Article

A Sub-Clustering Algorithm Based on Spatial Data Correlation for Energy Conservation in Wireless Sensor Networks

Ming-Hui Tsai and Yueh-Min Huang *

Department of Engineering Science, National Cheng-Kung University, No.1, University Road, Tainan City 701, Taiwan; E-Mail: coluxtsai@gmail.com

* Author to whom correspondence should be addressed; E-Mail: huang@mail.ncku.edu.tw; Tel.: +886-6-27-57575 (ext. 63300); Fax: +886-6-27-66549.

Abstract: Wireless sensor networks (WSNs) have emerged as a promising solution for various applications due to their low cost and easy deployment. Typically, their limited power capability, *i.e.*, battery powered, make WSNs encounter the challenge of extension of network lifetime. Many hierarchical protocols show better ability of energy efficiency in the literature. Besides, data reduction based on the correlation of sensed readings can efficiently reduce the amount of required transmissions. Therefore, we use a sub-clustering procedure based on spatial data correlation to further separate the hierarchical (clustered) architecture of a WSN. The proposed algorithm (2TC-cor) is composed of two procedures: the prediction model construction procedure and the sub-clustering procedure. The energy conservation benefits by the reduced transmissions, which are dependent on the prediction model. Also, the energy can be further conserved because of the representative mechanism of sub-clustering. As presented by simulation results, it shows that 2TC-cor can effectively conserve energy and monitor accurately the environment within an acceptable level.

Keywords: WSNs; energy conservation; data correlation; prediction model; virtual cluster

1. Introduction

As the rapid advances recently in wireless communications and sensor technologies, wireless sensor networks (WSNs) have emerged as the promising solution for various applications due to their low cost and easy deployment. A wireless sensor network consists of sensor nodes spatially deployed over a geographical area for monitoring physical environment. These sensor nodes are mainly composed of four components: power unit for power supply, sensing unit for data acquisition (sampling), processing unit for data processing, and radio unit for wireless communication. Typically, their limited power capability, *i.e.*, battery powered, make WSNs encounter the challenge of extension of network lifetime. Many interesting literatures [1–5] have proposed how to address this kind of problem. Therefore, it is a key issue in the design of systems based on WSNs that how to reduce the energy consumption and further prolong the network lifetime.

With regard to the energy conservation, Anastasi *et al.* [4] identify three main enabling techniques, namely, duty cycling, data driven, and mobility-based approaches. The first category exploits the power management (including mechanisms on MAC layer, network layer and cross layers) for energy saving by considering duty cycling [6–9]. To this end, various criteria can be used to decide which/when sensor nodes should be active. It is a well-known fact that the radio unit is the dominant part of energy consumption in a sensor node [10]. Hence, it has been shown in previous works [10,11] that an idle sensor node should turn its radio off to avoid unneeded energy waste, namely avoid idle listening. In [12], the authors adapt the scheduled rendezvous scheme for duty cycling. The basic idea is that each node should wake up based on a wakeup schedule. They utilize the data gathering tree structure to achieve both energy efficiency and low packet delivery latency. The scheme staggers the active/sleep schedule of the nodes in the data gathering tree according to its depth in the tree. The second category focuses on sensor readings and can be further be subdivided into two main classifications: data reduction and energy-efficient data acquisition. The former adopts in-network processing, such as aggregation [13,14], compression [15] or prediction [16,17], in order to reduce the amount of data readings that need to be transmitted. Clearly, it can perform energy conservation that is benefitting from cutting down the transmissions. However, it has to pay more attention to keep the sensing accuracy within an acceptable level for the application, since data reduction is designed to reduce the amount of communication. The other proposes adaptive sampling schemes [18] for energy saving while especially taking energy hungry devices into consideration. Furthermore, unneeded and reducible sampling results in useless energy consumption because they bring about unneeded communications. The last category considers the specific schemes that the mobility of sensors nodes or sink is available. By introducing mobility in WSNs, routing schemes for the collection of data reading should take the mobile pattern of sensors nodes into consideration.

As mentioned above, a lot of studies on the WSNs have been dedicated to energy conservation. WSNs using hierarchical protocols are partitioned into clusters with different concerns [19]. Many hierarchical protocols show their better ability of energy efficiency, stability, and scalability [20,21] in WSNs. The cluster head can not only be one of the coordinators among the inter-cluster communication, but also be a key role for in-network processing within the intra-cluster. Besides, data reduction based on the correlation of sensed readings can efficiently reduce the amount of required transmissions. Therefore, we use a sub-clustering procedure based on spatial data correlation to further separate

the hierarchical (clustered) architecture of a WSN. The representative mechanism (data reduction) is also engaged in these sub-clusters for more energy conservation. In our algorithm, all sensor nodes are first grouped into clusters (physical clusters) using hierarchical protocols. Subsequently, a sensor node generates a prediction model to forecast the data readings instead of reporting periodically data readings. The prediction mechanism is based on temporal data correlation demonstrated in [16]. The forecasting value is regarded as the concrete data reading. Hence, the transmissions of data reading reports are rationally reduced. Afterwards, sensor nodes within a physical cluster having strong spatial data correlations are then grouped into a sub-cluster (or say virtual cluster). Those chosen sensor nodes become the representative nodes for the sub-clusters and bear the responsibility of sensing environment around the whole virtual cluster.

