Optimizing Energy Consumption for Big Data Collection in Large-Scale Wireless Sensor Networks With Mobile Collectors

Kenneth Li-Minn Ang, *Senior Member, IEEE*, Jasmine Kah Phooi Seng, and Adamu Murtala Zungeru, *Member, IEEE*

***Abstract*—Big sensor-based data environment and the emergence of large-scale wireless sensor networks (LS-WSNs), which are spread over wide geographic areas and contain thousands of sensor nodes, require new techniques for energy-efficient data collection. Recent approaches for data collection in WSNs have focused on techniques using mobile data collectors (MDCs) or sinks. Compared to traditional methods using static sinks, the MDC techniques give two advantages for data collection in LS-WSNs. These techniques can handle data collection over spatially separated geographical regions, and have been shown to require lower node energy consumption. Two common models for data collection using MDCs have been proposed: data collection using data mule (MULE), and sensor network with mobile access point (SENMA). The MULE and SENMA approaches can be characterized as representative of the multihop and the single-hop approaches for mobile data collection in WSNs. Although the basic architectures for MULE and SENMA have been well studied, the emergence of LS-WSNs which require partitioning the network into multiple groups and clusters prior to data collection has not been particularly addressed. This paper presents analytical approaches to determine the node energy consumption for LS-WSN MDC schemes and gives models for determining the optimal number of clusters for minimizing the energy consumption. The paper also addresses the tradeoffs when the LS-WSN MULE and SENMA models perform well.**

***Index Terms*—Data collection using data mule (MULE), energy consumption, large-scale wireless sensor networks (LS-WSNs), mobile data collectors, sensor network with mobile access point (SENMA).**

I. Introduction

BIG data environment deal with a high volume of data, a high velocity of changing data, and a high variety of data types [1]. It is the main driver for the second economy (a concept proposed by economist W.B. Arthur which refers to the economic activities running on processors, connectors, sensors, and executors) [2]. It is estimated that by 2030, the size of the second economy will approach that of the current traditional physical economy. The current surge in big data research is driven by the needs of Internet-based businesses (e.g., Google, Twitter, Facebook, etc.), where real-time data from human sources (e.g., emails, tweets, purchase histories) are used to feed large-scale analytic engines to produce value-added services such as customer analytics [3], e-commerce [4], and fraud detection [5]. With the emergence of a new networked sensors technology (e.g., large-scale wireless sensor networks (LS-WSNs), body sensor networks, Internet/network/web/vehicle-of-things), it is highly anticipated that the next generation of big data systems will need to deal with machine-generated data from these forms of networked sensor systems. An estimate from IBM is that the volume of machine-generated big data sources will increase to 42% of all data by 2020, up from 11% in 2005 [6]. A significant amount of machine-generated data will be produced from sensor-based data sources such as LS-WSNs.

The emergence of LS-WSNs, which are spread over wide geographic areas and contain thousands of sensor nodes, requires new techniques for energy-efficient data collection. The LS-WSN applications generate a very large volume of sensor data which imposes formidable challenges for the efficient collection, transmission, storage, and processing (analytics) of the data. This challenge is further increased because each wireless sensor would have to operate under stringent hardware resource constraints in an energy-constrained environment. LS-WSNs require two important challenges to be addressed for data collection: 1) monitoring in large geographic areas with unconnected regions and 2) energy-efficient data collection using the sensor nodes. The first challenge is connected with the "divided networks problem" where nodes in a region cannot physically communicate with nodes in another region. In this case, a single collector or sink based in the first region cannot be used for data collection for the second region. The second challenge is connected with optimizing the life-time of battery-powered sensor nodes with a limited supply of energy and which cannot be recharged after deployment. These two challenges can be effectively addressed using mobile data collectors (MDCs). The MDC techniques give two advantages over traditional methods using static sinks. These techniques can handle data collection over spatially separated geographical regions where the MDC can be used to roam over different regions, and have been shown to require lower node energy consumption where the collection protocol overheads are effectively transferred from the sensor nodes to the MDC.

Manuscript received May 30, 2015; revised March 29, 2016 and July 31, 2016; accepted September 13, 2016. Date of publication April 28, 2017; date of current version March 23, 2018. This work was supported by Charles Sturt University Compact Funding (A541-2006-121-25411).

K. L.-M. Ang and J. K. P. Seng are with the School of Computing and Mathematics, Charles Sturt University, NSW 2678, Australia (e-mail: lang@csu.edu.au; kseng@csu.edu.au).

A. M. Zungeru is with the Botswana International University of Science and Technology, Palapye, Botswana (e-mail: zungerum@biust.ac.bw).

Digital Object Identifier 10.1109/JSYST.2016.2630691

In the literature, two common models for data collection using MDCs have been proposed: data collection using data mule (MULE) [7], [8] and sensor network with mobile access point (SENMA) [9], [10]. The MULE and SENMA approaches can be characterized as representative of the multihop and the single-hop approaches for mobile data collection in WSNs. Although the basic architectures for MULE and SENMA have been well studied, the emergence of LS-WSNs, which require partitioning the network into multiple groups and clusters prior to data collection, has not been particularly addressed. The main challenge to be solved is how to partition the network into groups and clusters to minimize the total energy consumption of the nodes. This is not a trivial problem and requires careful analysis of the wireless communications requirements at the different network layers. For data collection, the MDC patrols the centroids of the clusters and is equipped to transmit the collected data to the base station using a long distance and energy expensive radio link. Our focus is not on the energy consumption of the mobile collector or base station. We only focus on optimizing the energy consumption of the sensor nodes. Also, we take a systems approach to solving the problem and do not consider optimizations at the MAC layer (i.e., we assume a perfect MAC with no data loss and contention for transmission and reception. The research works in [11] and [12] showed using software simulations that increasing the number of clusters in a network may not always result in the best solution to reduce the node energy consumption because of the "data request (DREQ) flooding problem." The energy consumption due to DREQ flooding becomes higher with an increasing number of clusters in the group. On the other hand, the average transmission paths from nodes to cluster centroids become shorter with an increasing number of clusters and decrease the energy consumption for data collection. These conflicting requirements for node energy consumption require a careful consideration of the tradeoff involved with increasing the number of clusters in a group.

The main contributions of this paper are as follows: 1) we propose new analytical approaches to determine the node energy consumption for LS-WSN mobile data collection schemes (such as MULE and SENMA) which require partitioning the network into multiple clusters prior to data collection and 2) multicluster network models for determining the optimal number of clusters for minimizing the energy consumption in LS-WSNs. The motivation for analytical-based approaches compared to simulation-based approaches is discussed later in Section II. This paper also addresses the tradeoff and scenarios when the LS-WSN MULE and SENMA models perform well. Our analytical models and results in this paper can serve as a baseline for comparisons with simulation-based approaches.

This paper is organized as follows. Section II gives an overview of big data sensor-based environments. Section III reviews some related works on mobile data collection technologies in LS-WSNs. Section IV presents our proposed LS-WSN model and gives details for different data collection models. The energy consumption analysis for the MULE and SENMA models are presented in Sections V and VI. Section VI also discusses the tradeoff scenarios when MULE and SENMA perform well. Finally, some conclusions are given in Section VII.

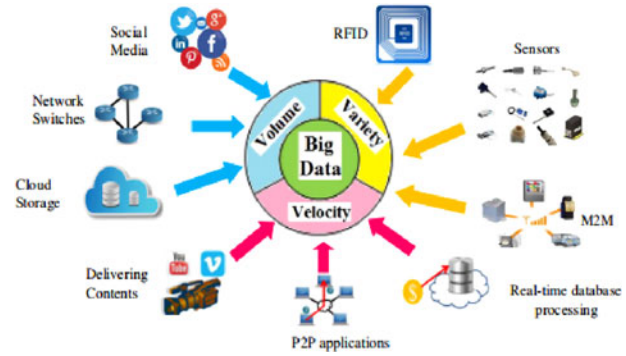


Fig. 1. Big data model showing main characteristics (volume, velocity, variety) and inputs from different data sources.

