Data Collection using Miniature Aerial Vehicles in Wireless Sensor Networks

Mathur, Prateek; Nielsen, Rasmus Hjorth; Prasad, Neeli R.; Prasad, Ramjee

Published in:
IET Wireless Sensor Systems

DOI (link to publication from Publisher):
10.1049/iet-wss.2014.0120

Publication date:
2016

Document Version
Accepted author manuscript, peer reviewed version

IET Wireless Sensor Systems

Research Article

IET Journals
The Institution of Engineering and Technology

Data collection using miniature aerial vehicles in wireless sensor networks

ISSN 2043-6386
Received on 9th March 2015
Revised on 15th July 2015
Accepted on 9th September 2015
doi: 10.1049/iet-wss.2014.0120
www.ietdl.org

Prateek Mathur[1] ✉, *Rasmus H. Nielsen*[1,2], *Neeli R. Prasad*[1], *Ramjee Prasad*[1]

[1]*Center for TeleInFrastruktur (CTIF), Aalborg University, Aalborg, Denmark*
[2]*Cisco Systems, San Jose, USA*
✉ *E-mail: in_pm@es.aau.dk*

Abstract: Energy constraints of sensor nodes in wireless sensor networks (WSNs) is a major challenge and minimising the overall data transmitted across a network using data aggregation, distributed source coding, and compressive sensing have been proposed as mechanisms for energy saving. Similarly, use of mobile nodes capable of relocating within the network has been widely explored for energy saving. In this study, the authors propose a novel method for using miniature aerial vehicles for data collection instead of actively sensing from a deployed network. The proposed mechanism is referred as *data collection fly* (*DCFly*). It is suitable for data collection from WSNs deployed in harsh-undulating terrain with the base station located far from the sensing region. The DCFly is compared with data collection based on multi-hop data aggregation and data collection with mobile sinks. The numerical results justify that the proposed data collection mechanism is effective and efficient for use in WSNs.

1 Introduction

Wireless sensor networks (WSNs) are deployed in an area to be monitored, referred to in this paper as the *sensing region*, for diverse applications such as habitat and environmental monitoring, battlefield surveillance, and industrial monitoring. Using active node mobility in the network, overall communication in the network can be reduced and specifically nodes close to the base station (BS) could be prevented from draining out. Various utilities that are possible using active node mobility in the network are: mobile relay, mobile data mule, and mobile BS. Alternatively, in-network data processing mechanisms such as data aggregation, network coding, and compressive sensing could be utilised for minimising the overall communication of the network. These two approaches, that is, node mobility, and in-network processing have been extensively explored individually and independently.

This paper presents a mechanism for data collection using a data collection fly (DCFly). The DCFly would aerially collect data from the network. This is the first work that explores the potential of using miniature aerial vehicles (MAVs) for collecting data, collectively involving node mobility and in-network data processing. The underlying framework adopted in this paper has been presented in our preliminary work [1]. The remaining paper has been structured as follows: in Section 2, previous work relating to node mobility and in-networking data processing has been presented. Section 3 provides an insight into factors governing optimal number of cluster heads (CHs) and inter-relation with position of BS, data collection operation by the DCFly; these aspects influence the network architecture. The factors governing the DCFly's possible use in the sensing region and the network operational conditions considered for evaluation are presented in Section 4. The DCFly's governing factors formulated as an optimisation problem to determine the feasible operational utility are presented in Section 5. The optimum number of clusters that can be covered by a single DCFly are presented in Section 6, along with the comparative evaluation of the proposed DCFly with a multi-hop data aggregation mechanism (referred as *DA mechanism*). The DCFly-based data collection is also compared with data collection from the network relying on mobile elements (MEs) in Sections 6.1 and 6.2. Finally, this paper is concluded in Section 7.

2 Related work

As stated earlier, the methods for minimising the energy consumption in the network can be broadly categorised into node mobility and in-network data processing-based mechanisms.

2.1 Node mobility

A mobile node can function as *mobile BS* (also referred as data collection agent and data mule) that would collect data from the nodes by traversing in their vicinity (communication range) [2]. In this paper, we refer mobile BS as ME for comparison with DCFly. Numerous studies have dealt with data collection from a deployed network where the ME visits some data collection points (nodes) that serve as sub-sink for other adjoining nodes (member nodes) [2–5]. The basic underlying objective being to determine the shortest path that can serve the maximum collection points within the permissible operational limits. Basically, they are variations of the travelling salesman problem that is known to be NP hard. Accordingly, the proposed approaches try to address the problem at hand of path planning, heuristically. The deployment conditions in terms of terrain and likelihood of obstacles/hindrances in the path of ME have not been taken into account. Ma and Yang [4] state that ME might encounter obstacles, but tackle it based on the consideration that a complete map of the sensing region is available that includes details of obstacles, so that the path of ME is adjusted to avoid them. The ME-based data collection approaches can be classified into predetermined path [4, 5] and autonomous movement [6]. Cluster-based path planning for the ME has been presented in [6], where the ME is expected to be guided by routing agents (CHs). The advantage of cluster-based guiding is that nodes and ME are not required to be location aware; however, the limitation of ground conditions has not been taken into account. In addition, the clustering is not done following the principles of optimal clustering (discussed in Section 3.1), and without optimal clustering there would be localised trees between collection points and adjoining nodes. The most commonly accepted mechanism for path planning involves shortest path – minimal spanning tree and its variations. Throughout this

paper for comparison a data collection round refers to one round by the DCFly or ME across the sensing region. The energy consumption of only the network nodes has been taken into account for the evaluations, energy consumption of mobile node is not considered, that is, DCFly and ME as it is case specific.

The actuation mechanism for movement of mobile nodes using springs, wings, and even passive mobility using water flow has been presented in addition to wheels [7–9]. The concept of *flying sensors* has been proposed by Dantu *et al.* [9]. Flying sensors are intended to operate as an aerial mobile sensor network for addressing application such as crowd monitoring, urban surveillance, and indoor emergencies [9, 10]. They are expected to sense for a given physical phenomenon, and report it to the BS. It is worth mentioning here that a major share of total energy in flying sensors is spent for actuation, leaving nominal energy resources for sensing operations. The DCFly would serve as an ME, and will not be involved with active sensing for any physical phenomenon; therefore, almost all energy resource would be available for actuation.

2.2 In-network data processing

In-network data processing refers to data operations that are intended to aggregate the raw data, and even eliminate surplus or undesired data; done with the intention of reducing the energy consumption in the network. Some nodes process the raw data based on the in-network data processing mechanism before transferring the data to the BS. Many data aggregation techniques for sensor networks have been proposed using different aggregation functions such as sum, average, and based on spatial/temporal factors. Similarly, different structures for data aggregation such as tree based, cluster based, and structure free data aggregation have been put forward [11–13]. Use of Slepian Wolf source coding technique for data collection has also been proposed, its main advantage being that it shifts computational complexity to the BS, compared with standard data aggregation techniques [12, 14]. However, this technique has an operation limitation that prior to commencing operations the spatial and temporal details of the sensing region are required. Compressive sensing-based data gathering in WSNs has been put forward in [15], as a mechanism to reduce inter-node communication required for data gathering in a network. In the proposed approach using DCFly, in-network data processing can be supported, as the CH could implement any appropriate data aggregation mechanism or even compressive sensing on the data.