The main contributions of this paper are twofold. First, we use two tiers hierarchical architecture to form physical clusters and virtual clusters as shown in Figure 1. Under this kind of architecture, the energy conservation benefits by the reduced transmissions, which are dependent on the prediction model. Also, the energy can be further conserved because of the representative mechanism by sub-clustering. Only the representative sensor node of the virtual cluster turns its sensing unit on. Oppositely, the others within a virtual cluster can just power off for energy saving. Secondly, the decrement of transmitted data readings results in the possible decreased accuracy in monitoring environment. Moreover, the representative sensor nodes need to be able to represent the monitoring area well. Therefore, we try to keep the accuracy of monitoring within an acceptable level. Note that the representative sensor needs to transmit to physical cluster head (abbreviated as *pch*) its readings, which differ from the predicted readings by more than a certain pre-specified error threshold.

Figure 1. Two tiers hierarchical architecture is composed of physical and virtual clusters.
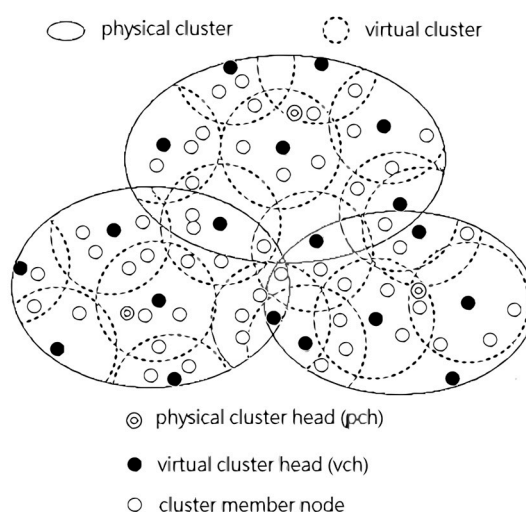
The remainder of this paper is organized as follows. Section 2 reviews several hierarchical protocols. Also, some data reduction algorithms, based on correlation and prediction proposed previously, are surveyed in a concise manner. In Section 3, we put forward the system model of our scheme. A sub-clustering algorithm under a hierarchical architecture is presented. Also, we investigate our data reduction mechanism based on data prediction and correlation. Then, the simulation results of the proposed algorithms are presented in Section 4. Finally, conclusion is given in the last section of the paper.

2. Related Work

A survey of energy conservation in WSNs is given in [4]. In order to efficiently reduce the energy consumption and further prolong the network lifetime, we focus on the literatures of energy conservation, especially those hierarchical protocols with concern on energy efficiency and data reduction schemes using data correlation and prediction.

Certainly, the hierarchical architecture is introduced to WSNs because it has proven to be an effective approach to provide better energy efficiency. As network nodes are organized in clusters, the router nodes are responsible for coordinating activities within the cluster and forwarding data readings between clusters. The use of routing hierarchy has many advantages, including better energy efficiency and scalability. LEACH (Low Energy Adaptive Clustering Hierarchy) [22] is the most famous algorithm that uses clustering to prolong the network lifetime. From the energy perspective, it uses randomized rotation of the cluster heads to distribute the energy consumption among the sensor nodes in a network. Many research efforts in energy efficiency follow LEACH with different concerns, such as cluster head selection, clustering cost, and cluster maintenance. In [23], the authors propose an improved routing algorithm of ring based multi-hop clustering (IRBMC) to prolong the lifetime of network. WSNs are divided into heterogeneous spacing rings to control the cluster number and build unequal size of clusters in different rings. The cluster heads selection is depending on the residual average energy estimation, which is taken into account to balance energy consumption of nodes in each ring. Also, in [24], two integer linear program (ILP) formulations are proposed for assigning sensor nodes to clusters in a two-tiered network where the relay nodes are used as cluster heads. The objective of ILP formulations is to load balanced clustering, that is, to maximize the lifetime of the relay node network by distributing the total load for data communication over the different relay nodes.

Data reduction techniques are designed to reduce the amount of transmissions, including data aggregation [25,26], compression, and prediction [16,17]. In [26], the authors propose a data correlation-based clustering (DCC). In each round of data acquisition, data readings of all nodes in one cluster are gathered at the cluster head, on which data readings will be compressed before send to the sink. The sink mines data correlations from historical data to construct clusters. However, the network is divided into several disjoint clusters by minimum degree first clustering, which is irrelevant with data correlation. That is, the data correlation between the cluster members can't be substantiated. In [16], PREMON paradigm trades increased computation (for deriving prediction models) for savings in number transmissions. In [17], the authors propose a buddy protocol that sensor nodes with high temporal correlation formed a buddy group. The nodes within the same buddy group (cluster) take turns keeping their radio on. The representative node responds on behalf of its buddies. However, GAF [27] is used for clustering in this protocol and the data readings between buddies have not been verified as being correlated. Therefore, the difference of data readings between the representative and buddies is not negligible.

