

General Framework for Network Throughput Maximization in Sink-based Energy Harvesting Wireless Sensor Networks

Abbas Mehrabi, *Student Member IEEE*, Kiseon Kim, *Senior Member IEEE*

Abstract—Due to the advancement in energy harvesting wireless sensor networks (EH-WSNs), the data collection from one-hop stationary sensor nodes using a path-constrained mobile element (ME) has become one of the challenging issues. Toward the throughput improvement, we propose a general framework for network throughput maximization (NTM) problem by optimizing practically feasible parameters. For each proposed scenario, a mixed integer linear programming (MILP) optimization model is introduced for the problem formulation. Due to the NP-Hardness of the MILP models, we design two efficient algorithms namely as ODSAA and ODAA for two practically implementable scenarios. Having a preknowledge about the deployed location of nodes, the proposed algorithms run centrally by sink and find the sub-optimal solutions within a reasonable computation time. Furthermore, under the uniform distribution of energy harvesting, we find out two threshold points on respectively energy harvesting mean and battery capacity of nodes after which the network throughput reaches a stable point. Finally, simulations are conducted on different set of node deployments, which the results confirm that the proposed algorithms significantly improve the data throughput collected by sink and also the theoretical thresholds provide a confidence interval of 90%.

Index Terms—Energy Harvesting Wireless Sensor Networks (EH-WSNs), Network Throughput, Mixed Integer Linear Programming (MILP), NP-Hardness, Energy Harvesting Mean, Battery Capacity Threshold.



1 INTRODUCTION

THANKS to several surveillance applications, data collection from one-hop stationary sensor nodes in wireless sensor networks (WSNs) using a mobile sink has attracted several research attentions, recently [7], [8], [11]. In the road condition surveillance, sensor nodes are stationary deployed to sense statistical data such as humidity or temperature and a mobile vehicle with path-constrained trajectory is employed to collect data from the sensors in an efficient manner. The collected data is forwarded to a main base station for further processing or aggregation and ultimately the improvement of road quality [2]. On the other side, from the energy harvesting point of view as a potential future development of surveillance applications, the unlimited lifetime and continues monitoring is guaranteed. In an energy harvesting wireless sensor network (EH-WSN), sensor node have the possibility of recharging their battery from the ambient energy resources in their surroundings such as solar, wind or vibrations [23]. Parallel to this advancement, the data collection from sensor nodes using mobile sinks has been also studied in EH-WSNs with most focus on throughput maximization [7], [8], [11]. From the mobility point of view, the existing works assume

a constant speed for the mobile sink [7], [8], [11]. However, in many emergency situations such as military applications, considering constant velocity for the mobile sink is not applicable since the data delivery delay must satisfy the application deadline [17]. Furthermore, the motion of sink with a constant speed can unnecessarily waste its energy. Therefore, within the time periods which sink has less available data in its vicinity to collect, it must speed up in order to conserve its energy for consecutive slots with the possibility of more available data.

The problem of maximizing total collected data from one-hop stationary sensor nodes which are deployed along a straight path using a mobile sink in EH-WSNs, namely as network throughput maximization (NTM) problem, was initially introduced in [7], [8]. Due to some unrealistic assumptions on their problem formulation, the authors in [11] reformulate the problem by taking into account the heterogeneity of transmission period and energy harvesting aspect of nodes over different time intervals. The problem scenario in [11] is based on the assumption that the speed of sink and its travelled distance at each time slot are fixed during whole data collection process. In other words, the system parameters are not fully utilized in [11] from the practical point of view. Motivated by this fact, one may further improve the network throughput by optimizing practically feasible parameters such as the travelled distance by sink at each time slot or even its speed. Furthermore, the assumption of constant energy harvesting in [11] does

• Abbas Mehrabi and Kiseon Kim are with the School of Electrical Engineering and Computer Science at Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea.
E-mail: mehrabi@gist.ac.kr, kskim@gist.ac.kr.