3 Preliminaries

As stated earlier, the DCFly would support the BS located far from the deployment region. This is contrary to the existing notion of WSN network architecture. Accordingly, the inter-relation between CH election (clustering) and BS is discussed in the following section. Subsequently, DCFly as an MAV and the network conditions under which the data would be collected are discussed.

3.1 Cluster head

It is stated in [16, 17] that the radius of a cluster (CH_{rad}) can be represented as $\sqrt{\text{Area}/\text{CH}_{opt}}$, where *Area* represents the size of sensing region and CH_{opt} represents the optimum number of CHs – clusters in the sensing region. The optimum number of clusters in the sensing region can be represented as a circle with the CH at the centre. The number of optimum CHs can be derived as [16]

$$\text{CH}_{opt} = \sqrt{\frac{N\epsilon_{fs}A}{2\pi\left(\epsilon_{amp}d_{toBS}^{n} - E_{elec}^{Rx}\right)}} \tag{1}$$

where N represents the total nodes in the sensing region, d_{toBS} is the distance to the BS from nodes, E_{elec}^{Rx} is the energy consumption of the receiver circuit, and ϵ_{amp} is the energy consumed by transmitter amplifier, A represents the area of the sensing region, and ϵ_{fs} represents the free space radio propagation model. The optimal number of clusters depends on the following network parameters [16]

$$\text{CH}_{opt} = f(N,\ M \text{ or } R,\ d_{toBS},\ E_{elec}^{Rx},\ \epsilon_{amp}) \tag{2}$$

where M or R represents the dimensions of the sensing region. The authors further state that the number of clusters reduces as the BS moves away from centre of the sensing region, and for a BS located far from the sensing region (more than the communication range of a CH), only one large cluster spanning the sensing region would be feasible. This would be necessitated, as data transmission cost with more than one cluster transmitting to the BS would be operationally infeasible. It can be inferred for a BS located far from the sensing region, all nodes would be subjected to multi-hop communication to reach the single CH that communicates with the BS. In the proposed approach, data would be collected aerially removing the necessity of multi-hop CH–CH communication with BS. This paper does not per se look into CH elections and clustering.

3.2 DCFly

In the proposed method, the DCFly traverses the sensing region collecting data from the clusters. For collecting data from the cluster, DCFly descends from its normal flight path and we refer to this as DCFly's *cluster-hop* (as shown in Fig. 1). The top view of the clusters in the sensing region is shown in Fig. 2. The DCFly is assumed to be an MAV capable of performing a vertical take off and landing (VTOL) at cluster-hops (represented as a mini helicopter in these figures). Environmental monitoring in a forest region is considered as an application in this paper. The DCFly is assumed to navigate at a height of 2 m above the tree canopy to stay clear of any obstruction, and make a cluster-hop on tree canopy for data collection (axial flight). In applications where it is infeasible for the DCFly to land (static position), it would still have axial flight to decrease the flying height to be within the communication range of CHs (shown in Fig. 1).

The DCFly would house a horizontally oriented antenna so that it has a good radiation pattern to communicate with the CHs on the ground. The DCFly and other nodes in the network are all considered to possess a whip antenna with a maximum radiating power of at least 0 dBm; this would comfortably suffice the vertical data communication between CHs and aerial DCFly. On the basis of field evaluations, it can be determined whether the nodes (on ground) can have a horizontal or vertically oriented antenna, and also the need for additional radiating power.

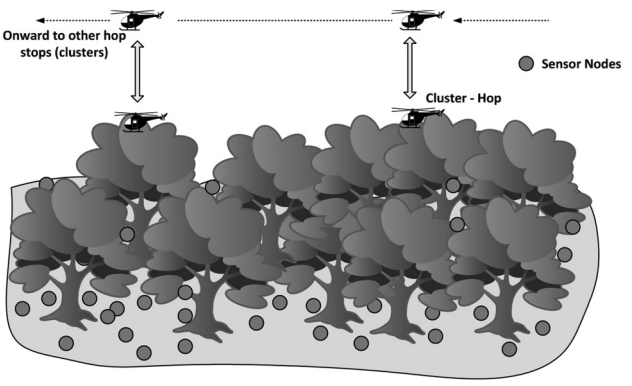

Fig. 1 *DCFly – forward and axial flight (cluster-hops) in a forest area [1]*

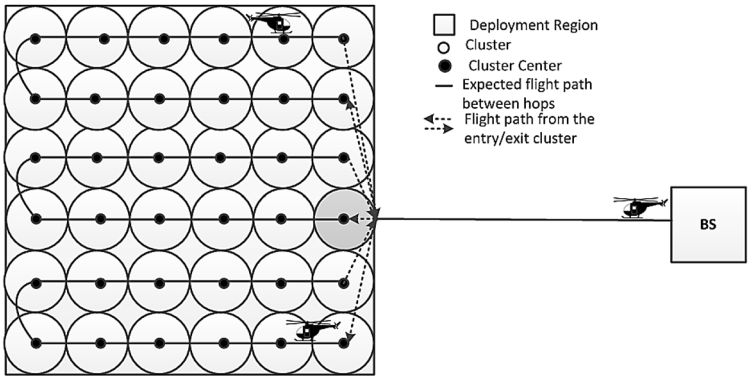

Fig. 2 *Sensing region from top, path of DCFly [1]*

4 DCFly: governing factors

For data collection by a DCFly, there are certain factors that would influence the optimum number of cluster-hops that could be covered. The size of the cluster would be of primary importance, as for small sized clusters there would be more cluster-hops and the DCFly would be required to change flight states from forward to axial for collecting data, imposing additional burden on the energy resources. Additionally, making cluster-hops would incur extra time for traversing a given sensing region compared with an MAV traversing it in a continuous forward flight. Large clusters would not be favourable either as they necessitate distant nodes within the cluster to communicate with the CH in multiple hops. There would also be a maximum serviceable limit (distance in metres) that the DCFly can traverse with a full charged battery. Therefore, the clustering of the nodes, that is, optimal number of clusters in the proposed mechanism are dependent on the operational capacity of the DCFly in comparison with the factors for determining optimal clusters stated earlier in (2). The above factors governing the DCFly's utility, derived in appropriate mathematical notation are formulated as an optimisation problem to determine the clusters that can be covered by a single DCFly, further elaborated in Section 5. Specific network conditions under which this data collection mechanism is expected to operate are as follows:

(i) The delivery deadline at the BS is high (10–15 min).
(ii) The nodes have unique IDs and the CHs have access to the IDs of the nodes in their cluster.
(iii) The DCFly has capacity to hook/cling to stay firmly at the cluster-hop.
(iv) Nodes in the network are unaware of their location, including the DCFly.
(v) The sensing region is square in shape so that it can be divided into even number of rows and columns, with one DCFly covering a pair of rows.
(vi) Beyond the entry/exit cluster the DCFly is able to navigate to the BS without path guiding assistance, and it is possible to recharge the DCFly at the BS between data collection rounds.
(vii) DCFly is assisted by the clusters in the first column from entry/exit cluster for navigating to its assigned rows (Fig. 2).
(viii) The DCFly is not impacted by external factors such as wind while traversing the sensing region.