In [25], the authors propose a distributed clustering-based aggregation algorithm for spatially correlated WSNs. When the clusters are constructed by the α-local spatial clustering, the data readings of cluster head and members have very high correlation. Only the cluster heads need to do the data sampling work. The data readings are then aggregated to a cluster head backbone and transmitted to the sink. Due to the lack of re-clustering mechanism, the constructed clusters based on weighted

dominant set cannot accommodate themselves to the changing environment. Further, the cluster head has to keep sensing in order to monitor environment. Therefore, the cluster head tends to energy depletion.

3. A Sub-Clustering Algorithm Based on Spatial Data Correlation

Our proposed algorithm (two tiers clustering based on data correlation) abbreviated as 2TC-cor is composed of two procedures: the prediction model construction procedure and the sub-clustering procedure. Of course, the sensor network is partitioned into physical clusters by the distributed hierarchical routing protocols before the proposed procedures. We can choose LEACH, DCC, IRBMC, or others as the underlying hierarchical routing protocol due to the independence from those procedures. However, cluster heads, gateway nodes, backbone nodes, or the connected dominating set (CDS) play the key roles for the preservation of networks reachability in the hierarchical routing protocols. These nodes, called as router nodes, generally have to always switch their power on according to the adapted routing protocols. Therefore, it is important that the energy saving (power on/off) at router nodes is dependent on the underlying routing protocol rather than resulted from our sub-clustering procedure.

In this section, we will first introduce the construction of prediction models that are used to predict the data readings of sensor nodes. Each node within a physic cluster then attempts to establish buddy relationship with its neighbors during the sub-clustering procedure. The sensor nodes with high data correlation (or say buddies) are then divided into a virtual cluster and the chosen representative nodes bear the monitoring responsibility.

3.1. Constructing Prediction Models

Perhaps the readings of sensor nodes show high spatial and/or temporal correlation in the monitoring environment. We can predict a sensor node's reading based on the readings of the sensor nodes around it (spatial correlation) or based on itself historical readings (temporal correlation) [16]. Therefore, at the beginning of this procedure, each sensor node in a physical cluster calculates its prediction model according to its historical readings. The prediction models specify the readings at a sensor node as a function of its readings in the past. We suppose that a sensor node has enough memory to store the historical readings. Accordingly, sensor nodes generate their predictions by linear regression [28].

A sensor node s measures it's data reading $R_s(t)$ at time t. The prediction model of node s is denoted as $\hat{R}_s(t)$. Over time, it collects a set of data readings $(R_s(t_1), R_s(t_2), \ldots, R_s(t_k))$. With consideration of the characteristics of the application, we may pre-specify a set of basis functions, $F = \{f_1(t), f_2(t), \ldots, f_n(t)\}$. In regression, we would like to fit the basis functions F to those k data readings, that is, to find basis functions coefficients, $W = (w_1, w_2, \ldots, w_n)^\top$. Such that:

$$\hat{R}_s(t) = w_1 f_1(t) + w_2 f_2(t) + \cdots + w_n f_n(t) = \sum_{i=1}^{n} w_i f_i(t) \approx R_s(t) \tag{1}$$

where, $t = t_1, t_1 + t^*, t_1 + 2t^*, \ldots, t_k$, t^* is the frequency of reports requested based on the application and $t_k = t_1 + (k-1)t^*$. To formulize the approximation, we use root mean squared error (RMSE) as the error metric. Hence, the parameters W^* are calculated in order to minimize the RMSE:

$$W^* = \arg\min_w \sqrt{\frac{1}{k}\sum_{i=1}^{k}\left(R_s(t) - \hat{R}_s(t)\right)^2} = \arg\min_w \sqrt{\frac{1}{k}\sum_{i=1}^{k}(R_s(t) - \sum_{i=1}^{n}w_i f_i(t))^2} \qquad (2)$$

As presented in [28], the basis function coefficients W can be obtained.

Subsequently, we use the generated model $\hat{R}_s(t)$ as the predictor for the next Δ readings where $t = t_{k+1}, t_{k+2}, \ldots, t_{k+\Delta}$. That is to say, during the period (from $t_1 + kt^*$ to $t_1 + (k + \Delta - 1)t^*$), the forecasting value $\hat{R}_s(t)$ is conditionally regarded as the concrete reading $R_s(t)$ of sensor node at the t^* interval. If the difference between $\hat{R}_s(t)$ and $R_s(t)$ is within a certain pre-specified error threshold ε, the sensor node doesn't need to transmit its concrete reading to the cluster head (pch). The predicted value $\hat{R}_s(t)$ is treated as the concrete reading of sensor node s. It means that the sensor node can reduce transmissions by using prediction model. On the contrary, if the difference is more than ε, the sensor node needs to send out its concrete reading to pch. Such a predicted value is termed as "violation" and the sensor node needs to transmit its concrete reading to pch. After Δ readings, the coefficients W are calculated once again using the last k readings, as illustrated in Figure 2. And the new obtained W is sent to pch and combined with F to forecast the next Δ readings.