II. Background and Motivation

Fig. 1 shows the three main characteristics (volume, variety, and velocity) for a big data model, and the inputs from different data sources. This paper is focused on mobile data collection for big sensor-based data systems (a term to conceptualize the application of the big data model toward networked sensor systems). The first generation of big data systems was concerned with processing human-generated data for applications such as website analytics, social media analytics, credit card fraud, etc. It is anticipated that the second generation of big data systems will focus on the collection, transmission, storage, and processing (analytics) of machine-generated data such as from physical sensors. It is highly anticipated that a significant amount of machine-generated data will be produced from sensor-based data sources such as LS-WSNs involving hundreds to thousands of nodes.

An important design consideration for LS-WSNs would be regarding the data collection methods from the sensor nodes. Compared to traditional methods using static sinks, mobile data collection (MDC) techniques give two advantages for data collection in LS-WSNs: 1) the ability to handle data collection over spatially separated geographical regions and 2) a lower node energy consumption requirement. In MDC, the mobile collector can be used to roam over different geographical regions to perform the collection task. The remaining review of work in this section will briefly discuss MDC schemes for WSN environment. A comprehensive survey can be found in [16]. In general, we can distinguish three research aims for MDC schemes in the literature which are related to optimizing the energy consumption, collection latency, and network reliability. In this paper, we focus on optimizing the node energy consumption. Examples of research works to optimize the data collection latency and reliability can be found in [17]–[20], respectively.

The current practice in analyzing algorithm designs for data collection in LS-WSNs is by the use of simulation and practical approaches, which are done by optimizing the necessary performance metrics using heuristics and intuition. The evaluation and validation of the LS-WSN system are mostly accomplished using simulation-based approaches. Despite their wide use and merits for network systems and algorithms validation, simulation studies have some drawbacks such as long simulation times and inconsistent results. We, therefore, see a need to further strengthen the algorithm designs through mathematical analysis and to also provide a basis for future algorithms to be analyzed. To this end, we developed our modeling framework based on energy consumption for data collection in LS-WSNs by considering MULE and SENMA.

Finally, we give some comments for our motivation. Our objective in this research is to develop analytical approaches which can be used to model the node energy consumption in multicluster-based LS-WSNs which contain thousands of sensor nodes. Many research works for WSNs have used a simulation-based approach by employing WSN simulators such as TOSSIM [21], NS-2 [22], and routing modeling application simulation environment (RMASE) [23]. Some analytical approaches for modeling WSNs which we have drawn upon in this work can be found in [9], [10], and [24]. The advantage of analytical-based approaches compared to simulation-based approaches is that we can model very large-scale sensor networks containing thousands of nodes (which may require high-computational power or time for simulations) and experiment with many different scenarios in a short amount of time. Li and Serpen [21] remarked that the simulation time for TOSSIM increases quickly to several weeks for larger scale sensor networks with around 1000 motes or nodes. Another work by Asim and Tixeuil [25] commented that most of the currently available WSN simulators work well for small- to medium-sized networks but do not scale to really large networks. To the best of our knowledge, current WSN simulators cannot easily handle simulations with 10 000 nodes or more (required by LS-WSNs) without resorting to using high-performance computing facilities or specialized computational architectures such as the distributed computing architecture proposed by Asim and Tixeuil [25] for their XS-WSNet simulator.

The disadvantage of analytical-based approaches is that they do not (easily) model lower-layer characteristics such as the MAC layer (i.e., sacrifices some accuracy) which can be incorporated into simulation-based approaches. For network designers, the analytical models and results in this paper will be very useful to serve as a baseline for an initial network design without requiring high-computational power or time, and can later be used for comparisons or fine tuning with other methods such as using software simulators. After knowing the general trend and design characteristics using our analytical models, the network designer can then focus on computing resources and use software simulators to explore a specific parameter range with more design considerations (e.g., MAC layer characteristics such as channel contention window and duty-cycling, various routing protocols, etc.) where the optimum behavior is expected. For example, if we know from the analytical models that the optimum node energy consumption is expected within a cluster range of 40–60, then we can focus our software simulations to explore the network behavior within this range only. In this way, the analytical-based and simulation-based approaches can work together to deliver a good solution in a shorter period of time. Another approach would be to actually deploy the network for purposes of testing and observing the network behavior, but this would incur a very high cost.

III. Related Works

The analysis of WSN systems has received much attention. However, fewer works in analyzing LS-WSNs have been specifically addressed. Hamida and Chelius [34] proposed a line-based data dissemination algorithm with a mobile sink and compared it with other dissemination architectures using an analytical approach based on the average distances between source and sink nodes. Shin *et al*. [35], [36] modeled the energy consumption of a typical grid-based sensor network. These works are similar to the work in [37], where the energy consumption of a grid-based sensor network with consideration of the information extraction on the network with the limited available energy of sensor nodes was taken into account.

In LS-WSN, the SENMA architecture was first proposed in [38]. The MAC in SENMA was considered in [39]. Most of the literature on sensor networks focuses on multihop ad hoc architectures (see [40] for a survey). In [41], unmanned aerial vehicles (UAVs) were used to deploy sensors and serve as communication hubs. The settings considered in [41] are different from the work presented in [42]. It is assumed that the distance among sensors is large, and a UAV is required to visit each sensor one at a time. An adaptive path-planning algorithm is proposed in [41] that finds the minimum-cost path for a UAV to visit every sensor. The energy-efficiency comparison between SENMA and the ad hoc architecture is first presented in [38], although the calculation there does not take into account the possibility of scaling the transmission radius according to the size of the network. In [42], the energy consumption analysis was calculated, but the drawback is that Mergen *et al*. did not explicitly account for energy consumed in both transmission and reception. Other LS-WSN deployments can be found in the research works [13]–[15]. The work in [13] presents VigilNet for field surveillance which has been designed to scale up to 1000 motes/nodes. The work in [14] presents GreenOrbs which has been deployed for forest surveillance with up to 330 nodes. The work in [15] described the design and deployment of an LS-WSN for monitoring power equipment (transformers, circuit breakers, and compressors) in a substation using 122 low-power nodes. In the future, we expect to see larger deployments of LS-WSNs containing thousands of nodes.

For mobile data collection involving LS-WSNs, two common models have been proposed: In the literature, two common models for data collection using MDCs have been proposed: MULE [7], [8] and SENMA [9], [10]. The MULE approach is usually associated with mobile data collection using ground means (e.g., using mobile robotic vehicles) whereas the SENMA approach is usually associated with mobile data collection using aerial means (e.g., using light planes or UAV). From the networking viewpoint, the MULE and SENMA approaches are also different in the way that data are transferred from the nodes to the collector. The MULE approach can be characterized as representative of the multihop approach for mobile data collection in WSNs. A sensor which has data to send to the collector will

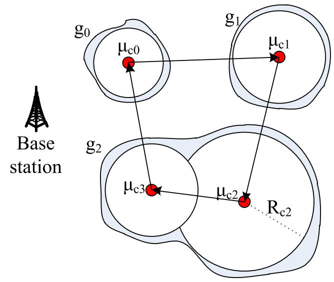


Fig. 2. System model of a large-scale wireless sensor network.

transmit its data to its neighboring nodes which will then perform the task of routing the data along a multihop path until the data reaches the collector. On the other hand, the SENMA approach can be characterized as representative of the single-hop approach. In SENMA, nodes can transmit directly to the mobile collector overhead [called a mobile access point (AP)]. The data collection schemes for MULE and SENMA will be discussed in detail in Section IV.

IV. Problem Definition

This section first presents our proposed LS-WSN model. Then, using the proposed network model, we discuss the data collection schemes for MULE and SENMA, with the objectives to identify the various tradeoff affecting the node energy consumption. The work presented in this section will be further extended in Sections IV and V to derive the energy consumption analysis for the LS-WSN MULE and SENMA models.

