the most residual energy (cell head) takes the responsibility for aggregating its own data collected by the other sensor nodes of the cell. The cell heads form a chain and aggregated data moves from head to head along the chain until it reaches the BS. Since all sensor nodes need to aggregate the collected data, the delay especially in the end nodes cannot be ignored. Furthermore, it cannot be a desirable solution for large networks or even for the very far distance sinks.

The dynamic errors or uncertainties related to insufficient or noisy data in many real world applications, can negatively affect the performance of the implemented signal processing systems. There are many techniques that includes statistical and Covariance Intersection (CI)-based methods are used in enhancing the QoS in WSNs. However, they are cannot cope with the uncertainty of the data

produced by WSNs. Moreover, the inflexibility of the methods prevents processing the data realistically [13]. In fact, using these methods requires a very complex and substantial computational effort to have optimal performance [14]. A Kalman Filter-based approach to correct the values measured by the sensors is discussed in [15]. Some of these methods are explicitly model-based, whereas others require tuning and training. In the general case, where *a priori* information is often not available, these approaches are typically deficient and can often lead to undesirable results.

Fuzzy logic-based approaches that utilise a fuzzy logic controller embedded in the sensor nodes have been used for cluster-head election [16], for reducing the deadline miss ratio associated with the real-time data transmission from the source node to the actuator maintained at a pre-determined desired level [17] and online routing algorithms to address the maximum lifetime routing problem in wireless sensor networks [18]. The distinguishing aspect of our work is the novel use of fuzzy membership functions and rules in the design of cost functions for the routing objectives considered in this work. In [15], a variable weight-based fuzzy data fusion algorithm is proposed. The main purpose is to enhance the accuracy of the data fusion process in WSNs. In this approach, each CH is assigned a different and unfixed fusion weight. The weights are changed using a fuzzy logic system, based on some factors such as delay, data amount and trustworthiness of the CHs. Lower fusion weighted CHs have lower influence. However, the authors did not take the message overhead and energy consumption into account. Shell *et al.* [13] presented a fuzzy logic system that is able to isolate instances of failure within the defined data set with fewer perceived instances of false positives and a higher degree of accuracy than classical methods alone. However, the authors did not consider the energy consumption properly.

In this paper, we use an interval type-II fuzzy logic system (T2FLS) as opposed to a type-1 fuzzy logic system (T1FLSs). The fuzzy membership functions in T2FLSs use membership degrees that are themselves fuzzy sets, whereas T1FLSs use fixed fuzzy memberships that cannot directly address variable conditions. Therefore, uncertain measured parameters would be neglected by a T1FLS and the performance will obviously be negatively influenced. This makes type-2 fuzzy sets very useful when there is difficulty in deciding the appropriate membership function with ambiguity. A recent study that compared the effects of the measurement noise in type-1 and type-2 FLSs concluded that the use of T2FLSs in real world applications that exhibit measurement noise and modelling uncertainties can be a better option than T1FLSs [17,19]. Furthermore, T2FLS technology has been regarded as a way to increase the fuzziness of a relation which implies an increased ability to handle inexact information in a logically correct manner [18].

Although there are works such as Z-slices [20], α -planes [21,22] or α -cuts [23] that use T2FLS or generalized T2FLS, interval T2FLS is preferred for use due to the computational complexity of the former [24]. The computations associated with IT2 fuzzy sets are very manageable, which makes interval T2FLS quite practical [25]. The basic concept of interval T2FLS is to consider a footprint of uncertainty, which can be described by two parameters bounding T1 fuzzy membership functions [18]. The interval T2FLS \bar{A} can be calculated as follows:

$$\bar{A} = \int_{\mathcal{J}} \mu(x, u) \, \mathcal{J} \in \mathcal{R}, 1 \leq \mathcal{J} \leq n \quad (1)$$

where x and u are the primary and the secondary variable and J_{μ} is the primary membership function of x . In case of IT2 fuzzy sets, all secondary grade of fuzzy set \bar{A} are equal to 1. The domain of the primary membership J_{μ} defines a footprint of uncertainty of fuzzy set \bar{A} , which can be described by its upper and lower membership functions. Hence, a footprint of uncertainty (\bar{A}) = $U_{\mu \in \mathcal{C}}(\underline{\mu}_{\mu}(x), \overline{\mu}_{\mu}(x))$.

3. System Model and Problem Overview

In this section, we present a brief description of our WSN system model and also present the problem overview. We consider a WSN with a large number of sensors distributed randomly in a certain deployment area. The sensors are organized into clusters based on their spatial proximity and each cluster has a CH [26]. There are three types of sensor nodes that collect temperature, humidity and smoke density data and then send it to its CH. The CHs are responsible for fusing and transferring the data to the BS. Then, each node, based on their current situation, uses FLC to assign a weight for their data. Next, the data are sent to the CH, which is connected to a BS wirelessly.

Uncertainty inherent in sensor measurements is one of the most fundamental issues to be addressed in a WSN environment. As sensor nodes are typically deployed densely, there exists significant redundancy in the data collected from sensor nodes. The duplicate data may lead to serious packet collisions, bandwidth waste and energy consumption. Similarly, sensor nodes can report erroneous readings for a number of reasons such as a possible manufacturing defects or environmental conditions. As a result, they can negatively affect making accurate decisions as well as the efficiency of energy consumption in the WSN. To solve the problem, it is necessary to address the unnecessary data in addition to redundant data transferred from sensor nodes.

