Data Aggregation with Sequential Detection in Wireless Sensors Network

Song Xiao-Ou
Engineer University of CAPF, Xi’an, 710086, China
Tel.: 0086+18691566981, fax: 0086+029-84563833
E-mail: e_miracle@163.com

Received: 10 July 2014 /Accepted: 30 September 2014 /Published: 31 October 2014

Abstract: As the advanced information acquiring technology, wireless sensor networks are being applied in more and more areas. Data aggregation is the effective information processing to solve the energy saving problem of WSN (Wireless Sensor Network). Most existing data aggregation algorithm discuss fusion problem in the scene that the number of nodes is fixed. However, the amount of data transmitted is directly proportional to the number of nodes. Hence the appropriate strategy should be that the number of nodes is alterable under the satisfaction of performance. In this paper, sequential detection is applied in data aggregation for wireless sensor network. And the mathematical model is established. Considering the fading problem of wireless channel and the characters of physical layer, WSN data aggregation with sequential detection (WDASD) is proposed. The characters of WDASD are tested by simulations. In the end, a cross-layer scheme is proposed and simulation test is made to validate it. Copyright © 2014 IFSA Publishing, S. L.

Keywords: Wireless Sensor Network, Data Aggregation, Sequential Detection, Cross-layer design, Energy consumption.

1. Introduction

In order to increase the sensing performance and the robustness, more redundant nodes are arranged in Wireless Sensors Network (WSN), which cause that the network load increases. Hence it is necessary to adopt data aggregation algorithm, to guarantee the energy consumption effective [1]. The data aggregation in Wireless Sensors Network is quite different from that in Wired Sensors Network, because wireless channel fading may cause serious affect to the transmitting data. The representative research results of data aggregation in WSN are illustrated as follow. The optimal decision fusion rule is researched in document [2], to propose a suboptimal algorithm with a compromise among performance, energy consumption and complexity. Utilizing the conditional probability density function, suboptimal data aggregation algorithms are proposed under different SNR conditions in document [3], which reduce the channel estimation complexity. Under the condition that transmitting bandwidth is limited, distributed aggregation in the cluster network is analyzed in document [4], which averages the received data at the leader node to make a decision. Supposing the wireless channel as Binary Symmetric Channel, distributed aggregation algorithm based on Weighted average is proposed in document [5]. Under the hypothesis that the number of nodes participating in the aggregation algorithm is constant, the above aggregation algorithms can gain some optimal performance indexes, in specified application scenarios.

When the number of nodes participating in the aggregation algorithm achieves a certain number, the
data aggregation system can meet the expected performance. In order to save energy consumption in WSN, the remaining nodes, not participating in the aggregation algorithm, should enter the sleep state and stop transmitting, which introduces sequential detection method of signal detection into the data aggregation algorithm. In this paper, data aggregation model based on sequential detection in WSN is constructed in the fading channel, which makes the number of nodes participating in the aggregation algorithm unfixed. Considering the modulation and demodulation characteristics of physical layer, WSN Data Aggregation base on Sequential Detection (WDASD) is further proposed. And its performance is proved by simulations. Then combining the thought of cross-layer scheduling, WDASD general scheme is finally proposed and tested by simulations.

2. Data Aggregation System Model

According to network topology structure, data fusion can be distinguished into cluster based data fusion, tree based data fusion, and cluster and tree based data fusion [6]. Whichever network topology structure are adopted, the elementary unite can come down to star network. It is supposed that there are one sink node and \( N^* \) sensor nodes in the elementary unite. Every sensor node doesn't communicate with other sensor nodes. At the sink node, the sensor data transmitted from different sensor nodes are collected and aggregated. Supposing that the node communication model is half-duplex communication, then communication between the sensor node and the sink node is established in turn. In order to specify the working mode of the elementary unite, it is supposed that TDMA is adopted as the network access mode, which means that the sink node allocates a time slot to every sensor node for sensor data transmission. The data transmitted from sensor node to sink node in turn can be described as \( \{x_1, x_2, \ldots, x_N\} \).