A. Large-Scale Wireless Sensor Network Model

Fig. 2 shows a model of an LS-WSN which is spatially separated into groups and clusters with the following parameters.

1) N—The total number of sensor nodes in the network. Each node has a label from n_0 to n_{N-1}.
2) G—The number of groups in the network. Each group has a label from g_0 to g_{G-1}. Each group is spatially separated and is not reachable by other groups. The nodes in each group are reachable by other nodes in the same group.
3) C—The number of clusters in the network. Each cluster has a label from c_0 to c_{C-1}. The number of clusters must be at least equal to the number of groups (i.e., $C \geq G$).
4) R_{c0}—The radius for cluster c_0.
5) μ_{c0}—The centroid of cluster c_0.
6) λ_{c0}—The average collection response rate of cluster c_0. This is the average proportion of nodes in cluster c_0 that responds to a data collection request from the mobile collector. A value of $\lambda_{c0} = 1$ means that 100% of the nodes in cluster c_0 has data to send in response to the request from the mobile collector. Conversely, a value of $\lambda_{c0} = 0$ means no nodes in the cluster has data to send in response to the query.
7) N_{g0}—The number of nodes in group g_0.
8) N_{c0}—The number of nodes in cluster c_0.
9) r—The transmission radius for each sensor node.
10) ε_N—The node energy consumption in the network.
11) ε_{g0}—The node energy consumption in group g_0.
12) ε_{c0}—The node energy consumption in cluster c_0.

Fig. 2 shows a network with three groups (g_0, g_1, g_2) and four clusters with centroids $(\mu_{c0}, \mu_{c1}, \mu_{c2}, \mu_{c3})$. Note that there may be shared nodes between clusters c_2 and c_3. The arrow paths show the movement of the mobile collector for data collection. For our analysis, we use a model where the sensor nodes are stationary after deployment. The only mobility allowed is by the mobile collector. The main challenge to be solved is how to partition the network into groups and clusters to minimize the total energy consumption of the nodes. We assume that the mobile collector has complete knowledge of the location of every sensor node. Although it may be possible to have multiple collectors to perform the data collection, we will use a single mobile collector in our analysis. We also use a "stop-and-collect" protocol [16] where data collection only occurs when the mobile collector is positioned stationary at the centroids of the clusters. We use a "wake up scheme" where nodes are equipped with an RF energy detector as described in [24]. Using the model in Fig. 2, we will next describe the data collection models for MULE and SENMA.

B. Data Collection Model Using Data Mule

The data mule is a mobile relay node that carries data from static sensor nodes to an infrastructure AP, in this case from the centroid of a cluster to the sink node. Because mules are power renewable due to the ability to move around the network, hence, their energy may not be exhausted. Static nodes exhaust their energy within a period of time, while mules help the network to balance up the energy.

1) Algorithm Description: The algorithm basically works with two packet types: DREQ packet and data reply (DREPLY) packet. The DREQ is responsible for sending available data from each source node to data mule on arrival at the centroid of a cluster. If data is not available for the mule on its arrival at the centroid of a cluster, a DREQ packet is flooded first around the neighbors of the centroid (i.e., nodes within its transmission range). If the nodes around the centroid have valid data, then data are sent to the data mule in a unicast form [8]. The working of the DREQ is as follows.

```
if
    DREPLY (data) = valid_data;
    {
    Construct the DREPLY packet and DREQ packet,
    send DREPLY packet to nodes PrevHop from which it
    has received DREQ
    }
    else
    {
    Forward the DREQ packet to the neighbors' nodes
    }
end if
```

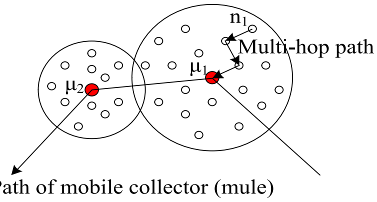


Fig. 3. Data collection model using data mule for a sensor group containing two clusters with centroids μ_1 and μ_2.

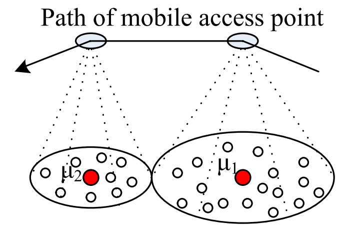


Fig. 4. Data collection model using SENMA for a sensor group containing two clusters with centroids μ_1 and μ_2.

At any node, when it receives a DREPLY (data), it does the following.

```
if
      current_node = data_mule;
      {
      Retrieve the data packet and wait for other DREPLY
      }
elseif
      current_node ≠ data_mule
         {
         Forward the DREPLY (data) to neighboring node
         following the trail content
         }
      else
         {
         Move to the next cluster
         }
end if
```

The data collection model using data mule [7], [8] is shown in Fig. 3. We consider the situation as shown in Fig. 3 where the mobile collector (mule) makes two stops to collect the sensor data from a group (i.e., there are two clusters in the group with centroids labeled as μ_1 and μ_2) and describe the data collection procedure.

On arriving at the centroid of the first cluster (location μ_1), the mobile collector (mule) will broadcast a DREQ message to initiate the data collection from the cluster. The sensor nodes in the neighborhood of the mobile collector receive the DREQ message and perform two tasks. The first task is to transmit their data packets (DATA) to the collector if the query request meets their sensing status. The average collection response rate of the cluster is given by λ_c. The data transmission from a sensor node to the collector uses a predetermined route and makes multiple hops over the network to reach the mobile collector. A sensor broadcasts its DATA packet using a transmission range to its neighbors and a predetermined node will perform the routing task and relay the packet toward the mobile collector. Fig. 3 shows a multihop path from a source node n_1 to the collector.

The second task is to broadcast the DREQ message to its neighboring nodes. The DREQ is repeatedly broadcasted from node to node until all nodes in the group have received the message. The mobile collector finishes the data collection at the first cluster and moves to the centroid of the second cluster (location μ_2). The data collection task is repeated for the second cluster and the mobile collector will broadcast a DREQ message to initiate data collection. It is important to note that although nodes may receive the DREQ message more than once, they only send data and broadcast the DREQ message once after the first time of receiving the message. The flooding of the group with DREQ messages each time the mobile collector arrives at a cluster centroid causes high energy consumption and this is described as the "DREQ flooding problem" [11], [12]. The node energy consumption due to DREQ flooding becomes higher with an increasing number of clusters in the group. On the other hand, the average transmission paths from source nodes to cluster centroids become shorter with an increasing number of clusters and decrease the energy consumption for data collection.

C. Data Collection Model Using SENMA

The data collection model using SENMA is shown in Fig. 4. In SENMA, the mobile collector called a mobile AP is usually associated with an aerial vehicle (e.g., UAV) to perform the data collection. In our model, we use SENMA as representative of the single-hop approach for mobile data collection where sensor nodes can communicate directly with the mobile AP. The mobile AP visits the centroids of the clusters in turn. On arriving at the centroid of each cluster (e.g., location μ_1), the mobile AP collector will broadcast a DREQ message (beacon) directly to all the nodes in the cluster to initiate the data collection. The communication from the mobile AP to the DREQ is received by all nodes in the cluster and nodes with data to send will respond and transmit their data directly to the mobile AP. In a real scenario, each sensor may be woken up in turn to transmit its data to the mobile AP to reduce contention at the MAC layer. We do not consider MAC layer optimizations in our model.

The energy consumption in SENMA when compared to MULE will be lower due to the elimination of the DREQ flooding for data initiation. On the other hand, the average transmission paths from nodes to the mobile AP become longer due to the single-hop approach. The sensor nodes need to increase their transmission range to reach the mobile AP and subsequently the node energy consumption will increase for data collection when compared to the multihop approach used in MULE. In the MULE (multihop) model, each node only needs to be able to transmit to its neighboring nodes and can use an optimal transmission range to accomplish this and reduce its energy consumption.

