EURASIP Journal on
Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH Open Access

Energy-efficient data acquisition for accurate signal estimation in wireless sensor networks

Xun Li[1*], Geoff V Merrett[2] and Neil M White[2]

Abstract

Long-term monitoring of an environment is a fundamental requirement for most wireless sensor networks. Owing to the fact that the sensor nodes have limited energy budget, prolonging their lifetime is essential in order to permit long-term monitoring. Furthermore, many applications require sensor nodes to obtain an accurate estimation of a point-source signal (for example, an animal call or seismic activity). Commonly, multiple sensor nodes simultaneously sample and then cooperate to estimate the event signal. The selection of cooperation nodes is important to reduce the estimation error while conserving the network's energy. In this paper, we present a novel method for sensor data acquisition and signal estimation, which considers estimation accuracy, energy conservation, and energy balance. The method, using a concept of 'virtual clusters,' forms groups of sensor nodes with the same spatial and temporal properties. Two algorithms are used to provide functionality. The 'distributed formation' algorithm automatically forms and classifies the virtual clusters. The 'round robin sample scheme' schedules the virtual clusters to sample the event signals in turn. The estimation error and the energy consumption of the method, when used with a generalized sensing model, are evaluated through analysis and simulation. The results show that this method can achieve an improved signal estimation while reducing and balancing energy consumption.

Keywords: Wireless sensor networks; Spatial correlation; Virtual cluster; Energy balance

1 Introduction

Wireless sensor networks (WSNs) are continuing to attract significant interest from the research community, with the promise of revolutionizing a wide range of application domains including environmental, building, and industrial process monitoring [1]. A WSN consists of multiple sensor nodes is deployed carefully or arbitrarily over a given field. A sensor node typically comprises four parts: one or more sensors, a microcontroller, a wireless transceiver, and a power source. Batteries are commonly used to power nodes in a WSN deployment but have a finite energy budget. When the battery is depleted, a node cannot perform its function or participate in packet routing, which can isolate large areas of the network [2]. Charging or replacing batteries may be expensive and is difficult or even impossible under many circumstances. While harvesting, environmental energy is becoming increasingly realizable, from which the power obtained is often unstable and typically provides an average power in the order of microwatts [3]. Hence, energy-efficient operation continues to be of considerable importance to the WSN design and operation, both conserving energy in an individual sensor node and also balancing energy consumption evenly across the networks. Most research on WSN energy conservation has focused on communication [4]. However, in some applications, the sensors themselves can consume more energy than that of the communication [5], and some techniques have been proposed to manage the energy in these situations [6].

Sensing applications in WSNs can be crudely classified into two categories: mapping the distribution of a parameter over an area (a common requirement of building control or environmental monitoring networks [7]) and monitoring a particular 'point-source' signal (for example, a seismic signal [8]). In this paper, we focus on the latter, where nodes are deployed at fixed positions and cooperate to sample and estimate the event signal. Before arriving at each sensor node, the signal is usually attenuated, delayed, and distorted by noise. In order to make accurate measurements, signal estimation error must be

*Correspondence: lixun@nudt.edu.cn
[1] School of Mechatronic Engineering and Automation, National University of Defense Technology, Changsha 410073, China
Full list of author information is available at the end of the article

taken into consideration in WSN design. Consistent with similar research [9,10], mean square error (MSE) is used as the index to illustrate the accuracy of estimation in this situation.

Arguably, the most effective method for conserving energy is to make nodes 'sleep' (where they enter a low-power sleep state) whenever possible. Signal estimation error can be reduced by only fusing data from a subset of the sensor nodes, but the estimation will be affected by spatial and temporal correlation between nodes. Through these two factors, at any one time, the selection of cooperation nodes will affect both the energy conservation and the estimation accuracy. In this paper, we propose a novel method to select and schedule suitable sensor nodes for an energy-efficient accurate signal estimation. The method uses a new concept, referred to as the virtual cluster (VC), which is a logical grouping of sensor nodes (irrespective of any communication clusters imposed by the network topology) with the same spatial and temporal properties. Two new algorithms, DF-VC (distributed formation of virtual clusters) and RRSS (round robin sample scheme), are proposed.

This paper is organized as follows. Section 2 discusses the related work in this area and highlights the improvements and benefits of the proposed method. Section 3 presents a model for data acquisition, which is subsequently analyzed in Section 4. Sections 5 and 6 provide details of the DF-VC and RRSS algorithms, respectively. Finally, Section 7 evaluates the proposed method through simulations, and Section 8 draws conclusions and identifies the prospects for future work.

2 Related work

Energy-efficient and accurate event signal estimation is an area of research that has drawn significant attention. Ribeiro and Giannakis [11] discussed distributed estimation in WSNs and introduced a class of maximum-likelihood estimators under a bandwidth constraint. In their method, each sensor transmits 1 bit per observation; hence, the energy consumed on communication is conserved while the estimation accuracy can be guaranteed through the Cramer-Rao lower bound. Zhi-Quan and Jin-Jun [12] presented a decentralized estimation scheme, where each sensor compressed its observation into a small number of bits with a length proportional to the logarithm of its local signal-to-noise ratio. Their scheme can guarantee a maximum estimation error. Due to the redundancy in sensor data, compression is another technique used to conserve energy. Based on the information theory, Pradhan et al. [13] reported a distributed compression framework to remove redundancy. The main task of the above works is to reduce the communication data while maintaining estimation accuracy. However, the selection and organization of suitable sensor nodes are ignored in the task. The behavior of sensor nodes has critical effects on the estimation and energy conservation.

One characteristic of WSNs is that nodes can be densely deployed. Adjacent sensor nodes are often highly spatially and temporally correlated. This correlation can explicitly affect the signal estimation error during signal fusion. By modeling the event and the noise as Gaussian random signals, Vuran et al. [9] analyzed spatial and temporal correlations in WSNs. Two separate distortion functions, representing the spatial and temporal errors, were derived. The spatial distortion function shows that the number of cooperation nodes and the spatial correlation among sensor data are the main factors to affect the estimation error. In the temporal distortion function, the sample frequency is the key factor to affect the estimation error. In fact, the estimation error cannot be independently separated into the spatial and temporal error, as some factors are coupled with each other. Vuran and Akan [14] extended the above work by presenting a uniform distortion function. Using an attenuated and delayed sensing model and incorporating signal propagation delays, their work closer resembled the properties of real environments and applications. Both of these papers conclude that a finite number of sensor nodes can cooperate to reduce the estimation error. However, techniques or algorithms to identify and manage these sets of suitable nodes are not addressed.

Karjee and Jamadagni [15] analyzed the estimation accuracy of clustered WSNs using the same correlation model and method presented by Vuran et al. [9]. They showed through simulations that a subset of sensor nodes could satisfy a given requirement of estimation accuracy. However, they found that the selected sensor nodes must be concentrated around the location of the event. In their later work [10], Karjee and Jamadagni further studied the selection of the sensor nodes for accurate estimation. Based on the distributed clusters, they developed a probabilistic model for each distributed cluster to pursue the estimation accuracy and reduce the energy consumption in the network. Vuran and Akyildiz [16] proposed a method to obtain a minimum number of sensor nodes, which satisfy a given estimation accuracy requirement. Each selected sensor node 'represented' a group of sensor nodes, while the remainder of nodes in the group were put into a sleep state. The representative nodes were selected by solving the distortion function at the sink node, which also performed the signal estimation. Sensor nodes in a group were adjacent to each other, and hence, representative nodes were spatially spread out (and hence not concentrated around the location of the event). Through this method, the authors reduced the amount of collected data and hence reduced the energy consumption of the network. In the methods presented above, the selected nodes will consume more energy than that of the unselected nodes. Hence, these methods do not provide a balanced approach.

Gedik et al. [17] adpoted an adaptive approach to data acquisition called ASAP, in which a cluster of sensor nodes is partitioned into smaller 'sub-clusters'. One or more nodes in each sub-cluster were selected to sense and report the event, while the other nodes were put into a sleep state. Data from the 'sleeping' nodes were predicted using probabilistic models. Willett et al. [18] presented an adaptive data acquisition scheme which was called 'backcasting'. The authors studied a scenario with n sensor nodes arranged regularly on a square lattice. A small subset of the nodes sample the environment and disseminate data to the sink node. This initial estimation is processed by the sink node and 'backcast' into the network. Additional nodes in the network then sample in order to refine the initial estimation and meet desired error targets. Both ASAP and 'backcasting' reduce the energy consumption of the network while maintaining a desired data accuracy. However, the methods are not balanced on energy consumption, and the dynamic adjustment of suitable sensor nodes will consume extra energy.