Figure 2. Using k concrete data readings to forecast the next Δ data readings.
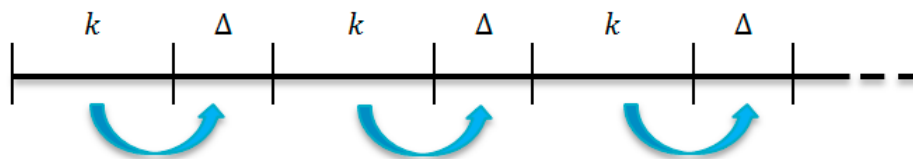
3.2. Sub-Clustering to be Virtual Clusters

In order for more energy savings to keep fewer sensor nodes radio on, or say to avoid idle listening, we further partition (sub-clustering) the physical cluster based on the spatial data correlation of sensor nodes. The sensor nodes with high spatial data correlation are formed into a virtual cluster and only the representative nodes keep their radio on. Here, our sub-clustering algorithm is composed of two procedures: virtual cluster head selection procedure (VCHS procedure) and virtual cluster construction procedure (VCC procedure). Even though large number of sensor nodes may turn their radio off, the network reachability in our algorithm is still preserved since the underlying hierarchical routing protocol is responsible for it.

At the specified sample time stamp t_k, pch has received the sets of concrete data readings $(R_s(t_1), R_s(t_2), \ldots, R_s(t_k))$ of sensor node $s \in S$ where S is the set of cluster members of pch. The communication hop counts between sensor nodes $s \in S$ are one or more which is depending on the underlying hierarchical routing protocol. Since we intend to separate the set S into several virtual clusters depending on the spatial data correlation, the calculation of correlation tries to identify in what degree a node $x \in S$ is correlated with its neighbors $y = \{y \mid y \in N(x)\}$ where $N(x)$ is denoted as the set of node x's neighbors belonging to S. The data readings of x and y are denoted as $(R_x(t_1), R_x(t_2), \ldots, R_x(t_k))$ and $(R_y(t_1), R_y(t_2), \ldots, R_y(t_k))$, respectively. We define d_{xy} as the difference (distance) between the data readings of node x and y by the Euclidean distance:

$$d_{xy} = \sqrt{\left|R_x(t_1) - R_y(t_1)\right|^2 + \left|R_x(t_2) - R_y(t_2)\right|^2 + \cdots + \left|R_x(t_k) - R_y(t_k)\right|^2} \tag{3}$$

Then, the expected value of d_{xy} is

$$E(d_{xy}) = \overline{d_x} = \frac{1}{|N(x)|} \sum_{y \in N(x)} d_{xy} \tag{4}$$

where $|N(x)|$ is the number of nodes in $N(x)$. Also, the deviation of d_{xy} is $D(d_{xy})$:

$$D(d_{xy}) = E\left\{\left[d_{xy} - E(d_{xy})\right]^2\right\} = \frac{1}{|N(x)|} \sum_{y \in N(x)} (d_{xy} - \overline{d_x})^2 = E(d_{xy}^2) - \left[E(d_{xy})\right]^2$$

$$= \frac{1}{|N(x)|} \sum_{y \in N(x)} d_{xy}^2 - \overline{d_x}^2 \tag{5}$$

In order to find out a measurement of spatial correlated weight for each node $x \in S$, we further define the weight scw_x of node x as

$$scw_x = \frac{\left[\sum_{y \in N(x)} |d_{xy} - \overline{d_x}|\right]^2}{|N(x)|^2 D(d_{xy})} = \frac{\left[\sum_{y \in N(x)} |d_{xy} - \overline{d_x}|\right]^2}{|N(x)| \sum_{y \in N(x)} (d_{xy} - \overline{d_x})^2} \tag{6}$$

According to Cauchy-Schwarz inequality, we have

$$\left[\sum_{y \in N(x)} |d_{xy} - \overline{d_x}|\right]^2 \le |N(x)| \sum_{y \in N(x)} (d_{xy} - \overline{d_x})^2 \tag{7}$$

As the result, $0 \le scw_x \le 1$. The definition of scw_x takes into consideration the average spatial distance deviation between node x and $N(x)$. Thus, the larger value of scw_x is, the smaller difference variation between node x and $N(x)$. That is, node x has higher spatial correlation with $N(x)$ as its scw_x is larger.

Based on the measurements, there are two possible situations when a node should be assigned as a determined/candidate one of virtual cluster head (vch):

1. A node has very low scw. ($lb - vch$, a determined vch satisfying the predefined lower bound scw^{lb})
2. A node has very high scw. ($ub - vch$, a candidate vch satisfying the predefined upper bound scw^{ub})

Obviously, nodes being in either of two situations should be conditionally chosen as the representative node for a virtual cluster, depicted in Figure 3a. For the first situation, a node becomes $lb - vch$ due to its low scw. This means that the data readings of the node are evidently different from its neighbors. Taking into consideration the application accuracy and integrity, a low scw node should be still awake in monitoring the environment. For the second situation, the upper bound of the weight is used to make sure that the nodes having higher correlation with their neighbors are chosen to be $ub - vch$s. However, it is possible that neighboring $ub - vch$s are highly correlated. They should be grouped into a virtual cluster, as depicted in Figure 3b. Noted that the scw_x^{max} is calculated at pch side and possible vchs are selected in a centralized manner. As to virtual cluster

construction (VCC) procedure, the virtual clusters are constructed in a distributed manner oppositely. If a node can join several clusters, it has to choose the nearest *vch* to join. The results can refer to Figure 3c. The details of VCC procedure are described as follows.

