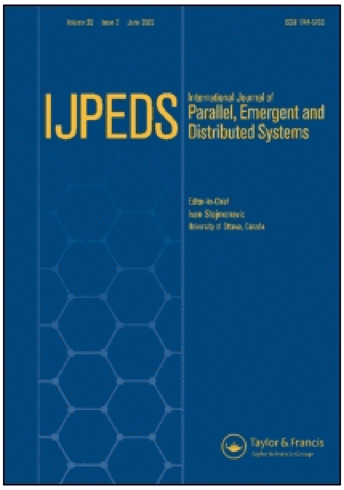

International Journal of Parallel, Emergent and Distributed Systems
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713729127

Energy-optimal grid-based clustering in wireless microsensor networks with data aggregation

Yanyan Zhuang [a]; Jianping Pan [a]; Guoxing Wu [a]
[a] University of Victoria, Victoria, BC, Canada

First published on: 16 February 2010

International Journal of Parallel, Emergent and Distributed Systems
iFirst article, 2010, 1–20

Taylor & Francis
Taylor & Francis Group

Energy-optimal grid-based clustering in wireless microsensor networks with data aggregation[1]

Yanyan Zhuang[a], Jianping Pan[a]* and Guoxing Wu[b]

[a]*University of Victoria, Victoria, BC, Canada;* [b]*Southeast University, Nanjing, Jiangsu, China*

(*Received 1 September 2009; final version received 8 September 2009*)

Wireless microsensor networks usually consist of a large number of small sensor nodes with limited onboard energy supply and deployed densely in a given area for information harvesting purposes. To reduce energy consumption and prolong network lifetime, clustering techniques are often used, among which the grid-based ones are very popular due to their simplicity and scalability. In this paper, we analyse and evaluate the energy-optimal grid size for a grid-based clustering and routing scheme proposed specifically for wireless microsensor networks. In addition, we also consider the effect of data aggregation on energy consumption and network lifetime. Through numerical and simulation results, we reveal the trade-off generic to all grid-based clustering schemes. Further, we propose a randomised technique to prolong the network lifetime and discuss other energy-saving opportunities. This paper provides some insights into the intrinsic limits of grid-based clustering schemes for wireless microsensor networks.

Keywords: wireless microsensor networks; data aggregation; grid-based clustering and routing; energy consumption; network lifetime

1. Introduction

Recent technology advancement has made sensor miniaturisation possible and affordable for real-world applications. Wireless microsensor networks, with a large number of small sensor nodes, have witnessed an increasing level of popularity in recent years and have revolutionised the way how information is collected and processed. Microsensor networks differ from other forms of wireless networks in their limited on-board energy supply and the large volume of data they are expected to transmit. Energy conservation therefore is of the primary concern in wireless microsensor networks for typical applications such as environment control and traffic monitoring. Due to this tight energy constraint, one major design challenge in wireless microsensor networks is to reduce the energy consumption or to increase the operational lifetime of a network as much as possible.

By dividing the entire sensor network into small clusters for easy management, and by putting the redundant sensor nodes in the same cluster into the sleep state to save energy, clustering schemes are promising for wireless microsensor networks due to their good scalability and energy conservation potentials. By using geographic coordinates for clustering and routing, grid-based schemes are particularly popular due to their simplicity. In fact, several clustering-based protocols have already been proposed for wireless

*Corresponding author. Email: pan@uvic.ca

ISSN 1744-5760 print/ISSN 1744-5779 online

DOI: 10.1080/17445760903548291
http://www.informaworld.com

microsensor networks, such as LEACH [5], two-tier data dissemination (TTDD) [8], EEDD [15] and our previous work [16].

However, one problem in grid-based clustering is how to determine a *suitable* grid size. Significant energy savings can be achieved when more nodes are put into the sleep state, therefore a larger cluster is preferable if the coverage and connectivity are still maintained. On the other hand, airborne radio transmissions are attenuated by a path loss factor scaling with the distance in a greater-than-linear fashion [6], and the total transmission energy can be reduced by dividing a long-distance transmission into several shorter ones. The problem is how to determine the optimal transmission range or grid size for energy efficiency, i.e. using the least amount of energy for data transmission while still allowing many nodes to go to sleep. Some work have been done in one-dimensional networks, such as [2,3], which give us the inspiration to model and optimise the energy consumption in two-dimensional networks. Recent work in two-dimensional networks [1] does not consider signal attenuation during wireless transmission. Moreover, it is based on a simple clustering and coordination scheme, which involves a flooding process after each successful grid head re-election. The same problem exists in [11].

A grid-based clustering scheme for two-dimensional microsensor networks has been proposed and implemented in our previous work [16]. In this paper, we further model and analyse the energy consumption of this scheme in a more general form, including both the radio and circuit energy consumption, as well as in the scenario of data aggregation. The optimal transmission range and grid size are deduced using this model, and through both numerical and simulation results, we evaluate better clustering strategies in terms of energy efficiency and discuss other energy-saving opportunities to further prolong the network lifetime.

The contribution of this paper is the analysis and evaluation of an energy-efficient clustering and routing scheme that totally eliminates the periodical flooding process and also considers the effect of data aggregation. By making fewer assumptions on the energy consumption and propagation loss models, our work reveals the energy trade-off generic to all grid-based clustering schemes, and also proposes a randomised technique to further prolong the network lifetime, as well as discussing other energy-saving opportunities. This paper therefore provides some insights into the intrinsic limits of grid-based clustering schemes for wireless microsensor networks, and helps determine a better clustering strategy for energy efficiency.

The rest of the paper is organised as follows. We introduce the background and related work in Section 2, as well as the problems in grid-based clustering. Section 3 describes the overall system design. A general modelling of energy consumption with the consideration of data aggregation, as well as the randomised technique, is given in Section 4, with the aim of achieving the optimal griding. Both numerical and simulation results are presented in Section 5, and in Section 6 we discuss some further improvements, followed by the conclusions in Section 7.

2. Background and related work

Clustering schemes that turn off unnecessary nodes within the transmission range of others can be of great benefit for energy conservation. With multi-hop routing, nodes can avoid long-range transmissions, and have no need to be active all the time due to dense deployment. Thus, many clustering schemes have been proposed in various contexts. In Mhatre et al. [9], the optimal node intensity is determined by Voronoi cells to guarantee a lifetime of at least certain units. Younis et al. proposed hybrid energy-efficient distributed clustering [14], which periodically selects cluster-heads (CHs) according to both their residual energy level and the node proximity to their neighbours. These clustering schemes are heuristic in nature, and demand time synchronisation or frequent message exchanges among nodes, which are not ideal in large-scale networks. Wang et al. [12] considers the data aggregation for grid-based sensor networks, but the forwarding tree is constructed by letting the sink node send query packets and flood the entire sensing area. This can lead to significant energy waste and communication overhead.

Grid-based clustering and routing schemes, in which clusters are equally sized square grids in a two-dimensional plane, have a simple structure with less routing management overhead, and all nodes in one grid are equivalent from the routing perspective. With the assistance of Global Positioning System (GPS) or localisation techniques [7], the square grid also provides easier coordination among all sensor nodes in the network. Therefore, it allows for a theoretical analysis while still being useful enough to incorporate all the important elements of a real network.

Extensive research work has been done in grid-based clustering. In the early work of Geographical Adaptive Fidelity (GAF) [13], the grid size s is chosen such that any two nodes in horizontally or vertically adjacent grids are within the transmission range, r, of each other, which is referred to as Manhattan walk in Figure 1(a). By investigating the worst-case scenario, the grid size should be $s \leqslant r/\sqrt{5}$. Recently, the work of [1] also uses this clustering structure. For the one-dimensional case [2], s should be less than $r/2$. More recent work of [15] and [16] used a smaller grid size, $s \leqslant r/\sqrt{8}$, allowing nodes in diagonal grids to be in the same transmission range as well, as shown in Figure 1(b). With the same transmission range r, there are fewer grids in Figure 1(a) to cover the field, but it may take more hops to reach the sink. Thus the trade-off between these two griding approaches is still an open question.