Except the spatial correlation, the selection of suitable sensor nodes should also consider the temporal correlation. The temporal correlation between nodes is affected significantly by sampling frequency which can be adjusted by the adaptive sampling schemes as the main approaches to address this case according to the applications [4]. Alippi et al. [19] presented an adaptive algorithm which could automatically adjust sampling frequency to a suitable value in real-time, based upon the Nyquist-Shannon theorem. By reducing over-sampling, the method can conserve the energy consumed by eliminating redundant data. However, the cooperation of sensor nodes is not considered.

For a group of sensor nodes, a comprehensive sampling scheme can reduce energy consumption of a WSN. Minglei and Hen [20] proposed a sampling scheme called collaborative sampling. In this scheme, the sensor nodes which broadcast their samples to the fusion center are selected according to the localized estimation error. Jing et al. presented a scheme referred to as asynchronous sampling [21], which shifted sampling times of sensor nodes in order to reduce the sampling rate for each sensor node. The above two methods showed that energy can be conserved by the cooperation of sensor nodes; however, these two papers need further consideration on both the accurate estimation and energy consumption and balance.

In our previous paper [22], we discussed the formation of sensor nodes for signal estimation; however, it also needs further research on the analysis and validation of the scheme, and the comparison with different schemes was not addressed. The novel aspects of the work presented in this paper, over the existing works, are the following:

- Firstly, unlike the existing works, we analyze the effect of time synchronization error and transmission delay on estimation error. The analysis shows that the section of suitable sensor nodes should consider these effects in order to get an accurate estimation.
- Secondly, an improved trade-off between the estimation accuracy and the energy conservation is obtained. By adjusting the algorithm's parameters, the estimation accuracy and the level of energy conservation can be adjusted to satisfy the user requirements.
- Thirdly, in addition to being conserved, energy is also balanced among nodes in the network. The balance of energy consumption is especially crucial in the dense sensor networks.
- Fourthly, the algorithms are low computation, and the majority of the process is completed at the start of application deployment. Hence, the dynamic adjustment of sensor nodes is avoided.

3 A model for data acquisition

3.1 Network architecture

A WSN may be deployed as a flat or hierarchical architecture [23]. In a hierarchical architecture adopted here, all sensor nodes are separated into communication clusters. All sensor nodes in a cluster communicate only with the cluster head (CH), which is usually performed as a resource-unconstrained node and able to aggregate data. A model for data acquisition, which forms the basis of the motivation and analysis of the work presented in this paper, forms the remainder of this section.

All of the network's sensor nodes, represented by a node set G, are in one of L communication clusters. Let C_i ($i = 1, 2, \dots L$) denote the i^{th} cluster. The CH of C_i is ch_i and the number of the sensor nodes (including ch_i) in C_i is L_i. Hence, C_i is depicted as

$$C_i = \{ch_i, n_j | n_j \in G, j \in [1, Z], j \neq i\}, \tag{1}$$

and

$$\sum_{i=1}^{L} L_i = Z,$$

where Z is the number of sensor nodes in G.

3.2 Spatial and temporal correlation among sensor data

A typical monitoring application, where a WSN is deployed over a field to sense a physical phenomenon, is considered. The event signal is denoted by $S(t)$ ($t \geq 0$). Signal $S(t)$ is modeled as a continuous signal with a limited frequency bandwidth and a normally distributed amplitude (with a zero mean and variance of σ_s^2). The position of $S(t)$ is $s\,:\,(x,y)$. The i^{th} sensor node, referred to as n_i $(i = 1, 2, \ldots, Z)$, is at the position $s_i\,:\,(x_i, y_i)$. At any observing location, the event signal is usually attenuated in relation to the distance between the observing location and the event.

A sensing model is defined to depict the observed signal at the position of each sensor node. Node n_i samples $S(t)$ as $S_i(t)$ after being attenuated and delayed,

$$S_i(t) = S(t - \delta_{is})/\rho_i. \tag{2}$$

Signal $S(t)$ propagates with velocity v; it will arrive at the position of node n_i after time lag $\delta_{is} = d_i/v$. The attenuation ratio, ρ_i, is calculated by

$$\frac{1}{\rho_i} = \frac{\theta_1^2}{\theta_2^2 + d_i^2}, \quad \theta_1, \theta_2 > 0, \tag{3}$$

where d_i is the Euclidean distance between the event and n_i. This quadratic model is commonly used to depict spatially correlated sensor data [24]. And as an attenuation model, it is widely used for attenuation of RF and acoustic signals [25,26].

$$d_i = ||s_i - s|| = \sqrt{(x_i - x)^2 + (y_i - y)^2}. \tag{4}$$

Signal $S_i(t)$ is a Gaussian random signal,

$$E[S_i(t)] = 0 \quad \text{and} \quad E[S_i^2(t)] = \sigma_s^2/\rho_i^2. \tag{5}$$

The correlation coefficient between $S_i(t)$ and $S_j(t)$ is

$$\begin{aligned}\mathrm{Corr}[S_i(t), S_j(t)] &= \frac{E[S_i(t)S_j(t)]}{\sqrt{\mathrm{Var}[S_i(t)]}\sqrt{\mathrm{Var}[S_j(t)]}}\\ &= \frac{E[S(t-\delta_{is})S(t-\delta_{js})]}{\sigma_s^2}.\end{aligned} \tag{6}$$

Equation 6 shows that the correlation coefficient between two sensor data is related to the time difference between them. The power exponential model [24] is used to depict the temporal correlation of two sensor data.

$$\begin{gathered}\mathrm{Corr}\{S(t), S(t+\Delta)\} = \frac{E[S(t)S(t+\Delta)]}{\sigma_s^2} = \vartheta(\Delta).\\ \vartheta(\Delta) = e^{-\Delta/\theta}, \quad \theta > 0.\end{gathered} \tag{7}$$

By combining (6) and (7), the coefficient is calculated as

$$\mathrm{Corr}[S_i(t), S_j(t)] = \vartheta(\delta_{is} - \delta_{js}). \tag{8}$$

Equation 8 shows that the correlation between two different sensor data is only related with temporal factors. Therefore, if the time difference is ignored, the correlation coefficient will be one. Hence, for a given sensing model, a spatial correlation model is not necessitated for sensor data.

The parameters, δ_{is} and ρ_i, are both determined by the distance d_i. In this study, it is assumed that an accurate measurement of distance is available at each node, and the effect of any error on the estimation accuracy is ignored. Many techniques and algorithms in previous study have been applied to make measurements for such distances [27,28].

3.3 Sampling of a sensor node

A sensor node samples the attenuated and delayed signal with noise. The noise at node i, depicted by $e_i(t)$, is modeled as an independent and identically distributed Gaussian variable with a zero mean and variance of σ_e^2. The signal sampled by n_i is depicted as shown in (9), where $e_i(t)$ is independent from the event signal.

$$X_i(t) = S_i(t) + e_i(t) \tag{9}$$

Each sensor node samples signal with same frequency. The sampling can be depicted by a sampling function $p(t)$.

$$p(t) = \sum_{k=0}^{+\infty} \delta(t - kT), \tag{10}$$

$$\delta(t - kT) = \begin{cases} 0 & t \neq kT, \\ 1 & t = kT, \end{cases} \tag{11}$$

where T is the sample period. By using $p(t)$ to sample $X_i(t)$, the sample sequence of $X_i(t)$ is derived as

$$\begin{cases} X_i^p(t) & \triangleq X_i(t)p(t), \\ X_i(k) & \triangleq X_i^p(kT) = S_i(k) + e_i(k), \\ S_i(k) & \triangleq S_i(kT), \\ e_i(k) & \triangleq e_i(kT). \end{cases} \tag{12}$$

The sensing and sampling process is illustrated by Figure 1.

At the observation location, the original signal $S(t)$ becomes $S_i(t)$ after delayed and attenuated through the coefficients d_i and ρ_i. In fact, the sensor node only senses the signal $X_i(t)$, i.e., $S_i(t)$ with noise $e_i(t)$. Signal $S_i(k)$ cannot be derived in the real environment. The sample of signal $X_i(t)$ is $X_i(k)$ at time instant kT. The sampling period T can be derived through adaptive sampling methods (for example, the algorithm proposed by Alippi et al. [19]). Signal $X_i(t)$, which is an approximation of $S_i(t)$, can be recovered from the discrete sample $X_i(k)$. A copy of signal $S(t)$ with time lag δ_{is} can be approximately derived by $X_i(t)\rho_i$.

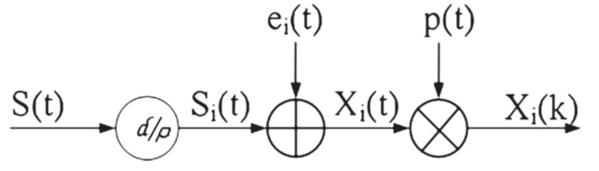

Figure 1 Sensing and sampling process of a sensor node.

