Research Article

Research on Reliability-Oriented Data Fusetaggregation Algorithm in Large-Scale Probabilistic Wireless Sensor Networks

Hai-xia Peng, Hai Zhao, Da-zhou Li, Shuai-zong Si, and Wei Cai

Institute of Information Science and Engineering, Northeastern University, No. 3 Wenhua Street, Lane 11, Shenyang 110819, China

Correspondence should be addressed to Hai-xia Peng; penghaixia@neuera.com

Received 8 December 2013; Accepted 22 April 2014; Published 21 May 2014

Academic Editor: Yung-Fa Huang

Copyright © 2014 Hai-xia Peng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A lot of facts show that many researches just place emphasis on data aggregation or data fusion, which is not beneficial to analyze the sensed data thoroughly and will lead to the aggregation results not being used fully; worse yet, the actual networks are always existed with lossy links; many now available aggregation algorithms are based on ideal network models and not any further analysis and fusion about aggregation results are done. Thus, we propose the concept of data fusaggregation so as to support processing sensed data while transmitting in large-scale probabilistic wireless sensor networks and propose a reliability-oriented data fusaggregation algorithm (RODFA) to assist users to get the monitoring information from the monitored geographic environment and measure the reliability of the information they get. RODFA also facilitates network administrator to improve the system sensing performance for large-scale probabilistic WSNs. In RODFA, the parameter $\eta$, which could reflect the reliability of aggregation result intuitively, is defined and calculated and it plays an important part in helping users to process aggregation result further. In our experiment, the validity of RODFA is verified by our simulation results, and the influence of network sizes and network performances on data fusaggregation is analyzed.

1. Introduction

Wireless sensor network (WSN) is a class of wireless ad hoc networks which consist of thousands of sensor nodes (SNs). Due to recent advance in microelectronics, wireless communications and sensor technologies have made the development of low-cost, low-power, multifunctional sensor nodes possible [1]. The capabilities of pervasive surveillance, sensor networks have attracted significant attention from many applications domains, such as habitat monitoring [2, 3], object tracking [4, 5], environment monitoring [6–8], military [9], traffic management [10], disaster management [11], and smart environments. In these applications, the aggregation and fusion of sensed data are very important for users to get the summary information from monitored areas. Many works about data aggregation and data fusion for WSNs emerged, which further promote the application domains of WSN [12–14]; however, firstly, the existing study rarely takes into consideration that SNs are restricted by resources due to low battery supply; secondly, most researches are aimed at studying the network capacity issue under ideal network model which does not match the actual network; thirdly, data aggregation and data fusion have different singular focus; for example, the focus of data aggregation is data transmitting while data fusion places emphasis on analyzing sensed data. Both data transmitting and analyzing are important to WSNs applications; lastly, there are few profound discussion and research on the network size’s and network performance’s influence on data fusaggregation result (which will be introduced in the next section). We will try to study these four issues in later sections.

Firstly, as energy resources provided for SNs are usually battery cells which are impossible to recharge during WSNs’ working process, SNs are restricted by resources due to low battery supply. Consequently, in this paper, energy saving technology is taken into consideration, to make the network model more practical. For better energy utilization, we choose cluster-based WSNs [15, 16]. In cluster-based WSNs, SNs resident in nearby area would form a cluster and SNs can select one of the clusters to be their cluster-head nodes.
(CHs). The CH organizes data pieces received from SNs into an aggregated result and then forwards the result to the base station (sink node) along with the regular routing paths.

Secondly, to the best of our knowledge, lots of the studies mentioned above are based on ideal deterministic network model (DNM), where any pair of nodes in a network is either connected or disconnected. If two nodes are connected, that is, there is a deterministic link between them, and then a successful data transmission can be guaranteed as long as there is no collision. Otherwise, if two nodes are disconnected, the direct communication between them is assumed to be impossible. However, for real application, this DNM assumption is too ideal and not practical on account of the “transitional region phenomenon” [17, 18]. Because of the transitional region phenomenon, a large number of network links (probably more than 90%) become unreliable, which is named lossy links [17]. Even without collisions, data transmission over a lossy link is successfully conducted with a certain probability rather than being completely guaranteed. Therefore, a more practical network model for WSN is the probabilistic network model (PNM) [17], in which data communications over a link are successful with a certain probability rather than always being successful or always unsuccessful. For convenience, the WSNs considered under the DNM/PNM are called deterministic/probabilistic WSNs. Hence, in order to make our research much closer to the real network and get more practical value than existing studies, our research is under probabilistic WSN shown as the network model in Section 2.

Thirdly, in order to truly achieve processing sensed data while transmitting among network, users, and network administrator, we first present the concept of data fusaggregation which is based on the definitions of data aggregation and data fusion. In many applications, WSNs are mainly used for gathering data acquired from the physical environment to an external base station [19]. During a data gathering process, if the raw data can be aggregated and only an aggregation value is transmitted to the sink node, it is called data aggregation [20]. And data fusion is a widely adopted signal processing technique that can improve the system sensing performance by jointly considering the measurements of multiple sensors [21]. As data aggregation and data fusion are both playing important functions in many applications of WSNs and they have different focuses, the realization of processing sensed data while transmitting has started to become important for WSNs. Thus, data fusaggregation is presented in our research as follows.

Definition 1 (data fusaggregation). The raw data can be aggregated; only one aggregation value will be transmitted to the sink node. And through analyzing the aggregation value, the sink node will get some other information which could measure the performance of network or be useful to users.

Lastly, on the basis of the concept of data fusaggregation, a data fusaggregation algorithm is proposed to comprehensively analyze and use sensed data to facilitate the improvement of system sensing performance. This algorithm is named reliability-oriented data fusaggregation algorithm (ROFDA). In this algorithm, firstly, the parameter $\eta$ (the lower limit value of reliability) is defined to measure the reliability of aggregation result. Then we obtain the formula for calculating the value of $\eta$ through theoretical derivation. Lastly, the aggregation result and the parameter $\eta$ will be sent to users through the sink which could be a reference for users’ handling the information they get. In our experiment part, we first study the influence of network size and network performance on RODFA result.

The rest of this paper is organized as follows. In Section 2, the network model is discussed; it includes the assumptions of this model and the problem definitions. In Section 3, we describe the mathematic foundations of RODFA. The calculation of $\eta$ for one $\text{Sum}(S)$ is given in Section 4. Section 5 shows the validity of RODFA. And we discuss factors affecting the RODFA (network size and network performance) in Section 6. In the last section, we conclude the paper.

2. Network Model

2.1. Assumptions of the Model. In this paper, we consider that a large-scale probabilistic WSN consists of $N$ sensors. Let $N_i$ be the number of active SNs in the monitored area at time $t$. $N_i$ is varying with time and unknown by the sink. Let $s_i$ ($1 \leq i \leq N_i$) be the sensed data of active sensor node $i$ at time $t$, and let $S_i(1, 2, 3, \ldots, s_i)$ be the set of all the sensed data in monitored area at time $t$. $s_i$ is stored in the active sensor node $i$ for $1 \leq i \leq N_i$. Since all sensed data are bounded, we use $\sup(S_i)$ to denote the upper bound of all sensed data.