In most existing work, energy consumption in electrical circuits has been ignored. Instead, communication-related energy consumption is usually assumed to take a major portion in the total energy consumption. For example, in [2] the energy consumption other than transmission is assumed to be a constant. This usually leads to a misleading notion that, to minimise energy consumption, it is preferable to send data with more relay nodes to avoid the greater-than-linear path loss penalty due to long-range transmissions. When

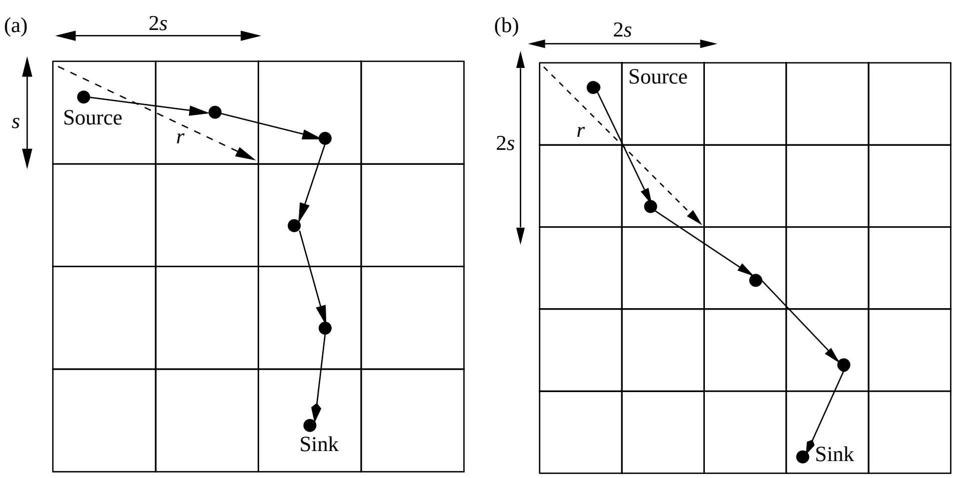

Figure 1. Manhattan walk ($s \leqslant r/\sqrt{5}$) and DF ($s \leqslant r/\sqrt{8}$).

taking the energy consumption in electrical circuits into account, however, more energy will be used if it takes more hops to reach the destination, and there will be more transmission attempts that lead to higher interference. Thus, there should be an optimal value in the number of transmissions which minimises the total energy spent in the network, or given a certain node density, the optimal transmission range to prolong network lifetime.

There is some effort in deriving the optimal communication range in one-dimensional networks [3], as well as in two-dimensional networks [1] and [11]. As mentioned in Section 1, the design and analysis in [1] and [11] are quite limited, while [3] studied a simple linear network and deduced the relationship between the optimal radio range and traffic load distribution. This work gave much insight into the relationship between network topology and energy efficiency. The simulation results, however, are obtained using the Friis free-space propagation model, which is only applicable in idealised conditions.

Based on the insights from the linear networks, our work focuses on the two-dimensional plane with a grid-based clustering and routing scheme. With fewer assumptions on energy consumption and propagation loss, our work is not merely a simple extension of the literature [3]. By first designing and implementing a multi-hop temperature monitoring system [16], we model the energy consumption in a general form to determine the minimum energy required to bring a unit of data from all nodes to the sink, as well as in the scenario where data aggregation is considered. Given the *working density* of sensor nodes, this model derives the upper and lower bounds of the energy consumption in wireless microsensor networks, which helps us evaluate better griding strategies and derive the optimal transmission range of sensor nodes in terms of energy efficiency.

3. System design

There are three modules in the proposed scheme: grid-based clustering, dynamic CH election and multi-hop routing between clusters [16]. The clustering process first divides the network into evenly sized grids, thus providing a location-based clustering topology to other modules. Based on the grid structure, dynamic election rotates the role of a node, either being a CH or a regular working node, according to its current energy level. Sensors change from one state to another according to the control packets they receive and their random backoff timer. Multi-hop routing selects a route among those elected CHs, which is also based on the spatially clustered structure.

Both grid-based clustering and CH election are designed for the purpose of routing maintenance, while multi-hop routing is for data forwarding between clusters. These three models therefore constitute the layered structure shown in Figure 2. With a grid-based scheme, energy can be further conserved by a predefined route between the data source and the sink.

3.1 Test bed configuration

In order to implement the above three modules for the prototype, we used a test bed that consists of 20 *Gainz* sensor nodes, designed by the Institute of Telecommunications and Integrated Circuits, Chinese Academy of Science (Ningbo Institute) (http://www.wsn.net.cn/cn/index.php). These sensor nodes, as shown in Figure 3, have both temperature and light sensing capabilities, an 8 bit ATmega128 RISC processor, and a single-chip 2.4 GHz IEEE 802.15.4-compliant RF transceiver (CC2420) (http://www.wsn.net.cn/cn/index.php).

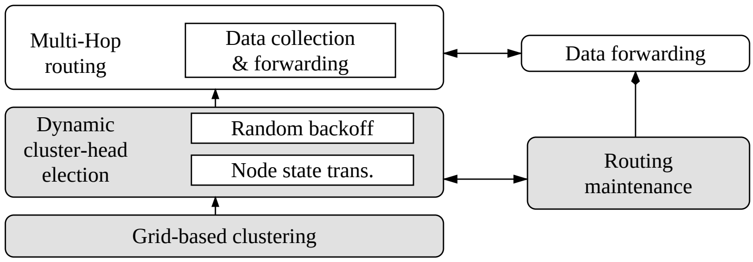

Figure 2. System design diagram.

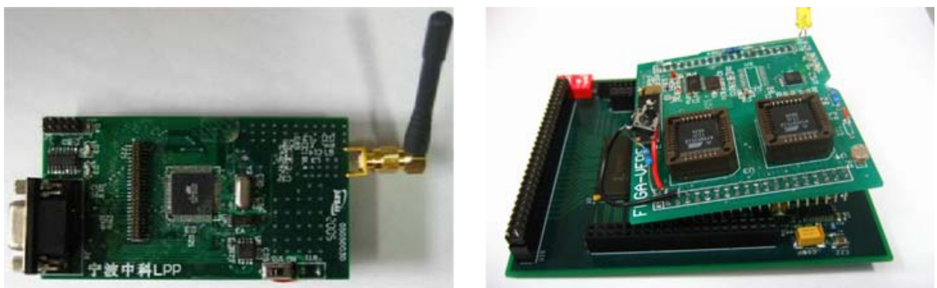

Figure 3. Gainz sensor node (http://www.wsn.net.cn/cn/index.php).

3.2 Grid formation

As mentioned in Section 2, there are several ways of dividing the sensing field into equal-sized grids. Once the grid size s is given, each node can calculate its grid coordinates (X, Y) according to its location (x, y):

$$X = \lceil x/s \rceil; \quad Y = \lceil y/s \rceil. \tag{1}$$

The $s \leqslant r/\sqrt{8}$ structure shown in Figure 1(b) is used for the grid formation in our scheme [16]. The choice between these different structures will be further discussed in Section 5 with numerical and simulation results. Node location (x, y) can be obtained by GPS devices or localisation techniques.

3.3 Packet design

3.3.1 Packet format

Figure 4 shows the basic packet format [16]. src_cl_no and dst_cl_no in bytes 0 and 2 are the identifiers of the source and destination cluster respectively, based on the grid coordinates calculated in (1). src_cl_ad and dst_cl_ad are the identifiers of sensor nodes inside a cluster, numbered in the order that sensors join the cluster. cl_no and cl_ad together distinguish an individual node in a certain cluster, just as network and host identifiers of IPv4 addresses.