Decision fusion [1] is adopted in this paper, which means that the binary decision results, "0" or "1", are transmitted from sensor nodes and aggregated at the sink node to get the global decision hypothesis \( H_0 \) or hypothesis \( H_1 \), described as follow:

1) \( H_0 \): There is no specified event happening.
2) \( H_1 \): There is the specified event happening.

The performance of data aggregation algorithm is evaluated by false alarm probability \( P(H_1 \mid H_0) \) and detection probability \( P(H_1 \mid H_1) \). Considering that WSN is energy limited network [7], energy consumption of the data aggregation system is also an important performance index under the satisfied false alarm probability and detection probability. It is illustrated in document [8] that energy consumption of the data aggregation system is proportional to the transmitting amount of data, when the transmitting power of node is fixed. Under the condition that the data packets transmitted by sensor node possess the same length, the transmitting amount of data is proportional to the number of sensor nodes participating in the aggregation algorithm, denoted by \( N \). Hence \( N \) is described as the performance index of energy consumption in this paper.

3. Aggregation Algorithm Based on Sequential Detection

Sequential detection belongs to statistic detection, which is usually used in signal detection. It is proved by Wold and Wolfwitz that sequential detection needs the least average number of observations to get a decision, with the fixed false alarm probability and missing detection probability [9]. It is supposed that all the sensor nodes work under ideal conditions without any noise interference. When the decision results transmitted from multiple sensor nodes through wireless channel, converge into a data sequence \( \{x_1, x_2, \ldots, x_N\} \) at the sink node, the noise introduced to the data aggregation system is just the noise of wireless channel. Then the decision process can be modeled as sequential detection, considering the number of data packets received by the sink node as the observation number. Hence the decision process of data aggregation based on sequential detection needs the least average number of participating nodes with the least energy consumption. The decision process of data aggregation based on sequential detection is specifically illustrated as follow.

Supposing that the data transmission between the \( N^* \)th sensor node and the sink node is completed, we can get an observation sequence with \( N \) decision results, described by \( (x_1, x_2, \ldots, x_N)^T \). The \( N \) dimensional random vector is then mapped to one point in the observation space \( \Phi \). As shown in Fig. 1, the observation space \( \Phi \) is divided into three decision domains [10], which satisfy the follow requirements:

\[
\Phi = \bigcup_{j=0}^{2} \Phi_j, \quad \Phi_i \cap \Phi_j = \emptyset, \quad i, j = 0, 1, 2, \quad (1)
\]

where \( \emptyset \) denotes empty set.

Fig. 1. Decision domain of sequential detection.
If the observation vector \( \mathbf{x}_n = (x_1, x_2, \ldots, x_N)^T \) is mapped into decision domain \( \mathbf{D}_1 \), we decide the hypothesis \( H_0 \). If the observation vector is mapped into decision domain \( \mathbf{D}_2 \), we decide the hypothesis \( H_1 \). And if the observation vector is mapped into decision domain \( \mathbf{D}_3 \), it indicates that the decision satisfying performance requirement cannot be made. At this moment, we must continue observation to receive the next sensor data \( x_{n+1} \). It is known from above analysis that there are two likelihood ratio detection thresholds, shown by \( \eta_0 \) and \( \eta_1 \), even \( \eta_1 > \eta_0 \). Hence the likelihood ratio detection can be expressed by

\[
\lambda(\mathbf{x}_n) = \frac{p(\mathbf{x}_n | H_1)}{p(\mathbf{x}_n | H_0)} \geq \eta_1 \quad (2)
\]

\[
\lambda(\mathbf{x}_n) = \frac{p(\mathbf{x}_n | H_1)}{p(\mathbf{x}_n | H_0)} \leq \eta_0 \quad (3)
\]

### 4. Decision Method with the Characteristics of Physical Layer and Wireless Channel

#### 4.1. Problem Description and Decision Method

Adopting 2FSK as the physical layer modulation mode, synchronization detection is chosen as the demodulation method, shown in Fig. 2.