Figure 3. The role transition of nodes during virtual cluster construction (VCC) process. (**a**) sensor nodes within a physical cluster; (**b**) some *ub-vchs* are absorbed by the *ub-vch* with max (scw); (**c**) vchs are determined after VCC procedure.

(a) sesor nodes within a physical cluster

(b) some ub-vchs are absorbed by the ub-vch with max (scw)

(c) vchs are dertermined after VCC procedure
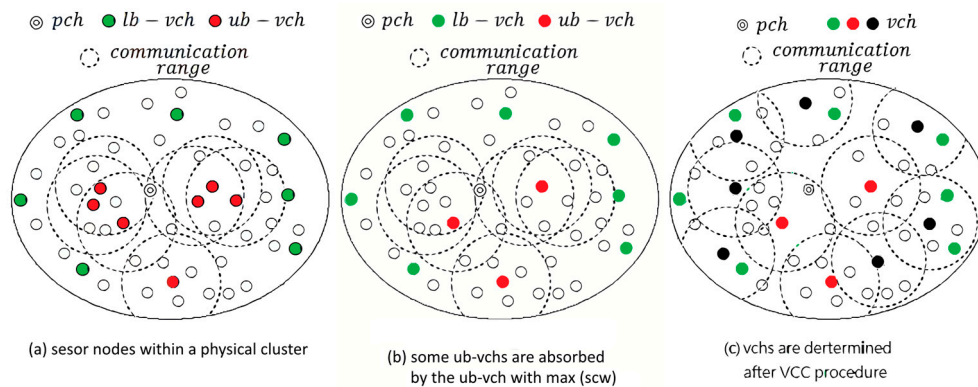
After the virtual cluster construction, the representative nodes have to bear the monitoring responsibility. Meanwhile, the energy consumption for readings transmission is effectively reduced because of the representative mechanism. That is to say, the rest of nodes simultaneously turn their radio and sensing unit off and the energy are further conserved due to the contribution of the representative nodes.

Virtual cluster heads selection (VCHS) procedure:

Input: $(R_s(t_1), R_s(t_2), ..., R_s(t_k))$ of sensor node $s \in S$, lower bound and upper bound of spatial correlated weight scw_x for any node $x \in S$

Output: the set $V^{'}, V^{''}$ of virtual cluster heads (vch) of S

Step1: $V^{'}, V^{''} = \emptyset$

Step2: calculation of spatial correlated weight scw_x for any node $x \in S$

Step3: on $lb - vch$ ($V^{'} = \emptyset$)

if $(scw_x \leq scw^{lb}) V^{'} = V^{'} + x;$

Step4: on $ub - vch$ ($V^{''} = \emptyset$)

for each node x, $scw_x^{max} = \max\{scw_y | y \in N(x)\}$

if $(scw_x \geq scw^{ub}$ and $scw_x > scw_x^{max}) V^{''} = V^{''} + x;$

Virtual cluster construction (VCC) procedure:

Step1: each node x $(x \in V^{''})$ broadcasts an *indicator* message to $N(x)$

Step2: each node x $(x \in \{S - V^{'} - V^{''}\})$ receives no *indicator* message after a predefined time period, x broadcasts an indicator message to $N(x)$

Step3: node y $(y \in \{S - V^{'} - V^{''}\})$ joins the virtual cluster of x (VC_x):

3-1: if y receives only one *indicator* message from x

3-2: if y receives more than two *indicator* messages from $N(y)$, then it joins VC_x if $d_{xy} = \min\{d_{xy} | x \in N(y)\}$

Step4: if node y decides to join VC_x, it sends a *join* message to node x

Step5: if node x receives a *join* message, it sends back an *acknowledge* message to y. Hence x is the virtual cluster head (*vch*) of VC_x and y is a member of VC_x.

4. Simulation Results

In this section, we develop a simulator in C++ for the evaluation of the proposed algorithm. To investigate the efficiency of energy conservation and the accuracy of monitoring environment, LEACH and 2TC-cor are compared in the experiments. HNA [29] denoted as the estimated value for the sensing intervals in which half of the nodes die is used as the metric to evaluate the efficiency of energy conservation. As to the accuracy of monitoring environment, some check points are predefined in order to examine the difference between the forecasting/representative readings and the concrete data readings. The difference is used as the metric of the accuracy of monitoring environment. Details of the simulation environment are given in Table 1. The nodes are randomly distributed in a square region and the sink node is located in the center of the region. Once the cluster members gather the data readings (sense the environment), they report to cluster head within the sensing interval. Particularly, the sensing interval is set to five seconds in order to speed up the observation of the simulation. Besides, the pre-specified basis function (F) is associated with k value to predict the sensing environment in our scenario. A scenario is simulated 30 times to produce more reliable results and the average of results is taken.