On the basis of the analysis in Section 1, for better energy utilization, we apply the dynamic minimal spanning tree routing protocol (DMSTRP) to our network model and our radio power model is similar to [22]. The DMSTRP is a cluster-based routing protocol for large-scale wireless sensor networks which is proposed in [23]. According to [23], in our network model, the monitored geographical area is divided into a number of clusters which are formed similar to [24], and the clusters are disjoined with each other. Here, assume that the monitored geographical area is fully covered by active sensors, and it is divided into $n$ clusters. Let $\omega$ be the maximum of network’s layers at time $t$, and let $m_1, m_2, m_3, \ldots, m_n$ be the numbers of nodes in the layer 1, layer 2, layer 3, and layer $\omega$ at time $t$, respectively. Let $S_{i1} = \{s_{i11}, s_{i12}, s_{i13}, \ldots, s_{i1m_i}\}$ be the set of all sensed data in layer 1 at time $t$, and let $S_{i2}, S_{i3}, \ldots, S_{in}$ be the set of all sensed data in layer 2, layer 3, and layer $\omega$ at time $t$, respectively.

When the sink node initiates one time of data collection at time $t$, we let $d$ be the distance of all lossy links, the lossy links are successfully conducted with a certain probability $q$, and let $B(q^t)$ be the set of all sensed data which are successfully sent to sink node from layer 1 at time $t$. Thus, the sensed data of layer 2, layer 3, layer $a$ will be sent successfully to the sink with a certain probability $q^2, q^3, \ldots, q^n$, respectively, and let $B(q^2), B(q^3), \ldots, B(q^n)$ be the set of all sensed data which are sensed successfully to sink node from the
layer 2, layer 3, . . . , layer 𝜔 at time 𝑡, respectively. And the function relationship between 𝑞 and 𝑑 will be presented in the next section.

2.2. Problem Definition. In this paper, when researching on data fusion aggregation, we take SUM operation, for example, and propose RODFA to comprehensively analyze and use sensed data for facilitating further promotion of WSNs’ application domains. The exact sum of the monitored area at time 𝑡 is defined as Sum(𝑆) = 𝜔 ∑𝑖=1 𝑆1𝑡𝑖 + ∑𝑖=1 𝑆2𝑡𝑖 + ∑𝑖=1 𝑆3𝑡𝑖 + . . . + ∑𝑖=1 𝑆𝜔𝑡𝑖. We will briefly introduce some related definitions of reliability as follows before introducing the steps of RODFA.

Definition 2 (ε-estimate). ̂ 𝐼, is called as an ε-estimate of 𝐼, if |̂ 𝐼 − 𝐼|/𝐼 ≤ ε for any ε (ε ≥ 0).

Definition 3 (reliability). For a given network and ε (ε ≥ 0), reliability is the probability of whether ̂ 𝐼, is the ε-estimate of 𝐼,.

Definition 4 (𝜂). For a given network and ε (ε ≥ 0), η (0 ≤ η ≤ 1) is the lower limit value of probability of whether ̂ 𝐼, is the ε-estimate of 𝐼, that is, η ≤ Pr(|̂ 𝐼 − 𝐼|/𝐼 ≤ ε) and we call it the lower limit value of reliability.

η is an important parameter which could measure the reliability of aggregation results. It will be sent to users for facilitating the next handling of aggregation result, should the users regard the aggregation result as a decision-making or just a reference point, even discard this aggregation result directly. The calculation of η will be completed in RODFA. The main steps of RODFA are shown as Figure 1, and they can be described as follows in detail.

Step 1. Sink node launches one time of data collection and sends the data acquisition command to every cluster, and the clusters will retransmit this command to their own child nodes.

Step 2. According to [22], every node will send its sensed data to its cluster node. And after analyzing and processing the sensed data, the cluster will retransmit the processing result to sink node.

Step 3. Sink node will get an approximate sum which will be discussed in Section 3: Sum(𝑆) = (1/𝑞)∑𝐵∈𝐵(3) 𝑆1𝑡1 + (1/𝑞) 𝑆1𝑡2 + . . . + (1/𝑞) ∑𝐵∈𝐵(5) 𝑆3𝑡𝑖 + . . . + (1/𝑞) ∑𝐵∈𝐵(𝜔) 𝑆𝜔𝑡𝑖.

Step 4. According to the formula of η ∆ 𝐿2 ≤ 𝑞−η inf(𝑁) inf(𝑆) exp(1 − 𝑞−η) sup(𝑆) (where δ = 1 − η, and the derivation process of this formula will be given in Section 4), sink node will calculate the value of η and turn it back to users with Sum(𝑆) together, or turn it back to the network administrator for facilitating the administrator to improve the system sensing performance. And the algorithm stops.

Combined with the above steps of RODFA, the key of RODFA is to obtain the value of η of Sum(𝑆). And the problem of computing the value of η is defined as follows.

Input:

(1) 𝑆1 = {𝑠1, 𝑠2, 𝑠3, . . . , 𝑠𝑖}, 𝑆2 = , 𝑆3 = , . . . , 𝑆𝜔 = ;

(2) ε (ε ≥ 0), 𝜔, and 𝑞;

(3) aggregation operator sum.

Output:

The value of η of the mathematic estimator of sum: Sum(𝑆).

For facilitating the later analysis, we will give the other two definitions of unbiased estimate and delay.

Definition 5 (unbiased estimate). ̂ 𝐼, is an unbiased estimate of 𝐼, if the mathematical expectation of ̂ 𝐼, is equal to 𝐼, that is, 𝐸(̂ 𝐼,) = 𝐼; otherwise ̂ 𝐼, is a biased estimator of 𝐼,.

Combining [23], DMSTRP connects nodes in clusters by MSTs. In each cluster, all nodes including the CH are connected by a MST and then the CH as a leader to collect data from the whole tree. All CHs are connected by another MST to route toward sink. And the processing of data fusing is handled along with the tree route. The average transmission distance of each node can be reduced by using MSTs as Figure 2 shows, and thus the energy dissipation of transmitting data is reduced.

In Figure 2, node 1, node 2, and node 6 are leaf nodes; node 3 and node 5 are father nodes; node 4 is the cluster head node. Because node 1 and node 2 have the same father, only node 1 can transmit in the first period. Thus the first period transmitting queue is {1, 6}; this means node 1 and node 6 can transmit their data at the same time. We can get the transmitting queue in the following period as {2, 5} and {3}. Based on the above analysis, we define delay as follows.

Definition 6 (delay). The amount of periods from the sink node initiates a data collection command to all sensed data from SNs being sent to the sink node, no matter how long it will last. This amount of time is called one delay.