Typical values of pkt_type are listed in the code column in Table 1. Further, bytes 0 and 1 together identify a source node of a packet, while bytes 2 and 3 identify the

byte 0	1	2	3	4	5	6~(L–1)
src_cl_no	src_cl_ad	dst_cl_no	dst_cl_ad	pkt_len(L)	pkt_type	payload

Figure 4. Packet format [16].

Table 1. Packet types.

Packet type	Code	Description
DATA	0*x*01	Sensed data
ACK	0*x*02	Acknowledgement
REQ_CH	0*x*03	Ask for info about CH
KEEP_ALIVE	0*x*04	Synchronisation packet with CH
CH_ACK	0*x*05	Response from an active CH
CH_NAK	0*x*06	Response from an inactive CH
ELE_CH	0*x*07	Declaration of becoming a CH

next hop. Sink node is a special 'next hop', whose identifier is known by all other nodes in the network during system initialisation or through control messages.

3.3.2 Packet type

Packet types are also listed in Table 1 (CH stands for cluster-head). DATA and ACK are typical packets for higher-level applications (such as environmental control). A regular node sends periodic KEEP_ALIVE packets, which include its identifier, to the CH, in order to keep itself synchronised. Whenever an active CH receives KEEP_ALIVE, it responds with a CH_ACK, otherwise it responds with CH_NAK, indicating that it is no longer the current CH. ELE_CH is the declaration message from a newly elected CH, which informs the regular working nodes to update their status and keep in touch with this new coordinator. Zhuang [16] describes the detailed format of data, control and diagnostic packets in our proposed clustering and routing scheme.

3.4 Dynamic CH election

The CH election process rotates the role of CH among all nodes in a cluster by making constant adaptation to the node energy level. There are two processes, *random backoff* and *node state transition*, that constitute the dynamic CH election (see Figure 2). Each time a CH finishes its duty cycle, it retires and all of the nodes in the cluster compete for this position by setting a backoff timer according to their residual energy level. Once the backoff timer fires, the node that first broadcasts a declaration message will become the CH in the next round. This *first-declare-wins* process continues until the energy in all nodes inside the cluster is depleted.

3.4.1 Random backoff

Suppose there are m energy levels in each sensor node, and the larger the value of i, the more the residual energy the node has in its on-board battery. Then nodes that are working in the ith energy level set their timeout value as a random number between $T(i)_{\text{start}}$ and $T(i)_{\text{end}}$

$$t_i = T(i)_{\text{start}} + k \times [T(i)_{\text{end}} - T(i)_{\text{start}}], \tag{2}$$

where $i \in \{1, 2, \ldots, m\}$, and k is a random number between [0,1]. $T(i)_{\text{start}}$ and $T(i)_{\text{end}}$ are chosen such that $T(m)_{\text{start}} < T(m)_{\text{end}} = T(m-1)_{\text{start}} < T(m-1)_{\text{end}} = \cdots = T(1)_{\text{start}} < T(1)_{\text{end}}$, i.e. the higher the energy level of a sensor, the shorter the backoff time it has (as the shaded area in Figure 5). The time interval $T = T(i)_{\text{end}} - T(i)_{\text{start}}$ is set to be a constant. Since a node with less residual energy has a longer backoff timeout value, it will be less likely to become the CH in the next duty cycle.

3.4.2 Node state transition

According to the packet types and backoff timer setting defined above, a sensor node changes its state as shown in Figure 6. Solid lines in the figure indicate that the node sends out a packet, or timeout occurs; dashed lines indicate that the node has received a packet. There are five possible node states.

(1) *Start Up*. A node joins the network.
(2) *Wait for CH*. A start-up node asks for information about the cluster it belongs to, by broadcasting a REQ_CH packet. Depending on whether there is an existing CH, the node will go to either (3) or (4).
(3) *Synchronisation with CH*. If there is an active CH around, CH_ACK will be received. The start-up node then synchronises with this CH by sending out KEEP_ALIVE messages periodically. If there is a CH re-election during a KEEP_ALIVE period, the inactive CH will send back CH_NAK to inform the regular node about the change in network status, making it go back to the start-up state.

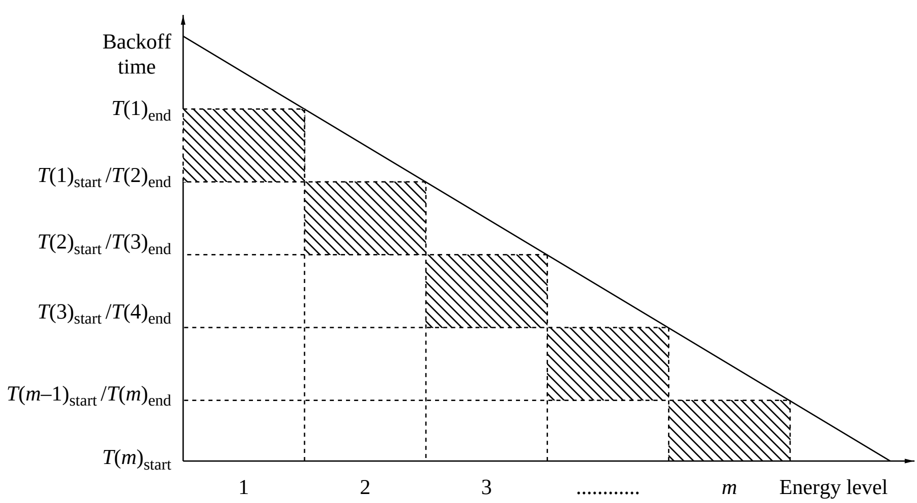

Figure 5. Energy level vs. backoff time.

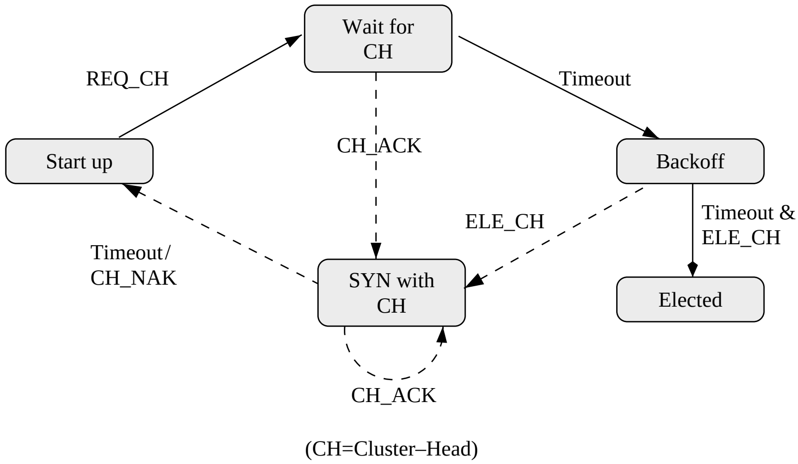

Figure 6. Node state transition diagram.

(4) *Back-off*. The node picks its backoff time according to (2), and then the backoff timer starts to count down. If the node receives an ELE_CH message during this backoff period, it stops its backoff timer, and then goes to (3).
(5) Otherwise, the node broadcasts ELE_CH on its own when the timer expires, announcing itself as the new CH. If two or more nodes simultaneously announce ELE_CH, then we break the tie by letting the node with a larger node address value win.

3.5 Properties of dynamic CH election

Property 3.1. With high probability there is exactly one CH in each cluster.