5 Optimisation problem formulation

Factors that decide the DCFly's total possible operation are as follows:

(i) Maximum serviceable limit of the DCFly on a full charged battery, and power consumed in forward/axial flight (cluster-hop).
(ii) Total time taken for traversing across the sensing region, and time spent per cluster-hop, in relation to the delivery deadline/data freshness requirement at the BS.
(iii) Size of the clusters and distance of the farthest non-CH node from the CH.

5.1 Power consumption – forward flight and axial flight (cluster-hop)

There is no direct relation stated between the forward flight and axial climb power consumption to the best of authors' knowledge, as it is case specific. However, both forward (P_{fwd}) and axial (P_{ax}) power are related to hover power (P_{hvr}) consumption and this is used in deriving the relation between forward and axial powers [18]. The calculations derived here would be applicable to a generalised representation of an MAV capable of performing VTOL. The power consumption for the forward flight is derived based on the power ratio of forward and hover powers, given as [18]

$$\frac{P_{\text{fwd}}}{P_{\text{hvr}}} = \frac{V_\infty \sin\alpha + v_{\text{ind}}}{v_{\text{hvr}}} \tag{3}$$

From this the power for forward flight can be represented as

$$P_{\text{fwd}} = \frac{V_\infty \sin\alpha + V_{\text{ind}}}{v_{\text{hvr}}} P_{\text{hvr}} \tag{4}$$

Here V_∞ represents the magnitude of free stream velocity and V_{ind} as the propeller induced velocity. The relation between the axial power (climb and descend) with hover power is given as follows (replaced with negative signs for descend) [18]

$$\frac{P_{\text{ax}}}{P_{\text{hvr}}} = \frac{V_{\text{ax}}}{2v_{\text{hvr}}} + \sqrt{\left(\frac{V_{\text{ax}}}{2v_{\text{hvr}}}\right)^2 + 1} \quad \text{for} \quad \frac{V_{\text{ax}}}{v_{\text{hvr}}} \geq 0 \tag{5}$$

On the basis of universal power curve, that is, a plot of $P_{\text{ax}}/P_{\text{hvr}}$ versus $V_{\text{ax}}/v_{\text{hvr}}$, we consider the ratio between hover V_{hvr} and climb velocity V_{ax}, within normal working stated as [18]

$$\frac{V_{\text{ax}}}{v_{\text{hvr}}} = 2 \tag{6}$$

For the aforesaid velocity ratio, ratio of $P_{\text{ax}}/P_{\text{hvr}}$ would also be 2 based on the universal power curve. Axial climb flight power consumption is always higher than hover power, while for certain velocity rates the descend power consumption could be lower than hover power [18]. However, for this evaluation power consumption for axial climb and descend is considered equal to compensate for any additional power consumption that may take place during conversion between flight modes: hover – descend and climb – hover. Forward flight power consumption (P_{fwd}) as

stated in (4) would be lower than hover flight power consumption (P_{hvr}) as there is a translational lift that assists forward flight with more air on the lower part of the rotor disk, leading to power saving. On the basis of this operational feature, we consider forward flight power as $P_{fwd} = 0.5P_{hvr}$ and then the relation between forward power and axial power would be

$$P_{ax} = 4P_{fwd} \tag{7}$$

These relations are applied for power consumption on a per unit basis. If the DCFly was supposed to cover the entire sensing region just in forward flight the total service limit represented by Dist_{serv} would be given as

$$\frac{E_{batt}}{P_{fwd}} = \text{Dist}_{serv}(m) \tag{8}$$

Here E_{batt} refers to the energy reserve of the DCFly's battery. For the DCFly to make cluster-hops for collecting data, the service limit would be lesser compared with an only forward flight. Considering that the DCFly traverses at a height 2 m higher than the tree canopy as stated earlier, taking into account both descend and ascend, axial power based on (7) would be

$$P_{ax} = 16P_{fwd} \tag{9}$$

Accordingly, the maximum service limit with only forward flight as given in (8), supposing the DCFly traverses covering only cluster-hops (CH_{hops}) would be

$$\frac{E_{batt}}{16P_{fwd}\text{CH}_{hops}} \tag{10}$$

In actuality, the DCFly's navigation would comprise of forward flight between cluster-hops and axial flights on the cluster-hops, and accordingly (10) would change to

$$\frac{E_{batt}}{16P_{fwd}\text{CH}_{hops} + 2\text{CH}_{rad}\text{CH}_{hops}} = \text{Dist}_{serv}(m) \tag{11}$$

where CH_{rad} represents the cluster radius. The above equation should have taken the second term in denominator as multiple with ($\text{CH}_{hops}-1$), but considered as CH_{hops} to take into account the distance navigable by the DCFly from the entry/exit cluster to the commencement row and vice versa from the termination row to the entry/exit cluster, and sideward flight operation while switching rows.

5.2 DCFly time for data collection round and cluster-hop

Total time required for one data collection round would comprise

$$T^{n}_{slots} + T^{DA}_{CH} + T^{DCFly}_{CH} + T^{round}_{DCFly} \tag{12}$$

In (12), it is the time for the DCFly data collection round T^{round}_{fly} and time for the DCFly to collect data from the CH T^{DCFly}_{CH} which are of significance, in comparison with time for the CH to collect data from all nodes T^{n}_{slots}, and time taken by the CH to carry-out the data aggregation based on the aggregating function (T^{DA}_{CH}). Especially, considering the fact that the T^{n}_{slots} and T^{DA}_{CH} tasks would run in parallel in all clusters, and would not influence total service time of the DCFly. The distance in relation to the time constraint for optimum delivery of the collected data based on T^{round}_{fly} can be represented as

$$16\text{CH}_{hops} + 2\text{CH}_{rad} \tag{13}$$

Twice the cluster radius (CH_{rad}) would be the distance between adjoining CHs organised in rows as shown in Fig. 2. It is assumed that the DCFly traverses between two clusters covering half the distance at uniform speed U_{DCFly}, and the remaining half decelerating to reach a standstill state (hover), followed by axial flight (cluster-hop). On the basis of this condition, the time t_1 for the DCFly to cover at least half the distance (CH_{rad}) between the clusters is as follows

$$t_1 = \frac{\text{CH}_{rad}}{U_{DCFly}} \tag{14}$$

Time for covering second half of the distance between the clusters t_2 is derived using *Equations of Motion* as follows

$$v^2 = u^2 + 2as \tag{15}$$

$$V^2_{DCfly} = U^2_{DCFly} + 2a\text{CH}_{rad} \tag{16}$$

$$a = -\left(\frac{U^2_{DCFly}}{2\text{CH}_{rad}}\right) \tag{17}$$

$$v = u + at : \text{substituting final velocity}(V_{DCfly} = 0) \tag{18}$$

$$t_2 = \left(\frac{2\text{CH}_{rad}}{U_{DCFly}}\right) \tag{19}$$

Accordingly, (12) can be represented based on the above calculations as

$$T^{n}_{slots} + T^{DA}_{CH} + T^{DCFly}_{CH}\text{CH}_{hops} + \text{CH}_{hops}\left(\frac{3\text{CH}_{rad}}{U_{DCFly}}\right) \tag{20}$$

where only the last two terms hold significance for the DCFly as stated earlier. Considering the delivery deadline set by BS as T_{delv}.

