國立交通大學

資訊工程學系

碩士論文

於無線微型感測網路減少重複資料之能源節省協定

A Redundancy-Aware Protocol for Power Saving in Wireless Sensor Networks

研 究 生：李卓育

指導教授：謝續平 博士

中華民國九十四年五月

於無線微型感測網路減少重複資料之能源節省協定

A Redundancy-Aware Protocol for Power Saving in Wireless Sensor Networks

研 究 生：李卓育　　Student: Zhuo-Yu Lee
指導教授：謝續平 博士　　Advisor: Dr. Shiuh-Pyng Shieh

國 立 交 通 大 學
資 訊 工 程 學 系
碩 士 論 文

A Thesis
Submitted to
Department of Computer Science and Information Engineering
College of Electrical Engineering and Computer Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master
in
Computer Science and Information Engineering

May 2005
Hsinchu, Taiwan, Republic of China

中 華 民 國 九 十 四 年 五 月

於無線微型感測網路減少重複資料之能源節省協定

研究生： 李卓育　　　　　　　　　　　指導教授： 謝續平 博士

國立交通大學資訊工程學系

摘要

在無線微型感測網路中，無線微型感測裝置為一資源有限的小型裝置，並且只擁有有限的能源。直接傳輸原始的資料會消耗大量的能源，因此常見的資料減少協定利用執行資料聚集程式或減少重複的資料來降低網路流量以達到節省能源的目的。然而，這些降低資料量的協定仍然有著需要大量運算、資料失真、應用程式的限制等等的缺點。在此論文中，我們提出一個適用於無線微型感測網路的減少重複資料協定，稱為REAP，來降低網路中的重複資料、節省能源，並且減緩上述的缺點。REAP有效地利用無線微型感測網路的特徵，空間與時間的關聯性，同時不會造成資料的失真。另外，在所觀測的環境有所大變動時，亦能保持極小的能源消耗。由我們所做的實驗顯示，REAP在一般的環境觀測下，能節省50%到60%的能源消耗。

A Redundancy-Aware Protocol for Power Saving in Wireless Sensor Networks

Student: Zhuo-Yu Lee Advisor: Shiuh-Pyng Shieh

Department of Computer Science and Information Engineering

National Chaio Tung University

Abstract

In *wireless sensor networks* (WSNs), sensor nodes are resource-constrained devices with limited energy. Transmitting raw data between sensor nodes consumes high energy, and therefore conventional data-efficient protocols reduce network traffic by performing intra-network aggregation functions or eliminating redundant data. However, some drawbacks still exist in these protocols, such as heavy computation overhead, data distortion, and application dependency. In this paper, we present a redundancy-aware protocol (REAP) for WANs to reduce redundant data, save energy savings, and relieve the drawbacks. REAP effectively utilizes the WSN characteristics of spatial and temporal correlations, and at the same time will not have data distortion. REAP can easily adapt to the change of monitored environment, while maintaining minimal energy cost. Simulation results showed that REAP can achieve 50% ~ 60% energy savings in monitoring physical phenomenon.

誌　謝

首先我要感謝我的指導教授謝續平教授兩年來的指導，在寫論文的過程中，不論是研究的方法、報告方式及寫作技巧，都給予我很大的協助。除了論文之外，也讓我在這兩年中學習到很多相關領域的知識。

我還要感謝實驗室的學長都樂意地跟我討論論文以及給我很好的建議。感謝李富源學長和黃士一學長，給我很多想法上的幫助，也感謝他們很有耐心的跟我討論我的論文。另外，我要感謝實驗室的同學，陳宗逸的 FTP 站讓我生活多采多姿，而且還讓我見識到如何成為一個好的推銷員。還有以後要理財，有吳孝展這位顧問可以幫忙，然後什麼事情他總是可以講很多。還有林亞正，因為每次他跟陳宗逸他們鬥嘴，我就覺得很搞笑。最後我們實驗室最有創業精神的邱建熹，以後發達了不要忘記我喔。最後最後，我們的助理 olga，她也是帶給我很多歡笑的一個寶。

最後，我要感謝我的家人，一路地支持我順利完成碩士的學業。此外，我也要感謝我女朋友在生活上的照顧以及她實驗室學長熱心的幫我解決一些相關的問題。還有 dsns lab 的成員，我永遠愛你們！

Table of Content

List of Figures

List of Tables

Chapter 1 Introduction

Wireless sensor networks (WSNs) are emerging technologies which is composed of several powerful base stations and hundred or thousand of small and autonomous device called sensor nodes. In WSNs, the deployed sensor nodes perform distributed observation locally and transmit their sense readings back to the base station over a tree structure. The sensor nodes are usually densely deployed to unattended environment to observe physical phenomenon such as temperature, humidity, light, precipitation, seismic intensity, and etc.

A sense node is resource-constrained with low computation power, limited storage, poor communication, low power capacity, and etc. Operated by battery makes power consumption an important issue for sensor nodes. The failure of a set of sensor nodes by energy depletion can lead to a partition of the sensor network and loss of important information for applications. There has been much research on the topic of energy-aware routing protocols in sensor networks [16], [17], [18], [19], [20], [21], and efficient data processing approaches [13], [14], [15] for reducing power consumption of sensor nodes and further increase the lifetime of the whole network.

1.1 Requirements

In this paper, we design a protocol which focuses on efficient data processing for reducing data volume rather than energy-efficient routing protocol. To design this protocol in wireless sensor networks may face some challenges:

- **Low computation overhead**

  In wireless sensor networks, all sensor nodes are resource-constrained devices. An energy-efficient protocol should try to minimize the computation overhead and save the maximal total energy in the network. Applying common data compression techniques to reduce data volume cost too much overhead for sensor nodes. Sensor nodes which perform encoding and decoding in these data compression techniques may drain their power quickly due to the complex computation. We give a discussion in section 2.3 to show that common data techniques are not suitable for wireless sensor networks. Therefore, designing an energy-efficient protocol should keep computation overhead in mind.

- **No data distortion**

  In some efficient-aware protocols [17], [21], [22], their schemes have some level of data distortion. Data distortion provides imprecise data for applications. These protocols with data distortion limited themselves to applications which need precise sensed data.

- **Fault Tolerance**

  One of the characteristics in wireless sensor network is unreliable links. Packet loss occurs more frequently than wired network due to limited bandwidth and interference. A mechanism is required for recovering the lost packet. Designing an energy-efficient protocol should take car of this factor.

1.2 Cluster-based wireless sensor networks

In wireless sensor networks, the resource constraints on sensor nodes make it an important challenge to develop efficient data processing techniques. It is inefficient to directly propagate the raw data which sensor nodes observed to the base station. Instead, raw data should be processed, aggregated locally, and reported back to the base station to avoid energy depletion. In recent research, a cluster-based network has been discussed in the literature to concretize this idea. In this architecture, the network is divided into small clusters and each cluster has one sensor node acting as *aggregator* which collects all sense readings of sensor nodes in the cluster and performs intra-network data aggregation. The sensor nodes which are the members of the cluster transmit their raw data to their own aggregator. Only the sensor nodes which are aggregators report their aggregated data to the base station. Therefore, the clustered-based networks can conserve energy depletion of the sensor nodes because that only aggregators need to propagate their collected data to the base station while other sensor nodes just transmit their raw data to their aggregator. Moreover, the aggregator nodes fuse the raw data into one single data by intra-network aggregation to reduce data volume and also benefit the power consumption for the whole networks. Note that if the aggregator node is chosen a priori and fixed throughout the system lifetime, it is clear that unlucky sensor node chosen to be the aggregator would die quickly. Thus the aggregator nodes should be rotated after a period of time in order not to drain energy of single sensor node.

1.3 Related Work

There have been many proposed energy-efficient schemes for wireless sensor networks in the literature. These schemes can be roughly classified into two categories. The first one is efficient-aware routing protocols, which can be further divided into three main categories, data-centric, hierarchical and location-based [3]. The other one category of energy-efficient schemes is efficient data processing which reduces the transmitted data through the networks to further reduce power consumption.

1.3.1 Energy-aware routing protocol

Directed diffusion [18] is a data-centric scheme using a naming scheme where data generated by a sensor node is named by attribute-value pairs. An interest is defined using a list of attribute-value pairs such as type of object, interval, duration, location, and etc. The interest is broadcast by a sink and each node receiving the interest performs caching for later use. The interest entry contains several gradients which are reply links to neighbors from which the interest was received. By using interest and gradients, directed diffusion enables sink and nodes to establish empirically good paths between them to achieve power saving.