![Fig. 2. 2FSK performance analysis using synchronization detection.](image)

As shown in Fig. 2, at the receiver two band pass filters are used to distinguish between two signals with the central frequencies \( f_1 \) and \( f_2 \). After the coherent demodulation, the output signal of two band pass filters are sent to the sampling decision device.

Supposing that the transmitting symbol is "1", the input waveform of the two branches in Fig. 2 are separately denoted by

- the upper branch \( x_i(t) = a + n_i(t) \), \( i = 1, 2, \ldots, N \) \( (4) \)
- the lower branch \( x_i(t) = n_i(t) \), \( i = 1, 2, \ldots, N \) \( (5) \)

where \( a \) is the signal component, \( n_i(t) \) and \( n_i(t) \) are low pass Gauss White Noise with the mean value 0 and the variance value \( \sigma^2 \).

Supposing that the transmitting symbol is "0", the input waveform of the two branches are separately denoted by

- the upper branch \( x_i(t) = n_i(t) \)
- the lower branch \( x_i(t) = a + n_i(t) \)

As shown in Fig. 2, the output of the sampling decision device is the synchronization detection result, denoted by \( z \).

\[ z = x_1 - x_2, \quad (8) \]

where \( z \) is Gaussian random variable. If the receiving signal is "1", the random variable \( z \) possesses the mean value \( a \) and the variance value \( 2\sigma^2 \). If the receiving signal is "0", the random variable \( z \) possesses the mean value \(-a\) and the variance value \( 2\sigma^2 \). The receiver SNR value is \( r = a^2 / 2\sigma^2 \).

Considering wireless channel fading, we define \( a_i \) by

\[ a_i = \sqrt{P_i} = \sqrt{P_i / P_{\text{tot}}}, \quad (9) \]

where \( P_{\text{tot}} \) is the path loss ratio of the \( i \)th channel, \( P_i \) is the transmitting power, and \( P_r \) is the receiving power. Supposing that the path loss between every sensor node and the sink node is constant, then the wireless channel can be considered as constant channel, which indicates that with the unchanged transmitting power of the \( i \)th sensor node, the parameter \( a_i \) is always constant.

According to the characters of physical layer, the receiving signal \( x_i \) at the sink node under the two hypotheses \( H_0 \) and \( H_1 \) is separately described by

\[
H_0 : x_i = -a_i + n_i, \quad k = 1, 2, \ldots, \quad (10)
\]

\[
H_1 : x_i = a_i + n_i, \quad k = 1, 2, \ldots, \quad (11)
\]

where \( n_i \) is the Gauss noise with the mean value 0 and the variance value \( 2\sigma^2 \); \( a_i \in \{a_1, a_2, \ldots, a_N\} \), and \( N \) is the number of the transmitting sensor nodes.

After the \( N \)th reception from sensor node, the likelihood ratio function is

\[
\lambda(x) = \frac{p(\mathbf{x}_1 | H_1)}{p(\mathbf{x}_1 | H_0)} \prod_{i=1}^{N} p(x_i | H_i) \prod_{i=1}^{N} p(x_i | H_i) \quad (12)
\]

\[ = \exp\left(\sum_{i=1}^{N} x_i a_i / \sigma^2 \right) \]
Then the log likelihood ratio function is

\[ \ln \lambda(\hat{x}) = \sum_{i=1}^{N} \frac{x_i a_i}{\sigma^2} \]  

(13)

Hence the decision expression is denoted by

\[ \sum_{i=1}^{N} \frac{x_i a_i}{\sigma^2} \geq \ln \eta_i \]  

(14)

\[ \sum_{i=1}^{N} \frac{x_i a_i}{\sigma^2} \leq \ln \eta_i \]  

(15)

The above decision method is WSN Data Aggregation base on Sequential Detection (WDASD).