4 Signal estimation and error analysis

During a signal estimation process, some factors affecting the estimation error are highlighted through the error analysis. In this section, we analyze the estimation error and present two lemmas to select suitable nodes. The impacts of communication delay and the time synchronization error are also discussed.

4.1 Estimating the event signal

Each sensor node may obtain an estimation of the event signal. For convenience, signal $S_i(k)$ is denoted by S_i. The estimation of S made by node n_i is given by

$$Y_i = X_i \rho_i. \tag{13}$$

The estimation Y_i is sent to the fusion node in which S is estimated in each period. The quantized sample mean estimator is used in fusion node.

$$\bar{S} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \frac{1}{N}\sum_{i=1}^{N}(S_i + e_i)\rho_i \tag{14}$$

The estimation accuracy is evaluated through the MSE.

$$E[(S - \bar{S})^2] = f(N, \sigma_s) + g(N, \sigma_e), \tag{15}$$

$$f(N, \sigma_s) = \sigma_s^2 + \frac{\sigma_s^2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\vartheta_{ij} - \frac{2\sigma_s^2}{N}\sum_{i=1}^{N}\vartheta_{is}, \tag{16}$$

$$g(N, \sigma_e) = \frac{\sigma_e^2}{N^2}\sum_{i=1}^{N}\rho_i^2, \tag{17}$$

where $\vartheta_{ij} = e^{-\delta_{ij}/\theta}$ is the correlation coefficient of nodes n_i and n_j, and $\vartheta_{is} = e^{-\delta_{is}/\theta}$ is the correlation coefficient of node n_i and the event signal.

Function $f(N, \sigma_s)$ can be treated as a distortion function of the systematic error which is affected by N, δ_{ij}, and δ_{is}. Parameters δ_{ij} and δ_{is} are determined by the position of the sensor nodes and represent the impact of the temporal factors.

$$\delta_{ij} = \delta_{is} - \delta_{js} = d_i/v - d_j/v = (d_i - d_j)/v \tag{18}$$

Function $g(N, \sigma_e)$ can be thought as a distortion function of the random error. The value of $g(N, \sigma_e)$ is determined by N and the attenuation ratio.

4.2 Analysis of estimation error

The estimation error includes two parts: systematic error and random error. Optimizing the distortion functions to derive the minimum error is a non-trivial task. However, through some approximations, either the systematic or random error can be controlled to a small value. The approximations are illustrated as Lemmas 1 and 2.

Lemma 1. *If the sensor nodes which join the estimation process are close to the event, the systematic error will not relate to the number of sensor nodes and can be ignored.*

Proof. According to the requirement of the Lemma 1, the following approximation can be derived.

$$\delta_{is} = d_i/v \approx 0,$$
$$\delta_{ij} = (d_i - d_j)/v \approx 0$$

Hence,

$$\delta_{ij} = e^{-\delta_{ij}/\theta} \approx 1,$$
$$\delta_{is} = e^{-\delta_{is}/\theta} \approx 1.$$

With (16),

$$f(N, \sigma_s) = \sigma_s^2 + \frac{\sigma_s^2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\vartheta_{ij} - \frac{2\sigma_s^2}{N}\sum_{i=1}^{N}\vartheta_{is} \approx \sigma_s^2 + \sigma_s^2 - 2\sigma_s^2 = 0.$$

□

Lemma 1 presents an important limit on the selection of sensor nodes. All selected sensor nodes should be close to the event. Let d_{th}^{p} denote a desired distance threshold, node n_i can join the estimation process under the condition of $d_i \leq d_{th}^{p}$.

The random error cannot be ignored; however, if the suitable sensor nodes are selected to participate in the estimation process, $g(N, \sigma_e)$ can be reduced.

Lemma 2. *Assume that N adjacent sensor nodes have joined the estimation process. If the $(N+1)^{th}$ sensor node satisfies*

$$\rho_{N+1} < \frac{\sqrt{2N+1}}{N} \times \sqrt{\sum_{i=1}^{N}\rho_i^2},$$

the random error will be reduced by adding the $(N+1)^{th}$ sensor node to join the estimation process.

Proof. According to (17),

$$\begin{aligned}
\Delta g_{N+1}^{N} &= g(N+1, \sigma_e) - g(N, \sigma_e) \\
&= \sigma_e^2\left[\sum_{i=1}^{N+1}\left(\frac{\rho_i}{N+1}\right)^2 - \sum_{i=1}^{N}\left(\frac{\rho_i}{N}\right)^2\right] \\
&= \sigma_e^2\left[\sum_{i=1}^{N}\left(\frac{\rho_i}{N+1}\right)^2 - \sum_{i=1}^{N}\left(\frac{\rho_i}{N}\right)^2 + \left(\frac{\rho_{N+1}}{N+1}\right)^2\right] \\
&= \frac{\sigma_e^2}{(N+1)^2}\left[\rho_{N+1}^2 - (2N+1)\sum_{i=1}^{N}\left(\frac{\rho_i}{N}\right)^2\right].
\end{aligned}$$

Under the restriction of Lemma 2,

$$\rho_{N+1}^2 - (2N+1)\sum_{i=1}^{N}\left(\frac{\rho_i}{N}\right)^2 < 0.$$

Hence,

$$g(N+1,\sigma_e) < g(N,\sigma_e).$$

□

From Lemma 2, two conclusions can be drawn. First, adjacent sensor nodes will always improve the accuracy. We can think that the attenuated coefficients of the adjacent sensor nodes are approximated as the same ρ.

$$\sqrt{2N+1}\sqrt{\sum_{i=1}^{N}\left(\frac{\rho_i}{N}\right)} \approx \rho\sqrt{\frac{2N+1}{N}} > \sqrt{2}\rho. \tag{19}$$

Second, a finite number of sensor nodes are sufficient for signal estimation. Due to the additional energy consumption, utilizing further sensor nodes to improve the accuracy is not worthwhile. The reduction of the error when an adjacent node is added is given by

$$|\Delta g_{N+1}^{N}| \approx \frac{\sigma_e^2}{(N+1)^2}\left|\rho^2-(2N+1)\sum_{i=1}^{N}\left(\frac{\rho}{N}\right)^2\right| = \frac{\sigma_e^2\rho^2}{(N+1)N}. \tag{20}$$

A parameter, d_{th}^{d}, is used to depict the adjacency of the sensor nodes. For any two nodes which join the estimation process, n_i and n_j, the distance between them should satisfy $d_{ij} \leq 2d_{th}^{d}$.

4.3 Effect of communication delay

Sensor data are transmitted to a fusion node, and the end-to-end communication delay will affect the estimation process. Witrant et al. [29] stated that the delay is random and increases with an increase number of hops between the sensor node and the fusion node. Their experiment was based on the Breath protocol using Tmote nodes. They found that the average end-to-end delay of Breath protocol is 200 ms over four hops and 60 ms over two hops. If the k^{th} sample of one sensor node cannot arrive the fusion node in the current sampling period, the sample of the sensor node will not join the k^{th} fusion process. Hence, the sample will be dropped, even if it can arrive at the fusion node after the current sampling period. Therefore, the impact of the end-to-end delay on the estimation process relates to the sampling period. If the delay is more than the period T, the sensor data cannot be gathered by the fusion node. The number of sensor data which are fused together according to (14) will not be enough and affects the estimation error.

Let $P_i(\Gamma_i \leq t)$ denote the probability that the data transmission time Γ_i of node n_i is less than t. The expected number of samples arriving at the fusion node from N sensor nodes in interval T is M.

$$M = \sum_{i=1}^{N} P_i(\Gamma_i \leq T) \tag{21}$$

The systematic error is calculated as

$$f(M,\sigma_s) = \sigma_s^2 + \frac{\sigma_s^2}{M^2}\sum_{i=1}^{N}\sum_{j=1}^{N} P_i(\Gamma_i \leq T)P_j(\Gamma_j \leq T)\vartheta_{ij} - \frac{2\sigma_s^2}{M}\sum_{i=1}^{N} P_i(\Gamma_i \leq T)\vartheta_{is}. \tag{22}$$

If the requirement of Lemma 1 is satisfied, then,

$$f(M,\sigma_s) \approx \sigma_s^2 + \frac{\sigma_s^2}{M^2}\times M^2 - \frac{2\sigma_s^2}{M}\times M = 0. \tag{23}$$

Therefore, the systematic error is not affected by the communication delay under the restriction of Lemma 1.