LEACH [6] is one of the most popular hierarchical routing protocols for wireless sensor networks. In LEACH, a sensor network is divided into several clusters. A cluster contains several nodes and one cluster head. In this architecture, local cluster heads act as routers to route data from the members of the clusters to base stations. It will save energy because that data propagation is only performed by cluster heads and all other nodes transmit their data to their own cluster heads which are only one-hop away from them. There have been several research based on LEACH to further improve the performance of power saving.

Lindsey et al. proposed a power-efficient data gathering scheme for wireless sensor networks [19]. The basic idea is that the nodes in the networks are organized to form a chain to the base station. The chain is constructed by the sensor nodes themselves using a greedy algorithm staring from some nodes. After the chain is constructed, each node on the chain receives data from its last neighbor, aggregate with its own data, and then transmits the fused data to its next neighbor on the chain. Although it reduces power consumption by decreasing the number of transmissions and reception by using data aggregation, knowing the topology for each node can introduce significant overhead especially for dense networks.

1.3.2 Efficient data processing

Other energy-efficient schemes are focus on data processing which reduce transmitted data for power saving. Some approaches proposed novel data compression schemes rather than common compression techniques to reduce data size due to heavy overhead of these common compression techniques.

Chou and Petrovic et al. [14] proposed an adaptive signal processing approach to reduce power consumption in wireless sensor networks. The base station constructs a tree-based codebook for compression with side information and broadcast to all sensor nodes in the networks. Each sensor node compresses its observation according to the given codebook to reduce data size, and transmits to the base station which then decodes these compressed data with side information. There are some drawbacks in their scheme. The base station needs to continuously track and exploit existing correlations in sensor data for decoding with side information. Moreover, it is possible for their scheme to make a decoding error due to the correlation noise. As a result, it would have some level of data distortion in their scheme.

PINCO [15] presented a power saving scheme by reducing redundancy in the data collected by sensor nodes. Each node receives the measurement from its neighbor and performs data fusion. The prefix of observations which the same are combined together to reduce the transmitted data size. TiNa [22] utilized temporal coherency to reduce the amount of data transmitted by all nodes. Their scheme define a user-specified parameter, called *tct*, which specifies the degree of which the user tolerant to the change of sense readings. The larger *tct*, the greater degree of distortion of sense readings would be. Similar approach is also proposed in [21] which utilized two thresholds, hard threshold and soft threshold. Only the nodes whose sense readings are beyond the hard threshold and equal or greater than soft threshold transmit their sense readings. However, this approach is not good for applications which need to periodically report the sense reading of sensor nodes. Applications may receive no data because no data reaches the thresholds.

1.4 Contribution

In this paper, we proposed a data-efficient protocol in the wireless sensor networks to reduce transmitted data volume as well as providing the same level of quality of sense readings. Each sensor node in our proposed protocol can save energy efficiently and further increasing lifetime of the whole network. Moreover, no data distortion will occur in our proposed protocol so that it is independent to upper applications in wireless sensor networks. Our proposed protocol is easy to implement and suitable for any cluster-based wireless sensor networks which perform intra-network data aggregation.

1.5 Synopsis

The rest of the paper is structured as follows. In the next chapter, we introduce the preliminaries needed in our proposed protocol. Chapter 3 describes the environment which our protocol discussed and gives a brief description of our proposed protocol which consists of three phases. In chapter 4, we analyze our protocol and give simulations to show that our performance on power saving. Finally, we conclude our proposed protocol in chapter 5.

Chapter 2 Preliminaries

In this chapter, we give a brief description to show the key point of power consumption on a sensor node and introduce the data redundancy which our proposed protocol motivated by. We also show that common data compression techniques are not suitable for wireless sensor networks. Finally, decimal data presentation would be introduced.

2.1 Power consumption on sensor nodes

A sensor node is a small device which can be made up of four basic components, sensing unit, processing unit, transceiver unit, and power unit [2]. The main task of a sensor node is to detect event in environment and report its observation. A sensor node uses sensing unit to observe the physical phenomenon, processing unit for processing observed data locally, and then transmits the data by transceiver unit. Therefore, power consumption on a sensor node falls on sensing unit, processing unit, and transceiver unit.

Hill and Szewczyk et al. [1] measured the current per hardware components of their sensor platform developed at UC Berkeley. Their measurement shows that transceiver unit consumes the most energy (17mA for transmitting data and 9.5mA for receiving data) while processing unit and sensing unit cost only a little energy (5mA for processing unit and 1mA for sensing unit). The current consumption for transmitting a single radio message was measured in [4]. Their result also shows that radio transmission costs the most power consumption. Therefore, the usage of transceiver unit is crucial for power consumption on sensor nodes. To reduce power consumption on sensor nodes, we need to diminish the usage of transceiver unit.

Hill and Szewczyk et al. [1] gave us a result that power consumption is proportion to the number of transmitted/received bits where 4317.89nJ for transmitting a bit and 2028.66nJ for receiving a bit. Moreover, Heinzelman and Chandrakasan et al. [6] proposed the first order radio model which is used to calculate transmission costs and receiving cost:

Transmitting

$$E_{Tx}(k,d) = E(k)_{Tx-elec} + E_{Tx-amp}(k,d)$$

$$\begin{aligned} E_{Tx}(k,d) &= k \times E_{elec} + k \times E_{amp} \times d^2 \\ &= k \times \left(E_{elec} + E_{amp} \times d^2\right) \end{aligned}$$

Receiving

$$E_{Rx}(k) = E(k)_{Rx-elec}$$

$$E_{Rx}(k) = k \times \left(E_{elec}\right)$$

where k is the number of transmitted/received bits and d is the transmitted distance. E_{elec} is the power consumption of running transmitter/receiver circuitry and E_{amp} for transmitting amplifier. In these equations, they also show us that the power consumption is proportion to the number of transmitted/received bits. Therefore, we should decrease the number of transmitted/received bits to achieve the goal of reducing power consumption on sensor nodes.

2.2 Data redundancy in wireless sensor networks

Data redundancy among the sense readings is a characteristic of the wireless sensor networks especially for densely deployed networks. It can be summarized as two kinds of redundancy, spatial redundancy and temporal redundancy.

- **Spatial redundancy**

  In order to achieve sufficient coverage and fault tolerance, applications in wireless sensor networks require dense sensor deployment which each sensor node is close to each other. As a result, the sense readings of the spatially proximal sensor nodes may be the same or different in a small bounds. These sense readings are spatial data redundancy and can be reduced to diminish the communicated data volume in wireless sensor networks.

- **Temporal redundancy**

  Some applications in wireless sensor networks may require sensor nodes to periodically perform observation and report their sense readings back to the base station. The nature of the radiated physical phenomenon constitutes the temporal correlation between each consecutive sense reading of a sensor node. The current sense reading of a sensor node may be the same as or less different to the sense reading at last time.

Motivated by the redundancy in wireless sensor networks, we proposed a protocol diminishing redundant data to further reduce power consumption for sensor nodes.

2.3 Common data compression techniques

In common data compression techniques, there are two main types, which include lossy data compression and lossless data compression. In this section, we give an overview of these techniques and show that it is not suitable for applying these compression techniques in wireless sensor networks.

As implied by the name of lossy data compression, after applying lossy data compression to message, the message can never be recovered exactly as it was compressed before. Therefore, using lossy data compression would lead to data distortion which is significant drawback we mentioned before. Lossless data compression has no data distortion and can be further divided into two categories, run-length coding and variable-length coding. The basic idea of run-length coding is that encoding continuous identical symbols into one such symbol and the length of the symbols. For example, if the message is “AAAAABBB”, after encoding by run-length coding, it becomes “***5***A***3***B”. The performance of run-length coding is not good for wireless sensor networks because the sense reading is usually digital data. Variable-length coding includes Huffman coding, arithmetic coding, LZW coding, and etc. These coding algorithms need long-term statistics or complex computation. They are not suitable for decreasing data volume in wireless sensor networks due to heavy overhead. To sum up, common data compression techniques are not suitable for

Lossy compression	***Lossless compression***	
Data distortion!	**Run-Length Coding**	**Variable-Length Coding**
	AAAAABBB→5A3B	*Huffman coding* *Arithmetic coding* *LZW coding*

Table 2-1. Common data compression techniques

being used by sensor nodes no matter lossy or lossless compression. In our protocol, we need to find out a method for decreasing data volume without data distortion, long-term statistic, and heavy computation overhead.