According to the concept of delay, we can get that the delay of Figure 2 is Figure 3.

3. Mathematical Foundations

The calculation of probability 𝑞 is the foundation of RODFA. Existing studies show that the probability 𝑞 will be a function of distance 𝑑 (the distance between any pair of nodes is connected). Next, we will briefly describe the derivation of this function and then do research on the estimator of sum.

3.1. Relationship between 𝑞 and Distance 𝑑. Let 𝑟 be the receiving signal-to-noise ratio (SNR) and let 𝑓 be the length
of one frame, and then the functional relationship between $q$ and $r$ can be described as [18]

$$q = \left(1 - 0.5 \times e^{-r/2}(1/0.64)\right)^{8f}.$$  (1)

In this paper, the propagation model is lognormal shadowing model [25], and the relationship between $r$ and the distance $d$ will be obtained in next section.

Let $PL(d)$ be the path loss for one certain location, $d_0$ the reference distance, $\beta$ the path loss exponent, and $\sigma$ Gaussian random variable available for zero mean. Then the relationship between $PL(d)$ and $d$ can be described by formula (2) and in units of dBm:

$$PL(d) = PL(d_0) + 10\beta\log_{10}\left(\frac{d}{d_0}\right) + X_\sigma.$$  (2)

On the basis of formula (2), we assume that $P_{\text{trans}}$ is the output power of the sender and $P_{\text{recv}}$ is the power that the receiver received. Then $P_{\text{recv}}$ can be obtained from the following:

$$P_{\text{recv}} = P_{\text{trans}} - PL(d).$$  (3)

Let $P_n$ be the platform noise and combined with the relationship (3) between the SNR and the power received by the receiver; we can see that

$$\text{SNR} = P_{\text{recv}} - P_n.$$  (4)

Once the model is determined, all of the parameters in the above formulas will be only given. As we set the propagation model to be lognormal shadowing model, the parameter settings for our experiment are similar to [26]. Combined with the above formulas (1)–(4), we can obtain the relationship between $q$ and $d$ that can be shown as

$$q = \left(1 - 0.5 \times e^{-\left(P_{\text{trans}} - PL(d) - 10\beta\log_{10}(d/d_0) - X_\sigma - N\right)/1.28}\right)^{8f}.$$  (5)

3.2. Estimator of Sum. Let $S_1, S_2, S_3, \ldots, S_\omega$ be the sets of sensed data in layer 1, layer 2, layer 3, …, layer $\omega$ at time $t$, respectively, and let $B^{(1)}, B^{(2)}, B^{(3)}, \ldots, B^{(\omega)}$ be the sets of all
sensed data which are sensed successfully to sink node from the layer 1, layer 2, layer 3, . . . , layer ω at time t, respectively. The mathematic estimator of sum is denoted by $\text{Sup}(S_t)$, and the $\overline{\text{Sum}}(S_t)$ can be computed by
\[
\overline{\text{Sum}}(S_t) = \frac{1}{q} \sum_{S_{t1i} \in B^{q(i)}} S_{t1i} + \frac{1}{q^2} \sum_{S_{t2i} \in B^{q(i)}} S_{t2i} + \cdots + \frac{1}{q^m} \sum_{S_{tmi} \in B^{q(i)}} S_{tmi},
\]
where $q$ is the probability in which lossy links are successfully conducted at time $t$. And according to the above definition of the unbiased estimate, Theorem 7 shows that $\overline{\text{Sum}}(S_t)$ is the unbiased estimator of the exact sum $\text{Sum}(S_t)$.

**Theorem 7.** Let $E(\overline{\text{Sum}}(S_t))$ be the mathematical expectation of $\overline{\text{Sum}}(S_t)$ and let $\text{Var}(\overline{\text{Sum}}(S_t))$ be the variance. Then,
\[
E\left(\overline{\text{Sum}}(S_t)\right) = \text{Sum}(S_t),
\]
\[
\text{Var}(\overline{\text{Sum}}(S_t)) \leq \sup(S_t) \text{Sum}(S_t) \left(1 - \frac{q}{q^m}\right). \quad (7)
\]

**Proof.** For any $i \leq m_k$, set random variables $X_{t1i}, X_{t2i}, X_{t3i}, \ldots, X_{tmi}$ satisfy the following equations, respectively:
\[
X_{t1i} = \begin{cases} 1 & \text{if } S_{t1i} \in B^{q(i)} \\ 0 & \text{if } S_{t1i} \notin B^{q(i)} \end{cases},
\]
\[
\vdots
\]
\[
X_{tmi} = \begin{cases} 1 & \text{if } S_{tmi} \in B^{q(i)} \\ 0 & \text{if } S_{tmi} \notin B^{q(i)} \end{cases}.
\]

Clearly, there are $\text{Pr}(X_{t1i} = 1) = q$, $\text{Pr}(X_{t1i} = 0) = 1 - q$, $\text{Pr}(X_{t2i} = 1) = q^2$, $\text{Pr}(X_{t2i} = 0) = 1 - q^2$, ..., $\text{Pr}(X_{tmi} = 1) = q^m$, and $\text{Pr}(X_{tmi} = 0) = 1 - q^m$. Meanwhile, according to the lognormal shadowing model, for $1 \leq i \neq j \leq m_k$, there exist the random variables $X_{tij}$ and $X_{tij}$ that are independent of each other ($1 \leq k \leq \omega$). Furthermore, according to the distribution of $X_{tij}$ ($1 \leq i \leq \omega$), there are $E(X_{t1i}) = q$, $E(X_{t2i}) = q^2$, ..., $E(X_{tmi}) = q^m$, $\text{Var}(X_{t1i}) = q(1 - q)$, $\text{Var}(X_{t2i}) = q^2(1 - q^3)$, ..., $\text{Var}(X_{tmi}) = q^m(1 - q^m)$.

Combining (6), we can get
\[
\overline{\text{Sum}}(S_t) = \frac{1}{q} \sum_{i=1}^{m_1} S_{t1i} X_{t1i} + \frac{1}{q^2} \sum_{i=1}^{m_2} S_{t2i} X_{t2i} + \cdots + \frac{1}{q^m} \sum_{i=1}^{m_m} S_{tmi} X_{tmi}.
\]

As $E(X_{t1i}) = q$, $E(X_{t2i}) = q^2$, ..., $E(X_{tmi}) = q^m$; there is
\[
E\left(\overline{\text{Sum}}(S_t)\right) = \sum_{i=1}^{m_1} S_{t1i} + \sum_{i=1}^{m_2} S_{t2i} + \cdots + \sum_{i=1}^{m_m} S_{tmi} = \text{Sum}(S_t). \quad (9)
\]