The random error is calculated as the following equation when the delay time is considered,

$$g(M,\sigma_e) = \frac{\sigma_e^2}{M^2}\sum_{i=1}^{N} P_i^2(\Gamma_i \leq T)\rho_i^2 \neq g(N,\sigma_e). \tag{24}$$

The error will increase when M decreases. The sampling period is determined by an application, and the increase of it is not suitable in many scenarios (particularly for energy efficiency). Reducing the delay is therefore the only method available to control error. As the number of hops is the predominant factor affecting the delay, the fusion node should be selected from adjacent sensor nodes. Again, this requirement shows the necessity of a distance threshold d_{th}^{d}.

The transmission time of adjacent sensor nodes can be taken equally, that is, $\Gamma_i = \Gamma$, $P_i(\Gamma_i \leq T) = P(\Gamma \leq T)$. Then,

$$\begin{aligned} g(M,\sigma_e) &= \frac{\sigma_e^2}{M^2}\sum_{i=1}^{N} P_i^2(\Gamma_i \leq T)\rho_i^2 \\ &= \frac{\sigma_e^2}{N^2}\frac{P^2(\Gamma \leq T)N^2}{M^2}\sum_{i=1}^{N}\rho_i^2. \end{aligned}$$

Due to

$$\frac{P^2(\Gamma \leq T)N^2}{M^2} = \frac{\left(\sum_{i=1}^{N} P(\Gamma \leq T)\right)^2}{\left(\sum_{i=1}^{N} P(\Gamma \leq T)\right)^2} = 1,$$

hence, we obtain

$$g(M,\sigma_e) = \frac{\sigma_e^2}{N^2}\sum_{i=1}^{N}\rho_i^2 = g(N,\sigma_e).$$

If the selection of sensor nodes satisfies the restrictions of Lemmas 1 and 2, then the impact of communication delay can be ignored.

4.4 Effect of time synchronization error

Sensor nodes sample a signal and add a timestamp of their local time, for which the fusion node uses the timestamp to process sensor data. The timestamp is

$$t_i^k = t_0 + \alpha_i + (k-1) \times T, \quad k = 1, 2, \ldots, \tag{25}$$

where α_i is the synchronization offset. In the previous analysis, we ignored the time synchronization error; hence, the time difference is calculated as depicted in (18). When the impact of time synchronization error is considered, (18) is revised as

$$\begin{aligned} \delta_{is} &= d_i/v + \alpha_i, \\ \delta_{ij} &= (d_i - d_j)/v + (\alpha_i - \alpha_j). \end{aligned} \tag{26}$$

The propagation delay is ignored if the sensor node is adjacent to the event location. Therefore, only the synchronization error is considered in the following discussion. The systematic error relates to the temporal correlation,

$$f(N, \sigma_s) = \sigma_s^2 + \frac{\sigma_s^2}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} e^{-(\alpha_i - \alpha_j)/\theta} - \frac{2\sigma_s^2}{N} \sum_{i=1}^{N} e^{-\alpha_i/\theta}. \tag{27}$$

Equation 27 shows that the relative and absolute time synchronization error all affect the estimation error. As the magnitude of the relative synchronization error doubles the magnitude of the absolute error, mitigating the relative error is more important.

Time synchronization error varies with different protocols, which usually correlate with the hops between the sensor node and the time base station. For the lightweight time synchronization protocol, the variance of synchronization error is $4h\sigma^2$, where h is the number of hops from the sensor node to the time base station and σ is the standard variance of point to point time difference. By using a testbed of 'COTS MOTES,' which is a narrow band radio and sensor platform developed by Warneke et al. [30], σ is estimated about 11.1 *us* [31]. The error will gracefully increase with an increase in the number of hops when the timing-synchronization protocol for sensor networks (TPSN) is used [32]. Ganeriwal et al. presented a prototype system that they builta round Berkeley Motes to implement TPSN. They reported that the average synchronization error was 16.9 μs over a single hop, while the error was 23 μs over five hops. If the selection of sensor nodes complies with Lemma 2 and a suitable synchronization protocol is implemented, the relative synchronization error is likely to be minimal.

5 Distributed algorithm to form virtual clusters

Based on the previous section, an algorithm is required to select suitable nodes in the network while maintaining energy efficiency and fulfilling the restrictions presented in the analysis. This section presents the concept of 'virtual cluster' and describes the proposed DF-VC algorithm which is used to set them up.

5.1 Definition of virtual cluster

The discussion in Section 4 presents two distance thresholds to limit the selection range of sensor nodes. By using the thresholds as radius, we can form two circles. Figure 2 illustrates the relationship of the two circles in a part of the monitoring field.

In Figure 2, there is one large circle and multiple smaller circles. The large circle depicts the selection range of suitable sensor nodes, and the radius of this circle complies with the requirement of Lemma 1. The smaller circles represent the range of the nodes grouped into a VC, and their radii are limited by Lemma 2. The sensor nodes in

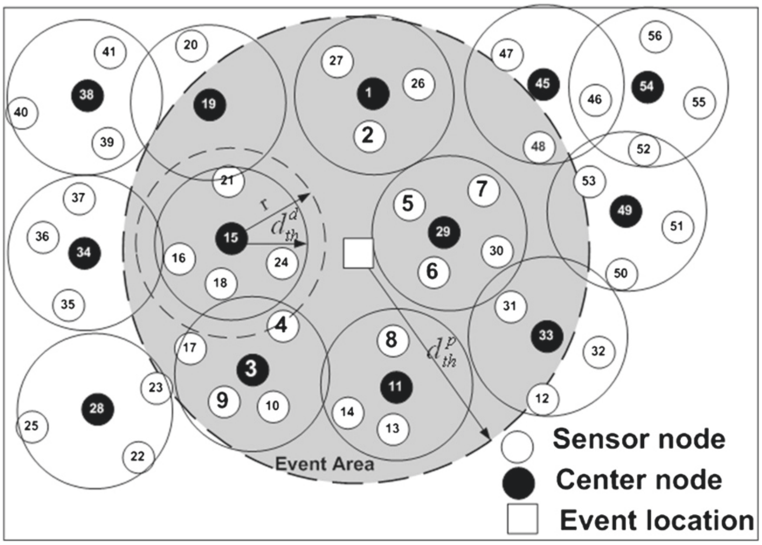

Figure 2 The topology of a network.

a small circle cooperate to sample the event signal. The fusion node in the same circle estimates the signal according to (14). A VC is a set of sensor nodes which satisfy the requirement of Lemmas 1 and 2. The j^{th} VC of communication cluster C_i, named VC_{ij}, satisfies on two conditions: if $n_k \in VC_{ij}$, then $n_k \in C_i$, $k \in [1, Z]$; and if n_k,$n_l \in VC_{ij}$, $k, l \in [1, Z]$, then $||s_k - s_l|| \leq 2d^d_{th}$.

The center node in each VC implements specific functions. Two special kinds of VCs are used in the paper: A usable VC is close enough to the event (i.e., the distance between the event location and the center node is less than d^p_{th}). A suitable VC is usable which contains the right number of nodes. Let parameters *Min_Num* and *Max_Num* denote the lower and upper limit of the number of sensor nodes, and N denote the number of sensor nodes, then, Min_Num $\leq N \leq$ Max_Num. The suitable VCs participate in sampling and estimating the event signal.

Because a suitable VC satisfies the requirement of Lemmas 1 and 2, the required accuracy can be obtained. The local fusion node in each suitable VC transmits the estimation to its CH which achieves the final estimation of the event signal. Owing to most of the transmission being located within the VCs, communication energy consumption is reduced.

5.2 Virtual clustering mechanism

The process of forming a VC is distributed. Once communication clusters have been formed, the formation of VCs begins within each communication cluster. This process should be completed before an WSN application begins.

The radius d^d_{th} and the parameter Max_Num are necessary when a VC is forming. These two parameters are derived at the sink node under the restriction of estimation error. It is assumed that the expected maximum error is D^*; accordingly, these two parameters should satisfy $E[(S - \bar{S})^2] \leq D^*$. The sink node sends these parameters to each CH, which is the first node to form a VC.

Threshold d^d_{th} should be less than r, which is the communication radius of sensor nodes. It is assumed that all sensor nodes are uniform and have the same communication radius. If $d^d_{th} > r$, a new value is set to be d^d_{th}, $d^d_{th} \leftarrow r - \Delta$, $\Delta > 0$.

Each center node knows which sensor node is in its VC according to the identification number (ID) of sensor node. The IDs may be unique to the whole network or can be local to the communication cluster.

Node n_k (at the beginning, this is CH, the communication cluster head) broadcasts a message with these parameters to other sensor nodes. After broadcasting, n_k will become a center node. Any sensor nodes from which the distance to n_k is less than r will receive the message. Node n_l, which does still not belong to any VC, will calculate the distance $||s_{n_k} - s_{n_l}||$ and compare the value with d^d_{th} after receiving the message. Node n_l will be in one of three possible states after the comparison. Three states are defined as follows:

- S1: $||s_{n_k} - s_{n_l}|| \leq d^d_{th}$,
- S2: $||s_{n_k} - s_{n_l}|| \leq r$, $||s_{n_k} - s_{n_l}|| > d^d_{th}$,
- S3: $||s_{n_k} - s_{n_l}|| > r$.