In our proposed protocol, we make use of the concept of DPCM (Differential Pulse Code Modulation) to reduce the transmitted data volume. DPCM is code modulation which only samples the difference between the actual sample value and its predicted value. JPEG also uses DPCM to compress image data. DCMP needs only light computation and very is suitable for compressing highly correlated data. Therefore, our proposed protocol employs the concept of difference on transmitting.

2.4 Decimal data presentation

For many applications in wireless sensor networks, the data they are interested is not only integer, but also including decimal data. For example, monitoring temperature in physical environment usually needs as precisely as decimal below one for accuracy. Therefore, we need to know how to represent decimal data in sensor nodes. A brief study will be introduced in this section.

Let's consider an n-bit ADC which outputs digital data with n bits, *M[n]*, and its resolution, *r*. A sensor node gets the n bits data from the ADC and the actual value is:

$$sense\ reading = M[n] \times r$$

The resolution is usually decimal number, so the sense reading would be decimal number. We take a temperature ADC, AD7416, as an example. This temperature sensor is a 10-bit ADC and the resolution is 0.25˚C. The temperature data format is shown in Table 3-2. This shows that the full theoretical range of AD7416 is from -128˚C to 127˚C. If a sensor node reads the data 0000101000 from the ADC, that is to say that the sense reading is 40×0.25˚C = 10˚C.

Temperature(˚C)	Digital output
-128˚C	10 0000 0000
-125˚C	10 0000 1100
-100˚C	10 0111 0000
-75˚C	10 1101 0100
-50˚C	11 0011 1000
-25˚C	11 1001 1100
-0.25˚C	11 1111 1111
0˚C	00 0000 0000
+0.25˚C	00 0000 0001
+10˚C	00 0010 1000
+25˚C	00 0110 0100
+50˚C	00 1100 1000
+75˚C	01 0010 1100
+100˚C	01 1001 0000
+125˚C	01 1111 0100
+127˚C	01 1111 1100

Table 3-2. Temperature data format of AD7416

Our proposed protocol is compatible to decimal data for sensor nodes with n-bit ADC. Computing the base value and difference operation are performed by using the digital output from the ADC and no other overhead in REAP to dealing with decimal data for sensor nodes.

Chapter 3 Proposed Scheme

In this section, we describe our basic idea of our proposed scheme, a REdundancy-Aware Protocol, called *REAP*, and the detail of it. We also discuss the usages in REAP and give a summary of it.

3.1 Environments

Our discussed environment is a cluster-based network, which can reduce data size by performing intra-network aggregation and conserve energy depletion for the whole network. Figure 3-1 (a) illustrates a random deployed sensor network and the network is formed into a cluster-based network in Figure 3-1 (b). In Figure 3-1 (b), each circle with dotted line is a cluster and has one aggregator node within it. The sensor nodes which are the members of a cluster are one-hop away from their aggregator node and the aggregator node is the neighbor of all members in the cluster. Our network clustering is based on [6]. Each sensor node has a probability to be an aggregator. When each node that has elected itself an aggregator broadcasts an advertisement message to the rest of the sensor nodes. The non-aggregator nodes join the cluster based on the signal strength of the advertisement. A non-aggregator node would choose the aggregator node to be its aggregator where the signal strength of the aggregator node's advertisement is the strongest among all received advertisements. By this decision mechanism, sensor nodes in clusters are closest to their aggregator node. Therefore, it would have highly spatial correlation between sensor nodes in the same cluster.

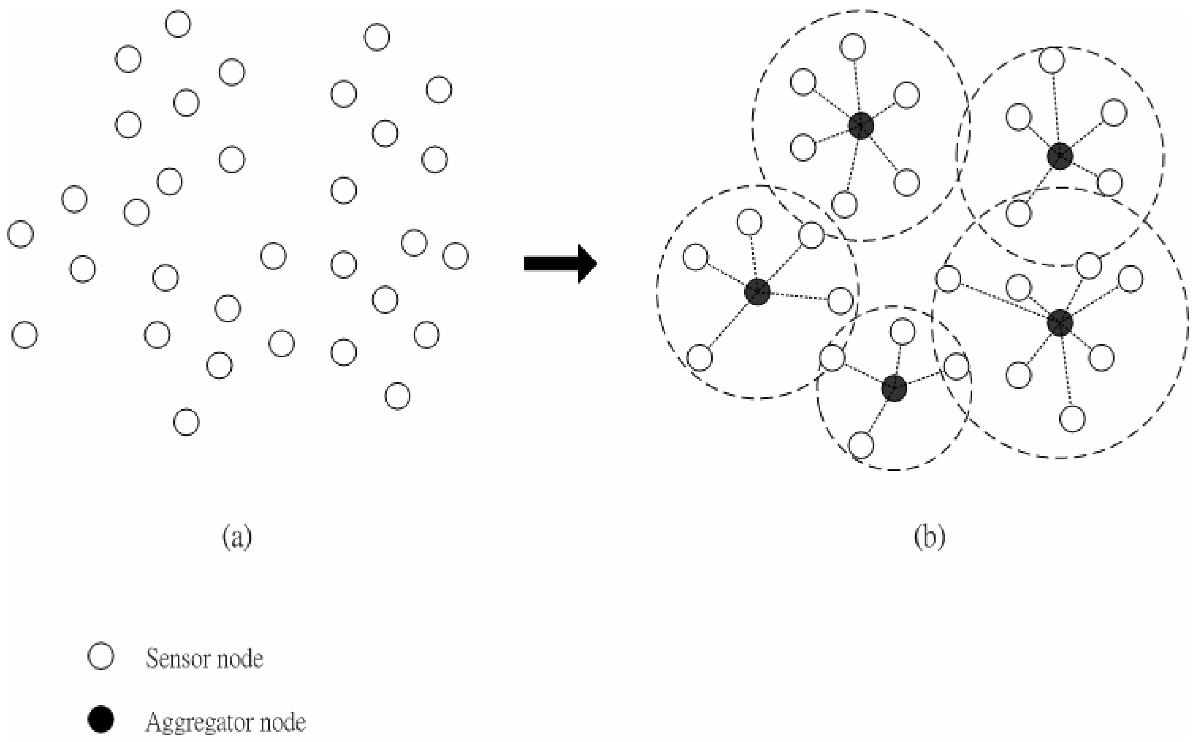


Figure 3-1. Cluster-based networks: (a) a random deployment sensor networks; (b) a cluster-based wireless sensor network after clustering

Each sensor node performs environment monitoring when receiving a query from the base station. The base station sends a query to each aggregator node and then the query is delivered to each sensor node in the cluster. The query may request sensor nodes to continuously sensing the environment and report the sense readings in intervals. Each aggregator node collects the sense readings from its members, performing intra-network data aggregation, and then reports the results to the base station. REAP is a protocol for applications to reduce transmitted data volume in each cluster, where the applications require continuously monitoring.

3.2 REAP: REdundancy-Aware Protocol

REAP is aiming at reducing transmitted data volume in the network to decrease the power consumption of each node. As we mentioned in chapter 1, the sense readings in densely deployed wireless sensor networks are likely to have both spatial and temporal redundancy.

We utilize data redundancy to reduce data volume.

In REAP, we focus on utilizing temporal redundancy rather than spatial redundancy. Spatial redundancy means that the sense reading of one sensor node may be the same as or less different to the sense readings of its neighbors. If a sensor node knows that its sense reading is the same as the sense readings of its neighbors, it should not transmit its sense reading because of the redundancy of the sense reading of its neighbors. To reduce such redundancy, sensor nodes need to overhear the sense readings of their neighbors and receive the sense readings for comparing with their own sense readings. Note that receiving data also consumes energy for sensor nodes. Receiving overhead may greater than the energy we saved by reducing spatial redundancy, so we do not concentrate on reducing spatial redundant data.

REAP makes use of the temporal redundancy, which the current sense reading of one sensor node is likely to be the same as the last sense reading or less different to it, especially in physical phenomenon. Motivated by this characteristic, a node only transmits the difference between the current sense reading and last sense reading in REAP. Because of temporal redundancy, the difference would be smaller than original data. Each time when a node needs to transmit data, it only transmits a small amount of data.

3.3 REAP in Detail

REAP consists of three phases, *base value setup phase*, *data report phase*, and *base value update phase*. First, we define some variables and give descriptions of them.