The objective function is formulated as

$$\text{Max}\ \text{CH}_{hops} + \text{CH}_{rad} + T^{DCFly}_{CH}$$

s.t.

$$16P_{fwd}\text{CH}_{hops} + 2(\text{CH}_{rad})\text{CH}_{hops} \le \text{Dist}_{serv} \tag{21}$$

$$T^{DCFly}_{CH}\text{CH}_{hops} + \left(\frac{3\text{CH}_{rad}}{U_{DCFly}}\right)\text{CH}_{hops} \le T_{delv} \tag{22}$$

$$\text{CH}_{rad} \le \text{radius}_{upper} \tag{23}$$

$$\text{CH}_{rad} \ge \text{radius}_{lower} \tag{24}$$

$$\text{CH}_{rad}T^{DCFly}_{CH} \le u\,\text{bound}_1 \tag{25}$$

$$\text{CH}_{hops} \in \mathbb{Z} \text{ and } \text{CH}_{hops},\ \text{CH}_{rad},\ T^{DCFly}_{CH} \ge 0 \tag{26}$$

The inequality constraints radius_{upper} and radius_{lower} refer to the upper and lower bounds on the cluster radius, that is, distance to farthest node from the CH. Constraint $u\text{bound}_1$ is an upper bound on the maximum product of CH_{rad} and T^{DCFly}_{CH} which leads to the desired maximisation of CH_{hops}, through the objective function requiring a cumulative product maximisation of the three terms. The optimisation problem comprises of non-linear inequalities, introduced due to the product of terms in (22) and (25). The variable CH_{hops} represents the cluster-hops, and can only be a positive integer. On the basis of these conditions, this is a mixed integer non-linear programming (MINLP) problem. We solved this optimisation problem using general algebraic modelling system, and a corresponding MINLP solver – discrete and continuous OPTimizer (DICOPT). It can be observed that the limitation imposed by the constraints can be changed as per the actual operational specification of a given MAV (DCFly).

Table 1 DCFly sensing region

Clusters	Maximum range, *m*	Delivery deadline time, *s*
4	224	520
6	336	780
8	448	1040
10	560	1300
12	672	1560

6 Results and discussion

As stated in the previous section by changing the values imposed by the constraints, and the upper and lower bounds on the variables we can obtain the optimum cluster-hops that can be serviced by a given DCFly in one data collection round. Accordingly, we have used the following values for the constraint limits and bounds: $U_{\text{DCFly}} = 0.5$ m/s, $\text{radius}_{\text{lower}} = 20$ m, $\text{radius}_{\text{upper}} = 25$ m, and $u\,\text{bound}_1 = 200$. Additionally, we are concerned with obtaining even number of cluster-hops so that the DCFly can cover half of them in one row and the remaining half in the other row (Fig. 2). It is intended to provide the number of clusters in a row beforehand to the DCFly so that it *counts* cluster-hops serviced in one row and then switches to the other row in reverse direction. Whether the DCFly turns right or left for completing the exiting row would be opposite in direction to that it flew while entering the sensing region as shown in Fig. 2. The optimum clusters covered by the DCFly along with applicable service limit and delivery time are shown in Table 1. CH_{rad} was 20 m through the applicable bounds stated earlier and the permissible time for the DCFly at each cluster-hop ($T_{\text{CH}}^{\text{DCFly}}$) at 10 s. The communication radius R_{c} and sensing radius R_{s} are considered as 10 and 5 m, respectively. Therefore, the farthest nodes could be maximum two hops away from the CH. Accordingly, the sensing region that is covered by DCFly can be determined subtracting twice the dimension (length) of square sensing region from the maximum range stated in Table 1. Since we want the DCFly to scan two rows in its forward and return flight. On the basis of the sensing region dimensions and width of two rows serviced by a single DCFly, the number of DCFly(s) that would be required to cover the complete sensing region can be determined before commencement of the network operations.

The proposed mechanism is compared with multi-hop data aggregation (CH–CH communication) in the network, referred here as *DA mechanism*. The mechanisms are evaluated considering two parameters, that is, the total energy consumed for completing one data collection round of computation and transmission error. The energy consumed in one computation round of data aggregation in a network considering CH and non-CH node energy consumptions is as follows [16, 17]

$$E_{\text{CH}} = lE_{\text{elec}}^{\text{Rx}}\left(\frac{N}{\text{CH}_{\text{opt}}} - 1\right) + l\epsilon_{\text{DA}}\frac{N}{\text{CH}_{\text{opt}}} + lE_{\text{elec}}^{\text{Tx}} + l\epsilon_{\text{amp}}d_{\text{toBS}}^{n} \tag{27}$$

where l is the packet data size, $E_{\text{elec}}^{\text{Rx}}$ is the energy consumed in order to receive a packet, and $E_{\text{elec}}^{\text{Tx}}$ is the energy consumed to transmit a data packet. N is the number of nodes, CH_{opt} is the optimum number of clusters, and ϵ_{amp} is the energy consumed by the transmitter amplifier. The energy consumed in a non-CH node is represented as [16, 17]

$$E_{\text{nonCH}} = lE_{\text{elec}}^{\text{Tx}} + l\epsilon_{\text{amp}}d_{\text{toCH}}^{n} \tag{28}$$

The transmission amplifier ϵ_{amp} is considered to operate based on a free space propagation model as $\epsilon_{\text{amp}} = \epsilon_{\text{fs}}$ for communication between nodes and the communication between CHs in (27), considered for deriving the energy consumed in CH and non-CH nodes. The total energy consumed in the network can be stated as [16, 17]

$$E_{\text{round}} = E_{\text{CH}}^{\text{Total}} + E_{\text{nonCH}}^{\text{Total}} \tag{29}$$