Table 1. Simulation environment.

| Parameter	Value
Network size	1000 m × 1000 m
Number of nodes	500, 750, 1000, and 1250
Signal range	150 m
Sensing interval (t^*)	5 (s)
Ratio of Δ to k ($k = 72$)	1/6
Pre-specified basis function (F)	$F = \{1, t, t^2\}$
Error threshold ε ($|\hat{R}_s(t) - R_s(t)|$)	$0.1 * R_s(t)$
Upper bound of spatial correlated weight (scw^{ub})	0.8
Lower bound of spatial correlated weight (scw^{lb})	0.2

4.1. Efficiency of Energy Conservation

We first evaluate the efficiency of energy conservation by comparing the HNA of the two algorithms (LEACH and 2TC-cor). Simply, the LEACH protocol is a hierarchical protocol in which nodes transmit their data readings to cluster heads. The mechanism of cluster head rotation leads to a balanced energy consumption and hence to a longer lifetime of the network. Obviously, the denser networks have the longer lifetime as shown in Figure 4. But the cluster members are continuously sensing the environment and reporting to cluster head. It leads to unnecessary energy consumption while the reported data are highly temporal and/or spatial correlated.

Our 2TC-cor algorithm uses prediction model to reduce the transmission in which the reporting data are highly temporal correlated. Also, some nodes can turn off their radio units (*i.e.*, avoid idle

listening) for more energy saving based on the prediction mechanism. Furthermore, the nodes with highly spatial correlated data are grouped into virtual clusters. In the context, only the representative node is working and the neighboring nodes within a virtual cluster rest instead. It is rational that the representative mechanism results in the reduction of energy consumption, especially under a denser network. However, the overhead of constructing virtual clusters is non-negligible. In the experiments of sparser network, the expended energy for constructing virtual clusters exceeds the conserved energy by the predicting and representative mechanism. Therefore, the performance of our algorithm is not as good as LEACH in the metric of energy efficiency under the sparser network condition (number of nodes are 500 and 750). But, the benefits of energy conservation reveal finally under the denser network condition (number of nodes are 1000 and 1250) as depicted in Figure 4.

Figure 4. The time at which half of the node are dead.
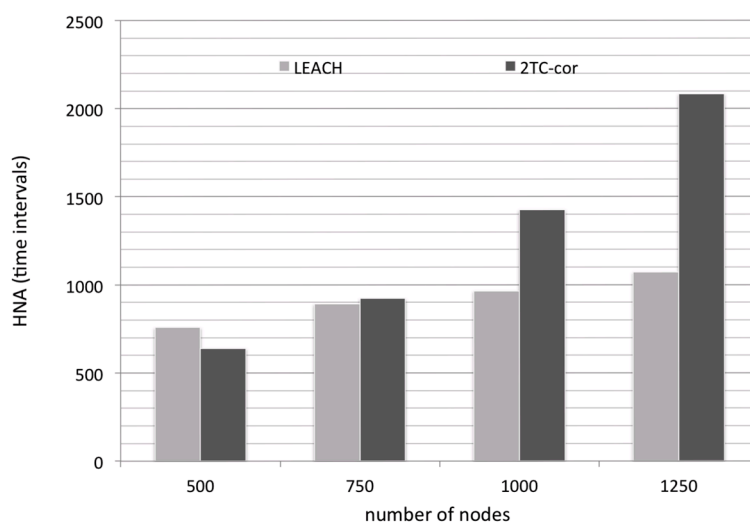
4.2. Accuracy of Monitoring Environment

Next, we focus on the accuracy of monitoring environment. Due to the prediction and representative mechanism, the possible difference between the forecasting/representative and concrete data readings comes from the reduced transmission. Therefore, in order to investigate 2TC-cor accuracy, we predefine 4 check points to examine the difference. The check points are set at each 82 time intervals. The aim is to make the check points set at the Δ period, seen in Figure 2.

In order to collect the concrete data readings, 2TC algorithm should be adaptively altered. It is worth noting that the representative mechanism should be ignored during the simulation of altered 2TC-cor. That is to say, not only the representative node (*vch*) works but all of the cluster members in a virtual cluster have to work (*i.e.*, sense the environment and report to *vch*). Besides, the prediction model should be ignored, too. Each *vch* report continuously its data readings to its *pch* instead of reporting only while "violations" appearing (reporting occasionally based on the prediction model).