If n_l is in state $S1$, it will join the VC, which n_k creates, and reply to n_k with its ID. If n_l is in state S2 after comparison, it will set a timer T1 within a random time. Before T1 expires, n_l keeps listening to new message. It stops the timer if it can change state to S1 after having received a message. When T1 expires, n_l will broadcast a message to claim a new VC and becomes a center node itself. Because $d^d_{th} < r$, n_l can identify these parameters from the message that it received.

Initially, all sensor nodes are in state S3 and set a timer T2 within a random time (longer than the time interval of T1). Before T2 expires, sensor nodes should keep listening to any message which may change its state to S1 or S2, T2 will stop if such a state change occurs. When T2 expires, the sensor node is unable to receive any broadcast message from other nodes and does not know the necessary parameters. It therefore sends a message to inquire of these parameters. An adjacent node will receive the inquiry and reply with these parameters. Figure 3 illustrates the changes of these states.

Besides the states S1, S2, and S3, other additional states are required. State S4 is a transition state which is entered when timer T2 expires; in this state, the node will broadcast a message as a center node after it receives a reply from its inquiry. State S5 is the final 'operational' state of each sensor node, where it does not reply to any broadcast message from other nodes claiming a new VC. The exception is that the node will respond to the inquiry messages from other nodes.

Once a VC is formed, the center node maintains an ID table of its members. Every sensor node in a VC will receive parameters and commands via center node. Each center node reports its ID table to CH and receives parameters and commands from its CH. Most communications are made within VCs.

The number of nodes in each VC is adjusted through parameter Max_Num. Ideally, each VC should contain almost the same number of nodes, with which should match the requirement of Lemma 2. After broadcasting a message to claim a VC, each center node receives and then counts the replies. Once the count of the reply is equal to Max_Num, the center node rejects the replies of other nodes using a NACK message, these nodes then join or create other VCs.

Algorithm DF-VC is illustrated in Algorithm 1. Note that, although VCs are limited to a single hop, communication clusters still utilize multi-hop networking.

Algorithm 1 DF-VC algorithm

```
Initialization:
    BMC: Broadcast message to claim a VC;
    BME: Broadcast message to inquire a distance;
    RMJ: Reply message to join a VC;
    RMA: Reply message to answer an inquiry;
    $M_{ij}$: The Max_Num of the $VC_{ij}$.
Set timer T2.
Keep listening until receive a BMC
if receive a BMC then
    if timer T2 is running then
        Stop timer T2.
    end if
    Extract the parameters from the message.
    Calculate real distance to the event.
    if real distance<distance threshold then
        Stop timer T1;
        Send a RMJ, Join the VC.
    else
        Set timer T1.
    end if
end if
Timer T1 expires:
Initialize $M_{ij} = 0$;
Send a BMC, become a new center node.
Keep listening until $M_{ij}$ = Max_Num.
if receive a RMJ then
    if $M_{ij}$ < Max_Num then
        Extract node ID from the message and store it;
        $M_{ij} = M_{ij} + 1$.
    else
        Send NACK message to the sensor node,
    end if
end if
Timer T2 expires:
Initialize $M_{ij} = 0$;
Send a BME.
Keep listening until $M_{ij}$ = Max_Num.
if receive a RMA then
    Extract the parameters from the message and store
    it;
    Send a BMC, become a new center node.
end if
if receive a RMJ then
    if $M_{ij}$ < Max_Num then
        Extract node ID from the message and store it;
        $M_{ij} = M_{ij} + 1$.
    else
        Send NACK to the sensor node.
    end if
end if
```

6 Round robin sampling scheme

After VCs are formed, the selection of suitable VCs is critical for accurate estimation when an event occurs. In this section, the criteria to select suitable VCs is discussed, and the algorithms to schedule these VCs and rotate local fusion nodes are presented in this section. Combining the suitable selection, the schedule, and the rotation, energy consumption is balanced with accurate estimation and energy conservation.

6.1 Selection of suitable VCs

Although the number of sensor nodes is adjusted through parameter Max_Num, some usable VCs still may contain too few nodes. A usable VC with too few nodes impacts the estimation accuracy; hence, it should not be included in the estimation process.

Figure 4 shows the simulation results of VC formation. Figure 4a shows the number of nodes in each VC, while Figure 4b shows the respective MSE of the estimated signal. Note that, although VCs 2, 4, 7, and 10 contain the maximum number of nodes (Max_Num = 9), they are not deemed to be usable. This is because these VCs are too far from the event location (i.e., $> d_{th}^{p}$), and hence, their estimation accuracies are too low (as can be seen by the large MSE in Figure 4b).

Furthermore, the impact of the number of nodes in a VC on the estimation accuracy can also be shown. VC 9 has less nodes than VC 10 does; hence, it obtains a greater MSE than that of VC 10. To improve the accuracy, VCs with too few nodes are eliminated from the estimation process using parameter Min_Num. If Min_Num is set to be 8, from Figure 4a, we can see that VCs 1, 3, 5, 6, and 8 will be selected as the suitable VCs.

6.2 Round robin sampling scheme

From Figure 4, multiple suitable VCs can be identified. A suitable VC can derive an estimation of the event signal with the required level of accuracy. If a suitable VC is selected to work in active mode, the sensor nodes in this VC will continuously acquire sensor data. The energy of these sensor nodes will be consumed more quickly than that of those sensor nodes in the sleep state. Therefore, a method is required to balance energy consumption among the suitable VCs.

The method proposed to balance energy consumption is to alternate the active VC. Each suitable VC takes part in estimating the event signal during its time slot and falls in a sleep state at other slots. When an event occurs, the CH identifies which VCs should participate in sampling the event signal. The suitable VCs may be in different clusters; hence, a CH cannot determine the time slots independently. Every CH sends the number of VCs which should be selected to sample the signal to the sink node, which calculates the sample time for each cluster. This

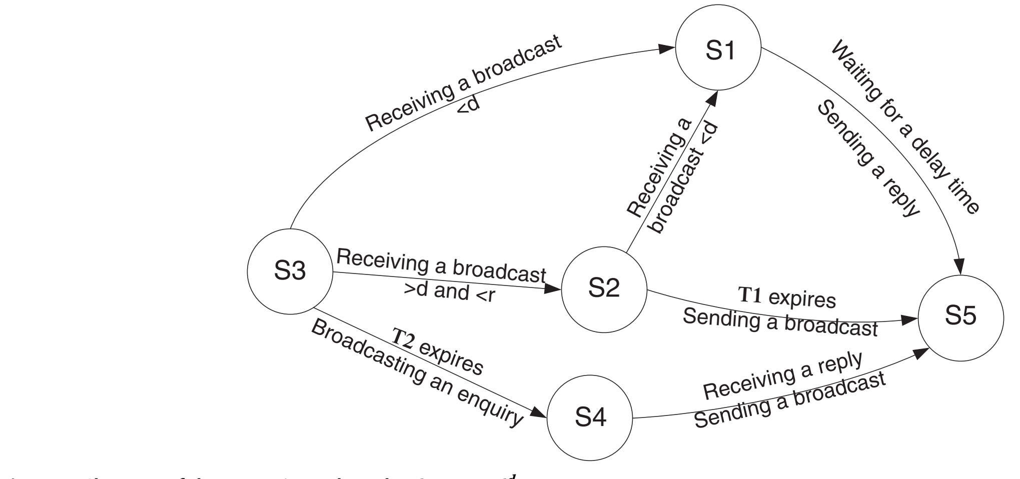

Figure 3 Changes of the states in each node, d means d^d_{th}.

information is returned to each CH, which informs the center nodes of their time slot.

Let R denote the number of suitable VCs. These VCs are ordered by their cluster and VC IDs and recorded in the sink node as $\{VC_{1,1}, VC_{1,2}, \ldots, VC_{2,1}, \ldots, VC_{ij}, \ldots\}$. Every element in the queue is indexed by a number, $index(VC_{ij}) = l,\ l \in [1, R]$. The time for VC_{ij} to begin sampling is calculated by each sensor node as

$$t_{ij} = t_0 + l \times T + T' \times \omega, \quad w = 0, 1, 2, \ldots. \tag{28}$$

T' is the period within which all suitable VCs sample and are calculated by the sink node as

$$T' = R \times T. \tag{29}$$

The sampling scheme is illustrated, as shown in Figure 5.