B	the base value used for reducing transmitted data volume
D	the difference between current sense reading and the base value
N_{cj}	The number of sensor nodes in a cluster, c_j
M[n]	the sense reading with n bits

f_b(M)	a function for computing the number of bits of M: $f_b(M) = \lfloor \log_2 M \rfloor + 1$

Table 3-1. Definitions for REAP

These variables will be used in the following sections. In the following, we give a detail description for each phase of REAP.

I. Base Value Setup Phase

In order to reduce transmitted data volume, we use the concept of difference for transmitting. One approach is that each sensor node transmits the difference between the current sense reading and last sense reading to reduce transmitted data, but this approach has four drawbacks. First, each sensor node transmits only the difference, so the aggregator node needs to record the last sense readings of all sensor nodes in the cluster and uses the difference and the last sense reading to recover the actual current sense reading for each sensor node. A large storage space is required for the resource-constrained aggregator node. Second, when the aggregator node receives the difference of one sensor node, the aggregator node needs to look up the table which records the last sensor readings of all sensor nodes in the cluster to find the corresponding sense reading of the sensor node. The table would be large if there are many sensor nodes in the cluster, so looking up table will be an overhead for the aggregator nodes. Third, in each cluster, the sensor node acting as the aggregator would rotates among the other high-energy sensor nodes in order to not drain the energy quickly [6]. The records of the last sense readings of all sensor nodes should be delivered to the new aggregator node to make this approach workable. Delivering the large-size records is also a heavy overhead and dissipates energy. Finally, packet loss may more usually occur due to the unreliable links in wireless sensor networks. Packet loss fatally influences on this approach because that the aggregator node can not recover the actual sense reading if the last sense reading was lost.

To avoid these drawbacks and also achieve the goal of reducing transmitted data volume, we use only one single value for all sensor nodes in the cluster, including the aggregator node, to reduce transmitted data volume. This value should be selected carefully to minimize transmitted data volume for all sensor nodes in the cluster. Let's consider that there are N sensor nodes which are the members of a cluster and their respective sense readings are, M_1, M_2, …, M_N. Assume the single value we want to find is x. By the concept of transmitting the difference, we can formulate the total amount of data which all sensor nodes need to transmit in the cluster:

$$f(x) = \sum_{i=1}^{N} \left| (x - M_i) \right| \tag{1}$$

Mathematically, equation (1) is called *mean absolute deviation*. We give a simple proof in appendix to prove that when $f(x)$ is minimal. The result shows that choosing x as median of $<M_1, M_2, \ldots, M_N>$ makes $f(x)$ is minimal. Consequently, we should choosing x as the median of $<M_1, M_2, \ldots, M_N>$. However, REAP chooses x as the *mean* of $<M_1, M_2, \ldots, M_N>$ and we call this value as *base value*, which is known for all sensor nodes in the cluster. Choosing the base value as mean is more meaningful than median for REAP. Although mean is not optimal solution for equation (1), REAP considers only about total number of bits of all differences. The total number of number would be close to the case which choosing the base value as median. Moreover, choosing the base value as mean helps REAP to update the base value more precisely than choosing median because that mean represents the average value of all sense reading and thus can further be the criterion for updating base value in base value update phase.

In REAP, the aggregator nodes are responsible for creating the base value. The base value is computed by the following equation:

$$B = \frac{\sum_{i=1}^{N_c} M_i[n]}{N_c} \qquad (2)$$

where N_c is the total number of sensor nodes in one cluster. Figure 3-2 illustrates how to setup the base value. When a cluster is formed, all sensor nodes transmit their current sense readings to the aggregator node. The aggregator node then computes the base value by the equation (2) and broadcasts the base value to all sensor nodes. Therefore, all sensor nodes in the cluster have the base value and further use it to reduce transmitted data volume by transmitting only the difference between the current sense reading and the base value.

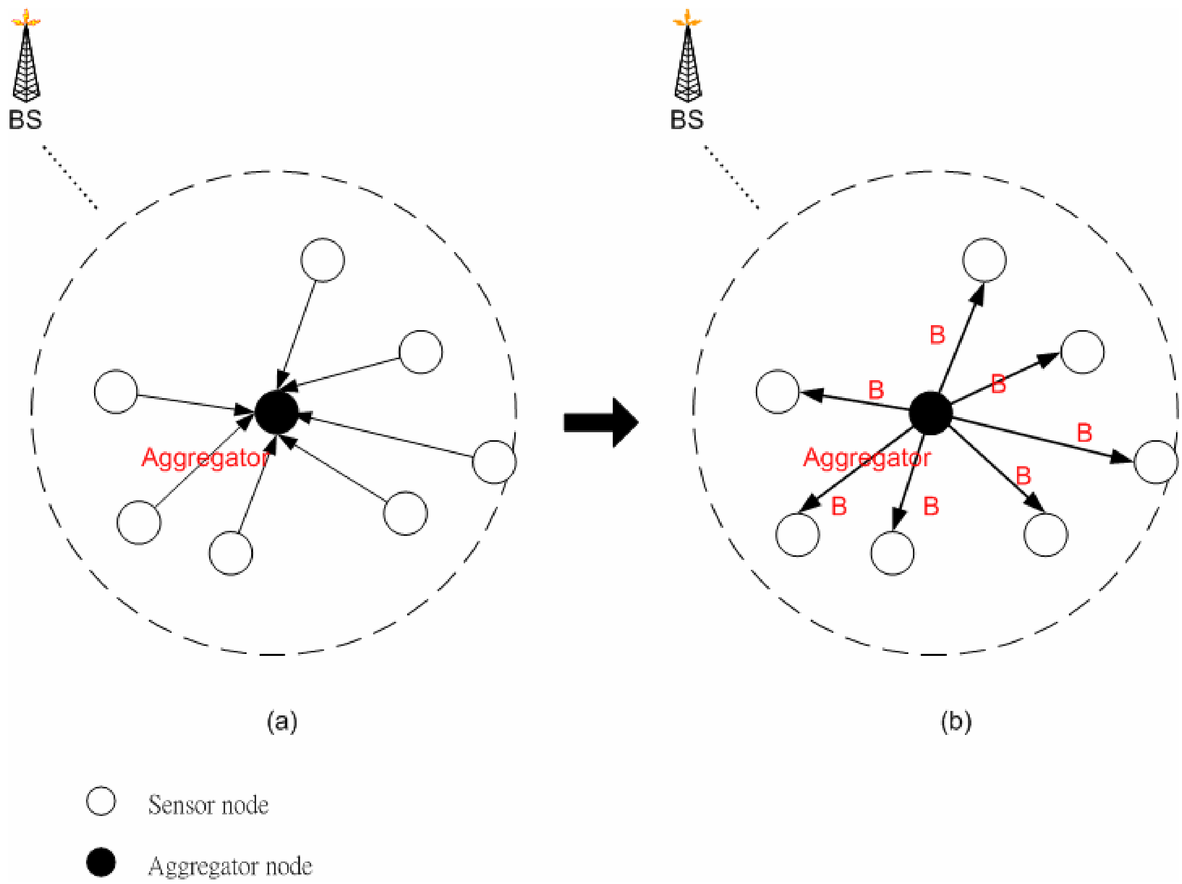


Figure 3-2. Update of the base value: (a)the aggregator node collect all sense readings; (b) The aggregator node broadcast the base value to all sense nodes in the cluster

Base value setup phase initiates the base value since the cluster is formed. Further data transmission will utilize the base value to reduce transmitted data volume. Under this design, only one single value is used for all nodes and avoids the overheads of storage space, table looking up, table delivering, and against error recovery for sense readings.

II. Data Report Phase

After the initialization of the base value, all sensor nodes in the cluster report their sense readings to the aggregator node for performing intra-network data aggregation when receiving a query from the base station. Figure 3-3 shows a procedure of one sensor node reporting its sense reading.