For two independent events $X$ and $Y$, let $D(X)$ and $D(Y)$ be the variances of event $X$ and event $Y$, respectively; then $D(X + Y) = D(X) + D(Y)$. Combine $\text{Var}(X_{t1i}) = q(1 - q)$, ..., $\text{Var}(X_{tmi}) = q^m(1 - q^m)$; then,
\[
\text{Var}\left(\overline{\text{Sum}}(S_t)\right) = \frac{1 - q}{q} \left( \sum_{i=1}^{m_1} S_{t1i}^2 + \sum_{i=1}^{m_2} S_{t2i}^2 + \cdots + \sum_{i=1}^{m_m} S_{tmi}^2 \right) \leq \frac{1 - q}{q} \left( \sup(S_t) \sum_{i=1}^{m_1} S_{t1i} + \sum_{i=1}^{m_2} S_{t2i} + \cdots + \sum_{i=1}^{m_m} S_{tmi} \right)
\]
\[
= \frac{1 - q}{q} \sup(S_t) \sum_{i=1}^{m_1} S_{t1i} \left(1 + q + q^2 + \cdots + q^{m_1 - 1}\right) + \frac{1 - q^2}{q^2} \left(1 + q^2 + q^4 + \cdots + q^{m_2 - 2}\right) \sum_{i=1}^{m_2} S_{t2i}
\]
\[
+ \cdots + \frac{1 - q^m}{q^m} \sup(S_t) \sum_{i=1}^{m_m} S_{tmi} \left(1 + q^m + q^{2m} + \cdots + q^{m_m - 1}\right).
\]

According to formula (11), as $\frac{1}{q^m - 1}$\(/(1 + q + q^2 + \cdots + q^{m_1 - 1}) \leq 1, \frac{1}{q^{m_2 - 2}}/(1 + q^2 + q^4 + \cdots + q^{m_2 - 2}) \leq 1, \frac{1}{q^{m_3 - 3}}/(1 + q^3 + q^6 + \cdots + q^{m_3 - 3}) \leq 1, \ldots$, then,
\[
\text{Var}\left(\overline{\text{Sum}}(S_t)\right) \leq \frac{1 - q^m}{q^m} \sup(S_t) \left[ \sum_{i=1}^{m_1} S_{t1i} + \sum_{i=1}^{m_2} S_{t2i} + \sum_{i=1}^{m_3} S_{t3i} + \cdots + \sum_{i=1}^{m_m} S_{tmi} \right]
\]
\[
= \left(1 - \frac{q^m}{q^m}\right) \sup(S_t) \times \sum_{i=1}^{m_1} S_{t1i} + \sum_{i=1}^{m_2} S_{t2i} + \sum_{i=1}^{m_3} S_{t3i} + \cdots + \sum_{i=1}^{m_m} S_{tmi} \right)
\]
\[
= \text{Sup}(S_t) \times \sum_{i=1}^{m_1} S_{t1i} + \sum_{i=1}^{m_2} S_{t2i} + \sum_{i=1}^{m_3} S_{t3i} + \cdots + \sum_{i=1}^{m_m} S_{tmi} \right)
\]
\[
= \text{Var}\left(\sum_{i=1}^{m_1} S_{t1i} + \sum_{i=1}^{m_2} S_{t2i} + \sum_{i=1}^{m_3} S_{t3i} + \cdots + \sum_{i=1}^{m_m} S_{tmi} \right) \leq \text{Var}(\text{Sum}(S_t)), \quad (12)
\]

Theorem 7 shows that the mathematic estimator of sum $\overline{\text{Sum}}(S_t)$ is the unbiased estimator of the exact sum $\text{Sum}(S_t)$. The upper limit of the variance value $\text{Var}(\overline{\text{Sum}}(S_t))$ is inversely proportional to the probability $q$. That is to say, with the probability $q$ increasing, the upper limit of $\text{Var}(\overline{\text{Sum}}(S_t))$ can be arbitrarily small. Thus, referring to [27], with the increase of $q$, the relative error between $\overline{\text{Sum}}(S_t)$ and $\text{Sum}(S_t)$ will gradually decrease, and if $q$ is sufficiently large, this relative error can be arbitrarily small. \qed
4. Calculating $\eta$ of $\hat{\text{Sum}}(S_i)$

According to the steps of RODFA in Section 2.2, the key step is to calculate $\eta$ of $\hat{\text{Sum}}(S_i)$. Thus, we will research on this issue next.

The steps of calculating the value of $\eta$ are (1) proving that $\hat{\text{Sum}}(S_i)$ obeys the normal distribution; (2) transforming the normal distribution into standard normal distribution; and (3) utilizing the characteristics of standard normal distribution to calculate the value of $\eta$.

For any $i$, let the variable $Y_{\omega}$ (1 ≤ $K$ ≤ $\omega$) be as follows:

$$
Y_{\omega} = \begin{cases} 
S_{1t} & \text{if } S_{1t} \in B^{(q)} \\
0 & \text{if } S_{1t} \notin B^{(q)}
\end{cases}
$$

$$
Y_{\omega i} = \begin{cases} 
S_{\omega ti} & \text{if } S_{\omega ti} \in B^{(q_\omega)} \\
0 & \text{if } S_{\omega ti} \notin B^{(q_\omega)}
\end{cases}
$$

There is $\hat{\text{Sum}}(S_i) = \frac{1}{(q_\omega)} \sum_{\omega i} S_{\omega ti}$, $S_{\omega ti} = \frac{1}{q_\omega} [S_{\omega ti} - q_\omega S_{\omega ti}]$.

Firstly, we need to prove that $\hat{\text{Sum}}(S_i)$ obeys normal distribution. In view of the linear combination of $n$ independent normal distribution functions still obey normal distribution through proving that the sum of each layer data obeys normal distribution to prove that $\hat{\text{Sum}}(S_i)$ obeys the normal distribution. Reference [28] shows that, if each layer data conforms to Lyapunov condition, the sum of each layer data will be in accordance with the application conditions of the central limit theorem; that is, the sum of each layer data will obey the normal distribution. And Theorem 8 proofs that the data in layer 1, layer 2, ..., layer $\omega$ conform to Lyapunov condition, respectively.

Theorem 8. The $\omega$ groups of sequence of random variables $Y_{\omega}$ (1 ≤ $K$ ≤ $\omega$) satisfy the Lyapunov condition; that is $\exists \xi_k > 0$ satisfy the following formula:

$$
\lim_{m_t \to \infty} \frac{1}{m_t} \sum_{i=1}^{m_t} E \left( |Y_{\omega i} - \mu_{\omega i}|^{2+\xi_i} \right) = 0.
$$

Among them, 1 ≤ $K$ ≤ $\omega$, $m_t$ is the number of data in layer $k$ at time $t$, $m_{\omega t} = \sum_{i=1}^{m_t} \sigma_{\omega i}$, and for $\forall i$ (1 ≤ $i$ ≤ $m_{\omega t}$), there are $\mu_{\omega i} = E(Y_{\omega i})$ and $\sigma_{\omega i} = \text{Var}(Y_{\omega i})$.