These uncertainties necessitate the development of sensor data fusion strategies that can combine information in a coherent and synergistic manner to yield a robust, accurate, and consistent description of the quantities of interest in the environment. A general block diagram of a data fusion mechanism is given in Figure 1. The sensor nodes S_1, S_2 and S_3 collect data d_1, d_2 and d_3 from the environment. In many situations that depend on the current conditions of the sensor nodes, d_1, d_2 and d_3 might not be exactly the true value. There are many useless data packets generated and transferred from each sensor node in each round of data collection. In many situations, a sensor node is not able to recognize the useless data while it generates data packets for further processing.

Rather than each sensor node sending the data to BS, they send the data to a fusion node. The fusion node creates a single internal representation of the environment from its inputs. The single representation is then forwarded to the BS. This means that the BS in general does not have access to the individual sensor measurements.

The main cause that negatively influences performance of a data aggregation mechanism is misbehaving sensor nodes. Different environmental factors such as a sudden change of temperature or humidity can influence a sensor node's behavior. These environmental factors decrease or increase the output signal of the sensors, which creates an ultralow frequency noise in the transferred signals. In addition, non-operating environmental limits such as a high or low temperature of air surrounding the sensor nodes usually influence the sensor's performance. The operating temperature range is the length of ambient temperatures given by their upper and lower extremes, within which the sensor nodes maintain their expected accuracy. In different WSN applications, it is impossible to confirm that the

collected data are true values of the events without taking samples or analyzing the data history. Sensors usually carry noise in their measurements, thus we cannot be 100% sure that the measured value is correct. It does not matter how an event (e.g., temperature) is measured by a sensor or how close the measurement is to the true value, we can never be sure that it is accurate. Equation (2) captures the error induced by a sensor node [27]:

$$\phi = t - t^* \tag{2}$$

where t is any individual measurement and t^* is the true value.

Let ρ be the percentage of the correct data received by BS. Equation (3) shows the percentage of the accurate data packets received by a BS:

$$\rho = \frac{L_c}{L} \times 100 \tag{3}$$

where L is the total number of events defined by the application and L_c is the number of messages that contain only the true value of the received data.

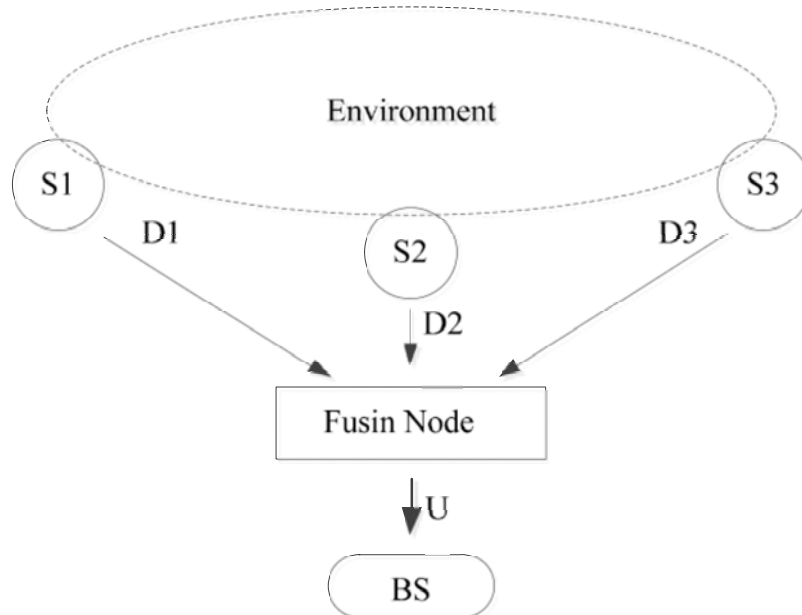


Figure 1. Fusion mechanism.

Let Z_q be the network lifetime, α be data redundancy that is satisfied by a user-defined value ($\alpha \in [0, 1]$) similar to the percentage of collected true value of data ($\rho \in [0, 1]$). The problem addressed in this paper is formulated as follows:

$$\text{Maximise } Z_q \tag{4}$$

$$\text{s.t. } \rho \geq \rho_{min} \text{ and } \alpha \geq \alpha_{min} \tag{5}$$

The objective is to maximize the network lifetime Equation (4) ensuring that the percentage of the true value of data and data redundancy are satisfied by a user-defined value (Equation (5)). Since it is impossible to confirm that the collected data are true values of the events without taking samples or analyzing data history [26], we suggest assigning a weight for each collected data. The weights are determined based on current storage conditions of the source nodes. If the sensor nodes are not in the

expected condition as revealed in their data sheets, the determined weights are changed. The better condition a sensor node is the higher weight is assigned to its collected data. In fact, the weights show the percentage of correctness of the data. Based on the calculated weights, the correct value of data can be distinguished and separated from others. As a result, a higher QoS can be received by the BS.