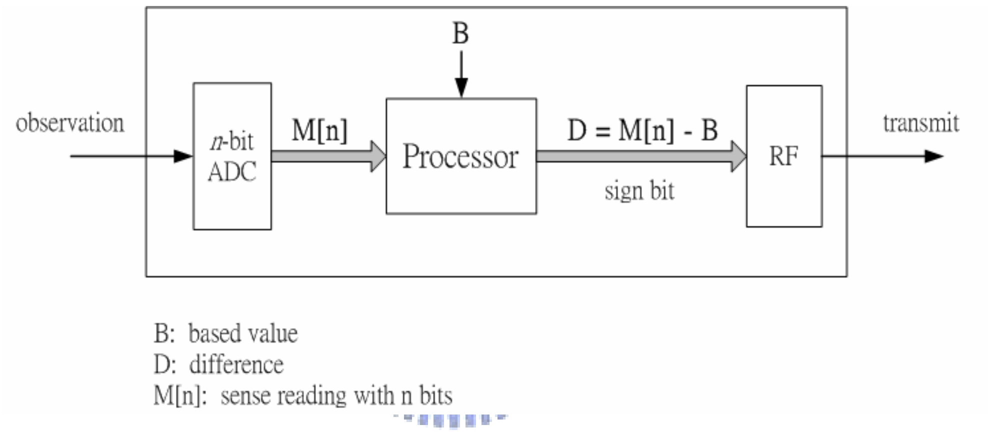


Figure 3-3. Procedure of transmitting a sense reading for a sensor node

A sensor node gets the *n*-bit sense reading, $M[n]$, from the *n*-bit analog-to-digital converter (ADC) and then computes the difference, D, by the *n*-bit sense reading and the base value, B, which is setup at base value setup phase.

$$D = M[n] - B \tag{3}$$

The sensor node transmits the value, D, as its sense reading to the aggregator node. Note that M may be bigger or smaller than B, so we need one bit to tell the aggregator node that D is positive or negative. We call this bit as sign bit, which 1 represents that D is negative and 0 for positive. Moreover, after performing the equation (3), D may be 0. If D is equal to 0, the sensor node only transmits the sign bit, 0, to tell the aggregator node that its sense reading is the same as the base value.

The aggregator node would receive all sense readings from all sensor nodes. When the aggregator node receives one packet from senor node, n_i, it extracts the sense reading, D_i, from the packet and checks the sign bit for recovering the actual sense reading. If the signed bit is 1, the aggregator node subtracts D_i from the base value to recover the actual sense reading while adding D to the base value if the signed bit is 0.

$$M_i = \begin{cases} B + D_i & if \ s_i = 0 \\ B - D_i & if \ s_i = 1 \end{cases} \tag{4}$$

The aggregator node only performs simple subtraction and addition instruction to recover the actual sense reading for each sensor node. This method is both lightweight for data reduction and data recovery, so it is appropriate for resource-constrained sensor nodes.

The links between sensor nodes are unreliable in wireless sensor networks. The data transmitted by sensor nodes may not be received by the aggregator node. We will give a discussion about this problem in section 3.4.

III. Base Value Update Phase

In REAP, all sensor nodes in the cluster keep the base value for reducing transmitted data volume. However, physical phenomenon may change as time. If the sense readings of all sensor nodes are changed dramatically after a long period of time, the differences between the sense readings and the base value are large if the base value is unchanged. If the sense readings of all sensor nodes in a cluster are increasing, the performance of our protocol is still better than transmitting raw sense readings. However, if the sense readings become less than half of the base value, the performance of REAP will be worse than it without our scheme because the difference between of the sense readings and the base value is larger than the original sense reading. Energy consumption will increase when the sense readings are far away from the base value. Therefore, the base value should be updated to fit to the current physical phenomenon. In base value update phase, the aggregator node will update the base value and broadcast the new base value to all sensor nodes in the cluster to maintain the performance of REAP.

A straightforward approach can be introduced, which is updating the base value at each round when the aggregator node collects all sensor readings from all sensor nodes. Note that it also consumes energy when updating the base value due to the transmission of the aggregator node and reception of all other sensor nodes in the cluster. The power consumption for updating the base value can be calculated by the following equation:

$$E_{overhead} = E_{Tx} \times (f_b(B_{new})) + E_{Rx} \times (f_b(B_{new})) \times (N_c - 1) \tag{5}$$

where E_{Tx} and E_{Rx} are the power consumption of transmission and reception for one bit. B_{new} is the new base value at current round collecting sense reading, and the $f_b()$ is the function for computing the number of bits. Let's consider the total energy saved at the next collection round after updating the base value, which is compared with the case without updating. The extra energy saving after updating the base value can be described as:

$$\begin{aligned} E_{saved} &= E_{Tx} \times f_b\left(\left|B_{original} - B_{next}\right|\right) \times (N_c - 1) - E_{Tx} \times f_b\left(\left|B_{new} - B_{next}\right|\right) \times (N_c - 1) \\ &= E_{Tx} \times \left(f_b\left(\left|B_{original} - B_{next}\right|\right) - f_b\left(\left|B_{new} - B_{next}\right|\right)\right) \times (N_c - 1) \end{aligned} \tag{6}$$

where B_{next} is the base value at the next collection round and $B_{original}$ is the original base value which is current used by all sensor nodes. If the overhead of updating the base value is larger than the extra energy saving at next collection round,

$$E_{overhead} > E_{saved} \quad (7)$$

the power consumption in the cluster will worse than it without updating the base valued at each collection round. The equation (7) would be easily true if high temporal correlation exists in the cluster (i.e. E_{saved} is very small). Therefore we should not update the base value at each collection round. In REAP, the aggregator node updates the base value only when the base value at next round is far away the current original base value. The updating criterion is:

$$\frac{\left| f_b(B_{new}) - f_b(B_{original}) \right|}{f_b(B_{original})} > T_{threshold} \quad (8)$$

If $T_{threshold}$ is small, the aggregator node would update the base value more frequently while larger $T_{threshold}$ makes the update of the base value less unusual. We give a simple analysis for $T_{threshold}$. If the base value at next round is larger than the current base value, the performance of REAP is stiller better because that the differences are still smaller than original raw data. On the other hand, if the base value at next round is smaller than half of the current base value, the performance would be worse than protocol without REAP because that the differences are larger than original raw data. Therefore, $T_{threshold}$ should be smaller than 1/2. Choosing $T_{threshold}$ is dependent on environments and we suggest that choosing $T_{threshold}$ between 0.2 ~ 0.3 would get better performance in most environments.

3.4 Error Reconstruction

In our framework, packet loss may occur due to the unreliability in wireless sensor networks. A packet transmitted from a sensor node in the cluster may not be received at the aggregator node. Losing packets makes it imprecise for applications because that this packet may impact the result of intra-network aggregation. Moreover, it also affects the computation of the base value. Imprecise base value may result in fewer energy savings. In such a case, the aggregator node should recover the corresponding sense reading of the lost packet for computing the base value precisely. In the dense sensor networks, it has highly spatial correlation between sensor nodes. The sense reading of a small set of sensor nodes which are neighbors may have some level of correlation. If we can find this correlation, the aggregator node can reconstruct the lost sense reading of the sensor node by this correlation between the neighbors of the sensor node. Equation (9) shows a model for spatial correlation in wireless sensor networks [11]:

$$K_{\vartheta}^{PE}(d) = e^{(-d/\theta_1)^{\theta_2}}; \quad \theta_1 > 0,\ \theta_2 \in (0,2] \qquad (9)$$

This model is a power exponential function which contains the exponential (θ_2=1) and squared exponential (θ_2=2) model. The parameter θ_1 controls the relation between the distance of the sensor nodes and the correlation model. The sensor nodes become more correlated with increasing θ_1 . Using this correlation model, the aggregator nodes need to know the distances of two neighbors, so localization techniques [25] are required for providing location information for aggregator nodes. It consumes energy for using these proposed location techniques so that it conflicts with the goal of energy saving. In REAP, we do not using the spatial correlation model for error recovery due to the overhead of localization. Instead, we ignore the lost packets when computing the base value. The aggregator node computes the base value base on equation (2) where only using the total received sense readings. We will show the influence of packet lost on REAP in chapter 4.