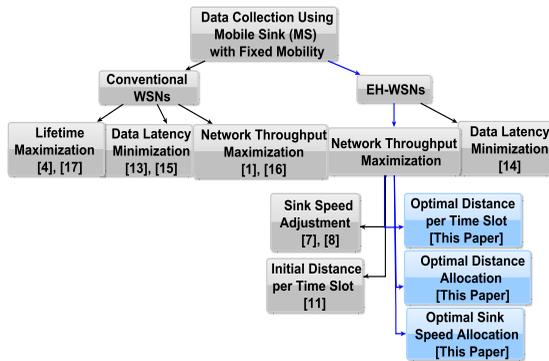


Fig. 1: Data collection in WSNs using mobile elements with fixed mobility pattern under different objectives.

not always meet in practical data collection scenarios since the pattern of energy harvesting from the solar resources is non-deterministic and vary due to periodically change in weather conditions [19], [20]. Consequently, from the prospective of EH-WSNs designers, the energy harvesting resources which affect the network throughput can be utilized in order to minimize the system cost in practical data collection scenarios.

The structure of this paper is organized as follows: In Section 2, the related works together with the main contributions of the paper are discussed. The data collection and energy harvesting models are detailed in Section 3 and Section 4 is devoted to the formulation of NTM problem under different scenarios. The problem approaches, proposed algorithms and related theoretical analysis are all discussed in Section 5 and 6. The analytical thresholds on energy harvesting mean and battery capacity of sensors are derived in Section 7. Simulations are conducted in Section 8 and finally Section 9 concludes the paper.

2 RELATED WORKS

Due to the advancement in WSNs architecture based on the mobile sinks, the data collection using a path-constrained mobile sink has been extensively studied by several researchers [1], [15], [4]. Fig. 1 shows a classification of works on data collection in WSNs using a sink with fixed mobility pattern under different problem objectives. Francesco *et al.* [1] survey thoroughly the data collection works in WSNs using the mobile sinks. In the prospective of sink mobility, the *Fixed Mobility* pattern where the mobile sink moves on a path-constrained trajectories has found many research attentions due to its wide range of practical applications [3], [15], [4]. Using a path-constrained sink for data gathering from one-hop sensor nodes not only improves the network throughput [4] but also reduces the data delivery latency [15].

On the other side, the replacement of conventional energy-limited sensors with energy harvesting nodes has introduced new challenges and opportunities for

data collection applications. In conventional WSNs, the data collection strategies follow mainly three objectives: *Network Lifetime Maximization*, *Data Latency Minimization* or *Network Throughput Maximization*. Due to the constrained imposed on the energy of nodes in conventional WSNs, the authors in [4] introduce the problem of data gathering from far away sensors using relay nodes close to the trajectory of sink with the objective of maximizing overall network lifetime. Y. Yun, et al. [17] introduce the problem of collecting data from sensor nodes with the objective of maximizing network lifetime under the constraint that data delivery delay must satisfy the network deadline. They formulate the problem as an Integer Linear Programming (ILP) optimization model and solve it directly using standard ILP solvers. In [15] and [13], the authors study the problem of determining some pre-specified sojourn locations in the sensors area such that by visiting those locations by sink temporarily and collecting data from the neighboring nodes, the overall data latency is minimized.

From the network throughput point of view in conventional WSNs, the algorithms aim to maximize the total volume of data which is collected by sink in one round of its trajectory. X. Xu, et al. [16] introduce the problem of environmental monitoring using a mobile vehicle. They devise a two stage heuristic which finds a near-optimal tour for the trajectory of sink such that by collecting data from the sensors in the neighboring of that trajectory, the total collected data is maximized.

On the other hand, due to the recent advancement in energy harvesting wireless sensor networks (EH-WSNs), the sensor nodes can harvest energy periodically from the ambient resources in their surroundings such as wind, solar or vibrations [6], [12], [22]. Therefore, the data collection algorithms in EH-WSNs do not take into account the lifetime maximization as long as the sensors have the possibility of energy replenishment. X. Ren, et al. [14] investigate the problem of data gathering from the energy harvesting nodes using a mobile sink under the constraint that the overall data delivery latency must meet the network deadline.