4. Proposed Data Fusion Approach

Algorithm 1 explains the proposed data fusion process. This algorithm is collecting data from the three different sensors until it detects the events. The inputs of the algorithm are node temperature, humidity ratio and signal to noise ratio. The output of is the fused collected sensory data.

Algorithm 1 Data Fusion Algorithm

```

1. INPUT: (T: Node temperature, H: Humidity ratio, N: Signal to noise ratio)
2. OUTPUT: Fused Data
3. BEGIN
4. WHILE (Event NOT Detected )
5. FOR all the cluster members in one kind
6.      $\mu_{FC}$  ← FLC (T, H, N)
7.     IF  $\mu_{FC} \geq \alpha$ 
8.         Data and  $\mu_{FC}$  will be sent CH
9.     ELSE
10.        Collected data will not be considered
11.    END IF
12. END FOR
13.  $\mu_{FC}$  ← Received data by CHs from one kind of node will be fused
14. Consequent ← FLC ( $\mu_{FC_1}, \mu_{FC_2}, \dots, \mu_{FC_n}$ )
15. IF the consequent was not changed
16.    Disregards the received data
17. ELSE IF
18.    Send the Consequent to BS
19.    IF the event detected by BS
20.    Report the event
21.    END IF
22. END IF
23. END WHILE

```

END Algorithm

Figure 2 shows the proposed data fusion and transferring process. As in previous works [15,27–29], the sensors will be embedded with a fuzzy logic controller (FLC). The purpose of FLC is to find a confidence level of the collected data considering the current condition of the sensor nodes. Each sensor node collects data and calculate a confidence factor (CF) using FLC for each collected data packet. To calculate CF, the FLC is developed based on the Sugeno method [30].

The FLC considers the non-operating temperature (T) and humidity (H) range to create the membership functions. A random membership function is also used for noise to signal ratio (N) of the sensors. FLC produces a confidence factor ($\mu_{FC_1}, \mu_{FC_2}, \dots, \mu_{FC_n}$) for each sensor data ($\mu_{FC_1}, \mu_{FC_2}, \dots, \mu_{FC_n}$). There are three inputs for the FLC: Temperature, Humidity Rate and Signal to Noise Ratio. For each

input synthetic data is used. For each data we use a Gaussian distribution with its mean and covariance matrix representing the expected value and its uncertainty (10% of the value). Then, the values are normalized to fit in the [0,1] as the inputs of the fuzzy system. Then, we extracted linguistic variables from the normalized data. The linguistic variables used to represent them are divided into three levels: Low, Medium and High. The consequent or the output of the FLC is divided into five levels: Very Low (VLow), Low, Medium, High and Very High (VHigh). 20% of the data is used to determine the membership functions and also the rules.

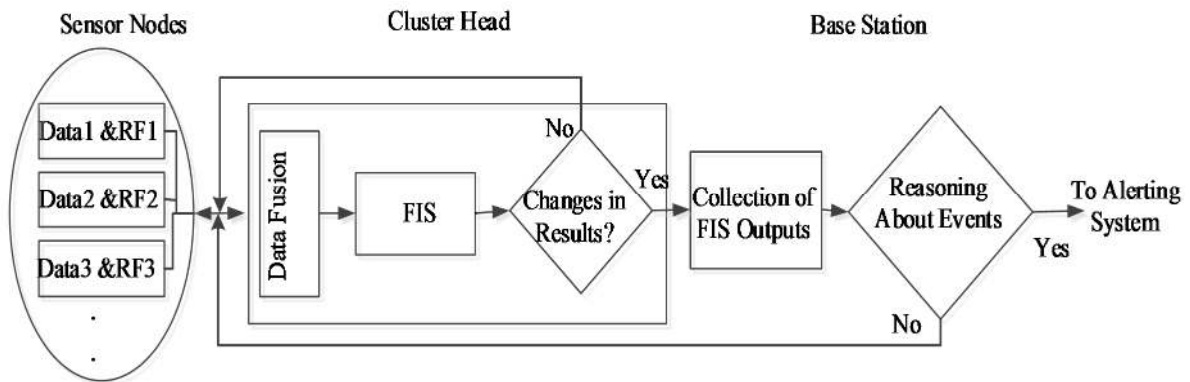


Figure 2. Proposed flow chart.

Based on the fuzzy variables shown in Figure 3, fuzzy rules are defined as shown in Table 1. Since each input variable has three fuzzy states, e.g., Low, Medium and High, thus, the total number of possible fuzzy inference rules for the developed system, is $3 \times 3 \times 3 = 27$.