3.5 Summary

In our proposed protocol, REAP, it consists of three main phases: *base value setup phase, data report phase*, and *base value update phase*. The base value setup phase is to initialize the base value for further being used in data report phase. In data report phase, sensor nodes transmit only the difference between the base value and their sense reading for each collection round. The base value may not suitable for reducing data volume when the difference between the base value and the sense

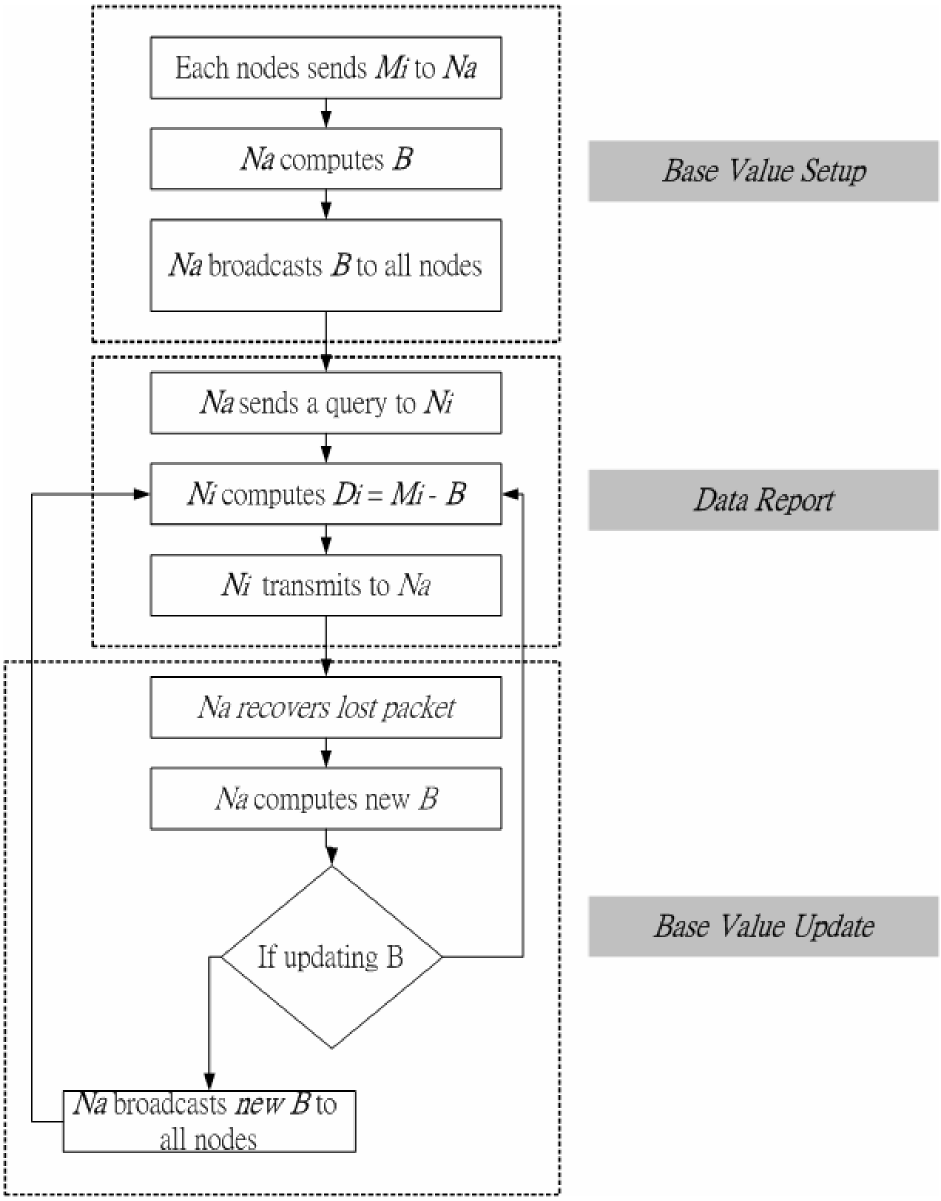


Figure 3-4. Protocol flow of REAP

readings. The aggregator nodes may update the base value in base value update phase to make the differences in a small range for maintaining the good performance in REAP. Figure 3-4 gives a review of REAP and shows a simple flow of the three main phases where N_a is the aggregator node and N_i is node i.

Chapter 4 Performance Analysis

In this chapter, we provide simulation results based on REAP. The simulations were performed for the measurements on temperature, humidity, and pressure. We measured the energy savings compared with original data gathering protocol without REAP. In the following section, we use the term of “original protocol” to represent the protocol where sensor nodes transmit their raw sense readings to aggregator nodes.

4.1 Simulation environment setup

In the simulations, we measured the performances on temperature, humidity, and pressure. The ranges of these data sets are shown in Table 4-1. The range of temperature is decreasing from 10˚C to 8˚C, 60% increasing to 99% for humidity, and 980hpa increasing to 1160hpa for pressure. The respective resolutions, which we used, of these data sets are 0.25, 1, and 0.25. The changes of the value in the data sets are smooth for temporal correlation between sense readings.

Data sets	Approximately range	Data resolution
Temperature	10˚C ~ 8˚C	0.25
Humidity	60% ~ 99%	1
Pressure	980hpa ~ 1160hpa	0.25

Table 4-1. Data set for the simulation

We simulated REAP over 100 samples of temperature, humidity, and pressure for each sensor node. We define it as a *round* when the aggregator collects all sense readings of sensor nodes one time. Each cluster has one aggregator and 10 sensor nodes to be its members, so the total number of samples in our simulation is 1,100. The power consumption for transmitting and receiving one bit are 4317.89nJ and 2028.66nJ that are measured by [1].

4.2 Simulation results

In the simulations, we measure the amount of totally consumed energy of each round. Note that reception also consumes energy, so consumed energy was calculated including the transmission and reception costs of sensor nodes. We measured the power consumption of each round in REAP with $T_{threshold} = 0.3$ compared with original protocol. In REAP, updating the base value also costs energy as equation (5) described, so the measurement of REAP contains the energy overhead of updating the base value.