Proof. Combining Section 2, the sensed data of the layer 1, layer 2, layer 3, ..., layer $\omega$ will be sent successfully to the sink with a certain probability $q, q^2, q^3, \ldots, q^{\omega}$, respectively. $S_{1t1}, S_{2t1}, S_{3t1}, \ldots, S_{\omega t1}$ are the set of all sensed data in layer 1, layer 2, layer 3, ..., layer $\omega$ at time $t$, respectively, and $B^{(q)}, B^{(q^2)}, B^{(q^3)}, \ldots, B^{(q^\omega)}$ are the sets of all sensed data which are sent successfully to sink node from layer 1, layer 2, layer 3, ..., layer $\omega$ at time $t$, respectively.

For Layer 1 ($k = 1$). There is $\mu_{1t} = E(Y_{1t1}) = qS_{1t1}$, and $\sigma_{1t} = \text{Var}(Y_{1t1}) = s_{1t1}^2 q (1-q)$.

Order $\xi_1 = 1$, according to the above analysis, for $\forall i$ (1 ≤ $i$ ≤ $m_{\omega t}$), there is

$$
E \left( |Y_{1t1} - \mu_{1t1}|^{2+\xi_i} \right) = E \left( |Y_{1t1} - \mu_{1t1}|^{3} \right)
$$

$$
= q(s_{1t1} - qS_{1t1})^3 + (1 - q)(qS_{1t1})^3
$$

$$
= s_{1t1}^3 q (1-q) (1-2q+2q^2).
$$

Meanwhile, we have

$$
S_{m_{1t}}^2 = s_{1t1}^3 = \sum_{i=1}^{m_{1t}} s_{1t1}^2 q (1-q) \sqrt{\sum_{i=1}^{m_{1t}} s_{1t1}^2 q (1-q)}.
$$

Combine formula (15) with formula (16), there is

$$
\lim_{m_{1t} \to \infty} \frac{1}{m_{1t}} \sum_{i=1}^{m_{1t}} E \left( |Y_{1t1} - \mu_{1t1}|^{3} \right) = q s_{1t1}^3 q (1-q) (1-2q+2q^2).
$$

Among them, $\text{inf}(S_{1t1})$ and $\text{sup}(S_{1t1})$, respectively, present the lower limit and upper limit of sensed data in layer 1. Thus, $\text{inf}(S_{1t1}) \leq \text{sup}(S_{1t1}) \leq \infty$; hence $\lim_{m_{1t} \to \infty} \frac{1}{m_{1t}} \sum_{i=1}^{m_{1t}} E \left( |Y_{1t1} - \mu_{1t1}|^{3} \right) = 0$. Meanwhile, $s_{1t1}^3 \geq 0$ and $E\left(|Y_{1t1} - \mu_{1t1}|^{3}\right) \geq 0$; therefore $\lim_{m_{1t} \to \infty} \frac{1}{m_{1t}} \sum_{i=1}^{m_{1t}} E\left(|Y_{1t1} - \mu_{1t1}|^{3}\right) \geq 0$. In conclusion, $\lim_{m_{1t} \to \infty} \frac{1}{m_{1t}} \sum_{i=1}^{m_{1t}} E\left(|Y_{1t1} - \mu_{1t1}|^{3}\right) = 0$; that is, when $k = 1$, there is $\xi_1 = 1$ to satisfy formula (14) in Theorem 8, and $Y_{1t1}$ satisfies the Lyapunov condition. According to [28], $\hat{\text{Sum}}(S_i) = \sum_{\omega i} S_{\omega ti}$ meets the application conditions of central limit theorem; that is, $\hat{\text{Sum}}(S_i)$ obeys normal distribution.

Omitting the analysis of other layers (the researching method for other layers is same as $k = 1$), we will then describe the analysis of layer $\omega$; that is, $k = \omega$.

For Layer $\omega$ ($k = \omega$). Same as the analysis of $k = 1$, when $k = \omega$, there are $u_{\omega i} = E(Y_{\omega i}) = q^\omega S_{\omega ti}$ and $\sigma_{\omega i} = \text{Var}(Y_{\omega i}) = q^\omega (1-q^\omega) S_{\omega ti}^2$. For $\forall i$ (1 ≤ $i$ ≤ $m_{\omega t}$), if we let $\xi_\omega = 1$, then there will be

$$
E \left( |Y_{\omega i} - u_{\omega i}|^{2+\xi_i} \right) = E \left( |Y_{\omega i} - u_{\omega i}|^{3} \right)
$$

$$
= q^\omega (S_{\omega ti} - q^\omega S_{\omega ti})^3
$$

$$
= (1-q^\omega) q^\omega S_{\omega ti}^3
$$

$$
= q^\omega S_{\omega ti}^3 (1-q^\omega) (1-2q^\omega+2q^2\omega),
$$

$$
S_{m_{\omega t}}^2 = s_{\omega ti}^3 = \sum_{i=1}^{m_{\omega t}} s_{\omega ti}^2 q (1-q) \sqrt{\sum_{i=1}^{m_{\omega t}} s_{\omega ti}^2 q (1-q)}.
$$


Combining formula (18) with formula (19), there is
\[
\lim_{m_\omega \to \infty} \frac{1}{m_\omega} \sum_{i=1}^{m_\omega} E \left( |Y_{\omega i} - u_{\omega i}|^3 \right) = \frac{(1 - 2q^o + 2q^o)q^o}{q^o - 1} \lim_{m_\omega \to \infty} \frac{1}{m_\omega} \sum_{i=1}^{m_\omega} S_{\omega i}^3 \leq \frac{\left( 1 - 2q^o + 2q^o \right) q^o }{q^o - 1} \frac{\sup(S_{\omega})^3}{\inf(S_{\omega})^3} \lim_{m_\omega \to \infty} \frac{1}{\sqrt{m_\omega}}.
\]

That is, for layer \( \omega \), \( \xi_\omega = 1 \) satisfies the formula (14) in Theorem 8 to make \( \lim_{m_\omega \to \infty} (1/s^3_{\omega i}) \sum_{i=1}^{m_\omega} E(|Y_{\omega i} - \mu_{\omega i}|^3) = 0 \), and \( Y_{\omega i} \) also satisfies the Lyapunov condition. As whether the sensed data in each layer can be sent successfully to sink node is independent of each other, so \( \sum(S_\omega) \) is the sum of these \( \omega \) independent variable normal distributions \( \sum(S_\omega) \). Thus \( \sum(S_\omega) \) obeys normal distribution. For a given relative error limit \( \epsilon \), Theorem 9 describes the calculation of \( \eta \).