It is considered in (27) that CHs are capable of communicating directly with the BS, in the current evaluation it is considered that CHs communicate with each other in a multi-hop manner, and that only CHs close to the BS are able to communicate directly. Furthermore, as stated earlier the DCFly's specific utility is for deployments in which the BS is anticipated to be far from the sensing region. The CHs consume energy in receiving the data from the farther neighbour, and transmitting their data along with the neighbour CH's data to the next CH toward the BS. The farthest CH would not be involved in receiving any data packets and accordingly the term CH_{opt} is represented as $\text{CH}_{\text{opt}} - 1$ (multiplied with $E_{\text{elec}}^{\text{Rx}}$). The total energy consumed by the CHs in one data collection round is represented as (applicable to *DA mechanism*) (see (30))

$$E_{\text{CH}} = l\left(\left(E_{\text{elec}}^{\text{Rx}}\left(\frac{N}{\text{CH}_{\text{opt}}} - 1\right) + \epsilon_{\text{DA}}\frac{N}{\text{CH}_{\text{opt}}} + \epsilon_{\text{fs}}d_{\text{nextCH}}^{2}\right)\text{CH}_{\text{opt}} + \text{CH}_{\text{opt}}E_{\text{elec}}^{\text{Tx}} + E_{\text{elec}}^{\text{Rx}}(\text{CH}_{\text{opt}} - 1)\right) \tag{30}$$

where ϵ_{DA} represents the energy consumption for data aggregation operation and d_{nextCH}^{2} is the distance to next CH. Energy consumption for a single non-CH as represented by (28) would result in energy consumption for all non-CH nodes collectively as

$$E_{\text{nonCH}} = l\left(E_{\text{elec}}^{\text{Tx}} + \epsilon_{\text{fs}}d_{\text{toCH}}^{2}\left(N - \text{CH}_{\text{opt}}\right)\right) \tag{31}$$

In the DCFly-based data collection the energy spent in one data collection round is given as

$$E_{\text{CH}} = l\left(E_{\text{elec}}^{\text{Rx}}\left(\frac{N}{\text{CH}_{\text{opt}}} - 1\right) + \epsilon_{\text{DA}}\frac{N}{\text{CH}_{\text{opt}}}\right)(\text{CH}_{\text{opt}}) + l(\epsilon_{\text{fs}}d_{\text{tofly}}^{2} + E_{\text{elec}}^{\text{Tx}})\text{CH}_{\text{opt}} \tag{32}$$

where d_{tofly} represents the distance of DCFly from the CH. The energy spent in the non-CH nodes would be same as in (31).

It should be noted here that we derived our even number of clusters, such that the DCFly covers half of them in one row and the remaining half in the other row. Therefore, half of the optimum clusters (clusters in one row) as stated in Table 2 are considered for the numerical evaluation. The total energy for completion of one data collection round at the BS is shown in Fig. 3*a*. The energy consumption per round for DCFly is marginally higher than the DA mechanism, but it is well compensated by the avoidance of CH–CH multi-hop communication, possibility to place BS away from the sensing region, and the network's capacity to operate on undulating terrain.

Significant packet losses occur in multi-hop communication that impact the overall reliability and specifically network lifetime due to the need for retransmissions. Especially, considering the fact that energy consumption for radio communication is manifold higher than power consumption for computation on the node [17]. The two mechanisms are compared for the possible transmission errors they may incur based on [19]

$$p = 1 - (1 - p_{\text{link}})^{N_{\text{hps}}} \tag{33}$$

where p is the probability of error on the entire path and p_{link} is the packet error rate of an individual link, and for the comparison two values are considered 0.05 or 0.20. N_{hps} represents the number of hops on the path. Using these values in (33), the difference of the

Table 2 Parameters used for numerical evaluation

Parameter	Value
N	50
E_{elec}^{Tx}	100 nJ/bit
E_{elec}^{Rx}	100 nJ/bit
CH_{opt}/N_{hps}	2,3,4,5,6
ϵ_{DA}	5 nJ/bit/signal
e_{fs}	10 pJ/bit/m^2
d_{nextCH}	40 m
d_{toCH}/d_{tofly}	20 m
l	256 bytes
p_{link}	0.05,0.2

two mechanisms in reliability of data communication for varying number of clusters is shown in Fig. 3*b*. It can be observed from this figure that with the DCFly mechanism, the total transmission error stays constant as there is just one hop between the CH and the DCFly on the tree canopy, that is, cluster-hop.

DCFly is compared with ME-based data collection approaches, where ME is expected to traverse across the sensing region moving on the ground and visiting data collection points. The two approaches have been compared in terms of data collection points covered (basically number of vertices on a graph), energy consumption, and network lifetime.

6.1 DCFly and ME: data collection points covered

The most appropriate and largely accepted method for determining the data collection path for an ME is basically determining the shortest path connecting data collection points. As stated earlier, determining the shortest path connecting a set of points has been done following heuristic approaches. Therefore, the appropriate comparison would rather be to determine the operational difference between DCFly and ME. As stated earlier, the most common method for determining the data collection path for ME rely on shortest path – minimum spanning tree and their variations. Accordingly, let the network consist of a complete graph $G = (V, E)$ where V is the set of data collection points and E is the set of edges between the data collection points. Let the set of nodes that serve as data collections points for the visiting ME be M and the set of other adjoining nodes (member nodes) that deliver their data to the data collection point be represented by N. The approaches are intended to be compared in terms of ratio of data collection points covered by the visiting ME to length of ME path. As stated earlier, in the DCFly approach the nodes are considered to be clustered with the CH located at the centre of a circle. The distance traversed per cluster would be the diameter of the circle (twice of CH_{rad}, i.e. 40 m) is considered in determining path length versus number of nodes served. The average node degree (d) is given as [20]

$$d = \frac{n\pi R_c^2}{A^2} = \mu\pi R_c^2, \tag{34}$$

where n is the number of nodes, A is the length of the sensing region, and μ is the density of nodes (n/A^2). For the numerical analysis following values are considered: $n = 100$, $A = 100$ m, and $\mu = 0.01$. The average number of nodes in a cluster is given as probability (P_c) that a node is inside the cluster given as [20]

$$P_c = \frac{\pi k^2 R_c^2}{A^2}, \tag{35}$$

where k is the number of hops from farthest node to CH (considered $k = 2$), substituting from (34) for d

$$P_c = \frac{dk^2}{n} \tag{36}$$

The average size of a cluster is therefore [20]

$$E(N_c) = nP_c = dk^2 \tag{37}$$

We are considering the minimum spanning tree and Steiner minimum tree as the tree approaches used for planning the ME path. Minimum spanning tree and Steiner minimum tree have been tackled based on approximation algorithms and heuristic approaches, and therefore determining the number of data collection points is not directly feasible given sets of vertices and edges.

Assuming the length served by the ME is L, and average length of the edge connecting two vertices as l_{edge}, then $L/l_{edge} = w$, where w is the number of edges, and number of vertices would be $w + 1$. In the case of the Steiner tree, the number of vertices covered is not known as some approximate points function as Steiner points [5]. However, for a given set of points P on the euclidean plane, length of Steiner minimum tree (L_s), and length of minimum spanning tree L_m are related as: $L_s(P) \geq (\sqrt{3}/2)L_m(P)$. The inequality is referred as Gilbert–Pollak conjecture [21]. Here, we consider both sides