4.2. Threshold Evaluation

It is supposed that the constraint value of \( P(H_1 | H_0) \) and \( P(H_0 | H_1) \) is separately \( \alpha \) and \( \beta \).

\[ \alpha = \int p(\hat{x}_y | H_0) d\hat{x}_y \]  

(16)

\[ 1 - \beta = \int p(\hat{x}_y | H_1) d\hat{x}_y \]

(17)

Under the hypothesis \( H_1 \), if the decision is \( H_1 \), it must be satisfied that \( \lambda(\hat{x}_y) \geq \eta_i \), which is brought into Eq. (17) to conclude

\[ 1 - \beta \geq \eta_i \int p(\hat{x}_y | H_1) d\hat{x}_y = \eta_i \alpha \]  

(18)

Hence

\[ \eta_i \leq \frac{(1 - \beta)}{\alpha} \]  

(19)

According to the sequential detection requirement that the decision is \( H_1 \) under the condition \( \lambda(\hat{x}_y) \geq \eta_i \), the theory upper limit of \( \eta_i \) is chosen to guarantee enough observation data to get the decision satisfying the performance index. Hence the threshold \( \eta_i \) is denoted by

\[ \eta_i = \frac{(1 - \beta)}{\alpha} \]  

(20)

Similarly

\[ \eta_s = \frac{\beta}{(1 - \alpha)} \]

When \( \ln \eta_s < \ln \lambda(\hat{x}_y) < \ln \eta_i \), the sink node does not make any decision. But it is supposed that there are only \( N^* \) sensor nodes in the Wireless Sensor Network. When number of the sensor nodes participating in the aggregation algorithm comes to \( N^* \), the compulsory decision result is made.

4.3. Simulation

It is supposed that the AWGN channel parameters from different sensor nodes to sink node are the same; a sequence is the all "1" sequence; SNR value is \( a^2 / 2\sigma_x^2 = 0.5 \), i.e. 0dB; there are 100 sensor nodes in the elementary unite; and the constraint values of \( P(H_0 | H_0) \) and \( P(H_1 | H_1) \) are all 0.01. The simulation results are shown in Fig. 3.

![Simulation Results](image_url)

As shown in Fig. 3(a), in 1000 decision processes, there are five wrong decision processes, which satisfy the constraint of false alarm. It is illustrated in Fig. 3(b) that the number of nodes participating in the aggregation algorithm is not confirmed in the range of 2 to 30, with the mean value 6.5610. The larger the noise variance is, the wider the fluctuation range of the participating node number is, with a larger mean value.

In the realistic environment, the channel parameters from different sensor nodes to sink node are always different. Hence it is supposed that the receiving signal strength is distributed evenly in the
range of \([0, 1]\); at the sink node the noise variance is \(\sigma_1^2 = 2\sigma_2^2 = 2\). Then the simulation result is shown in Fig. 4.

**Fig. 4.** WDASD performance with different channel parameters.

As shown in Fig. 4, with different channel parameters, the number of participating nodes is larger than the number of participating nodes with the same channel parameters. And the curve fluctuation is also larger when the channel parameters are different.

On this basis, the effect of different channel parameter arrangements to the average number of participating nodes is researched as follow. In the simulation, 50 \(a\) sequences are randomly generated; and 1000 fusion decision processes are carried on according to every \(a\) sequence to get the average number of nodes participating in the aggregation algorithm. As shown in Fig. 5, under the condition of different channel parameters arrangements, the average numbers of participating nodes are different.

**Fig. 5.** WDASD performance with random channel arrangement.

Because different channel parameters arrangements cause the SNR value different at the sink node in the same receiving order, which means that the receiving order of signals with different SNR values has a big affect on the algorithm performance.