Basically, most of the works on data collection using a mobile node in EH-WSNs are defined with the objective of network throughput maximization [7], [8], [11]. Due to the wide range of applications for data collection on a direct path, the authors in [7], [8] investigate the NTM problem in an EH-WSN using a path-constrained mobile sink. Due to some unrealistic assumptions considered in [7] and [8] and in order to further improve the network throughput, the authors in [11] propose a condition on the fixed distance travelled by sink at all time slots. Based on the proposed condition, they develop an online centralized algorithm which not only has low computational complexity but outperforms the previous works in term of network throughput. Motivated by the drawback

of considering fixed distance travelled by sink and its constant speed in [11], in order to further improve the throughput, we propose different scenarios for NTM problem for which practical parameters are optimally utilized. Furthermore, due to the infeasibility of assumption on constant energy harvesting in practical data collection scenarios [22], [10], in contrast to [11], we consider an uniform energy harvesting distribution in this work to statistically model the amount of harvested energy by nodes during different time intervals.

We note that the objective of NTM problem considered in this work is determining the optimal distance or sink speed together with the selection of one candidate node at each time slot. This makes our work different from traditional data forwarding approach [13] based on classical travelling salesman problem (TSP) in the sense that in classical TSP model, the obtainable throughput in each path is known and the objective is only the selection of optimal trajectory.

2.1 Contributions

The following are our main contributions in this paper:

- A general framework is proposed for NTM problem in sink-based data collection in EH-WSNs such that by optimizing practically feasible parameters, the network throughput significantly improves. Note that we refer to the total collected data by sink as the network throughput throughout the paper.
- A mixed integer linear programming (MILP) optimization model is proposed for each defined scenario of NTM problem which accommodates the problem-specific constraints. Due to the NP-Hardness of the optimization models, two efficient algorithms with polynomial run-time complexity are designed to cope with two practically implementable scenarios.
- Considering the uniform energy harvesting distribution, we show that the increase in the harvested energy by nodes does not guarantee the monotonic increase in total data collected by sink. Motivated by this fact and depending on the battery capacity of nodes, we derive a theoretical threshold on the amount of harvested energy such that by increasing the energy after that threshold, the network throughput reaches a stable point.
- Furthermore, given the mean of energy harvesting distribution, we find out an analytical threshold on the battery capacity such that when the capacity exceeds that threshold point, the network throughput is saturated. As an advantage, these established thresholds can help the designers of EH-WSNs to minimize efficiently the cost of energy harvesting resources in practical data collection scenarios.

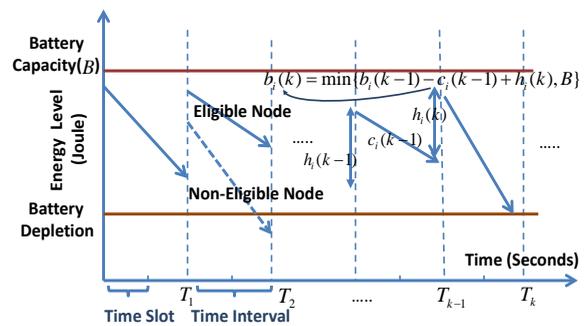


Fig. 2: The energy harvesting at the beginning and the energy consumption of nodes during each time interval.

3 DATA COLLECTION AND ENERGY HARVESTING MODELS

Following the IEEE 802.15.4 standard for WANs (Wide Area Networks), we assume that $|V|$ sensor nodes are uniformly deployed along both sides of a direct path with fixed length L . The sensors are remained stationary during the data collection. A mobile sink with large data buffer size moves on the path as its trajectory to collect data from the nodes in one-hop of its transmission range along the path. The total time duration of one round path traversal by sink is divided into several consecutive time slots. Under two different scenarios, the sink travels the path with either a constant speed v_m during whole of its trajectory or changes its speed at different time slots. With the given speed v and a time slot with duration τ , the mobile sink travels distance $l = v \cdot \tau$ on the path during that time slot. Due to the physical interference between nodes transmitting simultaneously, the mobile sink receives data successfully from at most one sensor node for the whole transmission period of node at each time slot [7], [8], [11]. Furthermore, for initiating the data collection process and in order to localize the duration of message exchange between the sink and sensor nodes, following the work in [11], we consider the continues duration of every two consecutive time slots as one time interval. Based on the sensors' information which is provided to the sink through message passing at the beginning of each time interval, the sink makes decision on which nodes must send their data at the current time interval.