1) Algorithm Description: The algorithm basically works on the one-hop transmission between sensor nodes and mobile agents with two packets: DREQ packet and DREPLY packet. The DREQ is responsible for sending available data from source node to mobile AP on arrival at the centroid of a cluster. On arrival at the centroid of a cluster by the mobile AP, a DREQ is flooded first to all the nodes in the cluster, which also serves as a synchronization or wake-up signaling. The sensor listens to the beacon and decides if it should transmit its available packet based on the condition that its channel is under the most favorable fading condition [9], [10]. If the nodes in the cluster have a valid data and have their channel under the most favorable condition, then data are sent to the mobile AP in a unicast form. The working of the DREQ is as follows.

```
if
    DREPLY (data) = valid_data;
    {
    Construct the DREPLY packet and send DREPLY
    packet to mobile access point (AP) from which it has
    received DREQ
    }
    else
    {
    Go to idle mode
    }
end if
```

At any node, when it receives a DREPLY (data), it does the following.

```
if
    current_node = AP;
    {
    Retrieve the data packet and wait for other DREPLY
    }
    else
    {
    Move to the next cluster
    }
end if
```

V. Energy Consumption Analysis for Data Mule Model

We will consider an analytical approach to model the energy consumption for the data collection using the MULE model with a single group. The energy consumption for multiple groups in the network can be obtained by the summation of the energy consumption for all the groups (i.e., $\varepsilon_N = \sum \varepsilon_g$). Fig. 5 shows a mathematical model for a network of *N* nodes which has been partitioned into one group and seven equal clusters (i.e., $G = 1$, $C = 7$, and $N_g = N$). The nodes are uniformly distributed and there are an equal number of nodes in each cluster. In the figure, the cluster centroids have been labeled as μ_1 to μ_7. The radius of the network is given by *R* and the radius of each cluster is given by R_c. In this case of seven clusters, R_c has a value of *R*/3. The number of nodes in each cluster is given by N_c, and in this case $N_c = N_g/7$. There are no shared nodes amongst the clusters.

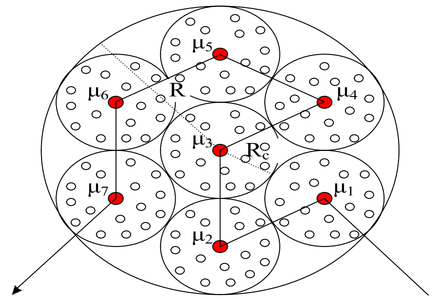


Fig. 5. Mathematical model for a network with seven clusters.

We have to consider two cases for the total energy consumption.

1) The energy consumption for data initiation for the mobile collector to broadcast the DREQ to neighboring sensor nodes and the subsequent flooding of the network. The length of the DREQ packet in bits is given by L_{DREQ}.
2) The energy consumption for data collection for the nodes to transmit their DATA to the mobile collector. The length of the DATA packet in bits is given by L_{DATA}.

In our model, we consider the situation when the sensor nodes are stationary after deployment. In the general case, the sensor nodes will also need to expend energy to determine routing paths to the mobile collector. However, if the sensor nodes do not move, this process only needs to occur once when the nodes are first deployed. Thus, in our model, we do not consider the energy costs for route discovery and only consider the energy consumption for data initiation and data collection. The energy consumption for a single cluster network using a multihop approach has been derived by Zhao and Tong [24]. We adapt the analysis in [24] for a multicluster network.

A. Energy Consumption for Data Initiation

The energy consumption for data initiation for the mobile collector to broadcast the DREQ packets and the subsequent flooding at a cluster centroid is determined by the number of DREQ packets generated in the flooding process. The number of DREQ packets generated in the flooding process is determined by the number of nodes in the group N_g. For a large N ($N \to \infty$), the energy consumption for data initiation for a cluster is given by

$$\varepsilon_{\mathrm{init},c}\ (r) = N_g\ L_{\mathrm{DREQ}} \times \left\{ E_{\mathrm{tx}}\,(r) + \frac{r^2}{{R_c}^2}\,(N_c - 1)\,E_{\mathrm{rx}} \right\} \tag{1}$$

where

1) $E_{\mathrm{tx}}(r)$ is the energy consumed by a node for a transmission that covers a neighborhood of radius *r* and is given by the physical model [26]

$$E_{\mathrm{tx}}\ (r) = e_{\mathrm{tx}}\ + \max\left\{e_{\min}, e_{\mathrm{out}} r^{\alpha}\right\} \tag{2}$$

where α is the path attenuation factor, e_{tx} is the energy consumed by the transmitter circuitry, e_{out} is the antenna

output energy, and $e_{\min}$ is the minimum energy radiated regardless of the transmission range.
2) E_{rx} is the energy consumed by a node for receiving.
3) *r* is the optimal node transmission radius and is given by $R_c \sqrt{\frac{\log N_c}{N_c}}$.

The total energy consumption for all clusters for data initiation is then given by

$$\varepsilon_{\mathrm{INIT}}\ (r) = \ C \times \varepsilon_{\mathrm{init},c}\,(r)\,. \tag{3}$$

B. Energy Consumption for Data Collection

The energy consumption for data collection from the sensor nodes to the mobile collector is determined by the number of nodes that respond to the DREQ query and have data to send in response to the request from the mobile collector. This is modeled by λ_c which is the average collection response rate of the cluster. A sensor with data to send broadcasts the DATA packet using a transmission range to its neighbors and a predetermined node will perform the routing task and relay the packet toward the mobile collector using a multihop path. For a large $N\ (N \rightarrow \infty)$, the energy consumption for data collection for a cluster is given by

$$\begin{aligned} &\varepsilon_{\mathrm{collect},c}\ (r) \\ &= \lambda_c N_c\ L_{\mathrm{DATA}}\,\frac{2R_c}{3r}\left\{E_{\mathrm{tx}}\,(r) + \frac{r^2}{{R_c}^2}\,(N_c - 1)\,E_{\mathrm{rx}}\right\}. \end{aligned} \tag{4}$$

The total energy consumption for all clusters for data collection is then given by

$$\varepsilon_{\mathrm{COLLECT}}\ (r) = \ C \times \varepsilon_{\mathrm{collect},c}\,(r)\,. \tag{5}$$

C. Total Energy Consumption

The total energy consumption for the network is then given by the summation of the energy consumption for data initiation and collection

$$\varepsilon_N\ \ (r) = \varepsilon_{\mathrm{INIT}}\ \ (r) + \varepsilon_{\mathrm{COLLECT}}\,(r)\,. \tag{6}$$

The important differences concerning the node energy consumption for the data initiation and the data collection processes can be seen from (1) and (4). The DREQ packets are flooded within the group and depend on the number of nodes within the group N_{g}. On the other hand, the transmitted DATA packets to the mobile collector only occur within a particular cluster and only depend on the number of nodes within the cluster N_c. Increasing the number of clusters *C* in a group will increase the "DREQ flooding problem" for data initiation but will also decrease the radius of a cluster R_c for data collection.

D. Energy Consumption Tradeoffs for Multiple Clusters

Equations (1)–(6) allow for an analytical investigation of the mobile data collection problem for multiple clusters and the different tradeoffs, which are involved. We first need to determine values for R_c for different values of *C*. We adopt the approach to use circles of equal size for modeling the clusters in the group. As shown in the mathematical model in Fig. 5, for $C = 7$, an appropriate value for R_c is *R*/3. This is known as the "circle packing problem" [27] to arrange as many equal size circles into a container circle with no overlapping and optimize the packing density. Some values of R_c for different values of *C* from 1 to 19 based on the known circle packing solutions which can be obtained from [28] are shown in Table I.