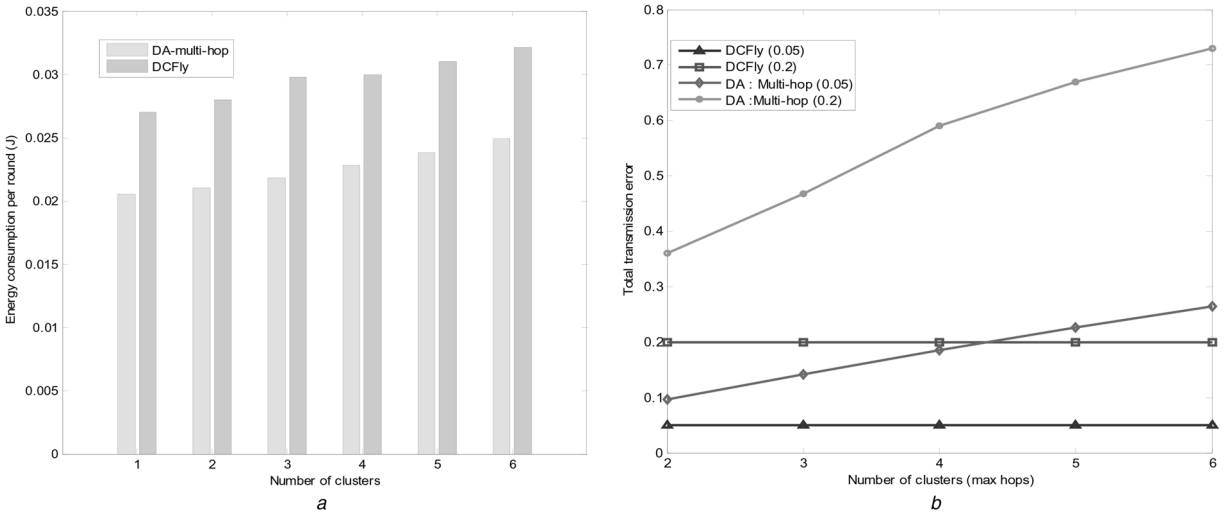

Fig. 3 *Comparison of the two mechanisms*

a Energy consumed
b Transmission error

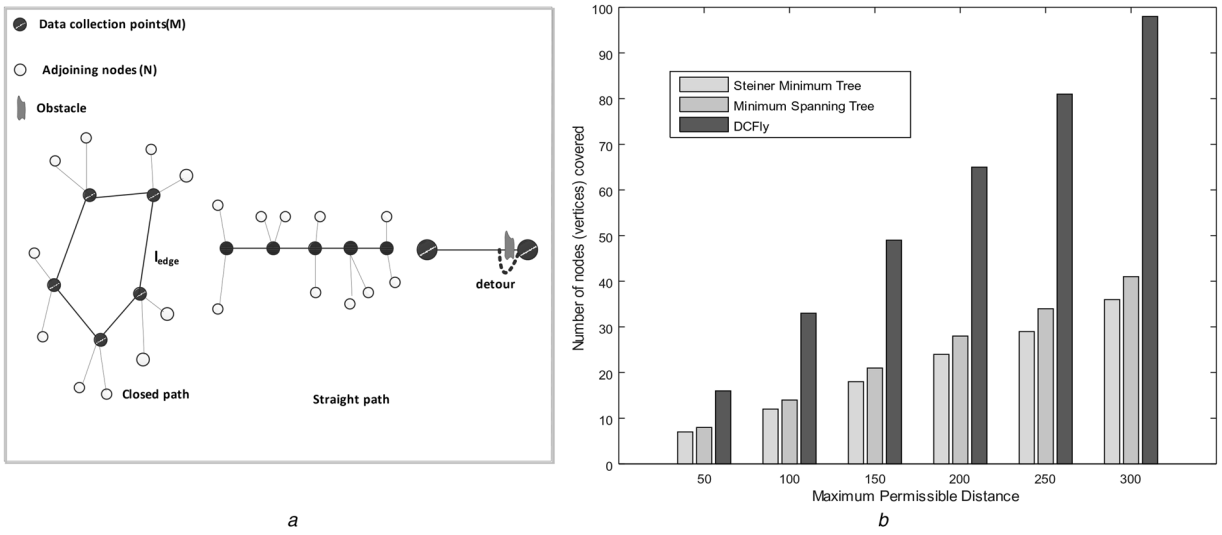

Fig. 4 *Comparison of DCFly and ME: data collection points covered*

a Data collection path and detour due to obstacle on l_{edge}
b Number of data collection points covered in reference to ME path length

equated removing the inequality

$$L_s(P) = \left(\frac{\sqrt{3}}{2}\right) L_m(P) \tag{38}$$

It would be appropriate to consider that number of vertices for both trees would be related in the same proportion as the total length. The average length of the edge (l_{edge}) is taken as $(R_s + R_c)/2 = 7.5$ m, considering that the minimum legitimate distance between two nodes is R_s (else they overlap) and the maximum distance is R_c (maximum communication range). The influence of changing the total length is shown in Fig. 4*b*, it can be inferred from this figure that with aerial data collection from an optimal cluster layout, the number of nodes covered is manifold higher than the Steiner minimal tree and minimum spanning tree. The higher the number of nodes covered with unit length of ME path the better it is, as each collection point can serve for numerous adjoining nodes, and accordingly larger sensing region could be covered.

6.2 DCFly and ME: energy consumption and network lifetime

To collect the data, the ME is expected to reposition itself in the proximity of data collection points. The energy consumption for an individual node can be given as [3]

$$p \simeq e(l_r + l_t), \tag{39}$$

where l_r and l_t represent the number of bits received and transmitted and e represents the energy consumption per bit transmitted. Let λ represent the data generation rate of an individual node j during one data collection round of the ME, then $l_t^j = l_r^j + \lambda$, where l_t^j and l_r^j represent the amount of data transmitted and received by node $j \in N$, all nodes are assumed to transmit their data to the nearest data collection point. Representing the total data in terms of hop is given as

$$\sum_{j=1}^{N} l_r^j = \sum_{j=1}^{N} h_j \lambda, \tag{40}$$

where h_j is the lowest number of hops from the node to the destination collection point. The total energy consumption of the nodes would be [3]

$$p = \sum_{j=1}^{N} p_j \simeq e(l_r^j + l_t^j) = \sum_{j=1}^{N} = e(2l_r^j + \lambda)$$
$$= \sum_{j=1}^{N} e(2h_j + 1)\lambda \tag{41}$$

However, in addition to transmitting the collected data, data collection points are assumed to be responsible for ME path planning. The data collection points are expected to at least coordinate with their neighbouring nodes, that is, M_i coordinates with M_{i+1} and M_{i-1}, this is based on an assumption that each collection point is responsible for its two edges, collectively expressed for all Ms as $2M(e)$. The data collection by ME is prone to encountering obstacles and hindrances on ground that can force the ME into taking a detour from the shortest path and necessitate additional communication overhead (shown in Fig. 4*a*). The possible obstacles that may be encountered are hard to predict and this could be the justification for non-existing work on this. However, the novelty of the proposed DCFly is centred on overcoming this underlying limitation of data collection by ground-based ME. Therefore, an appropriate model to take into account the likelihood of an obstacle and ground terrain conditions is necessitated. The probability of an obstacle P_{obs} can be based on the general profile of the sensing region; an obstacle/hindrance could be anything, e.g. boulder, log of wood, and pothole that could limit the movement on the shortest path between the collection points. The ground roughness, that is, the amount of smoothness or undulating nature of terrain is determined based on radio propagation path loss relying on finite difference time domain analysis [22]. Accordingly, we propose an expression that represents the amount of overhead due to obstacles/hindrances and the ground terrain (roughness/smoothness) as

$$e(P_{obs} + G_{rough}) l_{edge} w, \tag{42}$$

where G_{rough} represents the roughness of the ground terrain and w is the number of edges. G_{rough} would be zero only for an absolutely levelled ground and P_{obs} would be zero if there is no obstacle at all; therefore, the expression would be zero only under ideal conditions. Accordingly, the overall energy consumption by all