Proof. With the above dynamic CH election, we assume that there are n_i nodes in a particular cluster, each with m different energy levels. According to Figure 5, only nodes within the same energy level could possibly set the same backoff timer. If it takes Δt time to transmit a packet to other nodes, then the probability that there is only one node that broadcasts ELE_CH is

$$p \geq \left(1 - \frac{\Delta t}{T}\right)^{n_i/m-1}.$$

Typically, Δt is very small when compared with T, and $n_i/m - 1$ will not be a large number. For example, when $\Delta t = 10\,\text{ms}$, $T = 1\,\text{s}$, $n_i = 10$ and $m = 6$, then $p \geq 0.99^{0.25} = 0.993$. In our design, even if there is a collision, it can be solved by letting the node with a larger address value win eventually. □

Property 3.2. Energy consumption can be evenly distributed among all nodes in the network.

Proof. We start by analysing a simple case, and then generalise it to other cases. Assume that there are n_i nodes in a certain cluster, each with m different energy levels. The energy consumption of a regular node is r per unit time, and the CH spends an extra amount of R for coordinating each node in the cluster (i.e. $(n_i - 1)R$ in total). If the duty cycle of a CH is t, and in the simple case, every node has the same initial energy, then after every duty cycle t (i.e. after each re-election), the energy reduction is

$$\Delta Q = [n_i r + (n_i - 1)R]t.$$

Initially, all nodes have the same energy, $Q_{01} = Q_{02} = \cdots = Q_{0n_i}$, corresponding to nodes $\{N_1, N_2, \ldots N_{n_i}\}$.

Before the kth election, the residual energy of each node is $Q_{k1}, Q_{k2}, \ldots Q_{kn_i}$. They can be divided into m levels at most, i.e. $D_1, D_2, \ldots D_m$. During the kth election, suppose node N_j with $Q_{kj} \in D_m$ is elected, then it can either degrade to level $m - 1$ after t, or stay in level m. And $Q_{(k+1)j} = Q_{kj} - [r + (n_i - 1)R]t$, while other $Q_{(k+1)i} = Q_{ki} - rt$ (for all $i \neq j$).

(1) If N_j is degraded to level $m - 1$, then the next CH will be elected from nodes in $\{D_m - N_j\}$. After all the remaining nodes have been elected once, they will be degraded to the same level $m - 1$.
(2) If N_j stays in level m, it is possible that it will get re-elected. Suppose N_j gets degraded to level $m - 1$ after q elections, other nodes in $\{D_m - N_j\}$ will also get degraded after q rounds because being a CH has almost the same energy consumption in each duty cycle.

After degrading, these nodes will compete with all the nodes in level $m - 1$. Thus according to (1) and (2), all n_i nodes will be competing for CH and get elected in a periodic manner, if they have the same initial energy. Since nodes with higher energy levels get degraded first, energy consumption will be evenly distributed among all nodes in the same cluster.

If $Q_{01}, Q_{02}, \ldots, Q_{0n_i}$ are not the same, then nodes with more battery energy will be elected first, and get degraded first. Assuming that after k' elections, energy distribution goes back to the state *before the kth election*, then the situation will be the same again, making energy degrade and distribute evenly. □

3.6 Multi-hop routing

Data transmissions between different clusters only involve active CHs. In wireless microsensor networks, any node can be a potential data source. The grid structure allows packets to be forwarded in a predefined manner, as shown in Figure 1. Thus two routing strategies are possible, depending on the griding scheme.

If $s \leqslant r/\sqrt{5}$, packets always go horizontally or vertically until they reach the sink (i.e. Manhattan walk), since the transmission range of a node cannot cover all nodes in its diagonal grids. While if $s \leqslant r/\sqrt{8}$, packets can travel between diagonal grids. Only when packets are forwarded to the same row or column as the sink, will they go horizontally or vertically (i.e. diagonal-first (DF) routing). Due to the grid structure, whenever there is data to send, the sender can transmit without the need of setting up a route in advance.

4. System modelling

In this section, we model the energy needed for packet transmission and reception, and the optimal grid structure, as well as in the case of data aggregation. We assume that there is one sink node in an $L \times L$ m^2 sensing field, all other nodes are aware of its location, and all nodes use the same transmission range r.

4.1 Energy consumption

A general energy consumption model is used here. The transmitter dissipates energy to power up its *e*lectrical circuit, as well as the power *a*mplifier for transmission, i.e. $E^{\mathrm{tx}} = E_{\mathrm{e}}^{\mathrm{tx}} + E_{\mathrm{a}}^{\mathrm{tx}}$. The receiver only needs energy to power on the radio electronics, i.e. $E^{\mathrm{rx}} = E_{\mathrm{e}}^{\mathrm{rx}}$ [10]. Energy in electrical circuit, E_{e}, is determined by the built-in parameters of the microsensors, including factors such as the coding, modulation and filtering of the signal before it is sent to the transmit amplifier, while the communication-related power consumption mainly depends on the environment and the distance it transverses.

Electrical signals are subject to attenuation once they are sent out by the transmitter. The propagation of electromagnetic waves can be modelled by a decreasing power law function of the distance between the transmitter and receiver, d. If d is smaller than a crossover threshold[2] d_c, the propagation loss is proportional to d^2, or d^4 otherwise[3]. Power control, therefore, is used to invert this loss by setting the power amplifier E_{a} at the transmitter properly to ensure a certain power level at the receiver. Thus, to transmit a b-bit message over a distance d, the energy consumption by the transmitter is[4]

$$E^{\mathrm{tx}}(b,d) = E_{\mathrm{e}}^{\mathrm{tx}}(b) + E_{\mathrm{a}}^{\mathrm{tx}}(b,d) = bE_{\mathrm{e}} + E_{\mathrm{a}}(b,d), \tag{3}$$

$$E_{\mathrm{a}}(b,d) = \begin{cases} b\varepsilon_{\mathrm{Friis}} d^2 & \text{if } d \leqslant d_c, \\ b\varepsilon_{\mathrm{two-ray}} d^4 & \text{if } d > d_c. \end{cases} \tag{4}$$

And the energy for receiving a b-bit message is

$$E^{\mathrm{rx}}(b) = E_{\mathrm{e}}^{\mathrm{rx}}(b) = bE_{\mathrm{e}}. \tag{5}$$

If there are n nodes in the sensing area, the *deploying density* P is n/L^2. Since not all sensors in the field are actively sensing, we only need to use a portion of all the nodes for information harvesting. Define ρ as the *working density*, and assume that each active node senses one unit of data from the environment in each time slot, the total energy consumption for both CHs and regular working nodes are

$$E_{\mathrm{ch}} = n\rho[2E_{\mathrm{e}} + E_{\mathrm{a}}(d_{\mathrm{int}})]\cdot E[\mathrm{hop}], \tag{6}$$

$$E_{\mathrm{wk}} = n\rho[E_{\mathrm{e}} + E_{\mathrm{a}}(d_{\mathrm{inn}})], \tag{7}$$

where,

(1) d_{inn} and d_{int} are the distances between active nodes in the same cluster and between neighbouring clusters, respectively. d_{inn} and d_{int} thus determine the energy used by the power amplifier.

(2) $E[\mathrm{hop}]$ is the average number of hops from any data source to the sink. Given the location of the sink grid (m, n), any data transmission from grid (i, j) following the DF routing has to go through $H(i,j)_{\mathrm{diag}} = \max\{|i-m|, |j-n|\}$ hops; when the data transmission follows Manhattan walk, $H(i,j)_{\mathrm{Man}} = |i-m| + |j-n|$. Therefore,

$$E[\mathrm{hop}] = \sum_{i,j=0}^{k} H(i,j)/k^2. \tag{8}$$

4.2 Optimal griding

From (6)–(8), we see the relation between the grid size s and the total energy consumption. Whether a grid-based routing scheme can forward data more efficiently depends on the size of the grids in the network and the average number of hops

$$E_{\mathrm{total}} = n\rho\{[2E_{\mathrm{e}} + E_{\mathrm{a}}(d_{\mathrm{int}})]\cdot E[\mathrm{hop}] + E_{\mathrm{e}} + E_{\mathrm{a}}(d_{\mathrm{inn}})\}. \tag{9}$$

Intuitively, sensors can have a shorter transmission range if a smaller grid size is used, so the communication-related power to overcome the propagation loss will be reduced. However, the energy used in radio electronics is increased due to a larger number of transmission and reception attempts. On the other hand, there will be more nodes inside a cluster when using a larger grid size, providing sufficient energy and more forwarding opportunities. But larger grids can also lead to a longer separation between the transmitter and receiver. The trade-off between picking a small or large grid size to optimise energy consumption is evaluated in Section 5 with the model calculation and simulation results.