**Theorem 9.** Supposing \( \delta = 1 - \eta \), \( \phi_{\delta/2} \) is the \( \delta/2 \) quantile of standardized normal distribution, if \( \phi_{\delta/2} \) satisfies the following formula:
\[
\phi_{\delta/2}^2 \leq q^o \frac{\inf(N_\epsilon) \inf(S_\epsilon) \epsilon^2}{(1 - q^o) \sup(S_\epsilon)}.
\]

Then, the probability that the relative error between \( \sum(S_\epsilon) \) and \( \sum(S_\epsilon) \) satisfies the given error limit \( \epsilon \) will be equal or greater than \( \eta \); that is,
\[
\Pr \left( \left| \frac{\sum(S_\epsilon) - \sum(S_\epsilon)}{\sum(S_\epsilon)} \right| \leq \epsilon \right) \geq \eta.
\]

**Proof.** From formula (21), there is \( \inf(N_\epsilon) \inf(S_\epsilon) \geq \phi_{\delta/2}^2 \sup(S_\epsilon)(1 - q^o)/q^o \). As \( \inf(N_\epsilon) \) and \( \inf(S_\epsilon) \) are, respectively, the lower limit of \( N_\epsilon \) and the lower limit of the value of all sensed data, so there is \( \sum(S_\epsilon) = \sum_{i=1}^{N_\epsilon} s_{\epsilon i} \geq \inf(N_\epsilon) \inf(S_\epsilon) \). Hence,
\[
\epsilon^2 \sum(S_\epsilon) \geq \phi_{\delta/2}^2 \sup(S_\epsilon)(1 - q^o)/q^o.
\]

Theorem 7 shows that \( \text{Var}(\sum(S_\epsilon)) \leq (1 - q^o)/q^o \). \( \sup(S_\epsilon) \sum(S_\epsilon) \), \( E(\sum(S_\epsilon)) = \sum(S_\epsilon) \), and as \( \sum(S_\epsilon) \) obeys normal distribution, from formula (23) there is
\[
\Pr \left\{ \frac{\left| \sum\left( \frac{\sum(S_\epsilon) - \sum(S_\epsilon)}{\phi_{\delta/2} \sqrt{\text{Var}(\sum(S_\epsilon))}} \right) \right| \geq 1 \right\} = \delta.
\]

Combining the knowledge of standard normal distribution quantile [29], (23), (24), and \( \delta = 1 - \eta \), there is
\[
\Pr \left( \left| \frac{\sum(S_\epsilon) - \sum(S_\epsilon)}{\sum(S_\epsilon)} \right| \leq \epsilon \right) \geq \eta.
\]

**5. Validity of RODFA**

To evaluate RODFA, we use Matlab to simulate a sensor network of 5000 nodes. Using the above 5000 nodes and 285 cluster-heads randomly deployed in network area of 5000 m \( \times \) 5000 m, 10000 m \( \times \) 10000 m, 15000 m \( \times \) 15000 m, we make the DMSTRP protocol running in the simulation network system to connect all nodes and cluster-heads into a whole. For time \( t \) in ready phase, we can obtain the maximum of network's layers through counting the numbers of layers of all clusters. As all the nodes are randomly deployed in network area, and as far as possible to ensure uniformity in the process of deploying, we can randomly select one cluster from the network, and then to obtain the distance \( d \) of all lossy links through calculating the mean distance of all distances between every two linked nodes in this cluster. Plugging this distance \( d \) into formula (5), then we can get the probability \( q \) at time \( t \).

This group experiments are to investigate whether RODFA is valid. As data communications over a link is successful with a certain probability rather than always being successful or always fail in probabilistic WSNs [17], every time we calculate the reliability of \( \sum(S_\epsilon) \), we will get the different reliability value even if all relevant information of the network is fixed and immovable (i.e., the number of active nodes, the structure of the cluster, the number of clusters, the number of nodes in every cluster, the distance of all links, and the layout of all nodes are changeless).

We do 10000 times of calculations when all relevant information in the network is fixed and immovable and let the initial energy of sink node, cluster-heads, and nodes be large enough (no node will die in experiment process). For these 10000 calculations, we calculate the relative error between \( \sum(S_\epsilon) \) and \( \sum(S_\epsilon) \) obtained from our simulation network. In order to get the value of \( \sum(S_\epsilon) \), in this experiment, we assume that the CHs will not only retransmit the aggregation result to sink node but also relay all sensed data that are from other SNs or CHs and then do statistics analysis on cumulative probability under different relative error bounds \( \epsilon \), shown as the blue dashed lines in Figures 3, 4, and 5, and we call this reliability as statistical reliability. Furthermore, according to the above analysis, sink node can get the values of \( \omega \) and \( q \) for time \( t \) easily, and after putting \( \omega \) and \( q \) in \( \phi_{\delta/2}^2 \leq q^o \) and \( \phi_{\delta/2} \leq \epsilon \), we can calculate the value of \( \eta \) for every relative error bound \( \epsilon \) and then describe the line of \( \eta \) as the red forked lines in Figures 3, 4, and 5, and we call this reliability as RODFA reliability.

Figures 3, 4, and 5, respectively, illustrate the statistical results from simulation experiments (statistical reliability).
and the calculation results from RODFA (RODFA reliability) for network area of 5000 m × 5000 m, 10000 m × 10000 m, 15000 m × 15000 m. In Figure 3, the blue dashed line shows that the statistical reliability for ε = 0.0225 is 0.9999, and the red forked line shows that, when the RODFA reliability η of Šum(S₁) reaches 0.9999, the relative error bound ε between Šum(S₁) and Sum(S₁) will be 0.0907, and the difference between 0.0225 and 0.0907 is 0.0682. Furthermore, the blue dashed line also demonstrates that, for the network with 5000 m × 5000 m size, the maximum of the relative error bound ε between Šum(S₁) and Sum(S₁) from the network simulation results is 0.0234. In Figure 4, it can be found that when the statistical reliability of Šum(S₁) (shown as the blue dashed line) reaches 0.9999, ε is equal to 0.0367; and the calculation results of RODFA demonstrate that, when RODFA reliability η reaches 0.9999, the relative error bound ε will be 0.1379, and the difference between 0.0367 and 0.1379 is 0.1012. In addition, the blue dashed line also shows that the maximum of the relative error bound ε between Šum(S₁) and Sum(S₁) from the network simulation results is 0.0413. In Figure 5, when the statistical reliability of Šum(S₁) reaches 0.9999, the relative error bound ε is equal to 0.0508; and the red forked line shows that, when RODFA reliability η of Šum(S₁) reaches 0.9999, the relative error bound ε will be 0.1768, and the difference between 0.0508 and 0.1768 is 0.126. Furthermore, the blue dashed line also demonstrates that the maximum of the relative error bound ε in the network simulation results is 0.0513.

After describing the comparative analysis of each figure separately, we will investigate and analyze the relevant feature details shown in Figures 3, 4, and 5 later.