From the energy replenishment point of view, the mobile sink is assumed to have unlimited energy storage, while solar-based energy harvesting sensor nodes such as Heliomote based on MICA2 platform are deployed which periodically recharge their battery from the ambient solar energy resources [6], [10]. The amount of harvested energy varies at different time intervals therefore is non-deterministic. Fig. 2 illustrates a schematic view of energy harvesting by sensor nodes at different time intervals where the energy harvesting point corresponds to the beginning of the time interval. The amount of harvested energy

by nodes at the beginning of each time interval which for the sake of simplicity, we assume that follows a uniform distribution in this work, is provided to the mobile sink via message passing when sink is located at the beginning of interval.

Following the widely adopted energy model [12], the worst case is when the energy consumption rate of node is more than its energy harvesting rate. Therefore, the sensors must preserve their harvested energy during each time interval. Denoted by B_i the battery capacity of sensor node s_i , its energy budget at the beginning of time interval k is given by $b_i(k) = \min\{b_i(k-1) - c_i(k-1) + h_i(k), B_i\}$ where $c_i(k-1)$ is the consumed energy by node i at time interval $k-1$ and $h_i(k)$ represents its harvested energy at time interval k . From the energy consumption point of view, the sensor nodes are divided into two groups: A **non-eligible** node if either it has not slot opportunity for data transmission to the sink or its energy budget within a time slot is less than the required energy for data transmission and, **eligible** node if it has enough energy for transmitting its volume of data to the sink at that time slot. In Fig. 2, the energy consumption pattern of eligible and non-eligible nodes has been shown with bold and dashed lines, respectively. We remark that the possibility of energy replenishment compared to energy-limited sensor nodes and also finding the opportunity for data transmission in subsequent rounds of sink's trajectory on the path can also address the fairness issue.

4 PROBLEM STATEMENT

With the aforementioned data collection and energy harvesting models, the network throughput maximization (NTM) problem is defined as the problem of allocating time slots to sensor nodes considering their energy budget such that the total collected data by sink is maximized. To formulate the NTM problem, we need first to define the following system parameters: Let r_{ij} denote the data transmission rate of node s_i at time slot t_j . Variable τ_{ij} is used to denote the time duration in which the mobile sink can collect data from sensor s_i at time slot t_j . Corresponding to this time period, the mobile sink travels a distance d_{ij} on the path within the coverage range of s_i at time slot t_j . Note that sensor node s_i can have three possible deployed locations for its (x_i, y_i) -coordinate at time slot t_j which are when its communication range covers two consecutive time slots t_{j-1}, t_j , only time slot t_j or two consecutive slots t_j, t_{j+1} . With fixed value of l , it is straightforward to see that the value of $d_{ij}, 1 \leq i \leq |V|, 1 \leq j \leq |T|$, in all three deployed locations is obtained from the following relation:

$$d_{ij} = \begin{cases} \text{Max}\{\{\text{Min}(l(j, x_{end}) - \text{Max}(l(j-1), x_{start})), 0\} \\ 0, \end{cases} \begin{matrix} r_i > y_i \\ r_i \leq y_i \end{matrix} \quad (1)$$

Where x_{start} and x_{end} are respectively the starting and ending intersection points of node's projection with the path. Furthermore, the binary variable a_{ij} is defined to indicates the allocation of time slot t_j to sensor node s_i such that $a_{ij} = 1$ if time slot t_j is allocated to sensor node s_i and $a_{ij} = 0$, otherwise. With all system parameters defined in Table 1, the NTM problem which takes into account the effective transmission period together with the energy harvesting aspect of sensor nodes is formulated as the following ILP model presented in [11]. Here, $|TT| = |T|/2$ is the number of time intervals.

$$\text{Maximize}_{\{a_{ij}\}} \frac{1}{|T| \times \tau} \sum_{i=1}^{|V|} \sum_{j=1}^{|T|} a_{ij} \cdot r_{ij} \cdot \tau_{ij} \quad (2)$$