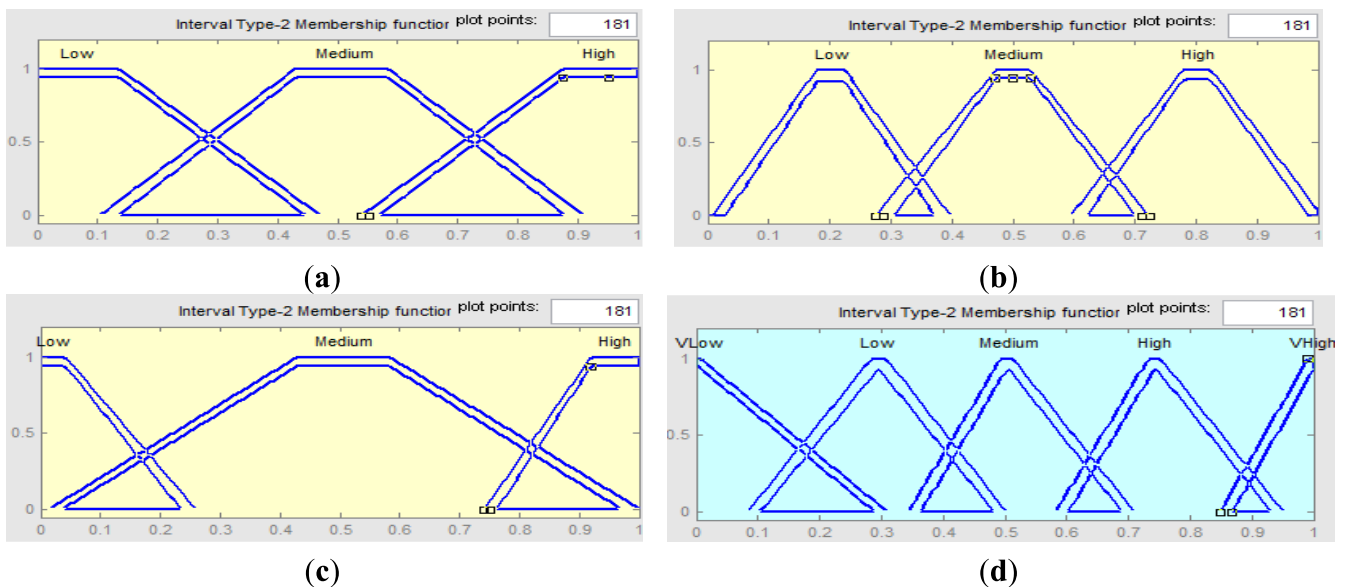


Figure 3. Membership functions of FLC. (a) Temperature; (b) Humidity Rate; (c) Signal to Noise Ratio; (d) Confidence Factor.

Table 1. Inference rules.

Rule NO	Input Variable			Output
	Temperature	Humidity Rate	Signal to Noise	Confidence Factor
1	High	High	High	VHigh
2	High	High	Medium	VHigh
3	High	High	Low	High
4	High	Medium	High	High
5	High	Medium	Medium	High
6	High	Medium	Low	Medium
7	High	Low	High	Medium
8	High	Low	Medium	Medium
9	High	Low	Low	Low
10	Med	High	High	Medium
11	Med	High	Medium	Low
12	Med	High	Low	Low
13	Med	Medium	High	Medium
14	Med	Medium	Medium	Medium
15	Med	Medium	Low	Low
16	Med	Low	High	Medium
17	Med	Low	Medium	Low
18	Med	Low	Low	VLow
19	Low	High	High	High
20	Low	High	Medium	Medium
21	Low	High	Low	Low
22	Low	Medium	High	Medium
23	Low	Medium	Medium	Low
24	Low	Medium	Low	VLow
25	Low	Low	High	Low
26	Low	Low	Medium	VLow
27	Low	Low	Low	VLow

The FLC determines whether the temperature and humidity rates of the sensor nodes and also the signal to noise ratio are in the acceptable range. To accomplish that, the FLC compares its input measurements with the desired range for each sensor. The desired range for each sensor can be found on its specified datasheet. The output of FLC for each sensor can be 100% only if the environmental factors are in the desirable range.

In case of being out of the range, FLC produces a confidence factor ($0\% \leq \mu_{CF} \leq 100\%$) for the collected data. Each node compares the calculated *CFs* against a cut-off value. In the proposed system, the value is provided by users to decide if the fuzzy output should be considered. If the created factor of each data is less than the considered value, the data will be disregarded. Otherwise, it will be sent to CHs in a data message. This prevents the corruption of the correct value of data by the other data that does not present the true values. For fusion purpose, the message also consists of a Node-ID of each source node.

CHs aggregate the data received from the cluster members. The data fusion process is started by CHs at the end of each round of data collection. Equation (6) is used to aggregate the entire data received from cluster members in the same kind with different locations:

$$FD = \frac{(CF_1 \times D_1) + (CF_2 \times D_2) + (CF_3 \times D_3) + \dots + (CF_m \times D_m)}{CF_1 + CF_2 + CF_3 + \dots + CF_m} \quad (6)$$

where D_i is the received data from the same kind of cluster members (e.g., light detector), CF_i is the calculated confidence factor of the collected data by each sensor node and FD is a combination of the data received from them. In fact, FD is a combination of the data with a higher certainty as it combines the received data based on their confidence factors. As a result of considering different sources of one kind of sensor node in different locations, FD provides a better view of the environment. The FD is also robust as data from multiple sensors with their own confidence factors mitigate the problem of sensor failure. That is because in case of any disorder of a sensor in a cluster, the environment still can be monitored as the entire received data is combined and then examined by the system. The FD is calculated for any kind of deployed sensor nodes individually. Therefore, that would be a set of FD s. Next, FD s are stored in a matrix with one row and m columns ($M \times 1$). Equation (7) presents $M \times 1$ matrix:

$$M \times 1 = \{FD_1, FD_2, \dots, FD_m\} \quad (7)$$

As an example, we consider three temperatures, four light and three smoke density sensors. They are deployed on a cluster-based method. Over a period of time, the temperature sensors sense the environment to be at 20 °C, 15 °C and 10 °C with the confidence factors of 0.75, 0.65 and 0.41 respectively. As a result, the fused temperature data (FD_{temp}) would be 15.93. The fused light detector (FD_{light}) and fused smoke density data (FD_{smoke}) are 49.8 and 33.2, respectively. Based on this $M \times 1$ matrix of the monitoring system is $M \times 1 = \{15.93, 49.8, 32.2\}$

Then, the vector $M \times 1$ will be fed and processed by the FLC with different inputs in the CHs. The output of the system is the consequent (fuzzy response) of events occurring in the monitored areas. FLC analyses the $M \times 1$ vector based on a provided fuzzy rules after data fusion process was accomplished. If the consequent of the events in the clustered nodes was not changed, the CHs do not forward the data packet to the BS. Otherwise, the CHs send the output of the system to the BS. However, the change in the system output does not guarantee a correct detection and it is only considered as a possible event in the monitored area. Therefore, to make sure that the detection is accurate enough, the BS needs to regularly monitor and process the received outputs that are generated over the time. In fact, all the received probabilities are constantly processed by the BS. If a constant change is noticed, the event will be reported to the alerting subsystem by BS. Otherwise, the algorithm continues collecting data.

5. Performance Analysis

In this section, we present the performance analysis of the proposed approach and compare it with two other approaches. In the experiment, we focus on the expected amount of collecting correct data with respect to QoS.

5.1. Experiment Setup

We model the MTS420/400 sensor board and an IRIS with ATmega1281 processor and a mib520 programming base as BS. The sensor board consists of humidity, temperature and light sensor as well as a communication component. Table 2 presents the information about MTS420/400. In order to simulate noise effects in real sensors, a random high-frequency noise signal is added to the sensor signal. MATLAB was used to analyse the performance of the network.

Table 2. Parameters of MTS420/400.

Parameters	MTS420/400
Temperature Range	−40 to +123.8 °C
Humidity Range	0 to 100% RH
Signal to Noise Ratio	0 to 1

We calculate the energy consumption in each round of data collection using the method proposed in [31]. We use Equation (8) to evaluate the traffic flow of each protocol:

$$E_{total} = \frac{E_{trans} + E_{rx} + E_{tx} + E_{idle}}{T} \tag{8}$$

The proposed approach is compared against forest fire detection (FIM) [32] and a variable weight-based fuzzy data fusion algorithm for WSNs (VWFFA) [15] algorithms. The main reason for choosing the FIM is because it is a distributed algorithm that uses MTS420/400C sensor boards for the proposed algorithm. The algorithm is based on a state machine with five states. The transition from one state to another is generated when a relevant change is detected, indicating the probable existence of an event. If the ratio is less than a defined value, the machine moves to another state. The authors proved that algorithm based on a defined cut-off value method can be implemented more easily and with better performance than an algorithm based on Dempster-Shafer theory [32]. VWFFA is also chosen because it is a distributed fuzzy-based data fusion algorithm that enhances the QoS in WSNs. It is a weight-based fuzzy data fusion algorithm for WSNs that improves the accuracy and reliability of the global data fusion. In this algorithm, the weight of each CH in the global fusion is not fixed. The time delay, amount of data and trustworthiness of each CH affect the final fusion weight. That means a CH with too small a data value or too low trustworthiness cannot be given a big fusion weight. As a result, CHs with deficiencies have a small influence in global fusion. For validation purposes, we used the root means square error (RMSE). Table 3 compares RMSE of the proposed approach via FIM and VWFFA. To calculate RMSE, the most optimum result (100%) is considered as the predicted result in each call.

Table 3. RMSE.

Approaches	RMSE
Proposed approach	3.67
FIM	5.13
VWFFA	5.9

5.2. Result and Discussion

Figure 4 presents a percentage of correct data collected by BS by the proposed approach as well as FIM and VWFFA. As can be seen from the figure, the proposed approach provides a higher percentage of correct data compared with the other approaches.