Firstly, the three figures show that the maximums of relative error bounds ε for blue dashed lines are all small. These numbers show that the data aggregation method in our paper has better approximating effect, and it is also corresponding to the proof that Šum(S₁) is an unbiased estimate of Sum(S₁). Secondly, there is a gap between blue dashed line and red forked line in each figure, and reasons for this phenomenon are as follows: (1) the zooming out in the formula derivation process of upper limit variance of Šum(S₁)/(Var(Šum(S₁))) ≤ ((1 − 𝑞^ω)/𝑞^ω) Sup(S₁) × Sum(S₁); (2) just considering the distance d and neglecting other factors which also affect the value of 𝑞 between two connected nodes. (3) In these three figures, due to the increase of network area, the statistical reliabilities go down (for a same relative error value) and the maximums of relative error bounds ε between Sum(S₁) and Šum(S₁) grow bigger (from 0.0234 to 0.0513, shown as the blue dashed lines in these three figures). (4) With the increase of network area, the RODFA reliability η is falling, but its changing rate is greater than that of the statistical reliability (0.0682 < 0.1012 < 0.126).

The above analysis indicates that the RODFA is validity, and η of Šum(S₁) which is calculated by RODFA has a greater likelihood to be the minimum of Šum(S₁) reliability.

6. Discussing the Factors Affecting RODFA

The aim of the former part of the experiment is to demonstrate the effectiveness of RODFA. Next we will turn RODFA embedded in our simulation network and investigate the factors affecting RODFA. We use Matlab simulator to implement a wireless sensor network. To ensure that the simulation results in this paper are correct, we use our simulator to do the same experiments (the network lifetime) of DMSTRP that use C/C++ in [23] and get the approximately same results. And
the optimal number of clusters in our simulative network is adopted according to [30].

In this section, the first group of experiments is to describe the changing trend of the value of $\eta$ of $\text{Sum}(S_j)$ (calculated by RODFA) with the increase of network running rounds. And the experiment results for different network areas are shown in Figure 6.

Due to these three lines belonging to the same type of experimental result, we will just analyze the blue curve shape in Figure 6. The blue line describes the changing trend of $\eta$ with the increase of network running rounds for network of 5000 m $\times$ 5000 m (when $\epsilon = 0.08$). As the establishment of clusters will take a certain amount of time (how long does it take is based on network area and active nodes number), that is why $\eta$ is 0 from the 0th round to the 9th round. After the sink node obtaining the values of $q$ and $\omega$ ($\omega = 3$ and $q = 0.9308$), RODFA will work out the value of $\eta$ and $\eta$ leaps to 0.9611 in the 9th round.

Based on the above analyses, the values of $q$ and $\omega$ mainly depend on network area, active nodes number, and internal structure of cluster and so on. As all of these three lines are describing the simulation results of network with certain area, thus we do not need to investigate the inference of network area on $q$ and $\omega$. From the blue line, we can find that the number of active nodes is 5000 in the first 610 rounds. But near the 400th round, clusters are remodeled, which makes the network structure changed; it also changes the values of $q$ and $\omega$. Thus, following the 409th round, $\eta$ suddenly reduces to 0.9457 and lasts for 210 rounds. Secondly, the active nodes number begins to decline from the 600th round. As setup phase and ready phase continuously cycle in the network, the values of $q$ and $\omega$ are different for different time. That leads to the result that $\eta$ has a sustained downward trend until its value becomes 0 from the 619th round. Lastly, we also find that $\eta$ is lower in the subsequent 123 rounds (from the 811th round to the 934th round). That is because the active nodes number remains at a low level in the subsequent rounds of network running, while less active nodes deploy in the network with a certain area, the value of $\omega$ will be larger and $q$ will be smaller, which makes $\eta$ small.

The second group of experiments is to investigate the influence of network size on RODFA. According to $q_{\theta/2}^2 \leq q^\omega \inf(N_j) \inf(S_j) e^2/(1 - q^\omega) \sup(S_j)$, the value of $\eta$ required given $\epsilon$, $q$, and $\omega$. Based on the above analysis, the values of $q$ and $\omega$ depend on network size when the routing protocol is certain. And as the value of $\epsilon$ is given by user’s application requirement, $\eta$ will change with the difference of network area. As shown in Figure 7, when $\epsilon$ is, respectively, set to be 0.01, 0.03, and 0.05, $\eta$ decreases with the growth of network size.

The last group of experiments is to analyze the network delay. According to the definition of delay in Section 2, we can get the delays of simulation network models with size of 5000 m $\times$ 5000 m, 10000 m $\times$ 10000 m, and 15000 m $\times$ 15000 m and describe them in Figure 8. Figure 8 shows that the network delay fluctuates up and down in 524 when the number of alive nodes is more than 1500; while when the number of alive nodes is less than 1500, the value of network delay is 0.9611 in the 9th round.

![Figure 6: $\eta$ of $\text{Sum}(S_j)$ in network area of 5000 m $\times$ 5000 m (blue), 10000 m $\times$ 10000 m (red), and 15000 m $\times$ 15000 m (green). From the blue line, we can find that the value of $\eta$ is 0 from the 0th to the 9th round, and in the 9th round, the value of $\eta$ leaps to 0.9611 and lasts for 400 rounds. In the 409th round, the value of $\eta$ suddenly reduces to 0.9457 and lasts for 210 rounds, and then there has a sudden drop of $\eta$ in the 619th round. Then $\eta$ has a sustained downward trend until its value becomes 0. The red line shows the $\eta$ of $\text{Sum}(S_j)$ in network area of 10000 m $\times$ 10000 m. We can find that, among the first 11 rounds, the value of $\eta$ is 0, and then it leaps to 0.9133 in the 11th round and lasts for 400 rounds. In the next 200 rounds (from the 311th round to the 511th round), $\eta$ is 0.8742, and in the 511th round, the value of $\eta$ reduces to 0.8504 again. After a sudden drop in the 631th round, $\eta$ has a sustained downward trend until its value becomes 0. The green line describes the $\eta$ of $\text{Sum}(S_j)$ in network area of 15000 m $\times$ 15000 m. $\eta = 0$ for the first 13 rounds, and in the 413th round, $\eta$ has a sudden drop, and then its value continues to fall until it becomes 0.

![Figure 7: The influence of network size on RODFA, where $\epsilon$, respectively, is set to be 0.01, 0.03, and 0.05. The number of nodes is 5000. The values of $\eta$ for these three $\epsilon$ are calculated while the network size increases from 5,000 to 15,000.](image-url)
delay begins to decrease until its value becomes 0. Probably this is due to the numbers of links and clusters in the network, and the value of network $\omega$ decreases with the decrease of alive nodes in the network, which lead to the changing trend of network delay shown in Figure 8.