4.3 Data aggregation

One unique characteristic of the information that sensor networks collect is its high spatial and temporal correlation and redundancy between sensor nodes. As a result, only a high-level description of the events is needed. Data aggregation (or data fusion) therefore can be used to combine correlated data into a smaller set of data that contains the refined information. In our proposed scheme, sensor nodes organise themselves into clusters. Since the energy used for communication is usually greater than that for computation, we therefore can further combine the energy-efficient grid-based clustering with application-specific data aggregation at the CHs to achieve even better performance.

Heinzelman [4] introduced the concept of *a*ggregation ratio γ: for every γ bits that must be sent to the base station when no data aggregation is performed, only 1 bit must be sent to the base station when local data aggregation is performed. By considering applications that use a simple aggregation operator, such as average (AVG), MAX, SUM or percentile (PERCT), multiple input packets can be aggregated into fewer output packets with a ratio of γ. Define the data aggregation function, $g(b)$, that gives the data volume after aggregating b bits of data, and the energy consumption due to aggregation operation E_{D} as

$$g(b) = \frac{b}{\gamma}; \quad E_{\mathrm{D}} = c\gamma,$$

where c is the energy consumption coefficient for data aggregation. Then the energy to perform both data aggregation on b bits and transmit the aggregated data is $bE_{\mathrm{D}} + E^{\mathrm{tx}}(g(b))$. Applying this to (6), we get

$$E_{\mathrm{ch}} = n\rho[E_{\mathrm{e}} + c\gamma + E_{\mathrm{a}}(d_{\mathrm{int}})/\gamma]\cdot E[\mathrm{hop}]. \tag{10}$$

5. Performance evaluation

Both numerical and simulation results are presented in this section to evaluate the energy-optimal grid size for a grid-based clustering scheme. We first analyse the average number of hops from all data sources to the sink located at any location, and then reveal the performance bounds and the trade-off generic to all griding schemes. A randomised technique is used to further prolong the network lifetime, as well as enhance the effect of data aggregation.

The results in this section are averaged over 80 simulations. Analysis and simulation parameters are given in Table 2. All working nodes send data to their CH, while all CHs do data gathering and aggregation when applicable, as well as data forwarding. The time interval during which sensor nodes are sending data depends on application requirements. For the environmental monitoring application used as an example in [16], this time interval is set to every 0.8 s. Shorter time intervals should be used whenever a smaller initial delay is required by the application.

5.1 Average hop count

E[hop] and node separation determine the energy spent by all the CHs. E[hop] is, in turn, determined by grid size and sink location. For a sink located at grid (m, n) in the sensing area of $k \times k$ grids, we divide the field into four blocks of $u \times v$ grids each, and obtain the total number of hops in each block as if the sink is at the field corner. First, we define that for each block $j = 1, 2, 3, 4$, we have different values of u and v, as shown in Table 3.

In DF routing, the sum of hops in each block is

$$\text{sum(hop)}_j = \sum_{1}^{\min\{u,v\}} i(2i-1) + \min\{u,v\} \times \sum_{\min\{u,v\}+1}^{\max\{u,v\}} i + \sum_{1}^{v} i.$$

In Manhattan walk,

$$\text{sum(hop)}_j = \sum_{1}^{\min\{u,v\}} i(i-1) + \min\{u,v\} \times \sum_{\min\{u,v\}+1}^{\max\{u,v\}} i + \sum_{\max\{u,v\}+1}^{u+v} i(u+v+1-i) + \sum_{1}^{v} i.$$

E[hop] is therefore given by

$$E[\text{hop}] = \frac{1}{k^2} \sum_{j=1}^{4} \text{sum(hop)}_j. \tag{11}$$

There are two extreme cases for the average number of hops, i.e. when the sink is in the centre of the field (best case) and in the corner (worst case). Figure 7 shows these two cases: DF routing always has fewer hops than Manhattan walk. It is also obvious that in the worst case, the average number of hops will be much larger than that in the best case.

Table 2. Analysis and simulation settings.

Parameter	Meaning	Value
L	Length of sensing area	200 m
n	Number of nodes	500
ρ	Working density	0.5
E_e	Electronics energy	50 nJ/bit
$\varepsilon_{\text{Friis}}$	Friis free-space coefficient	10 pJ/bit/m^2
$\varepsilon_{\text{two-ray}}$	Two-ray-ground coefficient	0.0013 pJ/bit/m^4

Table 3. *u* and *v* in each block.

Block index (j)	u	v
1	$m - 1$	$k - n$
2	$n - 1$	$m - 1$
3	$k - m$	$n - 1$
4	$k - n$	$k - m$

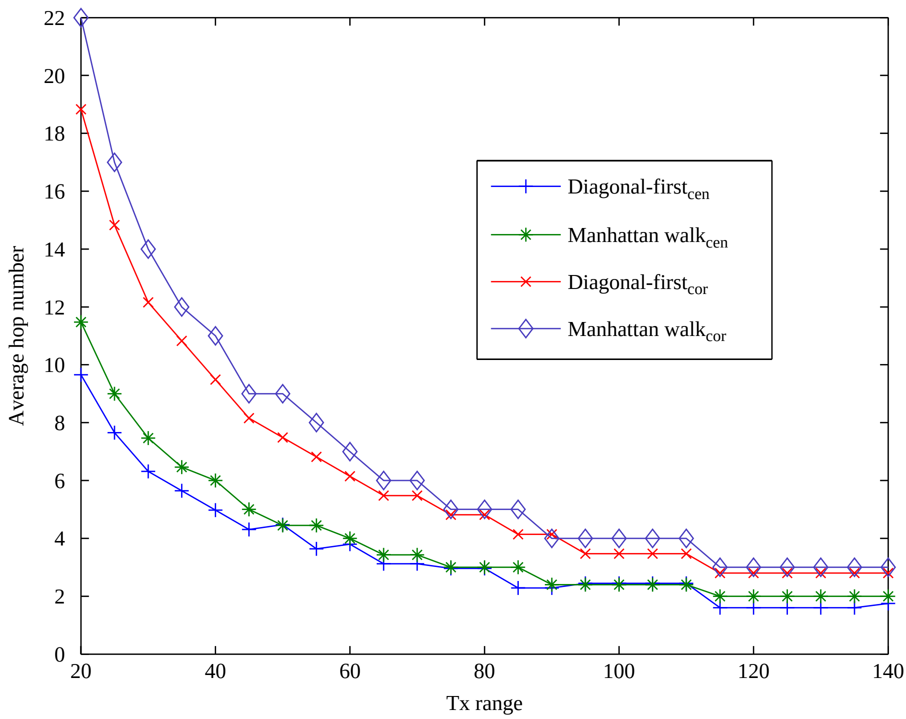

Figure 7. Average hop count (sensing area 200 × 20 m).

5.2 *Griding structure and energy consumption*

In Equation (9), E_{total} is determined by radio electronics, power amplifier and the average number of hops. To gain some insights into how the grid size affects total energy consumption, we use both numerical analysis and simulation to see whether there is an optimal grid size that minimises E_{total}.