TABLE I
VALUES FOR CLUSTER RADII FOR DIFFERENT NUMBER OF CLUSTERS

C	R_c	Density
1	R	1.0
3	0.46410R	0.646
5	0.37019R	0.685
7	0.33333R	0.777
9	0.27676R	0.689
11	0.25485R	0.714
13	0.23606R	0.724
15	0.22117R	0.733
17	0.20867R	0.740
19	0.20560R	0.803

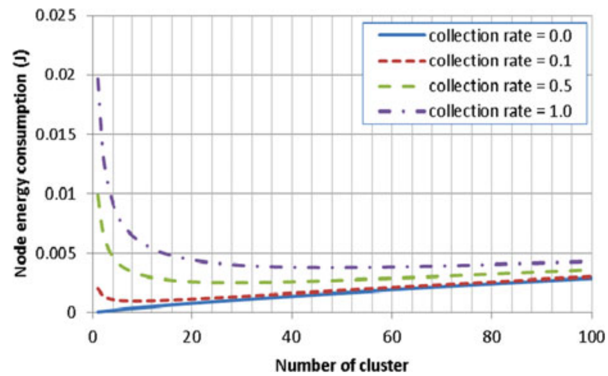


Fig. 6. Node energy consumption for varying number of clusters at different collection response rates λ_c (0.0, 0.1, 0.5, 1.0).

We are interested in answering the following questions. How does the node energy consumption vary with the number of clusters for:

1) different collection response rates λ_c?
2) different network radii *R*?
3) different packet size ratios of $L_{\mathrm{DATA}}/L_{\mathrm{DREQ}}$?

Fig. 6 shows the node energy consumption for different number of clusters at four different collection response rates λ_c (0.0, 0.1, 0.5, and 1.0). The following parameters were used: $N = 10\ \ 000$, $R = 100$ m, $e_{\mathrm{tx}} = 81\mathrm{e}-9\mathrm{J}, e_{\mathrm{out}} = 0.1\mathrm{e}-9\mathrm{J}$, $\alpha = 3$, $\mathrm{E}_{\mathrm{rx}} = 180\mathrm{e}-9\mathrm{J}, \mathrm{L}_{\mathrm{DREQ}} = 32$ bits, and $\mathrm{L}_{\mathrm{DATA}} = 512$ bits. At $\lambda_c = 0.0$, the node energy consumption is only expended for the data initiation process. There is no energy spent for data collection and the node energy consumption increases linearly for the number of clusters due to the DREQ flooding. At $\lambda_c = 0.1$ (i.e., 10% of the sensor nodes respond to the query), an optimal number of clusters were found at $C = 9$ and then the node energy consumption continues increasing with the number of clusters. Similarly, optimal number of clusters were found at $C = 28$ and 46 for $\lambda_c = 0.5$ and 1.0, respectively.

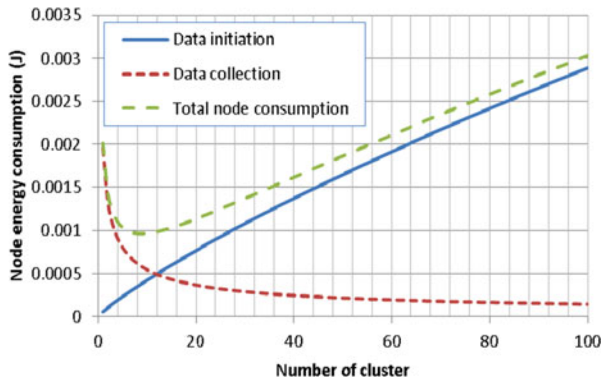


Fig. 7. Breakdown for the node energy consumption for varying number of clusters for data initiation and data collection ($\lambda_c = 0.1$).

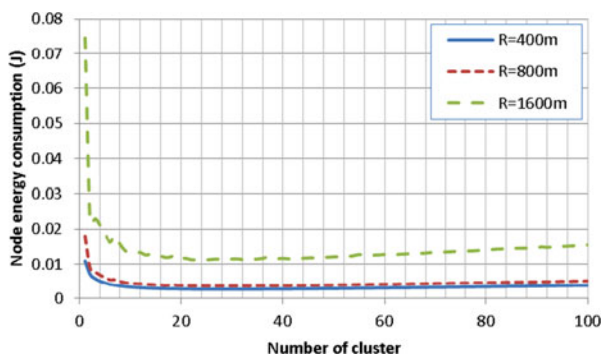


Fig. 8. Node energy consumption for varying number of clusters at different network radii *R* (400, 800, 1600 m).

Fig. 7 shows a breakdown of the node energy consumption at $\lambda_c = 0.1$. When the number of clusters increases, the energy required for data collection decreases to a very small value and the node energy consumption is bound by the energy required for data initiation. This analysis shows that increasing the number of clusters may not always give the best solution to reduce the node energy consumption. Takaishi *et al.* [11] obtained a similar conclusion using a simulation-based approach. As shown in Fig. 7, increasing the number of clusters may incur a much higher node energy consumption compared to just using a single cluster. This factor and tradeoffs need to be carefully taken into consideration in the design of the data collection scheme using the mobile collector.

Next, we are interested to investigate how the node energy consumption varies with different sizes of network radii and packet size ratios. Fig. 8 shows the average node energy consumption for different number of clusters at three different network radii *N* (400, 800, and 1600 m). The collection response rate was set to $\lambda_c = 0.5$. For $C = 1$, the node energy consumption obtained was 0.011J, 0.018J, and 0.075J for $R = 400$, 800, and 1600 m, respectively. The energy consumption at $R = 1600$ m is almost seven times the energy consumption at $R = 400$ m. The optimal number of clusters were found at $C = 29$ (0.0027J), 25 (0.0036J), and 24 (0.011J) for $R = 400$, 800, and 1600 m, respectively. For $C = 24$, the node energy consumption for $R = 1600$ m can be reduced by almost seven times from 0.075J to 0.011J. This analysis shows the usefulness for using a multiple number of clusters to perform the data collection using the mobile collector.

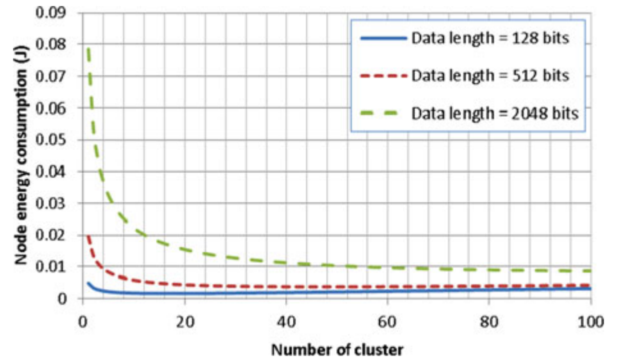


Fig. 9. Node energy consumption for varying number of clusters at different data lengths L_{DATA} (128, 512, 2048 bits).

Fig. 9 shows the average node energy consumption for different number of clusters at three different packet size ratios of $L_{\mathrm{DATA}}/L_{\mathrm{DREQ}}$ (128/32, 512/32, and 2048/32 bits). The collection response rate was set to $\lambda_c = 1.0$ (maximum data collection) and *R* was set to 100 m. The optimal number of clusters were found at $C = 17$ and 46 for $L_{\mathrm{DATA}} = 128$ and 512 bits, respectively. In this case, an optimal number of clusters were not found for $L_{\mathrm{DATA}} = 2048$ bits for up to 100 clusters. This may require increasing the number of clusters further to find the optimal point. However, even without an optimal point, the graph shows that the node energy consumption can be reduced by almost eight times using 100 clusters instead of one, thus, again validating the effectiveness of the multicluster approach for energy consumption reduction in LS-WSNs.

We performed some additional experimental results using the RMASE simulator [30]. Two protocols were used for the simulations: 1) message-initiated constraint-based routing (MCBR) and 2) termite-hill. The full details for MCBR and termite-hill can be found in [31] and [32], respectively. MCBR is a proactive routing protocol and represents a well-known constrained-based flooding scheme. Termite-hill is a state-of-the-art reactive routing protocol utilizing the swarm behavior of termites to give very low node energy consumption. The node energy consumption was modeled according to the Waspmote-802.15.4 sensor node from Libelium [33] with a transmission range of 35 m. For a first experiment, we simulated the collection rate from 0.0 to 0.9 for 120 nodes at a data rate of 250 kbps, and a data length of 512 bits. Other simulation parameters used are shown in Table II.