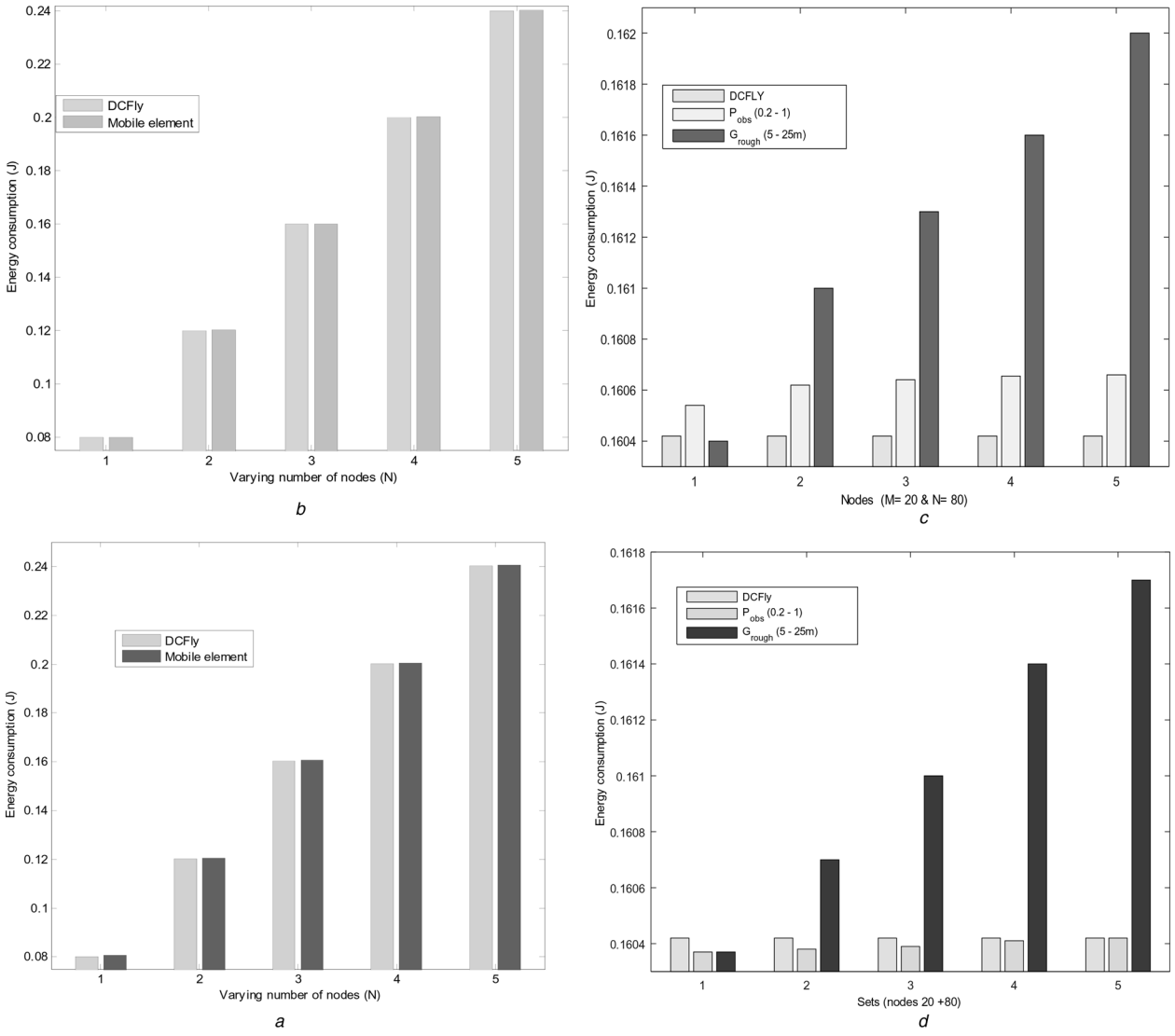

Fig. 5 *Comparison of DCFly and ME: energy consumption*

a DCFly and ME in close route path
b DCFly and ME on a straight line path
c Varying P_{obs} and G_{rough} influence on ME compared with DCFly (constant value) on closed path
d P_{obs} and G_{rough} based ME and DCFly (constant) on straight path

data collection points *M*s

$$Me(2h_j + 1)\lambda + 2M(e) + e(P_{obs} + G_{rough})l_{edge}w \quad (43)$$

The aerial data collection for the same route would however not encounter any obstacle/hindrance or impact of non-levelled ground. It is, however, assumed that the path planning would be the same for ME and DCFly [especially covering a closed irregular path ($c-$path) in the sensing region]. For the sake of comparison, the data collection points organised in a straight line are also considered, assuming that movement across a straight line path ($s-$path) would require no path planning. The ME-based data collection would however still be impacted by the obstacles and hindrances. Energy consumption for all ME(s) presented in (43) would change for straight line path as

$$Me(2h_j + 1)\lambda + e(P_{obs} + G_{rough})l_{edge}w \quad (44)$$

For DCFly (44) would be applicable in both *c-path* and *s-path*.

For the numerical analysis, we are considering $M = 20$ (constant throughout) and number of adjoining nodes (N) (varying between 2 and 6). All collection points are considered to have equal number of nodes. The P_{obs} is varied between 0.2 and 1 (increments of 0.2) and the G_{rough} between 5 and 25 m (increments of 5). The data generation rate of node $\lambda = 200$ bps and the nodes are expected to transfer data for 10 s for every data collection round. The radio energy factor $e = 0.5$ µJ, l_{edge} is 7.5 m, and w is 19. It can be observed from these figures that there is almost no difference between data collection between DCFly and ME (Figs. 5*a* and *b*), as $P_{obs} = 0.2$ and $G_{rough} = 5$ m are low, that is, 0.00037 J. In Fig. 5*c* the value for DCFly-based data collection is kept constant ($M = 20$ and $N = 80$) and is compared influence of change in P_{obs} (keeping G_{rough} constant) and vice versa for G_{rough} on ME. In comparison with Figs. 5*a* and *b*, the influence of increasing P_{obs} and G_{rough} is evident in Figs. 5*c* and *d*. The difference is more pronounced in the case of G_{rough} in comparison with P_{obs}. On the basis of these figures, it can be inferred that data collection using ME on obstacle ridden and undulating terrain is manifold higher than aerial data collection. The impact on the network lifetime has been determined considering the total initial energy of all nodes ($M+N$, initial energy of all nodes 20 J), collectively, divided by the energy consumed by the nodes in one data collection round, shown in Fig. 6. The advantage of data collection by DCFly is amply evident in comparison with ME-based approaches, and it can be concluded that with a larger