The algorithm used to implement this functionality is referred to as RRSS. RRSS requires only four parameters: t_0 (start time), l (index of each VC), T (sampling period), and R (number of suitable VCs). Parameters T and t_0 are derived by the sink node according to the application requirements, and R is derived by counting the number of suitable VCs. The sink node creates an initial index for each cluster, while each CH calculates the index for each of its VCs. The algorithm has three stages to implement its operation:

- *Stage 1: Collection.* Once a signal is to be sampled, a center node joining the estimation process sends a message to CH which counts the number of suitable VCs.
- *Stage 2: Allocation.* The sink node calculates the period T' and the index of each cluster and sends these parameters to each CH. Each CH calculates the index of each VC and sends this value along with T' to each center node, which broadcasts these values to each node in its VC.
- *Stage 3: Start.* The sink node sends a start command with the start time t_0 to each CH, and each CH sends this command to each center node which broadcasts

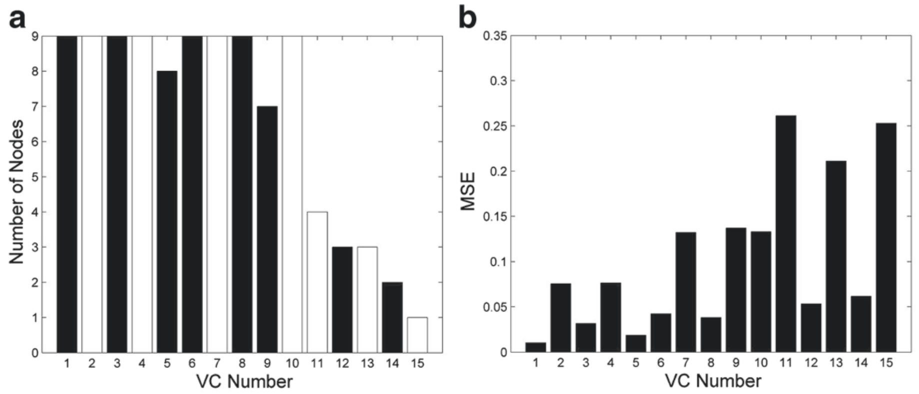

Figure 4 Results of the VC formation process. (a) Number of sensor nodes in each VC, the usable VCs are black. **(b)** MSE of each VC.

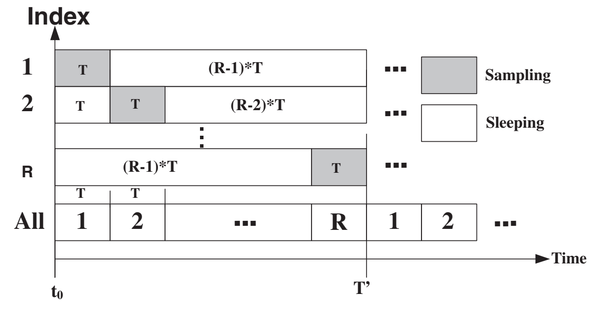

Figure 5 Sampling scheme.

the command to each sensor node. Nodes sample the signal according to these parameters.

When combined with VCs, RRSS guarantees the accuracy of the signal estimation. By adjusting the sample time of each VC, RRSS can conserve the energy of each sensor node. In each period T', a sensor node samples the signal only once. For the majority of the time in period T', i.e., $(R-1) \times T$, each node operates in its sleep state.

RRSS does not require accurate time synchronization. As long as the synchronization error is less than the sampling period, the scheme operates successfully.

6.3 Rotation of local fusion node

In each sampling and fusion process, a sensor node will receive the samples from other sensor nodes and estimate the event signal as a local fusion node in each suitable VC. Other sensor nodes only send their samples to the fusion node and hence consume less energy compared with the fusion node. After a VC is formed, the center node knows the whole IDs of the members in its VC. It will broadcast the IDs as a queue to each member. The center node is the first node to fuse the samples as a local fusion node. Subsequently, the local fusion node is selected dynamically in turn. A simple algorithm is used to select the dynamical fusion node based on the ID queue. The node with the next ID in the queue is selected to be the next fusion node. Because each node stores the ID queue, the selection process is performed by each node independently. Therefore, no addition energy is consumed on the rotation process.

6.4 Analysis of energy balance

To evaluate the ability of the proposed method to balance energy consumption among nodes in the network, the energy consumption of a selected sensor node n_i is modeled as:

$$E_i = E_S \times k_i^S + E_R \times k_i^R + E_T \times k_i^T, \tag{30}$$

where E_S, E_R, and E_T are the energy consumption of sampling once, receiving, and transmitting a packet, respectively, and k_i^S, k_i^R, and k_i^T are the number of the samples, the received, and transmitted packets, respectively. The energy consumed for computation is ignored. Furthermore, while the algorithms incur additional communication overhead for setup, this is minimal and infrequent, and only required at the beginning of the monitoring applications. Therefore, we assume that the energy consumption of this overhead can also be ignored.

The ratio (depicting the balance of energy consumption) between the energy consumption of two nodes, n_i and n_j, is defined as:

$$r_{ij} = \frac{E_i}{E_j} \tag{31}$$

If n_i and n_j are in the same suitable VC, due to the same behavior of these sensor nodes, n_i and n_j will almost consume the same energy from the average perspective, that means, $k_i^S = k_j^S$, $k_i^R = k_j^R$, and $k_i^T = k_j^T$, therefore, $r_{ij} = 1$.

If n_i and n_j are not in the same suitable VC, due to RRSS, the samples and the transmission number are the same, i.e., $k_i^S = k_j^S$ and $k_i^T = k_j^T$. However, k_i^R and k_j^R are related to the number of members in each VC when the nodes are fusion nodes. It is assumed that the VC, to which node n_i belongs, has m members. And the VC, to which node n_j belongs, has n members. In the time period $m \times n \times T'$, each sensor node, within the same VC with n_i, samples n times; the local fusion node receives $n-1$ packet each sample period, therefore, $k_i^R = m(n-1) \approx mn$. And hence, $k_j^R = n(m-1) \approx mn$. Therefore, the ratio is also 1 in this period. This means that the energy consumption of the nodes in different suitable VCs is also balanced.

6.5 Analysis of energy conservation

Each sensor node samples the signal once during period T'. Hence, for each period T', $k_i^S = 1$, $k_i^T = 1$, and $k_i^R = 0$ if the selected sensor node is not a local fusion node (else k_i^R equals to the number of sensor nodes in its VC). Compared to a baseline scheme where all sensor nodes continuously sample the signal during each sample period, it is obvious that the energy consumed in our scheme is reduced by a factor R.

The total energy consumed by the WSN during a period T' is given by:

$$E_{\mathrm{WSN}} = E_s \times N + (E_T + E_R) \times N + (E_T + E_R) \times H, \tag{32}$$

where N is the number of suitable sensor nodes and H is the number of hops from the local fusion nodes to the sink node. This highlights how energy is conserved through the two methods. First, during each sampling period, only a small subset of the WSN participates in sampling. Second, only the local fusion node transmits the data to the sink node while other sensor nodes transmit data in VCs.

7 Simulation results

In the above sections, a scheme of data acquisition is designed, and its performance is theoretically analyzed. In this section, the scheme is evaluated through simulations. The benefits of the proposed scheme are illustrated by comparing it with other state-of-the-art schemes.

7.1 Simulation environment

The simulated network contains 100 sensor nodes covering an area of 50×50 m^2. Sensor nodes are deployed at $x = i\times$ 5 m, $y = j\times$ 5 m, $(i, j = 1, 2, \ldots, 10)$, and the communication radius is 30 m. Timers 1 and 2 are set to be random numbers between $[1, 100]$ and $[101, 200]$ seconds, respectively. The expectation and variance of the event signal are $(0, 1)$, and the noise $(0, 0.1)$. The event to be monitored is at location (24 m, 26 m).

The Telos sensor node (containing the CC2420 radio transceiver) is considered. The transmit rate of the CC2420 is 250 kbps, which consumes 62.04 and 57.42 mW in receiving and transmitting modes, respectively. Assuming that communicated packets are 100 bits long, the energy consumption to receive or transmit a packet is $E_R = 24.82$ μJ and $E_T = 22.97$ μJ. The ADI accelerometer is used to measure the vibration of the seism, which has a power consumption of 1.12 mW and hence an energy consumption E_s, when sampling at a frequency of 20 times per second, of 5.6 μJ for a seismic signal with a maximum frequency 10 Hz.

For simplicity, the 100 sensor nodes belong to the same communication cluster. The sink node is located at (5 m, 5 m). Owing to the randomization in the formation process of VCs, each simulation is repeated ten times, and the average results are presented in this section.

7.2 Effects of parameter variation

Four parameters affect the selection of suitable VCs. Two parameters, d_{th}^{d} and Max_Num, determine the formation of VCs. The parameter d_{th}^{p} determines which VCs are usable, while Min_Num selects which VCs are suitable. These four parameters affect both the signal estimation accuracy and the energy consumption.