Fig. 10 shows the energy consumption for different collection rates for the two protocols, which increased as the collection rate increased. Termite-hill (being a reactive routing protocol) performed better than MCBR.

Next, we performed a large-scale experiment with 10 000 sensor nodes with a varying number of clusters. The nodes were randomly distributed in a topology of 100 $\times$ 100. We used a collection rate of 0.5, a cluster radius of 800 m, at a data rate of 250 kbps, and a data length of 512 bits. Fig. 11 shows the node energy consumption for different number of clusters for the two protocols. Both protocols showed the same

TABLE II
SIMULATION PARAMETERS

Parameters	Values
Routing protocol	Termite-hill, MCBR_Flooding
Size of topology (A)	10×10 km^2 (10K nodes)
Nodes distribution	Random distribution
Maximum no. of retransmission (n)	3
Source type, Destination type	Static, Mobile
Propagation model	Two-ray ground reflection
Transmission distance (R)	35 m
Speed of mobile sink (Collector)	10 m/s
Data traffic	Constant bit rate (CBR)
Data rate	250 kbps
Propagation model	Probabilistic
Energy consumption	Waspmote-802.15.4
Time of topology change	2 s
Simulation time, Average simulation times	360 s, 10
Number of nodes (N)	10 000

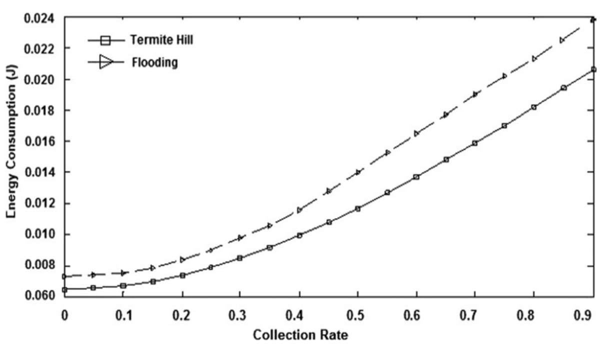


Fig. 10. Energy consumption for different collection rates for MCBR flooding and termite-hill.

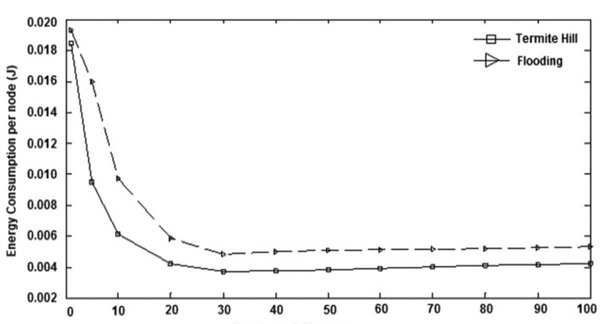


Fig. 11. Node energy consumption for varying number of clusters for MCBR flooding and termite-hill.

trend. Initially, as the number of clusters was increased, the node energy consumption decreased. An optimal point was reached at about 30 clusters. Increasing the number of clusters after 30 resulted in a slight increase in the energy consumption. Again, termite-hill performed better than MCBR. The experimental results for the routing protocols in Fig. 11 show a similar trend to the analytical models in Fig. 8, thus, validating our approach.

VI. Energy Consumption Analysis for SENMA Model

In this section, we will consider an analytical approach to model the energy consumption for the mobile data collection

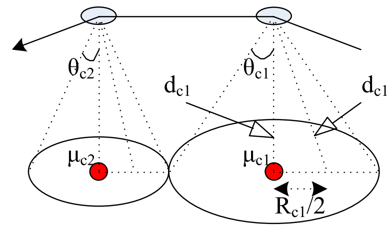


Fig. 12. Mathematical model for SENMA.

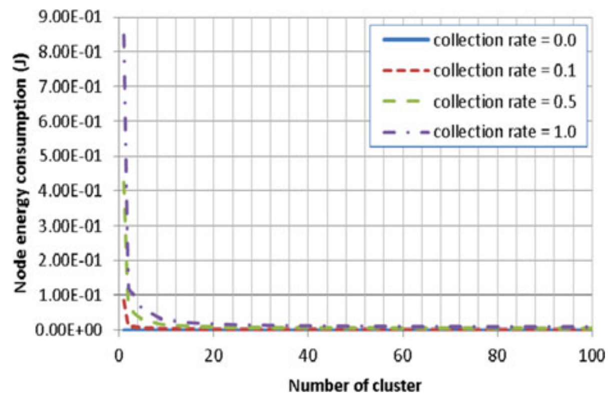


Fig. 13. Node energy consumption for varying number of clusters at different collection response rates λ_c (0.0, 0.1, 0.5, 1.0).

using the SENMA model with a single group. The energy consumption for multiple groups in the network can be obtained by the summation of the energy consumption for all the groups (i.e., $\varepsilon_N = \sum \varepsilon_g$). The node energy consumption for SENMA is only expended by nodes with data to send to the mobile AP on reception of the DREQ message from the mobile AP. Each sensor node has to transmit directly to the mobile AP as shown in Fig. 12. The energy consumption for a single cluster network for the SENMA model has been derived by Mergen *et al.* [9], [10]. We adapt the analysis in [9] and [10] for a multicluster network. For a large N ($N \to \infty$), the energy consumption for data collection for a cluster is given by

$$\varepsilon_{\mathrm{SENMA},c} = \frac{{d_c}^2 \tan^2\theta_c}{{R_c}^2} N_c E_{\mathrm{rx}} + \lambda_c N_c L_{\mathrm{DATA}} E_{\mathrm{tx}}\left(d_c{}^*\right) \quad (7)$$

where the cluster angle θ_c, and cluster distances d_c and d_c^* for the clusters are as shown in Fig. 12. The cluster distance d_c^* is used as an approximation for the average distance that a sensor node in the cluster has to transmit to the mobile AP. The total energy consumption for all clusters for data collection is then given by

$$\varepsilon_{\mathrm{SENMA}} = C \times \varepsilon_{\mathrm{SENMA},c}. \quad (8)$$

Fig. 13 shows the node energy consumption for different number of clusters for SENMA at four different collection response rates λ_c (0.0, 0.1, 0.5, and 1.0). The following parameters were used: $N = 10\,000$, $R = 500$ m, $d = 50$ m, $e_{\mathrm{tx}} = 81\mathrm{e} - 9\mathrm{J}$, $e_{\mathrm{out}} = 0.1\mathrm{e} - 9\mathrm{J}$, $\alpha = 3$, $E_{\mathrm{rx}} = 180\mathrm{e} - 9\mathrm{J}$, and $L_{\mathrm{DATA}} = 512$ bits. At $\lambda_c = 0.0$, the node energy consumption is only expended for listening for the mobile AP. There is no energy spent for

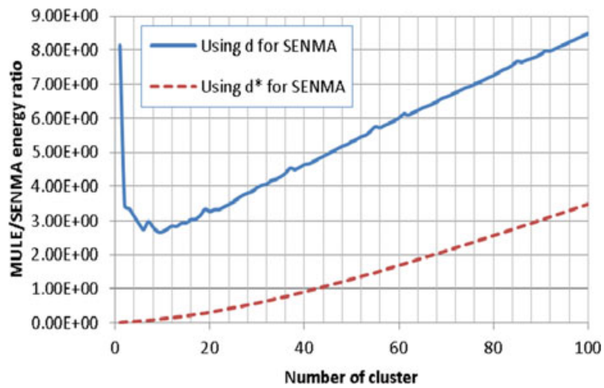


Fig. 14. Node energy consumption comparison ratio for $E_{\mathrm{MULE}}/E_{\mathrm{SENMA}}$ for varying number of clusters.

transmission to the mobile AP. This is the minimum energy that would have to be expended by the cluster for each time that the mobile AP appears and is the same for varying number of clusters. Fig. 13 shows the effectiveness of the multicluster approach to rapidly decrease the node energy consumption by increasing the number of clusters. This approach is effective even for higher collection rates.