First, solid and dash–dot lines in Figure 8 are the numerical results in DF routing, for the lower and upper bounds of the total energy consumption when each node generates one bit of data for each time slot. Simulation results are shown in dots. Analytical bounds in Figure 8 are not smooth because the number of grids is discrete, while the grid size changes continuously. When grid size is around 40–50 m, the total energy consumption reaches its lowest level for both the best and worst cases. This corresponds to an energy-optimal transmission range of about 110–130 m. Simulation results show a similar behaviour, except they always have a higher energy cost than analytical bounds due to the realistic network environment.

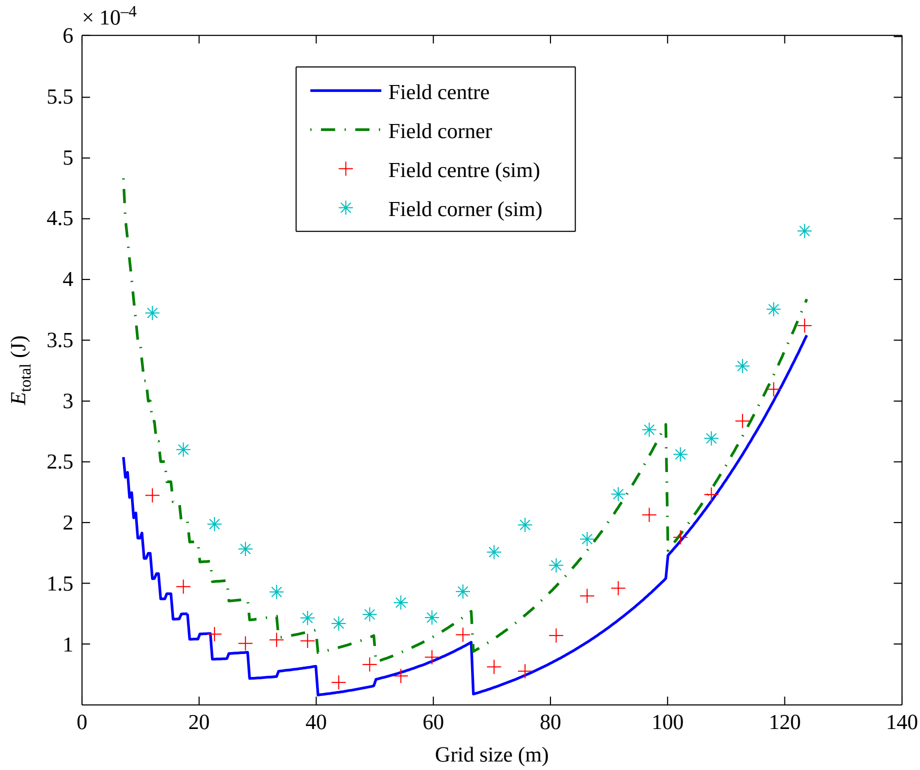

Figure 8. Energy consumption with DF routing.

Similarly, Figure 9 shows the results for Manhattan walk. When node transmission range is about 50–60 m, the energy consumption level reaches its minimal level. The optimal transmission range is also 110–130 m due to the same system parameter setting, although the total energy consumed in this case is higher than DF routing given the same grid size.

Thus, when the grid size is small, data transmission follows the Friis free-space model. Although signal attenuation is not significant, there are more nodes actively working and the average number of hops is larger. With the grid size increasing, some data transmissions are subject to d^4 attenuation, but more redundant nodes can be put to sleep, and less energy is spent in electrical circuit. At this stage, an optimal grid size is achieved. If the grid continues to grow, the energy consumption associated with transmission increases super-linearly with the radio range, so the total energy consumption grows exponentially with node separation.

Therefore, DF routing is used in our multi-hop system [16], since it is more energy efficient for the same parameter setting compared with Manhattan walk. Additionally, data traffic is more balanced due to more freedom of choice in transmission direction. The grid size should be between 40 and 50 m in order to optimise the total energy in the network.

5.3 *Network lifetime*

We define network lifetime as the time when the first grid in the network consumes all the energy of its nodes. Therefore it is determined by the grid that expends the largest amount

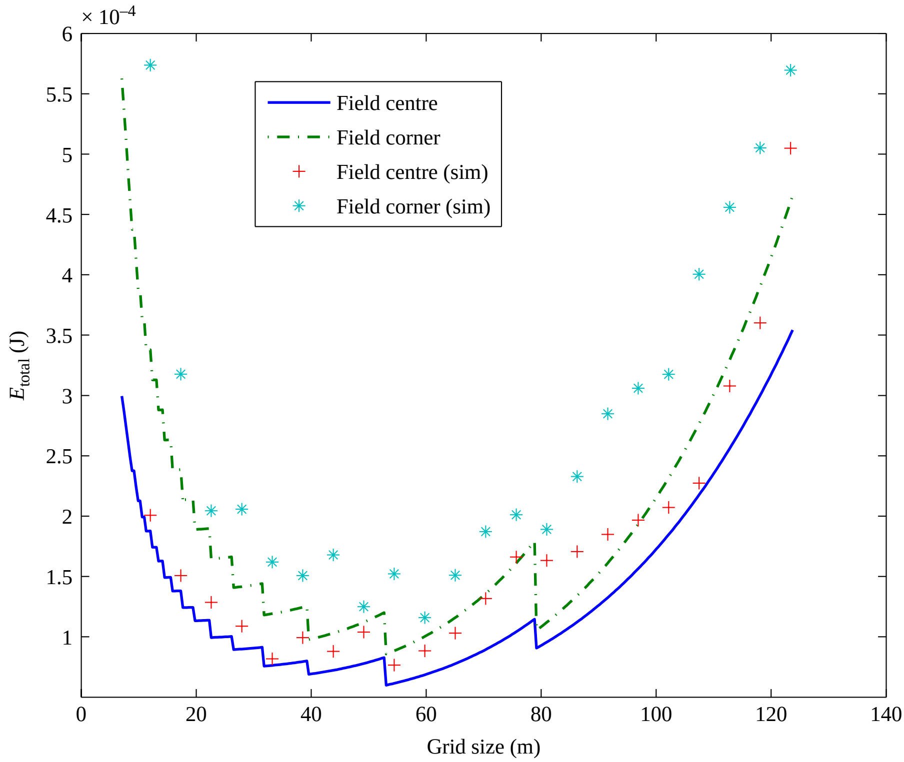

Figure 9. Energy consumption with Manhattan walk.

of energy in the network. In either DF routing or Manhattan walk, traffic is crowded in the grids of the same row or column coordinate as the sink grid (we call them *c*ross-band grids). Manhattan walk suffers more from this uneven distribution of energy due to the limited choices in data forwarding directions. Therefore, we need to make the *c*ross-band area less crowded with other energy-saving techniques.

The main reason for the relatively crowded area is, in both approaches, data traffic is always forwarded to a neighbouring grid with the preferred direction, and is eventually routed to the *c*ross-band area. Thus, this area is always crowded with data traffic that is in their final hops to the sink. To balance the energy distribution, routing decisions should be less constrained in choosing a forwarding direction. If we choose a direction randomly towards the destination, then the data traffic will be more balanced among nearby grids. As a result, the *cross-band* area will be less crowded and the network will have a longer lifetime.

Figure 10 shows the results of network lifetime with different routing techniques, with the same parameter settings in Section 2. In most cases, DF routing is better than Manhattan walk. With a randomised routing, the constraint in routing direction is further relaxed. Thus, the randomised DF routing works even better, although the curve of network lifetime fluctuates irregularly due to the randomness. In Figure 10, the optimal grid size is in the range of 50 m, which conforms to our previous results.