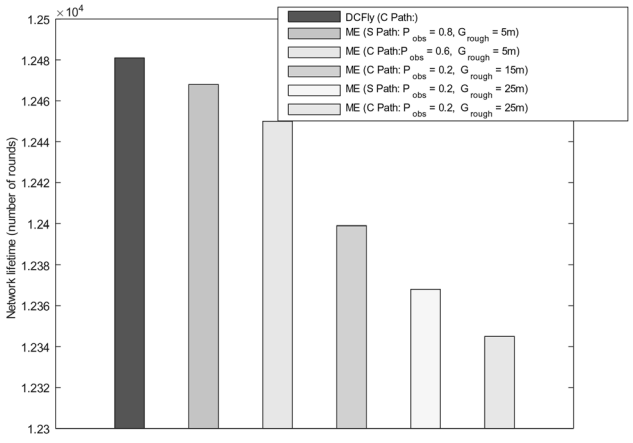

Fig. 6 *Network lifetime for the different approaches*

number of data collection points and adjoining nodes the effect of ground roughness and obstacle/hindrance-based overhead would become even more pronounced.

7 Conclusions

On the basis of the comparative evaluation of DCFly(s) with multi-hop data collection, and data collection by ME(s), it can be concluded that data collection using DCFly(s) is very promising and practical. Utility of MAVs for data collection instead of active sensing as proposed in this paper justifies the enormous unexplored utility of node mobility in the form of MAVs. Previously, the research on MAVs has been centred on two aspects, that is, efficient actuation and enhanced sensing capacity, but through the proposed mechanism, attention can be centred on achieving efficient actuation alone as it functions as a data collection agent (no sensing). The proposed mechanism is well suited for deployment of WSNs in harsh and inhabitable terrain. Multi-hop communication between CHs (a major energy consuming activity in the network) is completely avoided. This feature proves the phenomenal utility of the proposed mechanism. Usually it is considered that the BS is stationed close to the sensing region, but with the proposed mechanism the BS could be located at a distance from the sensing region, which is realistic for sensor network deployments.

8 References

1 Mathur, P., Nielsen, R., Prasad, N., *et al.*: 'Novel framework for data collection in wireless sensor networks using flying sensors'. 2014 IEEE Int. Conf. on Advanced Networks and Telecommunications Systems (ANTS), 2014

2 Xing, G., Wang, T., Xie, Z., *et al.*: 'Rendezvous planning in wireless sensor networks with mobile elements', *IEEE Trans. Mob. Comput.*, 2208, **7**, pp. 1430–1443

3 Gao, S., Zhang, H., Das, S.K.: 'Efficient data collection in wireless sensor networks with path-constrained mobile sinks', *IEEE Trans. Mob. Comput.*, 2011, **10**, pp. 592–608

4 Ma, M., Yang, Y.: 'Sencar: an energy-efficient data gathering mechanism for large-scale multihop sensor networks', *IEEE Trans. Parallel Distrib. Syst.*, 2007, **18**, pp. 1476–1488

5 Xing, G., Li, M., Wang, T., *et al.*: 'Efficient rendezvous algorithms for mobility-enabled wireless sensor networks', *IEEE Trans. Mob. Comput.*, 2012, **11**, pp. 47–60

6 Ching-Ju Lin, C.-F.C., Chou, P.-L.: 'Hcdd: Hierarchical cluster based data dissemination in wireless sensor networks with mobile sink'. IWCMC '06: Proc. of the 2006 Int. Conf. on Communications and Mobile Computing, 2006

7 Chellappan, S., Bai, X., Ma, B., *et al.*: 'Mobility limited flip-based sensor networks deployment', *IEEE Trans. Parallel Distrib. Syst.*, 2007, **18**, pp. 199–211

8 Lai, T.-t.T., Chen, Y.-h.T., Huang, P., *et al.*: 'Pipeprobe: a mobile sensor droplet for mapping hidden pipeline'. Proc. of the Eighth ACM Conf. on Embedded Networked Sensor Systems, 2010

9 Dantu, K., Kate, B., Waterman, J., *et al.*: 'Programming micro-aerial vehicle swarms with karma'. Proc. of the Ninth ACM Conf. on Embedded Networked Sensor Systems, 2011

10 Purohit, A., Sun, Z., Mokaya, F., *et al.*: 'Sensorfly: controlled-mobile sensing platform for indoor emergency response applications'. 2011 Tenth Int. Conf. on Information Processing in Sensor Networks (IPSN), 2011

11 Harb, H., Makhoul, A., Tawil, R., *et al.*: 'Energy-efficient data aggregation and transfer in periodic sensor networks', *IET Wirel. Sens. Syst.*, 2014, **4**, pp. 149–158

12 Luo, C., Wu, F., Sun, J., *et al.*: 'Efficient measurement generation and pervasive sparsity for compressive data gathering', *IEEE Trans. Wirel. Commun.* , 2010, **9**, pp. 3728–3738

13 Hoang, D., Kumar, R., Panda, S.: 'Optimal data aggregation tree in wireless sensor networks based on intelligent water drops algorithm', *IET Wirel. Sens. Syst.*, 2012, **2**, pp. 282–292

14 Huang, Z., Zheng, J.: 'A slepian-wolf coding based energy-efficient clustering algorithm for data aggregation in wireless sensor networks'. 2012 IEEE Int. Conf. on Communications (ICC), 2012

15 Chong, L., Feng, W., Jun, S., *et al.*: 'Compressive data gathering for large-scale wireless sensor networks'. Proc. of the 15th Annual Int. Conf. on Mobile Computing and Networking, ser. MobiCom '09, New York, NY, USA, 2009, pp. 145–156

16 Amini, N., Vahdatpour, A., Xu, W., *et al.*: 'Cluster size optimization in sensor networks with decentralized cluster-based protocols', *Comput. Commun.*, 2012, **35**, (2), pp. 207–220

17 Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: 'An application-specific protocol architecture for wireless microsensor networks', *IEEE Trans. Wirel. Commun.*, 2002, **1**, pp. 660–670

18 Leishman, J.G.: 'Principles of helicopter aerodynamics', in Rycroft, M., Stengel, R., (EDs.) (Cambridge University Press, 2005)

19 Banerjee, S., Misra, A.: 'Minimum energy paths for reliable communication in multi-hop wireless networks'. Proc. of the Third ACM Int. Symp. on Mobile Ad Hoc Networking and Computing, 2002

20 Youssef, M.A., Youssef, A., Younis, M.F.: 'Overlapping multihop clustering for wireless sensor networks', *IEEE Trans. Parallel Distrib. Syst.*, 2009, **20**, pp. 1844–1856

21 Du, D.-Z., Hwang, F.: 'A proof of the Gilbert–Pollak conjecture on the steiner ratio', *Algorithmica*, 1992, **7**, (1–6), pp. 121–135. [Online]. Available at http ://www.dx.doi.org/10.1007/BF01758755

22 Ohs, R., Schuster, J., Fung, T.: 'Full wave simulation of radiowave propagation in unattended ground sensor networks'. Radio and Wireless Symp., 2009. RWS '09, 2009