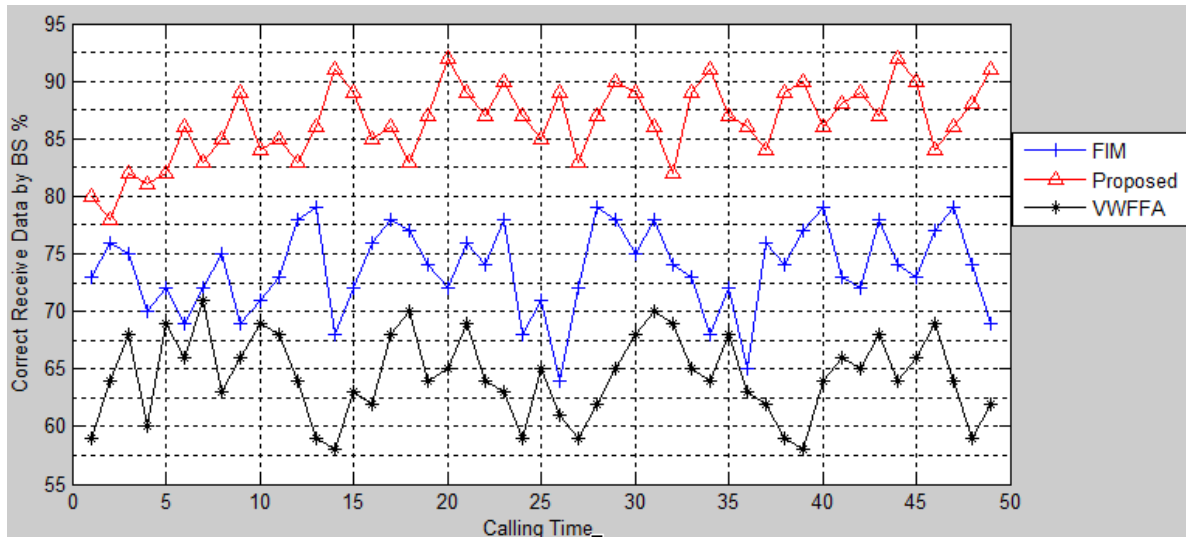


Figure 4. Data collection.

That is because the proposed approach eliminates the incorrect data to prevent corrupting the true value data in the fusion process. The result also shows that FIM has better performance than VWFFA. This is due to the consideration of uncertainty in the detection process. VWFFA provides the worse performance. In VWFFA the quality of collected data is considered based on their assigned weights. However, the weights are considered on all the data that includes true values and incorrect values of the data collected. Since the entire received data are aggregated by CHs, the true values might be influenced by the incorrect data in the fusion process.

Figure 5 shows the average number of the data packets that are transferred. The proposed approach has a lower average number of the data packets that are transferred than that the FIM and VWFFA. The proposed approach has better performance than the other approaches because the entire collected data do not need to be transferred as only the true valued data are transferred. Moreover, CHs are not required to transfer the fused data as they are able to find the consequent result of the occurring events. The CHs send the calculated consequent only if it shows an abnormality based on the history of the data, so the data that is sent to the BS is going to be only the possibility of the detected fire. In contrast to the proposed approach, FIM has the highest number of data packets that need to be transferred in the network. The reason behind that is the entire incorrect and correct collected data and also redundant data need to be transferred. VWFFA has a lower transferred data number in the network than FIM. The reason behind that is the developed network is cluster-based and the data are fused based on the assigned weights on clusters. In VWFFA, the assigned weights reduces the influence of the incorrect data on the true values in the fusion process.

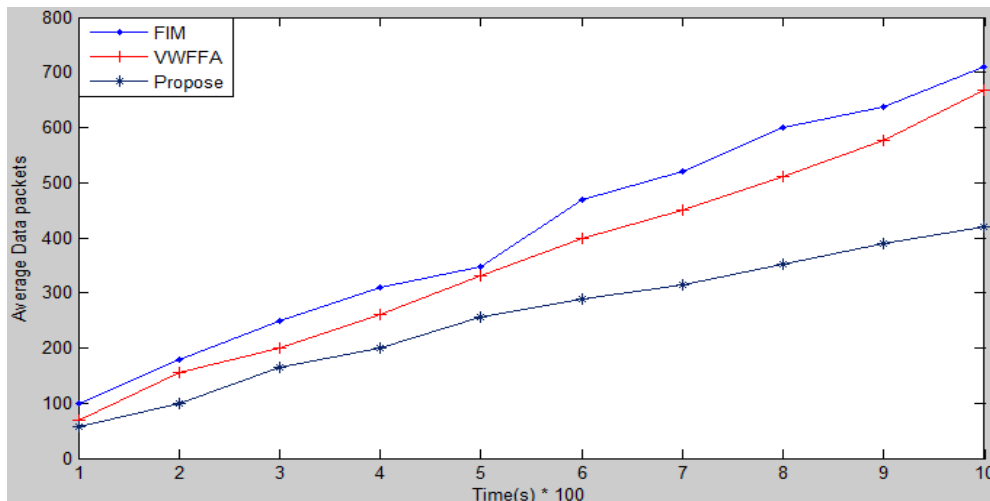


Figure 5. Transferred data packets.

Energy consumption is also a critical feature of a WSN. The study by Hill *et al.* [32] proves that each bit transmitted in a WSN consumes about as much energy as executing 800–1000 instructions. That means any reduction of the data transmitted in the WSN can significantly decrease the energy consumption. Figure 6 illustrates the energy consumption in the networks. From the figure, it can be seen that the proposed scheme minimizes the energy consumption in WSN. That is due to the lower transfer data packets in the network (Figure 5). Moreover, data packets similar to VWFFA are sent to their CHs consequently, lower energy is required as compared to FIM.