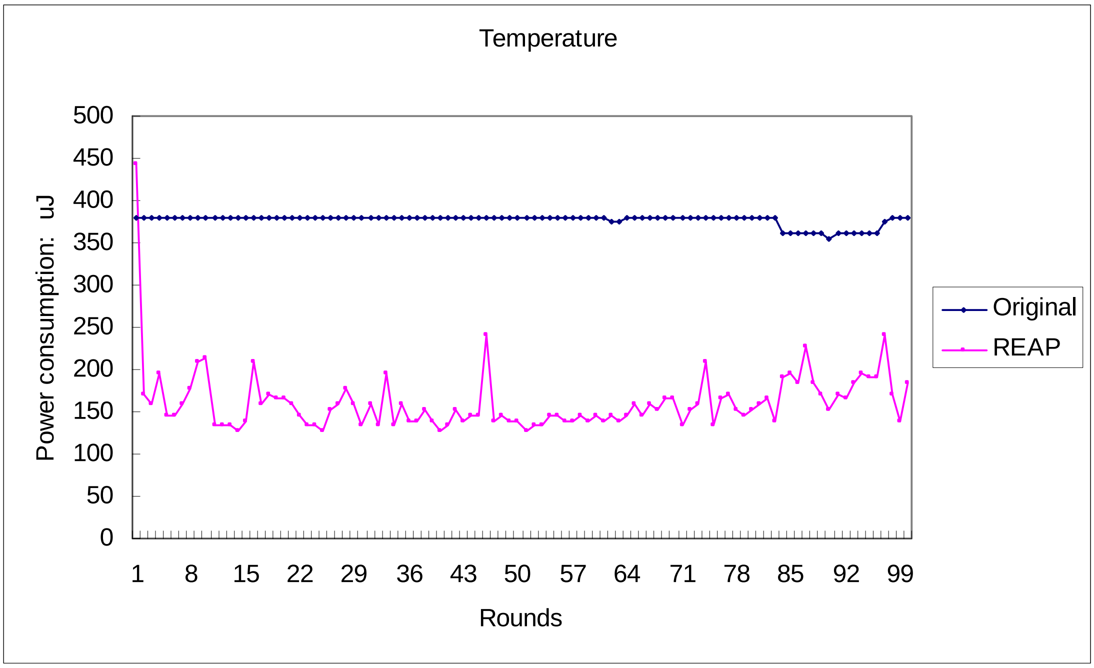


Figure 4-1. Energy consumption for temperature

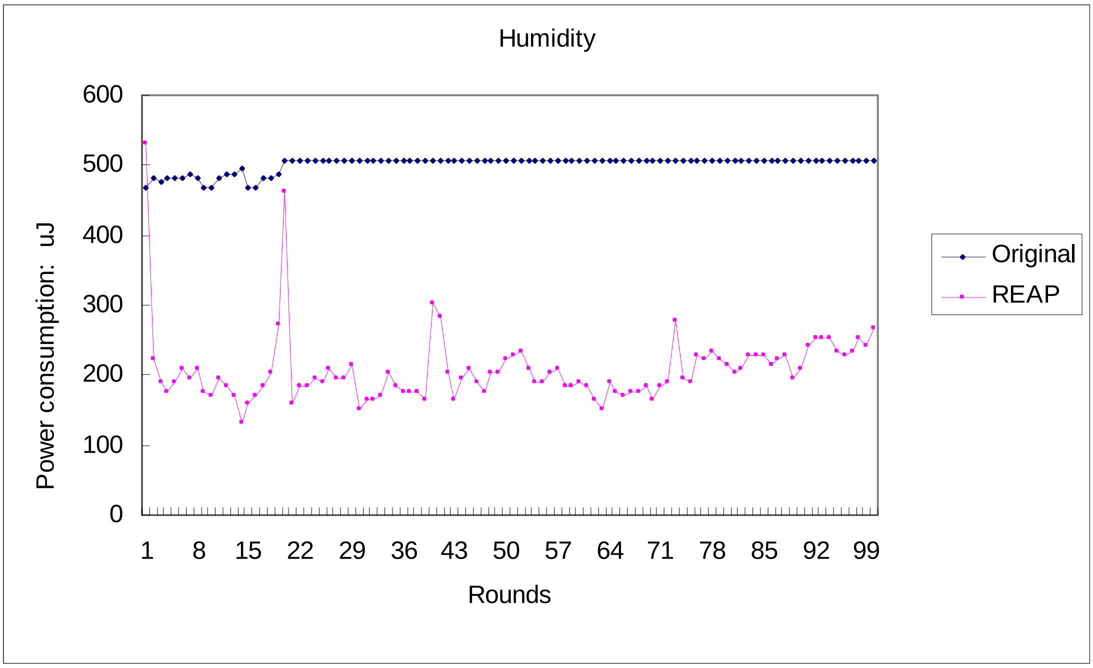


Figure 4-2. Energy consumption for humidity

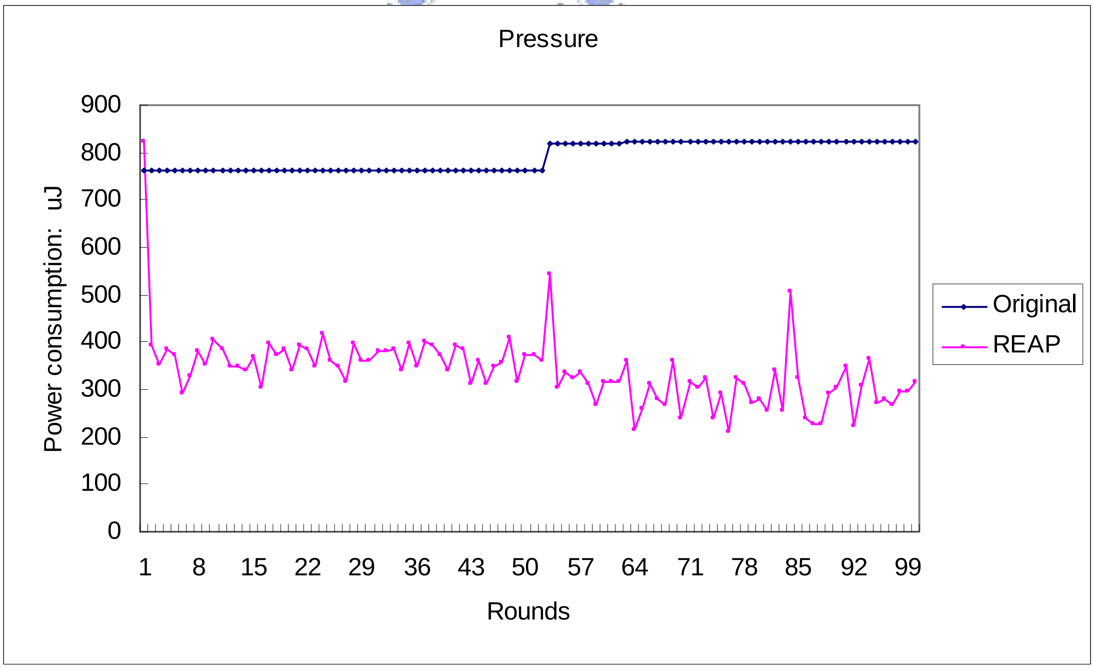


Figure 4-3. Energy consumption for pressure

Figure 4-1, 4-2, and 4-3 show the results of our simulations. It is clear that power consumption for each round in our proposed protocol is much less than original protocol. In our simulations, the aggregator node needs to calculate the base value and broadcast it to all sensor nodes in the cluster for initializing the base value, so the power consumption at the first round would larger than original protocol. However, after the first round, REAP substantially reduced power consumption at each round.

The following table, Table 4-2, shows total energy savings for REAP measuring temperature, humidity, and pressure. From our simulations, we achieved between 50% ~ 60% energy savings depending on the data set. We believe that REAP can achieve

Data Set	Temperature	Humidity	Pressure
Total energy savings	57.34%	58.59%	57.12%

Table 4-2. Energy savings of REAP for different data sets

more energy savings in more highly temporal and spatial correlated data sets where the difference can be smaller for reducing much power consumption.

4.3 REAP with packet loss

In this simulation, we measured the performance of REAP with packet loss. Packet loss may impact the computation of the base value and further influence the performance. Figure 4-4 shows the result of our simulation with packet loss.

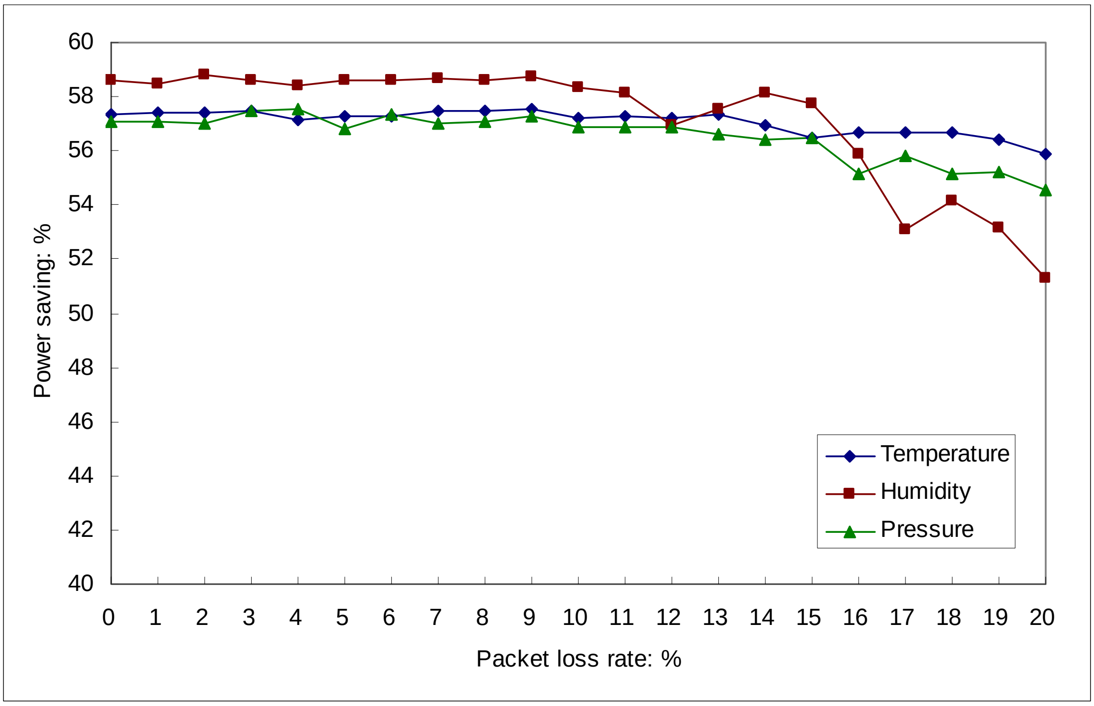


Figure 4-4. REAP with packet loss for different data sets

We measured the performance with packet loss rate from 0% to 20%. We can see that the performance is even better than it without packet loss in some conditions. This is because that ignoring the lost packets when computing the base value may make the base value closer to the sense readings. Due to the larger variation of data set, the performance for humidity decreases strikingly when packet loss rate increasing. The result shows that the performance is decreasing notably when packet loss rate is greater than 15%. However, the performance is strongly against to packet loss when the rate is less than 15% for the data sets. We believe that applying more precise interpolation techniques on REAP makes REAP more robust to packet loss.

Chapter 5 Conclusions

In this paper, we have proposed a data-efficient protocol called REAP. REAP reduces power consumption in wireless sensor networks by eliminating redundant transmitted data volume and provides no data distortion for upper applications. Since transmitted data consumes the most energy, using REAP results in significant energy savings. REAP mitigates the drawbacks of existing data-efficient protocols, including heavy computation overhead, data distortion, and application dependency. Our simulations have shown that REAP achieved about 50% ~ 60% energy savings for different data sets and provide a updating mechanism in REAP to maintain the performance even if physical phenomenon changes as time. REAP is suitable for any cluster-based networks and we believe that applications using REAP would save more energy in highly spatially and temporally correlated physical phenomenon.