The difference between reported readings from virtual cluster member nodes (*i.e.*, concrete data reading) and sensed readings of *vch* (*i.e.*, the representative node) is used to examine the 2TC-cor accuracy. We denote the average difference ($\bar{d}(t)$) at *vch* side as

$$\bar{d}(t) = \frac{|VC| - \sum_i^{VC} \frac{|R(t) - R_i(t)|}{R(t)}}{|VC|} \qquad (8)$$

where $R(t)$ is the sensed readings of vch, $R_i(t)$ is the concrete data reading from virtual cluster member node i and $|VC|$ is the number of virtual cluster member nodes. Finally, the accuracy index of our algorithm is defined as:

$$accuracy\ index = \sum_x^N \frac{\frac{\sum_y^S \bar{d}}{|S|}}{|N|} \qquad (9)$$

where x and y are pch and vch, $|N|$ and $|S|$ are the number of physical and virtual cluster heads, respectively.

Therefore, the accuracy index of our algorithm is obviously lower than LEACH because of the prediction and representative mechanism. As shown in Figure 5, the accuracy index is progressively increasing as the times goes by (or say nodes are dying). Naturally, nodes are gradually draining their energy. In order to accurately monitor the whole environment, more nodes are expected to work while more nodes are dead. In other words, as the working nodes are getting fewer, the accuracy index is getting higher due to the efficiency of representative mechanism is not as active.

Figure 5. The accuracy of 2TC-corr in monitoring environment.
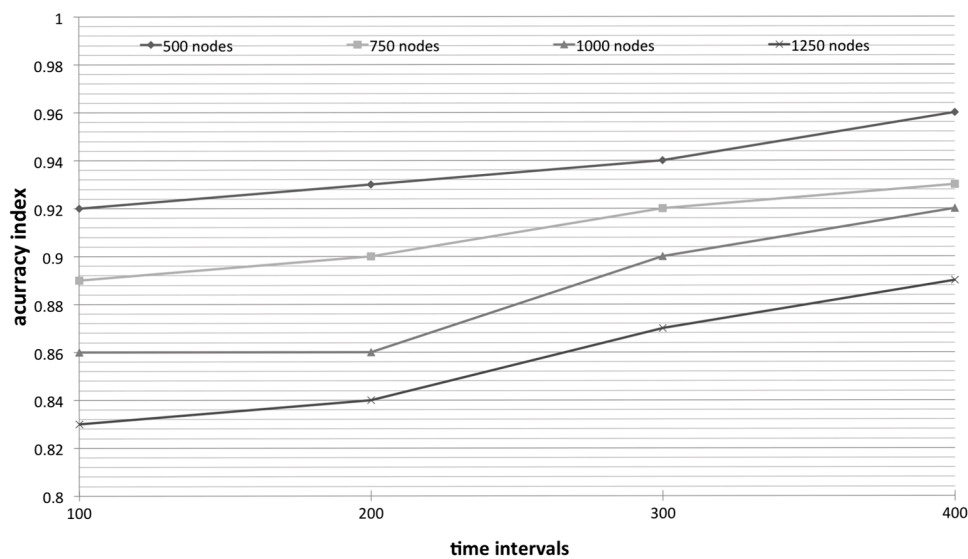
5. Conclusions

While saving energy, there are three key characteristics of WSNs that must be carefully maintained: quality of monitoring, reachability in the network, and communication delay. Due to the underlying hierarchical routing protocol, the reachability and communication delay can be disregarded in our algorithm rationally. Therefore, we just focus on the efficiency of energy conservation and accuracy of monitoring environment. In this article, we propose 2TC-cor with consideration of the spatial and temporal correlation of data readings. As presented by simulation results, it can effectively conserve energy by the prediction and representative mechanisms. However, the accuracy of monitoring

environment must be carefully investigated. Besides, with consideration of various applications, the representative mechanism or virtual clusters formation should take the residual energy of nodes into account in our future work.

Author Contributions

Ming-Hui Tsai did the research and conducted the experiments, analyzed the data and drafted the manuscript. Yueh-Min Huang acted as advisor.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Dietrich, I.; Dressler, F. On the lifetime of wireless sensor networks. *ACM Trans. Sen. Netw.* **2009**, *5*, 1–39.
2. Soro, S.; Heinzelman, W.B. Prolonging the lifetime of wireless sensor networks via unequal clustering. In Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05), Denvor, CO, USA, 4–8 April 2005.
3. Bagci, H.; Yazici, A. An energy aware fuzzy approach to unequal clustering in wireless sensor networks. *Appl. Soft Comput.* **2013**, *13*, 1741–1749.
4. Anastasi, G.; Conti, M.; di Francesco, M.; Passarella, A. Energy conservation in wireless sensor networks: A survey. *Ad Hoc Netw.* **2009**, *7*, 537–568.
5. Anisi, M.H.; Abdullah, A.H.; Razak, S.A.; Ngadi, M.A. Overview of data routing approaches for wireless sensor networks. *Sensors* **2012**, *12*, 3964–3996.
6. Demirkol, I.; Ersoy, C.; Alagoz, F. Mac protocols for wireless sensor networks: A survey. *IEEE Commun. Mag.* **2006**, *44*, 115–121.
7. Paruchuri, V.; Basavaraju, S.; Durresi, A.; Kannan, R.; Iyengar, S.S. Random asynchronous wakeup protocol for sensor networks. In Proceedings of the First International Conference on Broadband Networks, 2004 (BroadNets 2004), San Jose, CA, USA, 25–29 October 2004; pp. 710–717.
8. Langendoen, K. Medium access control in wireless sensor networks. *Medium Access Control Wirel. Netw.* **2008**, *2*, 535–560.
9. Ye, W.; Heidemann, J.; Estrin, D. Medium access control with coordinated adaptive sleeping for wireless sensor networks. *IEEE/ACM Trans. Netw.* **2004**, *12*, 493–506.
10. Tae, P.; Kyung-Joon, P.; Lee, M.J. Design and analysis of asynchronous wakeup for wireless sensor networks. *IEEE Trans. Wirel. Commun.* **2009**, *8*, 5530–5541.
11. Xue, Y.; Vaidya, N.F. A wakeup scheme for sensor networks: Achieving balance between energy saving and end-to-end delay. In Proceedings of the 10th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS 2004), King Edward, Toronto, ON, Canada, 25–28 May 2004; pp. 19–26.