5. Data Aggregation Cross-layer Design Scheme Based on WDASD

5.1. Data Aggregation Scheme

From the above discussion, it is known that if the sensor node with high SNR value transmits the sensor data first, namely in the descending order of receiving SNR value at the sink node, the performance of data aggregation algorithm can be optimized to reduce the number of anticipating nodes. And in WSN, SNR value of the receiving signal can be tested by RSSI in the physical layer. Considering that TDMA is adopted in the data aggregation system, the sink node can allocate the time slots to sensor nodes in the descending order of SNR value. Then the data aggregation cross-layer design scheme is illustrated as follow:

1) The data packets are transmitted by sensor nodes with the same fixed power, which denotes the signal strength by \(i_n\). At the sink node, the RSSI values of the receiving signals transmitted through different channels are measured, which is indicated by \(i_n(n = 1,2,\ldots,N')\). Then the background noise is measured, indicated by \(i_{\text{noise}}\). \(i_n\) is denoted by

\[
i_n = a_n^2
\]

2) The transmission order of the sensor nodes, i.e. \(\{\text{MID}_1, \text{MID}_2, \ldots, \text{MID}_n\}\), are sorted by the sink node in the descending order of receiving SNR value. Namely the sensor node with good link quality transmits first. First the sink node transmits the communication order information. Then every sensor node confirms its distributed communication time slot and measures environmental parameters. Lastly some strategy is adopted to wait for the communication time slot.

3) The sensor nodes transmit the data packets in order to the sink node. And the sink node receives the data packets and measures the RSSI values of the receiving signals. Meanwhile, WDASD is carried on to get a satisfied decision result. And then transmits a stopping fusion signal which makes the remaining nodes stop this data packets transmission.

4) The next data aggregation process is started. Then the above procedures are repeated.

5.2. Simulation

The simulation conditions are the same as that described in 3.3. Two cases are considered in this section:

1) Case 1: cross-layer scheduling is not adopted, which means that the sensor nodes transmit the data packets randomly;

2) Case 2: cross-layer scheduling is adopted, which means that the sensor nodes transmit the data
packets in the descending order of receiving SNR value. The simulation results are shown in Fig. 6.

![Fig. 6. WDASD performance adopting and not adopting cross-layer scheduling.]

As shown in Fig. 6, when cross-layer scheduling is adopted, the number of participating nodes is much decreased. Under the condition that there are 100 sensor nodes in the elementary unite, the performance of Median fusion algorithm [5] and WDASD algorithm with cross-layer scheduling is compared as the follow. It is supposed that through the AWGN channel the signals transmitted from 100 sensor nodes are distributed evenly in the range of [0, 1]. The noise variance increase progressively from one observation to the next. In one observation, 100000 fusion simulations are carried out to get the false decision probability and the average number of participating nodes, shown by Fig. 7.

As shown in Fig. 7(a), when the noise variance is less than 2, the performance of WDASD algorithm is worse than the performance of Median fusion algorithm in the aspect of false alarm decision probability. But with the noise variance larger than 2, the performance of WDASD algorithm is better than the performance of Median fusion algorithm. It is illustrated by Fig. 7(b) that the average numbers of participating number in WDASD algorithm with different noise variance are also different, which can reduces the number of participating nodes.

6. Conclusions

In order to reduce the number of nodes participating in the aggregation algorithm in the elementary unite of WSN, sequential detection method is applied into data aggregation, to propose WSN Data Aggregation base on Sequential Detection (WDASD) algorithm and prove its performance by simulations. To further improve the performance of WDASD algorithm, cross-layer scheduling scheme suitable for WDASD is proposed. The WDASD algorithm adopting cross-layer scheduling is simulated, which indicates that the number of participating nodes can be further reduced with lower energy consumption than that of WDASD algorithm not adopting cross-layer scheduling.

![Fig. 7. Performance comparison.](a) false decision probability. (b) the average number of participating nodes.

Acknowledgements

The research work was supported by National Natural Science Foundation of China under Grant No. 61101238.

References


2014 Copyright ©, International Frequency Sensor Association (IFSA) Publishing, S. L. All rights reserved.
(http://www.sensorsportal.com)