Figure 6 shows the effect of variation in d_{th}^{d} and Max_Num on estimation accuracy and energy consumption, while d_{th}^{p} is set to be 20 m and Min_Num is set to be 1 (i.e., all usable VCs are suitable).

Because the duration of sampling periods will be different in the simulation with the different parameters for a VC, in Figure 6, the average energy consumption, i.e., E_{WSN}/R, is found out through comparison.

Owing to a constant d_{th}^{p}, MSE is determined by the number of nodes in a suitable VC. When d_{th}^{d} is small (such as 5 to 10 min the results shown here), the number of nodes in each suitable VC is determined by d_{th}^{d} as Max_Num is not exceeded. However, when d_{th}^{d} is large, the number of nodes is determined by Max_Num. Hence, these parameters are coupled in determining the number of nodes in a VC. An increase of the number of nodes in a suitable VC (hence an increase in d_{th}^{d} and/or Max_Num) results in a decrease of MSE, as more nodes are contributing toward the estimation process. This is shown in Figure 6a. However, as the number of nodes in a suitable VC increases, the number of suitable VCs decreases. This means that T' decreases, and hence, the energy saving factor R decreases. This is shown in Figure 6b.

Figure 7 shows the impacts of varying Min_Num and d_{th}^{p} on MSE and energy consumption, while d_{th}^{d} and Max_Num are set to be 15 m and 15, respectively.

As d_{th}^{p} and Min_Num do not affect the VC formation process, VCs are the same regardless of the two parameters varied in these results. VCs are deemed to be suitable; however, they are affected by these parameters.

The results in Figure 7 show that MSE decreases when the number of nodes in a suitable VC increases; hence, an increase of Min_Num leads to a decrease of MSE. Also, as

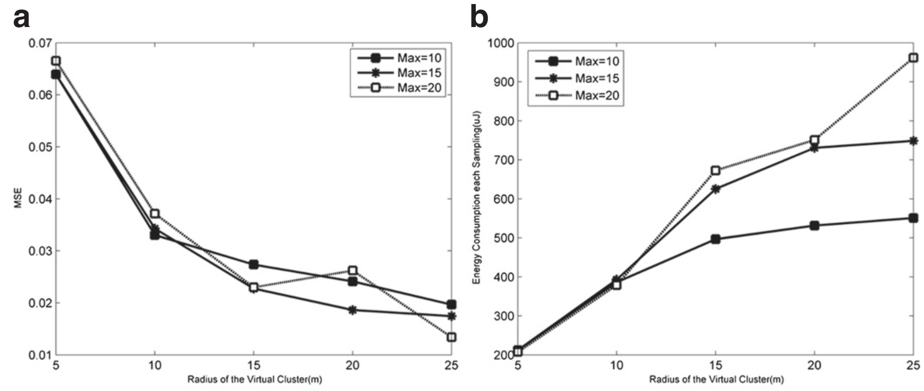

Figure 6 Impact of the radius of VC(d_{th}^{d}) and Max_Num. (a) Impacts on MSE; **(b)** impacts on energy consumption.

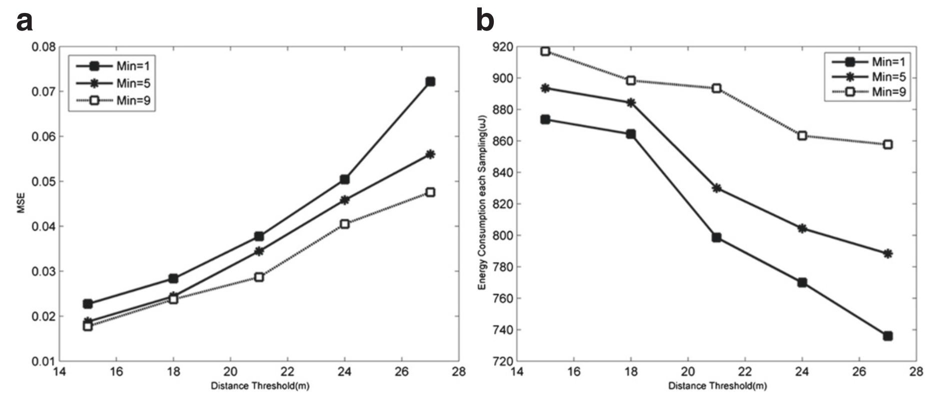

Figure 7 Impacts of distance threshold (d_{th}^{p}) and Min_Num. (a) Impacts on MSE; **(b)** impacts on energy consumption.

expected, a smaller value d_{th}^{p} leads to a lower MSE (nodes closer to the event can reconstruct the signal with greater accuracy, as presented by Vuran et al. [9]). This is shown in Figure 7a.

The results of Figure 7 also show that an increase of the number of nodes in a VC leads to an increase of the energy consumption. Therefore, an increase of Min_Num leads to an increase of the energy consumption as each suitable VC contains more nodes. The impact of d_{th}^{p} on energy consumption, however, is not so clear. An increase of d_{th}^{p} causes an increment of the number of the suitable VCs, which reduces energy consumption as R and hence T' increases. However, the number of suitable nodes also increases, which can cause an increase of the average energy consumption.

The simulation results shown in Figures 6 and 7 comply with Lemmas 1 and 2. When the proposed algorithm is used in real deployments, d_{th}^{d} can be set to be a larger value which will improve energy efficiency. And Max_Num is set to be the expected value which can assure the estimation signal. The selection of d_{th}^{p} and Min_Num should satisfy the requirement of estimation accuracy. The VC formation process can apply to both dense and sparse networks.

7.3 Comparison of different schemes

Different sampling schemes (such as those presented in Section 2) use different criteria to select the active nodes, which ultimately have an effect on the estimation error and energy efficiency. Two state-of-the-art schemes were presented in [15,16,18], which have similar objectives with the scheme presented in this paper. To illustrate the benefits of the scheme presented in this paper, these three schemes are made comparison. In [16], the network is split into a number of circular areas (not dissimilar from that shown in Figure 2). A single node in each of these areas is selected as a 'representative node' (referred to as a sampler node in [18]) which samples the event signal and transmits data to a sink node. In [15], the authors presented a different scheme, whereby selected sensor nodes are those close to the event location. The authors concluded that only 15 to 20 sensor nodes can accurately estimate a signal in a 900 m^2 field with 5 m × 5 m grid sensor topology.

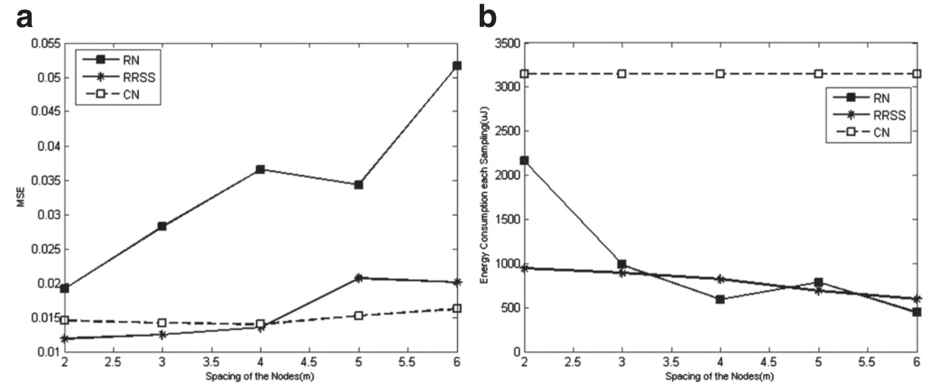

Figure 8 Comparison of the different schemes with a changing topology. (a) Comparison of MSE; **(b)** comparison of energy consumption.

For convenience, we name the scheme in [16] as RN (representative nodes) and the scheme in [15] as CN (concentrated nodes). The estimation method in CN is the same as that of in RN when the channel noise is ignored.

To implement RN, we apply the concept of VCs. From the perspective of VC, a center node is a representative node. All center nodes which are adjacent to the event join in the estimation process as representative nodes.

The average energy consumption of each sample process is used to make comparison. For RN and CN, energy consumption is

$$E_{\mathrm{WSN}} = E_s \times N \times R + (E_T + E_R) \times H \times R, \qquad (33)$$

where N is the number of sensor nodes which join in the sampling process and H is the hops sum from these sensor nodes to the sink node. Figure 8 compares MSE and energy consumption of RRSS, RN, and CN in different WSNs. In the simulation, $d_{th}^{p} = 15$ m, $d_{th}^{d} = 15$ m, Max_Num = 15, and Min_Num = 1. The number of nodes in CN is set to be 19. Each WSN is deployed with from 2 m × 2 m to 6 m × 6 m grid sensor topology.