12. Lu, G.; Krishnamachari, B.; Raghavendra, C.S. An adaptive energy-efficient and low-latency mac for data gathering in wireless sensor networks. In Proceedings of the 18th International Parallel and Distributed Processing Symposium, Santa Fe, NM, USA, 26–30 April 2004; p. 224.

13. Krishnamachari, B.; Estrin, D.; Wicker, S. The impact of data aggregation in wireless sensor networks,. In Proceedings of the 22nd International Conference on Distributed Computing Systems Workshops, Vienna, Austria, 2–5 July 2002; pp. 575–578.

14. Al-Karaki, J.N.; Kamal, A.E. Routing techniques in wireless sensor networks: A survey. *IEEE Wirel. commun.* **2004**, *11*, 6–28.

15. Pattem, S.; Krishnamachari, B.; Govindan, R. The impact of spatial correlation on routing with compression in wireless sensor networks. *ACM Trans. Sens. Netw.* **2008**, *4*, doi:10.1145/ 1387663.1387670.

16. Goel, S.; Imielinski, T. Prediction-based monitoring in sensor networks: Taking lessons from mpeg. *ACM SIGCOMM Comput. Commun. Rev.* **2001**, *31*, 82–98.

17. Goel, S.; Passarella, A.; Imielinski, T. Using buddies to live longer in a boring world (sensor network protocol). In Proceedings of the Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops 2006), Pisa, Italy, 13–17 March 2006; pp. 342–346.

18. Alippi, C.; Anastasi, G.; Di Francesco, M.; Roveri, M. An adaptive sampling algorithm for effective energy management in wireless sensor networks with energy-hungry sensors. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 335–344.

19. Abbasi, A.A.; Younis, M. A survey on clustering algorithms for wireless sensor networks. *Comput. Commun.* **2007**, *30*, 2826–2841.

20. Li, C.; Zhang, H.; Hao, B.; Li, J. A survey on routing protocols for large-scale wireless sensor networks. *Sensors* **2011**, *11*, 3498–3526.

21. Lee, S.J.; Belding-Royer, E.M.; Perkins, C.E. Scalability study of the ad hoc on-demand distance vector routing protocol. *Int. J. Netw. Manag.* **2003**, *13*, 97–114.

22. Heinzelman, W.R.; Chandrakasan, A.; Balakrishnan, H. Energy-efficient communication protocol for wireless microsensor networks. In Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2000; p. 10.

23. Huang, L.; Wang, M.; Wu, J. Hierarchical routing algorithm for wireless sensor network. *Appl. Mech. Mater.* **2013**, *427*, 2416–2419.

24. Bari, A.; Jaekel, A.; Bandyopadhyay, S. Clustering strategies for improving the lifetime of two-tiered sensor networks. *Comput. Commun.* **2008**, *31*, 3451–3459.

25. Ma, Y.; Guo, Y.; Tian, X.; Ghanem, M. Distributed clustering-based aggregation algorithm for spatial correlated sensor networks. *IEEE Sens. J.* **2011**, *11*, 641–648.

26. Jianhua, L.; Baili, Z.; Li, X. A data correlation-based wireless sensor network clustering algorithm. In Proceedings of the 2010 International Conference on Computer Application and System Modeling (ICCASM), Taiyuan, China, 22–24 October 2010; pp. V8-61–V68-65.

27. Xu, Y.; Heidemann, J.; Estrin, D. Geography-informed energy conservation for ad hoc routing. In Proceedings of the 7th Annual International Conference on Mobile Computing and Networking, Rome, Italy, 16–21 July 2001; pp. 70–84.

28. Guestrin, C.; Bodik, P.; Thibaux, R.; Paskin, M.; Madden, S. Distributed regression: An efficient framework for modeling sensor network data. In Proceedings of the Third International Symposium on Information Processing in Sensor Networks (IPSN 2004), Berkeley, CA, USA, 26–27 April 2004; pp. 1–10.
29. Handy, M.; Haase, M.; Timmermann, D. Low energy adaptive clustering hierarchy with deterministic cluster-head selection. In Proceedings of the 4th International Workshop on Mobile and Wireless Communications Network, Stockholm, Sweden, 9–11September 2002; pp. 368–372.