Next, we wish to compare the node energy consumption for SENMA and MULE for varying number of clusters in a very large-scale WSN scenario. The following parameters were used.

1) MULE ($N = 10\,000$, $R = 1000$ m, $e_{\mathrm{tx}} = 81\mathrm{e}-9\mathrm{J}$, $e_{\mathrm{out}} = 0.1\mathrm{e}-9\mathrm{J}$, $\alpha = 3$, $E_{\mathrm{rx}} = 180\mathrm{e}-9\mathrm{J}$, $L_{\mathrm{DREQ}} = 32$ bits, and $L_{\mathrm{DATA}} = 512$ bits).
2) SENMA ($N = 10\,000$, $R = 1000$ m, $d = 50$ m, $e_{\mathrm{tx}} = 81\mathrm{e}-9\mathrm{J}$, $e_{\mathrm{out}} = 0.1\mathrm{e}-9\mathrm{J}$, $\alpha = 3$, $E_{\mathrm{rx}} = 180\mathrm{e}-9\mathrm{J}$, and $L_{\mathrm{DATA}} = 512$ bits).

for a collection response rate of $\lambda_c = 0.1$. Fig. 14 shows the node energy consumption ratio $(E_{\mathrm{MULE}}/E_{\mathrm{SENMA}})$ for varying number of clusters. We used two cluster distances (d_c and $d_c{}^*$) for calculating the energy consumption. The graph for "using d^* for SENMA" used the calculation as shown in (7). The graph for "using d for SENMA" used (7) where the second term used $E_{\mathrm{tx}}(d_c)$ instead of $E_{\mathrm{tx}}(d_c^*)$. Mergen *et al.* [9] used the vertical distance d for their energy consumption analysis for SENMA and concluded that SENMA offers significant improvement over the multihop (MULE) model. This can be seen in the "using d for SENMA" graph which is always above the ratio of one. However, when we use d^* for the E_{tx} calculation, we find that the MULE approach is better when the number of clusters is fewer. This is because in a very large-scale WSN scenario, when the number of clusters is fewer (i.e., the cluster radius is substantially large), then the sensor nodes need to expend a substantial amount of node energy to transmit to the mobile AP for SENMA. With an increasing number of clusters, the cluster radii for SENMA decrease and the distance required to reach the mobile AP also decreases. For this scenario, a turning point (i.e., when the MULE/SENMA energy ratio becomes greater than one) was found for 43 clusters when the SENMA model becomes more efficient than the MULE model. Another factor to consider for selecting the SENMA (single-hop) model over the MULE (multihop) model is that the MULE model may also suffer from the "funneling effect problem" [16], [29] when nodes closer to the mobile collector will consume energy much faster than nodes further away from the collector.

VII. Conclusion

In this paper, we investigated the challenging issues pertaining to big data collection in LS-WSNs containing several thousands of sensor nodes using mobile collectors. While the use of mobile collectors (compared to static sinks) can result in lowering the node energy consumption and reduce the "funneling effect problem," they lead to a number of additional challenges such as the partitioning of the network into multiple groups and clusters prior to data collection. To address these challenges, we proposed analytical approaches to calculate the node energy consumption and to determine the optimal number of clusters for two representative mobile data collection models: MULE (multihop model) and SENMA (single-hop model). Our analytical models allow the network designer to model very large-scale sensor networks (10 000 nodes and above) and experiment with different network scenarios in a short amount of time. For MULE, our model showed that increasing the number of clusters may not always give the best solution to reduce the node energy consumption due to the "DREQ flooding problem." For SENMA, our model showed that for LS-WSN scenarios, the sensor nodes would need to expend a substantial amount of node energy to transmit to the mobile AP when the number of clusters is fewer.

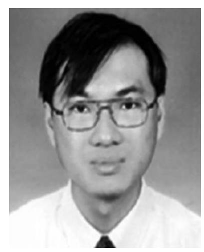
Acknowledgment

The authors would like to thank the anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper.

References

[1] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: A technology tutorial," *IEEE Access*, vol. 2, pp. 652–687, 2014.
[2] W. B. Arthur, "The second economy," *McKinsey Quart.*, Oct. 2011.
[3] N. Sun, J. G. Morris, J. Xu, X. Zhu, and M. Xie, "iCARE: A framework for big data-based banking customer analytics," *IBM J. Res. Develop.*, vol. 58, nos. 5/6, pp. 4:1–4:9, 2014.
[4] C. Sun, R. Gao, and H. Xi, "Big data based retail recommender system of non E-commerce," in *Proc. Int. Conf. Comput., Commun. Netw. Technol.*, 2014, pp. 1–7.
[5] U. Srinivasan and B. Arunasalam, "Leveraging big data analytics to reduce healthcare costs," *IT Prof.*, vol. 15, no. 6, pp. 21–28, Nov./Dec. 2013.
[6] [Online]. Available: http://www.ibm.com/big-data/au/en/big-data-and-analytics/operations-management.html
[7] G. Anastasi, M. Conti, and M. Di Francesco, "Data collection in sensor networks with data mules: An integrated simulation analysis," in *Proc. 13th IEEE Symp. Comput. Commun.*, 2008, pp. 1096–1102.
[8] R. C. Shah, S. Roy, S. Jain, and W. Brunette, "Data MULEs: Modeling and analysis of a three-tier architecture for sparse sensor networks," *Ad Hoc Netw.*, vol. 1, nos. 2–3, pp. 215–233, 2003.
[9] G. Mergen, Q. Zhao, and L. Tong, "Sensor networks with mobile access: Energy and capacity considerations," *IEEE Trans. Commun.*, vol. 54, no. 11, pp. 2033–2044, Nov. 2006.
[10] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," in *Proc. Int. Symp. Mil. Commun.*, 2003, pp. 688–693.
[11] D. Takaishi, H. Nishiyama, N. Kato, and R. Miura, "Toward energy efficient big data gathering in densely distributed sensor networks," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 388–397, Sep. 2014.