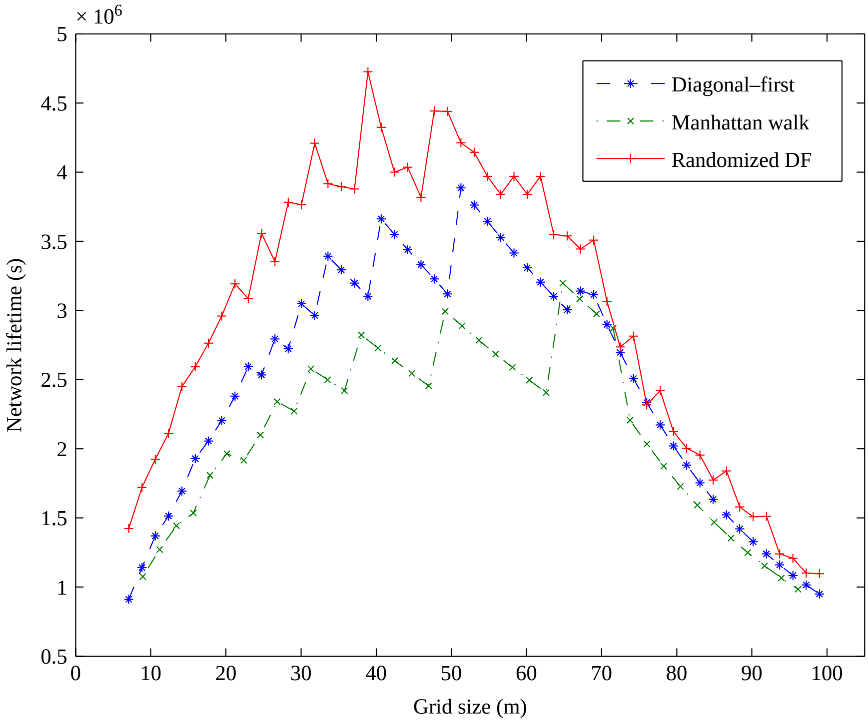

Figure 10. Network lifetime.

5.4 Data aggregation

Figure 11 shows a reduction in energy consumption with data aggregation in DF routing, and the energy consumption coefficient c is set to 2 nJ per aggregation operation. $\gamma = 1$ corresponds to the case without data aggregation. Clearly, with data aggregation, less energy will be consumed, and the optimal grid size has also changed from 50 to 70 m when $\gamma = 30$. Choosing an appropriate aggregation ratio is also important. In the figure, although a higher level of aggregation can reduce the actual volume of data being sent, it also adds the energy consumption due to data aggregation to the total energy consumption.

6. Further discussion

In this section, we discuss further opportunities for energy-saving in grid-based clustering schemes, which is our focus in the ongoing research and future work.

6.1 Energy–throughput trade-offs

So far our work has been in the energy domain – the *minimum energy* required to transmit data from all nodes to the sink – but has not considered the time, i.e. the *minimum time* to move the same amount of data from all nodes to the sink. This problem is equivalent to maximising network throughput: the maximum number of concurrent transmissions. Maximising throughput and lifetime, however, often conflicts with each other. Higher throughput leads to faster energy dissipation, which reduces the network lifetime.

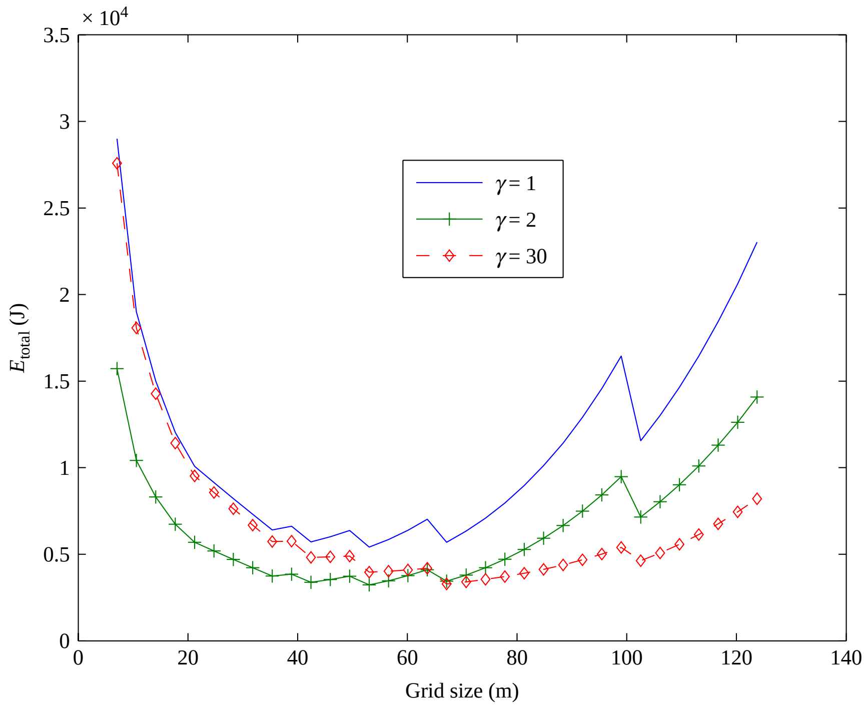

Figure 11. Energy consumption with data aggregation.

In general, to identify the optimal trade-off between throughput and lifetime is more interesting and practical than optimising either of them individually.

Transmissions from any node within a given range of the transmitter (referred to as the interference range) will cause a collision and result in packet error. In Figure 12, transmission on link *ac* and *bd* can be scheduled at the same time. Although the interference range of *a* and *b* (dashed lines) overlaps, it will not affect destination *c* and *d* that are inside the transmission range (solid lines). Therefore, determining the maximum number of concurrent transmissions also depends on the node position and data forwarding direction.

6.2 Opportunistic forwarding and opportunistic griding

We assume that nodes are uniformly distributed in all grids. Given the DF routing, it is guaranteed that one transmission will cover all neighbouring grids; however, depending on the location of the CH in the tagged grid, the transmission may reach CHs in some non-neighbouring grids in the forwarding direction. Therefore, there is a chance of *opportunistic forwarding*. When the current CH s is in (X, Y), data can be opportunistically forwarded to the stroked area (see Figure 13(a)). The area of possible opportunistic forwarding is

$$A(x,y) = \frac{r^2}{2}\left(\frac{\pi}{2} + \arcsin\frac{y}{r} + \arcsin\frac{x}{r}\right) + xy + \frac{x\sqrt{r^2 - x^2}}{2} + \frac{y\sqrt{r^2 - y^2}}{2} - 4s^2. \quad (12)$$

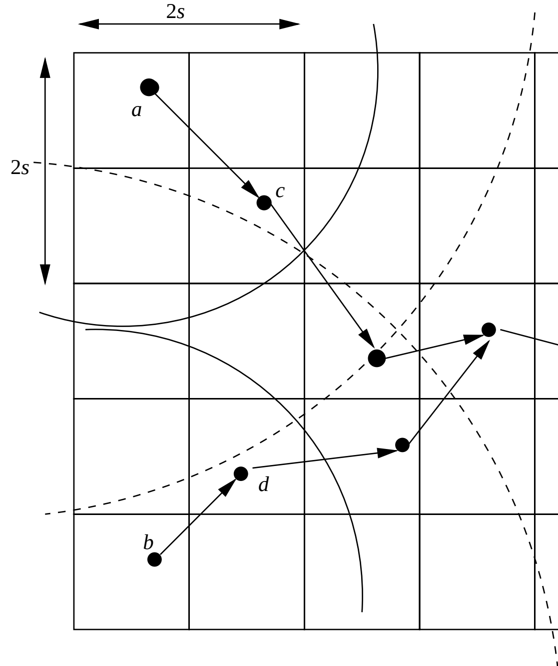

Figure 12. Concurrent transmission.