Figure 8a shows that RRSS and CN provide a similar estimation error. The estimation error of RN is greater. Figure 8b shows that CN will consume more energy. RRSS and RN consume similar energy, and both schemes consume less energy than that of CN. Therefore, a conclusion can be drawn that RRSS consumes less energy while providing an accurate estimation. RN and CN are only able to provide promising results for either in the field of energy consumption or in field estimation error. While RRSS provides a mechanism for balancing energy consumption (as analyzed theoretically in the previous sections), RN and CN do not consider this. Hence, energy balancing is not considered in this section.

8 Conclusions

The estimation error and energy conservation are of major importance to data acquisition in a WSN. In this paper, based on the concept of VC, we present a novel framework to accurately estimate the event signal while maintaining the energy-efficient operation being balanced across the network. The novelty of the scheme presented in this paper is the utility of VC. A VC clusters sensor nodes with the same spatial and temporal properties for signal estimation. Based on the analysis, one suitable VC can guarantee the estimation accuracy of the event signal. Through the scheduling of suitable VCs, energy consumption of sensor nodes in each suitable VC is balanced. Finally, as most communication is limited in each VC, the energy consumed through communication is also conserved. The upper and lower limits of the number on members in each VC are used in the forming process. The upper limit is used to average the scale of VCs, a better balance of energy consumption to be provided. The lower limit guarantees that there are enough sensor nodes in each VC; hence, the estimation accuracy can be guaranteed. Through adjusting these two parameters, the algorithm acquires better flexibility for adapting to different WSNs. At present, the proposed scheme assumes fixed-point events without any movement. This will be addressed in our future work, where we will consider the dynamic selection and scheduling of suitable nodes.

Competing interests

The authors declare that they have no competing interests.

Author details
[1]School of Mechatronic Engineering and Automation, National University of Defense Technology, Changsha 410073, China. [2]Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK.

Received: 19 January 2013 Accepted: 15 August 2013
Published: 14 September 2013

References

1. T Arampatzis, J Lygeros, S Manesis, in *Proceedings of the 13th Mediterranean Conference on Control and Automation*. A survey of applications of wireless sensors and wireless sensor networks (Limassol, 27–29 June 2005), pp. 719–724
2. L An-Feng, W Xian-You, C Zhi-Gang, G Wei-Hua, Research on the energy hole problem based on unequal cluster-radius for wireless sensor networks. Comput. Commun. **33**(3), 302–321 (2010)
3. J Yick, B Mukherjee, D Ghosal, Wireless sensor network survey. Comput. Netw. **52**(4), 2292–2330 (2008)
4. G Anastasi, MD Francesco, M Conti, A Passarella, Energy conservation in wireless sensor networks: a survey. Ad Hoc Netw. **7**, 537–568 (2009)
5. V Raghunathan, S Ganeriwal, M Srivastava, Emerging techniques for long lived wireless sensor networks. IEEE Commun. Mag. **44**(4), 108–114 (2006)
6. C Alippi, G Anastasi, M DiFrancesco, M Roveri, Energy management in wireless sensor networks with energy-hungry sensors. IEEE Instrum. Meas. Mag. **12**(2), 16–23 (2009)
7. LM Oliveira, JJ Rodrigues, Wireless sensor networks: a survey on environmental monitoring. J. Commun. **6**(2), 143–151 (2011)
8. S Savazzi1, L Goratti, U Spagnolin, M Latva-aho, in *Proceedings of the Wireless World Research Forum*. Short-range wireless sensor networks for high density seismic monitoring (Paris, 5–7 May 2009)
9. MC Vuran, OB Akan, IF Akyildiz, Spatio-temporal correlation: theory and applications for wireless sensor networks. Comput. Netw. **45**(3), 245–261 (2004)
10. J Karjee, HS Jamadagni, Energy aware node selection for cluster-based data accuracy estimation in wireless sensor networks. Int. J. Adv. Netw. Appl. **3**(5), 1311–1322 (2012)
11. A Ribeiro, G Giannakis, Bandwidth-constrained distributed estimation for wireless sensor networks–Part I: Gaussian case. IEEE Trans. Signal Proc. **54**(3), 1131–1143 (2006)
12. L Zhi-Quan, X Jin-Jun, Decentralized estimation in an inhomogeneous sensing environment. IEEE Trans. Inf. Theory. **51**(10), 3564–3575 (2005)
13. SS Pradhan, J Kusuma, K Ramchandran, Distributed compression in a dense microsensor network. IEEE Signal Proc. Mag. **19**(2), 51–60 (2002)
14. MC Vuran, B Akan, in *Proceedings of the IEEE International Conference on Communications*. Spatio-temporal characteristics of point and field sources in wireless sensor networks (Istanbul, 11–15 June 2006), pp. 234–239
15. J Karjee, HS Jamadagni, in *Proceedings of the IEEE International Conference on Information System and Computational Intelligence*. Data accuracy estimation for cluster with spatially correlated data in wireless sensor networks (Harbin, January 2011), pp. 284–291
16. MC Vuran, IF Akyildiz, Spatial correlation-based collaborative medium access control in wireless sensor networks. IEEE/ACM Trans. Netw. **14**(2), 316–329 (2006)
17. B Gedik, L Liu, PS YU, ASAP: an adaptive sampling approach to data collection in sensor networks. IEEE Trans. Parallel Distributed Syst. **18**(12), 1766–1783 (2007)

18. R Willett, A Martin, R Nowak, in *Proceedings of the Third International Symposium on Information Processing in Sensor Networks*. Backcasting: adaptive sampling for sensor networks (Berkeley, CA, 26–27 April 2004), pp. 124–133
19. C Alippi, G Anastasi, MD Francesco, M Roveri, An adaptive sampling algorithm for effective energy management in wireless sensor networks with energy-hungry sensors. IEEE Trans. Instrum. Meas. **59**(2), 335–344 (2010)
20. H Minglei, HY Hen, in *Proceedings of the IEEE Global Telecommunications Conference*. Collaborative sampling in wireless sensor networks (Miami, 6–10 December 2010), pp. 1–5
21. W Jing, L Yongghe, SK Das, Energy-efficient data gathering in wireless sensor networks with asynchronous sampling. ACM Trans. Sensor Netw. **6**(3), 1–37 (2010)
22. L Xun, T Shiqi, GV Merrett, NM White, in *Proceedings of the 2011 IEEE International Conference On Mechatronics and Automation*. Energy-efficient data acquisition in wireless sensor networks through spatial correlation (Beijing, 7–10 August 2011), pp. 1068–1073
23. AA Abbasi, M Younis, A survey on clustering algorithms for wireless sensor networks. Comput. Commun. **30**(14–15), 2826–2841 (2007)
24. JO Berger, V Oliviera, B Sanso, Objective Bayesian analysis of spatially correlated data. J. Am. Stat. Assoc. **96**(456), 1361–1374 (2001)
25. X Jiuqiang, L Wei, L Fenggao, Z Yuanyuan, W Chenglong, Distance measurement model based on RSSI in WSN. Wireless Sensor Netw. **2**, 606–611 (2010)
26. H Karl, A Willig, *Protocols and Architectures for Wireless Sensor Networks*. (Wiley, New York, 2005)
27. G Mao, B Fidan, BD Anderson, Wireless sensor networks localization techniques. Comput. Netw. **51**(10), 2529–2553 (2007)
28. W Yiyin, M Xiaoli, G Leus, Robust time-based localization for asynchronous networks. IEEE Trans. Signal Proc. **59**(9), 4397–4410 (2011)
29. E Witrant, P Park, M Johansson, in *Wireless Networking Based Control*, ed. by SK Mazumder. Time-delay estimation and finite-spectrum assignment for control over multi-hop WSN, vol. 1, 1st edition (Springer, Dearborn, 2011), pp. 135–152
30. B Warneke, B Atwood, KSJ Pister, in *Proceedings of the Fourteenth The 14th IEEE International Conference on Microelectromechanical Systems*. Smart dust mote forerunners (Interlaken, 21–25 January 2001), pp. 357–360
31. JV Greunen, J Rabaey, in *Proceedings of the 2nd ACM International Conference on Wireless Sensor Networks and Applications*. Lightweight time synchronization for sensor networks (San Diego, CA, 19 September 2003), pp. 11–19
32. S Ganeriwal, R Kumar, M BSrivastava, in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems*. Timing-sync protocol for sensor networks (Los Angeles, CA, 05–07 November 2003), pp. 138–149

doi:10.1186/1687-1499-2013-230
Cite this article as: Li *et al.*: Energy-efficient data acquisition for accurate signal estimation in wireless sensor networks. *EURASIP Journal on Wireless Communications and Networking* 2013 **2013**:230.