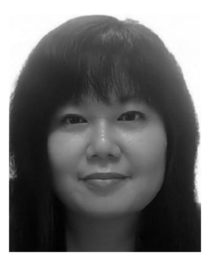
[12] D. Takaishi, H. Nishiyama, N. Kato, and R. Miura, "On the effect of data request message flooding in dense wireless sensor networks with a mobile sink," in *Proc. IEEE 78th Veh. Technol. Conf.*, 2013, pp. 1–5.
[13] T. He *et al.*, "Achieving long-term surveillance in VigilNet," in *Proc. 24th Annu. IEEE Conf. Comput. Commun.*, 2006, pp. 1–12.
[14] Y. Liu, Y. He, M. Li, J. Wang, K. Liu, and X. Li, "Does wireless sensor network scale? A measurement study on GreenOrbs," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 10, pp. 1983–1993, Oct. 2013.
[15] A. Nasipuri, R. Cox, J. Conrad, L. Van Der Zel, B. Rodriguez, and R. McKosky, "Design considerations for a large-scale wireless sensor network for substation monitoring," in *Proc. 5th IEEE Int. Workshop Pract. Issues Build. Sens. Netw. Appl.*, 2010, pp. 866–873.
[16] M. Di Francesco, S. K. Das, and G. Anastasi, "Data collection in wireless sensor networks with mobile elements: A survey," *ACM Trans. Sens. Netw.*, vol. 8, no. 1, pp. 1–31, 2011.
[17] R. Xu, H. Dai, F. Wang, and Z. Jia, "A convex hull based optimization to reduce the data delivery latency of the mobile elements in wireless sensor networks," in *Proc. 2013 IEEE Int. Conf. High Perform. Comput. Commun.*, 2013, pp. 2245–2252.
[18] L. He, J. Pan, and J. Xu, "A progressive approach to reducing data collection latency in wireless sensor networks with mobile elements," *IEEE Trans. Mobile Comput.*, vol. 12, no. 7, pp. 1308–1320, Jul. 2013.
[19] S. Vupputuri, K. K. Rachuri, and C. S. R. Murthy, "Using mobile data collectors to improve network lifetime of wireless sensor networks with reliability constraints," *J. Parallel Distrib. Comput.*, vol. 70, no. 7, pp. 767–778, 2010.
[20] Y.-F. Xiao, S.-Z. Chen, X. Li, and Y.-H. Li, "Reliability evaluation of wireless sensor networks using an enhanced OBDD algorithm," *J. China Univ. Posts Telecommun.*, vol. 16, no. 5, pp. 62–70, 2009.
[21] J. Li and G. Serpen, "Simulating heterogeneous and larger-scale wireless sensor networks with TOSSIM TinyOS emulator," *Procedia Comput. Sci.*, vol. 12, pp. 374–379, 2012.
[22] L. Tian, H. Du, and Y. Huang, "The simulation and analysis of LEACH protocol for wireless sensor network based on NS2," in *Proc. 2012 Int. Conf. Syst. Sci. Eng.*, 2012, pp. 530–533.
[23] A. M. Zungeru, L. M. Ang, and K. P. Seng, "Termite-hill: Performance optimized swarm intelligence based routing algorithm," *J. Netw. Comput. Appl.*, vol. 35, no. 6, pp. 1901–1917, 2012.
[24] Q. Zhao and L. Tong, "Energy efficiency of large-scale wireless networks: Proactive versus reactive networking," *IEEE J. Select. Areas Commun*, vol. 23, no. 5, pp. 1100–1112, May 2005.
[25] A. Asim and S. Tixeuil, "XS-WSNet: Extreme scale wireless sensor network simulation," in *Proc. 2010 IEEE Int. Symp. World Wireless Mobile Multimedia Netw.*, 2010, pp. 1–9.
[26] E. Shih *et al.*, "Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks," in *Proc. 7th Annu. Int. Conf. Mobile Comput. Netw.*, 2001, pp. 272–286.
[27] R. L. Graham, B. D. Lubachevsky, K. J. Nurmela, and P. R. J. Ostergard, "Dense packings of congruent circles in a circle," *Discrete Math.*, vol. 181, pp. 139–154, 1998.
[28] E. Specht, The best known packings of equal circles in a circle (complete up to $N = 2600$). [Online]. Available: http://hydra.nat.uni-magdeburg.de/packing/cci/cci.html
[29] L. Bagheri and M. D. T. Fooladi, "A rendezvous-based data collection algorithm with mobile sink in wireless sensor networks," in *Proc. 2014 4th Int. Conf. Comput. Knowl. Eng.*, 2014, pp. 758–762.
[30] Y. Zhang, G. Simon, and G. Balogh, "High-level sensor network simulations for routing performance evaluations," in *Proc. 3rd Int. Conf. Netw. Sens. Syst.*, 2006, pp. 1–4.
[31] Y. Zhang and M. P. J. Fromherz, "Message-initiated constraint-based routing for wireless ad-hoc sensor networks," in *Proc. 1st IEEE Consum. Commun. Netw. Conf.*, 2004, pp. 648–650.
[32] A. M. Zungeru, L. M. Ang, and K. P. Seng, "Termite-hill: Performance optimized swarm intelligence based routing algorithm for wireless sensor networks", *J. Netw. Comput. Appl.*, vol. 35, no. 6, pp. 1901–1917, 2012
[33] [Online]. Available: http://www.libelium.com/development/waspmote/documentation.
[34] E. B. Hamida and G. Chelius, "Analytical evaluation of virtual infrastructures for data dissemination in wireless sensor networks with mobile sink," in *Proc. 1st ACM Int. Workshop Sens. Actor Netw.*, 2007, pp. 3–10.
[35] J. Shin, J. Kim, K. Park, and D. Park, "Railroad: Virtual infrastructure for data dissemination in wireless sensor networks," in *Proc. 2nd ACM Int. Workshop Perform. Eval. Wireless Ad Hoc, Sens., Ubiquitous Netw.*, 2005, pp. 168–174.
[36] Z. A. Khan, C. Belleudy, and M. Auguin, "Analytical model for energy consumption analysis in grid based wireless sensor networks," in *Proc. Int. Conf. New Technol., Mobility Secur.*, 2009, pp. 1–5.
[37] A. M. Zungeru, L.-M. Ang, and K. P. Seng, "A formal mathematical framework for modeling and simulation of wireless sensor network environments utilizing the hill-building behavior of termites," *Simulation: Trans. Soc. Model. Simul. Int.*, vol. 89, no. 5, pp. 589–615, 2013.
[38] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," in *Proc. Int. Symp. Mil. Commun.*, vol. 1, 2003, pp. 688–693.
[39] P. Venkitasubramaniam, S. Adireddy, and L. Tong, "Sensor networks with mobile agents: Optimal random access and coding," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1058–1068, Aug. 2004.
[40] A. M. Zungeru, L.-M. Ang, and K. P. Seng, "Classical and swarm intelligence based routing protocols for wireless sensor networks: A survey and comparison," *J. Netw. Comput. Appl.*, vol. 35, no. 5, pp. 1508–1536, 2012.
[41] C. Cunningham and R. Roberts, "An adaptive path planning algorithm for cooperating unmanned air vehicles," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 4, 2001, pp. 3981–3986.
[42] G. Mergen, Q. Zhao, and L. Tong, "Sensor networks with mobile access: Energy and capacity considerations," *IEEE Trans. Commun.*, vol. 54, no. 11, pp. 2033–2044, Nov. 2006.



Kenneth Li-Minn Ang (SM'09) received the B.Eng. and Ph.D. degrees from Edith Cowan University, Joondalup, WA, Australia.

He is currently a Senior Academic Staff with the School of Computing and Mathematics, Charles Sturt University (CSU), Bathurst, NSW, Australia, and the Leader of the Intelligent Analytics and Sensing research group at CSU. He was previously an Associate Professor with Nottingham Malaysia, Selangor, Malaysia. He has authored or co-authored more than 100 papers in journals and international refereed conferences, and authored *Wireless Multimedia Sensor Networks on Reconfigurable Hardware.* His research interests include visual information processing, embedded systems and wireless sensor networks, reconfigurable computing, the development of real-world computer systems, large-scale data gathering in wireless multimedia sensor systems, big data analytics for sensor networks, and multimedia Internet of Things.

Dr. Ang is a Fellow of the Higher Education Academy, U.K.



Jasmine Kah Phooi Seng received the B.Eng. and Ph.D. degrees from the University of Tasmania, Hobart, TAS, Australia.

She is currently an Adjunct Professor with the School of Computing and Mathematics, Charles Sturt University, Bathurst, NSW, Australia. Before returning to Australia, she was a Professor and the Department Head of Computer Science and Networked System, Sunway University. Before joining Sunway University, she was an Associate Professor with the School of Electrical and Electronic Engineering, Nottingham University. She has authored or co-authored more than 230 papers in journals and international refereed conferences. Her research interests include intelligent visual processing, multimodal signal processing, artificial intelligence, multimedia wireless sensor network, affective computing, the development of intelligent systems, and multimodal big data analytics.

Adamu Murtala Zungeru (M'10) received the B.Eng. degree from the Federal University of Technology, Minna, Nigeria, the M.Sc. degree from Ahmadu Bello University, Zaria, Nigeria, and the Ph.D. degree from Nottignham University, Nottingham, U.K.

He was a Research Fellow with the Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge, MA, USA, where he also obtained a Postgraduate Teaching Certificate. He is currently a Senior Lecturer with the Botswana International University of Science and Technology, Serowe, Botswana. He has authored or co-authored more than 40 papers in journals and international refereed conferences. His research interests include swarm intelligence, wireless sensor networks, embedded systems, and design of analog and digital electronic circuits.

Dr. Zungeru is a member of the ACM and IEEE AP-S.