Reference

[1] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. Culler, and K.Pister, “System architecture directions for network sensors,” in *proceedings of the 9th International Conference on Architectural Support for Programming Languages and Operating Systems*, Cambridge, MA, USA, 2000.

[2] Akyildiz, I.F., W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A Survey on Sensor Networks,” *IEEE Communications Magazine* ,102-114, 2000.

[3] K. Akkaya and M. Younis, “A Survey on Routing Protocols for Wireless Sensor Networks,” in *Elsevier Ad Hoc Network Journal*, 2003.

[4] V. Shnayder, M. Hempstead, B. rong Chen, G. W. Allen, and M. Welsh, “Simulating the power consumption of large-scale sensor network applications,” in *proceedings of the 2nd International Conference on Embedded Networked Sensor Systems (SenSys 2004). ACM Press*, 2004.

[5] G. J. Pottie and W. J. Kaiser, “Wireless integrated network sensors,” in *Comm. ACM*, volume 43, pages 551-558, 2000.

[6] W. Heinzelman, A.P. Chandrakasan, and H. Balakrishnan, “Energy-Efficient Communication Protocol for Wireless Sensor Networks,” in proceedings of Hawaiian International Conference on Systems Science, 2000.

[7] A. Jindal and K. Psounis, “Modeling spatially-correlated sensor network data,” *Technical report, CENG-2004-10, University of Southern California, 2004*.

[8] S. Pattem, B. Krishnamachari, and R. Govindan, “The impact of spatial correlation on routing with compression in wireless sensor networks,” in *ACM/IEEE International Symposium on Information Processing in Sensor Networks (IPSN)*, (Berkeley, CA), 2004.

[9] S. Yoon and C. Shahabi, “An Energy Conserving Clustered Aggregation Technique

Leveraging Spatial Correlation," in *IEEE SECON poster*, 2004.

[10] Y. Gao, K. Wu, and F.Li, "Analysis on the Redundancy of Wireless Sensor Networks," in *proceedings of the Second ACM International Workshop on Wireless Sensor Networks (WSNA)*, pages 108 – 114.

[11] M. C. Vuran, O. B. Akan, and I. F. Akyildiz, "Spatio-Temporal Correlation: Theory and Applications for Wireless Sensor Networks," in *Computer Networks Journal (Elsevier)*, vol. 45, no. 3, pp. 245-261, 2004. .

[12] S. Tilak, N. B. Abu-Ghazaleh, and W. Heinzelman, "A Taxonomy of Wireless Micro-sensor Networks Models," in *ACM Mobile Computing and Communications Review (MC2R)*, 2002.

[13] J. Chou, D. Petrovic and K. Ramchandran, "Tracking and exploiting correlations in dense sensor networks," in *proceedings of Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2002.

[14] J. Chou, D. Petrovic and K. Ramchandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks," in *proceedings of the Conference on Computer Communications (INFOCOM)*, San Francisco, CA, April 2003.

[15] T. Arici, B. Gedik, Y. Altunbasak, and L. Liu, "PINCO: a Pipelined In-Network COmpression Scheme for Data Collection in Wireless Sensor Networks," in *proceedings of the 12th International Conference on Computer Communications and Networks (IEEE ICCCN'03)*, 2003.

[16] W. Heinzelman, J. Kulik and H. Balakrishman, "Adaptive protocols for information dissemination in wireless sensor networks," in proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'99), 1999.

[17] A. Manjeshwar and D. P. Agrawal, "APTEEN: A Hybrid Protocol for Efficient Routing and Comprehensive Information Retrieval in Wireless Sensor Networks," in *Proceedings of the 2nd International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile computing*, Ft. Lauderdale, FL, 2002.

[18] C. Intanagonwiwat, R. Govindan, and D. Estrin, “Directed diffusion: A scalable and robust communication paradigm for sensor networks,” in proceedings of 6th ACM/IEEE Mobicom Conference, 2000.

[19] S. Lindsey, C. S. Raghavendra, “PEGASIS: Power Efficient GAthering in Sensor Information Systems,” in proceedings of IEEE Aerospace Conference, 2002.

[20] S. Lindsey, C. S. Raghavendra and K. Sivalingam, “Data Gathering in Sensor Networks using the Energy*Delay Metric,” in *proceedings of the IPDPS Workshop on Issues in Wireless Networks and Mobile Computing,* San Francisco, CA, April 2001.

[21] A. Manjeshwar and D. P. Agrawal, “TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks,” in *1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing (IPDPS)*, 2001.

[22] M. A. Sharaf, J. Beaver, A. Labrinidis, and P. K. Chrysanthis, “TiNA: A Scheme for Temporal Coherency-Aware in-Network Aggregation,” in *proceedings of the 3rd ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE)*, San Diego, California, pages 69-76, 2003.

[23] B. Krishnamachari, D. Estrin, and S. Wicker, “Impact of data aggregation in sensor networks,” in *International Workshop on Distributed Event-Based Systems*, 2002.

[24] N. Bulusu, J. Heidemann and D. Estrin, “GPS-less Low Cost Outdoor Localization for Very Small Devices,” *IEEE Personal Communications, Special Issue on "Smart Spaces and Environments,"* Vol. 7, No. 5, pp. 28-34, 2000.

Appendix

We give a simple proof to show that the following equation, called *mean absolute deviation*, gets the he minimum value when x is the *median*.

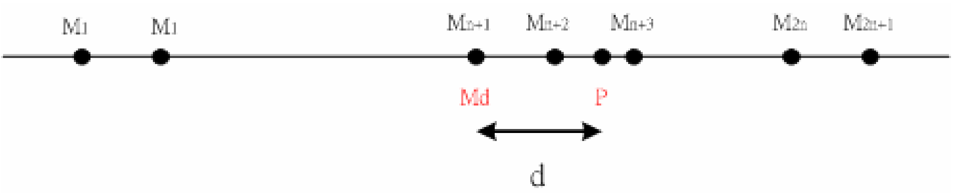
$$f(x) = \sum_{i=1}^{n} \left| (x - M_i) \right|$$

Proof:

We assume that M_d is the median of <$M_1, M_2, \ldots, M_n$> and P is the any value of <$M_1, M_2, \ldots, M_n$> in where $P \neq M_d$. We want to prove that for any P in <$M_1, M_2, \ldots, M_n$>, the following equation is always true.

$$\sum_{i=1}^{n} |M_i - M_d| < \sum_{i=1}^{n} |M_i - P|$$

We first prove the case when n is odd, where $N = 2n+1$. Thus, we have the sequence, <$M_1, M_2, \ldots, M_n, M_{n+1}, M_{n+2}, \ldots, M_{2n}, M_{2n+1}$ >, where $M_{i+1} > M_i$, $M_d = M_{n+1}$, $P \neq M_d$, and considering the case of $P > M_d$ as the following showing.



We assume that the distance between M_d and P is d and $P = M_d + d$ ($P = M_d - d$ can be proved similar by the following), then we have:

$$\left.\begin{aligned}
&|M_1 - M_d| = |M_1 - P| - d \\
&|M_2 - M_d| = |M_2 - P| - d \\
&\ldots\ldots \\
&|M_n - M_d| = |M_n - P| - d \\
&|M_{n+1} - M_d| = |M_{n+1} - P| - d
\end{aligned}\right\} n+1 \ \ items$$

$$\left.\begin{aligned}
&|M_{n+2} - M_d| < |M_{n+2} - P| + d \\
&|M_{n+3} - M_d| = |M_{n+3} - P| + d \\
&\ldots\ldots \\
&|M_{2n} - M_d| = |M_{2n} - P| + d \\
&|M_{2n+1} - M_d| = |M_{2n+1} - P| + d
\end{aligned}\right\} n \ \ items$$

Summing up the above equations:

$$\sum_{i=1}^{n}\left|M_i - M_d\right| < \sum_{i=1}^{n}\left|M_i - P\right| - (n+1)d + nd$$

$$= \sum_{i=1}^{n}\left|M_i - P\right| - d$$

$$\therefore \sum_{i=1}^{n}\left|M_i - M_d\right| < \sum_{i=1}^{n}\left|M_i - P\right|$$

The same principle can be also proved when *n* is even.