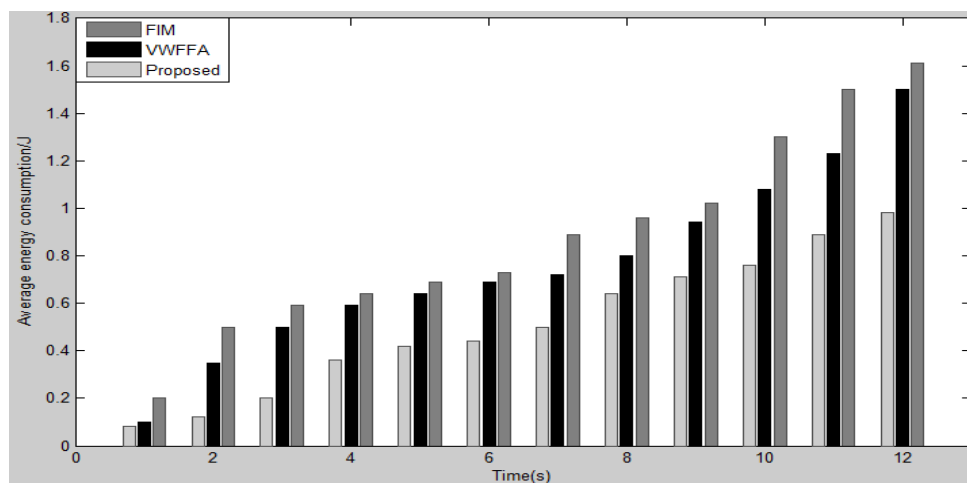


Figure 6. Energy consumption.

6. Conclusions

Handling sensing data errors and uncertainties in WSN while maximizing network lifetime are important issues in the design of applications and protocols for wireless sensor networks. In this paper, we have presented a fuzzy-based method for data fusion. Through experiments, we show that the proposed approach is able to distinguish and aggregate only true values of the collected data and is able to eliminate redundant data and consequently reduce energy consumption thus increasing the network lifetime. Performance analysis of the proposed approach shows that the algorithm is effective

in the quality of the data fusion. The proposed algorithm is compared against two baseline approaches experimentally in terms of data collection, number of transferred data packets and energy consumption. The results of the experiment show that the proposed approach achieves better results than the baseline approaches. In the future we will study the channels and related issues that might occur in transferring data from the CHs to the BS. We will also consider over the air programming to dynamically update the deployed nodes in the area of interest.

Acknowledgments

The author would like to thank Oscar Castillo, for providing them with Type-2 fuzzy logic toolbox. This work is supported by University of Malaya High Impact Research Grant no vote UM.C/625/HIR/MOHE/SC/13/2 from Ministry of Higher Education Malaysia.

Author Contributions

D.I. has contributed in writing the paper and editing the paper organization; J.H.A. has contributed in problem and model clarification, editing the paper, addressing reviewer comments, correcting typos and grammar; S.G. has contributed in adding theoretical background of IT2FLS; T.H. has contributed in writing the paper and editing the paper organization.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Abdelgawad, A.; Bayoumi, M. Data Fusion in WSN. In *Resource-Aware Data Fusion Algorithms for Wireless Sensor Networks*; Lecture Notes in Electrical Engineering Series 118; Springer: Berlin/Heidelberg, Germany, 2012; pp. 17–35.
2. Li, L.; Li, W. The analysis of data fusion energy consumption in WSN. In Proceedings of 2011 International Conference on System Science, Engineering Design and Manufacturing Informatization (ICSEM), Guiyang, China, 22–23 October 2011; pp. 310–313.
3. Dhasian, H.R.; Balasubramanian, P. Survey of data aggregation techniques using soft computing in wireless sensor networks. *Inf. Secur.* **2013**, *7*, 336–342.
4. Durrant-Whyte, H.F.; Henderson, T.C. Multisensor data fusion. In *Springer Handbook of Robotics*; Siciliano, B., Khatib, O., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 585–610.
5. Collotta, M.; Pau, G.; Scata, G. A fuzzy system to reduce power consumption in Wireless Sensor Networks: A comparison between WirelessHART and IEEE 802.15.4. In Proceedings of 2014 IEEE International Energy Conference (ENERGYCON), Cavtat, Croatia, 13–16 May 2014; pp. 766–771.
6. Collotta, M.; Messineo, A.; Nicolosi, G.; Pau, G. A Dynamic Fuzzy Controller to Meet Thermal Comfort by Using Neural Network Forecasted Parameters as the Input. *Energies* **2014**, *7*, 4727–4756.

7. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976.
8. Jung, W.-S.; Lim, K.-W.; Ko, Y.-B.; Park, S.-J. Efficient clustering-based data aggregation techniques for wireless sensor networks. *Wirel. Netw.* **2011**, *17*, 1387–1400.
9. Koutsonikolas, D.; Das, S.M.; Hu, Y.C.; Stojmenovic, I. Hierarchical geographic multicast routing for wireless sensor networks. *Wirel. Netw.* **2010**, *16*, 449–466.
10. Pandey, V.; Kaur, A.; Chand, N. A review on data aggregation techniques in wireless sensor network. *J. Electron. Electr. Eng.* **2010**, *1*, 01–08.
11. Guo, X.; Deng, F. Complete ternary tree-based data aggregation routing algorithm for wireless sensor networks. In Proceedings of International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC), Shengyang, China, 20–22 December 2013; pp. 578–581.
12. Wang, N.-C.; Chiang, Y.-K.; Hsieh, C.-H.; Chen, Y.-L. Grid-Based Data Aggregation for Wireless Sensor Networks. *J. Adv. Comput. Netw.* **2013**, *1*, doi:10.7763/JACN.2013.V1.66.
13. Shell, J.; Coupland, S.; Goodyer, E. Fuzzy data fusion for fault detection in wireless sensor networks. In Proceedings of 2010 UK Workshop on Computational Intelligence, Colchester, England, 8–10 September 2010; pp. 1–6.
14. Tafti, A.D.; Sadati, N. Novel adaptive Kalman filtering and fuzzy track fusion approach for real time applications. In Proceedings of 3rd IEEE Conference on Industrial Electronics and Applications, Singapore, 3–5 June 2008; pp. 120–125.
15. Wang, Q.; Liao, H.; Wang, K.; Sang, Y. A variable weight based fuzzy data fusion algorithm for WSN. In Proceedings of Ubiquitous Intelligence and Computing—8th International Conference, Banff, AB, Canada, 2–4 September 2011; pp. 490–502.
16. Díaz-Ramírez, A.; Tafoya, L.A.; Atempa, J.A.; Mejía-Alvarez, P. Wireless sensor networks and fusion information methods for forest fire detection. *Proced. Technol.* **2012**, *3*, 69–79.
17. Khanesar, M.A.; Teshnehlab, M.; Kayacan, E.; Kaynak, O. A novel type-2 fuzzy membership function: Application to the prediction of noisy data. In Proceedings of 2010 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, Taranto, Italy, 6–8 September 2010; pp. 128–133.
18. Kay, R.U. Fundamentals of the dempster-shafer theory and its applications to system safety and reliability modelling. *Reliab. Theory Appl.* **2007**, *3*, 173–185.
19. Izadi, D.; Abawajy, J.; Ghanavati, S. Fuzzy logic optimized wireless sensor network routing protocol. *J. High Speed Netw.* **2013**, *19*, 115–128.
20. Wagner, C.; Hagrass, H. Toward general type-2 fuzzy logic systems based on zSlices. *IEEE Trans. Fuzzy Syst.* **2010**, *18*, 637–660.
21. Mendel, J.M.; Liu, F.; Zhai, D. α -Plane Representation for Type-2 Fuzzy Sets: Theory and Applications. *IEEE Trans. Fuzzy Syst.* **2009**, *17*, 1189–1207.
22. Manic, M.; Frincke, D. Towards the fault tolerant software: Fuzzy extension of crisp equivalence voters. In Proceedings of The 27th Annual Conference of the IEEE Industrial Electronics Society, Denver, CO, USA, 29 November–2 December 2001; pp. 84–89.
23. Hamrawi, H.; Coupland, S.; John, R. A novel alpha-cut representation for type-2 fuzzy sets. In Proceedings of 2010 IEEE International Conference on Fuzzy Systems (FUZZ), Barcelona, Spain, 18–23 July 2010; pp. 1–8.

24. Liang, Q.; Mendel, J.M. Interval type-2 fuzzy logic systems: theory and design. *IEEE Trans. Fuzzy Syst.* **2000**, *8*, 535–550.
25. Mendel, J.M. On the importance of interval sets in type-2 fuzzy logic systems. In Proceedings of IFSA World Congress and 20th NAFIPS International Conference, Vancouver, BC, Canada, 25–28 July 2001; pp. 1647–1652.
26. Heinzelman, W.R.; Chandrakasan, A.; Balakrishnan, H. Energy-efficient communication protocol for wireless microsensor networks. In Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2000; doi:10.1109/HICSS.2000.926982.
27. Wang, P.; Su, G.D.; Li, C.B.; Chen, B.J.; Chen, B. Cluster-Head Election Mechanism for Wireless Sensor Networks. *Sens. Transducers* **2013**, *157*, 129–135.
28. Minhas, M.R.; Gopalakrishnan, S.; Leung, V.C. Fuzzy algorithms for maximum lifetime routing in wireless sensor networks. In Proceedings of Global Telecommunications Conference, 2008, New Orleans, LO, USA, 30 November–4 December 2008; pp. 1–6.
29. Hongge, S.; Weisheng, W.; Yi, C.; Shenxue, H.; Xin, Y. Application of fuzzy data fusion in multi-sensor environment monitor. In Proceedings of International Conference on Computational Intelligence and Security, Beijing, China, 11–14 December 2009; pp. 550–553.
30. Castillo, O.; Melin, P. 3 Type-2 Fuzzy Logic. In *Type-2 Fuzzy Logic: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 29–43.
31. Fraden, J. *Handbook of Modern Sensors: Physics, Designs, and Applications*, 4th ed.; Springer: Berlin/Heidelberg, Germany, 2004.
32. Biglarbegian, M.; Melek, W.; Mendel, J. On the robustness of Type-1 and Interval Type-2 fuzzy logic systems in modeling. *Inf. Sci.* **2011**, *181*, 1325–1347.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