7. Conclusions

Our proposed algorithm, RODFA, belongs to a data fusion aggregation algorithm which mainly includes aggregating sensed data from large-scale wireless sensor networks and doing fusion analysis of the aggregation results. We choose SUM to design RODFA in this paper and the main idea of RODFA is to calculate the parameter $\eta$ of $\sum(S_i)$ through auto analysis and synthesizing of $\sum(S_i)$. It can provide $\eta$ to users and facilitate them to do the next handling of $\sum(S_i)$; also it can facilitate the network administrator to improve system sensing performance. Experiments in Section 5 indicate that the parameter $\eta$, which is calculated by RODFA, has a greater likelihood to be the minimum of $\sum(S_i)$ reliability and RODFA is valid. In Section 6, we first investigate the influences of network size and network performance on data fusion aggregation to guide the network administrator on improving the system sensing performance.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The project is supported by The National Key Technology Research and Development Program of the Ministry of Science and Technology of China (Grant no. 2012BAH82F04).

References

[1] G. Huang, A. Y. Zomaya, F. C. Delicato, and P. F. Pires, "Long term and large scale time synchronization in wireless sensor networks," Computer Communications, vol. 37, pp. 77–91, 2014.
[2] A. Rozvyev, H. Hasbullah, and F. Subhan, "Indoor child tracking in wireless sensor network using fuzzy logic technique," Research Journal of Information Technology, vol. 3, no. 2, pp. 81–92, 2011.
[3] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, and D. Estrin, "Habitat monitoring with sensor networks," Communications of the ACM, vol. 47, no. 6, pp. 34–40, 2004.
[4] Z. B. Wang, W. Lou, Z. Wang, J. C. Ma, and H. L. Chen, "A hybrid cluster-based target tracking protocol for wireless sensor networks," International Journal of Distributed Sensor Networks, vol. 2013, Article ID 494863, 16 pages, 2013.
[5] O. García, A. Quintero, and S. Pierre, "A global profile-based algorithm for energy minimization in object tracking sensor networks," Computer Communications, vol. 33, no. 6, pp. 736–744, 2010.
[6] J. Zhang, G. M. Song, G. F. Qiao, Z. Li, and A. M. Wang, "A wireless sensor network system with a jumping node for unfriendly environments," International Journal of Distributed Sensor Networks, vol. 2012, Article ID 568240, 8 pages, 2012.
[7] N. Sabri, S. A. Aljunid, B. Ahmad, A. Yahya, R. Kamaruddin, and M. S. Salim, "Wireless sensor actor network based on fuzzy inference system for greenhouse climate control," Journal of Applied Sciences, vol. 11, no. 17, pp. 3104–3116, 2011.
[8] D. Kumar, "Monitoring forest cover changes using remote sensing and GIS: a global prospective," Research Journal of Environmental Sciences, vol. 5, pp. 105–123, 2011.
[9] I. Bekmezci and F. Alagoz, "Delay sensitive, energy efficient and fault tolerant distributed slot assignment algorithm for wireless sensor networks under convergecast data traffic," International Journal of Distributed Sensor Networks, vol. 5, no. 5, pp. 557–575, 2009.
[10] Y. L. Chen and H. P. Lai, "A fuzzy logical controller for traffic load parameter with priority-based rate in wireless multimedia sensor networks," Applied Soft Computing, vol. 14, pp. 594–602, 2014.
[11] Y.-C. Tseng, M.-S. Pan, and Y.-Y. Tsai, "Wireless sensor networks for emergency navigation," Computer, vol. 39, no. 7, pp. 55–62, 2006.
[12] A. Deligiannakis, Y. Kotidis, and N. Roussopoulos, "Processing approximate aggregate queries in wireless sensor networks," Information Systems, vol. 31, no. 8, pp. 770–792, 2006.
[13] Y.-C. Fan and A. L. P. Chen, "Efficient and robust sensor data aggregation using linear counting sketches," in Proceedings of the 22nd IEEE International Parallel and Distributed Processing Symposium (IPDPS ’08), pp. 1–12, April 2008.
[14] R. Tan, G. L. Xing, Z. H. Yuan, X. Liu, and J. G. Yao, "System-level calibration for data fusion in wireless sensor networks," ACM Transactions on Sensor Networks, vol. 28, no. 93, pp. 1–28, 2013.
[15] J. Elias, "Optimal design of energy-efficient and cost-effective wireless body area networks," Ad Hoc Networks, vol. 13, part B, pp. 560–574, 2014.
[16] Y.-L. Chen and J.-S. Lin, "Energy efficiency analysis of a chain-based scheme via intra-grid for wireless sensor networks," Computer Communications, vol. 35, no. 4, pp. 507–516, 2012.
[17] Y. Liu, Q. Zhang, and L. Ni, “Opportunity-based topology control in wireless sensor networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 3, pp. 405–416, 2010.

[18] M. Zuniga and B. Krishnamachari, “Analyzing the transitional region in low power wireless links,” in *Proceedings of the 1st Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (IEEE SECON ’04)*, pp. 517–526, October 2004.

[19] T. C. Wang, X. L. Qin, and L. Liu, “An energy-efficient and scalable secure data aggregation for wireless sensor networks,” *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 843485, 11 pages, 2013.

[20] P.-J. Wan, S. C.-H. Huang, L. Wang, Z. Wan, and X. Jia, “Minimum-latency aggregation scheduling in multihop wireless networks,” in *Proceedings of the 10th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc ’09)*, pp. 183–194, May 2009.

[21] P. K. Varshney, *Distributed Detection and Data Fusion*, Springer, 1996.

[22] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-efficient communication protocol for wireless microsensor networks,” in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, p. 223, January 2000.

[23] H. Guangyan, L. Xiaowei, and H. Jing, “Dynamic minimal spanning tree routing protocol for large wireless sensor networks,” in *Proceedings of the 1st IEEE Conference on Industrial Electronics and Applications (ICIEA ’06)*, pp. 1–5, Singapore, May 2006.

[24] S. D. Muruganathan, D. C. F. Ma, R. I. Bhasin, and A. O. Fapojuwo, “A centralized energy-efficient routing protocol for wireless sensor networks,” *IEEE Communications Magazine*, vol. 43, no. 3, pp. S8–S13, 2005.

[25] P. Fazio, F. de Rango, and C. Sottile, “A new interference aware on demand routing protocol for vehicular networks,” in *Proceedings of the International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS ’11)*, pp. 98–103, June 2011.

[26] S. M. He, *Research on the Key Technology of Opportunistic Routing in Multi-Radio Multi-Channel Wireless Mesh Network*, Hunan University, Hunan, China, 2013.

[27] S. Bernstein and R. Bernstein, *Elements of Statistics II: Inferential Statistics*, McGraw-Hill, Columbus, Ohio, USA, 1st edition, 2004.

[28] H. Fischer, *A History of the Central Limit Theorem: From Classical to Modern Probability Theorem*, Springer, New York, NY, USA, 1st edition, 2011.

[29] Z. Shen, S. Q. Xie, and C. Y. Pan, *Probability Theory & Mathematical Statistics*, Higher Education Press, Beijing, China, 2005.

[30] S. Lindsey, C. Raghavendra, and K. M. Sivalingam, “Data gathering algorithms in sensor networks using energy metrics,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, no. 9, pp. 924–935, 2002.
Submit your manuscripts at http://www.hindawi.com