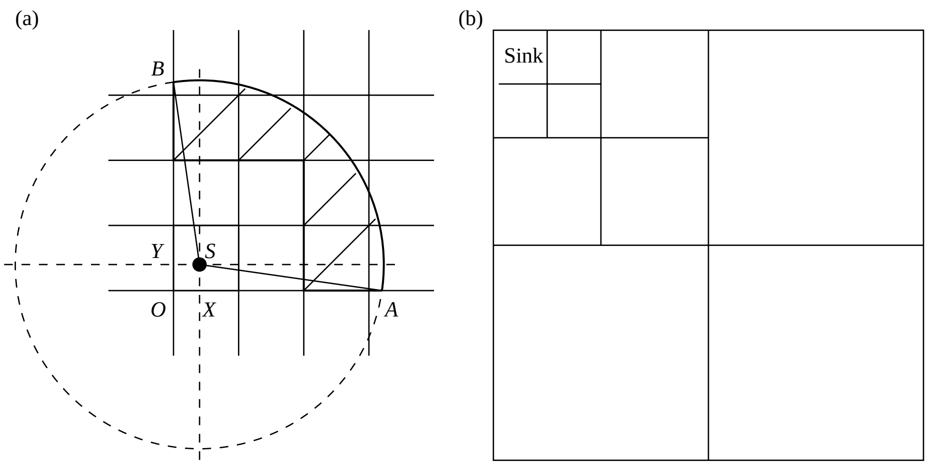

Figure 13. Opportunistic forwarding and opportunistic griding.

With the average CH density of $1/s^2$, the extra opportunistic coverage that a transmission can achieve is

$$\frac{1}{s^2}\int_0^s\int_0^s A(x,y)\mathrm{d}x\,\mathrm{d}y = 5.33\,s^2. \tag{13}$$

Further, since data traffic is crowded in the area close to the sink, *opportunistic griding* is therefore advantageous in smoothing energy distribution. Grids close to the sink, which have heavy traffic load, will have a smaller size compared with those that are farther away. Dividing the network into unequal grids (see Figure 13(b)) will also lead to different transmission range adjustment in a two-dimensional plane.

7. Conclusions

In wireless microsensor networks, energy consumption is the most important factor affecting network lifetime. Grid-based clustering organises sensor nodes into clusters and puts nodes not involved in forwarding into sleep. In this paper, we investigated energy-optimal grid-based clustering for sensor networks by modelling, analysis and simulation, as well as in the case of data aggregation. Both analytical and simulation results show that there is an optimal grid size that leads to the minimal energy consumption in a two-dimensional sensing field. In addition, randomised and opportunistic techniques further prolong the network lifetime. Our work provides insights into the intrinsic limits of grid-based clustering schemes, and helps determine a better clustering strategy for energy efficiency. The discussions in Section 6 constitute our ongoing and further work, which will lead to a more in-depth exploration of energy efficiency in wireless microsensor networks.

Acknowledgements

This work is supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), Canada Foundation for Innovation (CFI) and British Columbia Knowledge Development Fund (BCKDF).

This work is based on a grid-based clustering and routing scheme proposed in [16], which is done at the Key Laboratory of Computer Network and Information Integration (CNII) in Southeast University, Nanjing, China. Yanyan Zhuang would also like to thank Prof. Guanqun Gu for his guidance, Tyler Cadigan for proof reading, and Qin Xia, Lin Niu and Linfeng Liu for their support at CNII.

Notes

1. An early short version of this paper has been presented at the 6th International Workshop on Wireless *Ad hoc* and Sensor Networks (WWASN 2009) in conjunction with the 29th IEEE International Conference on Distributed Computing Systems (ICDCS 2009), Montreal, Que., Canada, June 22–26, 2009.
2. d_c is determined by system parameters such as the height of antenna, the wavelength of carrier signal, etc.
3. d^2 attenuation and d^4 attenuation correspond to Friis free space model and two-ray ground propagation model, respectively.
4. $\varepsilon_{\text{Friis}}$ and $\varepsilon_{\text{two-ray}}$ depend on the required receiver sensitivity.

References

[1] R. Akl and U. Sawant, *Grid-based coordinated routing in wireless sensor networks*, in *4th IEEE Consumer Communications and Networking Conference, (CCNC'07)*, Las Vegas, NV, USA, 2007, pp. 860–864.

[2] P. Cheng, C.-N. Chuah, and X. Liu, *Energy-aware node placement in wireless sensor networks*, in *IEEE Global Telecommunications Conference, (GLOBECOM'04)*, Dallas, TX, USA, Vol. 5, 2004, pp. 3210–3214.

[3] Q. Gao, K.J. Blow, D.J. Holding, I. Marshall, and X.H. Peng, *Radio range adjustment for energy efficient wireless sensor networks*, J. Ad-hoc Netw. 4(1) (2006), pp. 75–82. Available at http://www.cs.kent.ac.uk/pubs/2006/2194.

[4] W.B. Heinzelman, *Application-specific protocol architectures for wireless networks*, Ph.D. thesis, Massachusetts Institute of Technology, 2000.

[5] W.R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, *Energy-efficient communication protocol for wireless microsensor networks*, in *Hawaii International Conference on System Sciences (HICSS)*, Maui, HI, USA, 2000.
[6] W.B. Heinzelman, A.P. Chandrakasan, and H. Balakrishnan, *An application-specific protocol architecture for wireless microsensor networks*, IEEE Trans. Wireless Commun. 1(4) (2002), pp. 660–670. Available at http://dx.doi.org/10.1109/TWC.2002.804190.
[7] L. Lazos and R. Poovendran, *Serloc: Robust localization for wireless sensor networks*, ACM Trans. Sen. Netw. 1(1) (2005), pp. 73–100.
[8] H. Luo, F. Ye, J. Cheng, S. Lu, and L. Zhang, *TTDD: Two-tier data dissemination in large-scale wireless sensor networks*, Wireless Netw. 11 (2005), pp. 161–175.
[9] V. Mhatre, C. Rosenberg, D. Kofman, R. Mazumdar, and N. Shroff, *A minimum cost heterogeneous sensor network with a lifetime constraint*, IEEE Trans. Mobile Comput. 4(1) (2005), pp. 4–15.
[10] S. Sheng, L. Lynn, J. Peroulas, K. Stone, I. O'Donnell, and R. Brodersen, *A low power CMOS chipset for spread spectrum communications*, in *Proceedings of IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA, 1996, pp. 346–347.
[11] R. Vidhyapriya and P.T. Vanathi, *Energy efficient grid-based routing in wireless sensor networks*, Int. J. Intell. Comput. Cybern. 1(2) (2008), pp. 301–318.
[12] N. Wang, Y. Huang, J. Chen, and P. Yeh, *Energy-aware data aggregation for grid-based wireless sensor networks with a mobile sink*, in *Wireless Personal Communnications*, 2007, pp. 1539–1551.
[13] Y. Xu, J. Heidemann, and D. Estrin, *Geography-informed energy conservation for* ad hoc *routing*, in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking,* Rome, Italy, 2001, pp. 70–84.
[14] O. Younis and S. Fahmy, *Distributed clustering in* ad hoc *sensor networks: A hybrid, energy-efficient approach*, in *Proceedings of the IEEE INFOCOM*, Vol. 1, 2004, pp. 629–640.
[15] Z. Zhou, X. Xiang, X. Wang, and J. Pan, *An energy-efficient data-dissemination protocol in wireless sensor networks*, in *WOWMOM'06: Proceedings of the 2006 International Symposium on World of Wireless, Mobile and Multimedia Networks*, Buffalo, NY, USA, 2006, pp. 13–22.
[16] Y. Zhuang, *Research and implementation of energy-aware routing protocol in wireless sensor networks*, Master's thesis, Southeast University, Nanjing, China, 2007.