Subject to:

$$\begin{aligned} a_{ij} &= 0, \quad \forall 1 \leq i \leq |V|, j \notin A(i) & (3) \\ \sum_{i=1}^{|V|} a_{ij} &\leq 1, \quad \forall 1 \leq j \leq |T| & (4) \\ c_i(k) &\leq b_i(k) & (5) \\ \forall 1 \leq i \leq |V|, 1 \leq k \leq |TT| & \\ b_i(k) &= \text{Min}\{b_i(k-1) - c_i(k-1) + h_i(k), B_i\} & (6) \\ \forall 1 \leq i \leq |V|, 1 \leq k \leq |TT| & \\ c_i(k) &= \sum_{2(k-1) < j \leq 2k} a_{ij} \cdot P_{ij} \cdot \tau_{ij}, & (7) \\ \forall 1 \leq i \leq |V|, 1 \leq k \leq |TT| & \\ b_i(1) &= I_i + h_i(1), 1 \leq i \leq |V| & (8) \\ h_{min} &\leq h_i(k) \leq h_{max}, \forall 1 \leq i \leq |V|, 1 \leq k \leq |TT| & (9) \end{aligned}$$

In the above ILP problem, the only decision variable is a_{ij} and since the constant speed v_m is considered for the sink in data collection model of [11], the quantities τ_{ij} are derived by $\tau_{ij} = d_{ij}/v_m$. For all nodes $s_i, 1 \leq i \leq |V|$, the quantities B_i and I_i are known in advance and with the given values of $h_i(k)$ from the uniform interval $[h_{min}, h_{max}]$, the $b_i(k)$ quantities are determined from the relations (6), (7) and (8). Relation (6) also ensures that the energy budget of a node at each time interval does not exceed its battery capacity. The values of other parameters are known in advance. Constraints (3), (4) specify the allocation possibilities and constraint (5) ensures that the total energy consumption of a sensor node for data transmission to sink at the time slots allocated to it within a time interval does not exceed its energy budget at that interval.

As a drawback, the parameters of NTM problem (2)-(9) are not optimally utilized. Inspired by this fact in order to further improve the network throughput, we propose different scenarios for NTM problem by optimizing the practically feasible parameters and formulate the proposed scenarios using mixed integer programming (MIP) optimization models.

TABLE 1: The list of defined parameters in network throughput maximization (NTM) problem.

System Parameter	Parameter Description
$ V , T $	The number of sensor nodes and time slots
L, l, α	Fixed path length and fixed distance traveled by sink at all time slots, path loss exponent normally between 2 and 5
τ, v_m	Time slot period and the constant speed of the sink
v_j, l_j, τ_j	The speed of sink at, the traveled distance by sink in and the time duration of time slot t_j
τ_{ij}, d_{ij}	The period of time that node s_i observes sink in its communication range at time slot t_j and its corresponding distance on the path
(x_i, y_i)	Cartesian deployed location of sensor node s_i on \mathbb{R}^2 plane
R_i, x_{start}, x_{end}	Transmission range of sensor node s_i , starting and ending points where the projection of sensor node on \mathbb{R}^2 plane intersects the path
r_{ij}, P_{ij}	Transmission rate and power of sensor node s_i at time slot t_j
$b_i(k), h_i(k), c_i(k)$	The energy budget, the amount of harvested energy and the energy consumption of sensor node s_i at time interval k
I_i, B_i	Initial energy budget and battery capacity of sensor node s_i
a_{ij}	The allocation decision variable

It is noteworthy to mention that although other mathematical models such as techniques from game theory can capture the formulation of the proposed scenarios, due to the complication of such tools, we utilize MIP optimization models in this work to formulate the NTM scenarios with the following motivations:

- Since the node selection and the distance or speed of the sink at each time slot take values from binary and real domains, respectively, MIP is the most suitable approach to mathematically state the optimization problems in which the objective function and constraints involve two different types of decision variable [24].
- MIP optimization models have been well accepted for problems which explore the use of mobile sink in WSNs targeting at different objectives due to their relatively simple structure to incorporate the objective function and constraints involving various types of decision variables [25].

4.1 Optimal Distance per Time Slot

As the first scenario, the objective is to find an optimal distance that sink travels all time slots with that distance such that the network throughput is maximized assuming that sink maintains a constant speed during its trajectory on the path. Since the time duration of each time slot is not known in advance, with constant sink speed, the optimal distance subsequently determines the time duration and the number of time slots for the trajectory of sink on the path. Considering l here as a real decision variable which its value determines the distance for which sink travels all time slots, the first constraint of this problem scenario is $Max\{R_i, 1 \leq i \leq |V|\} \leq l \leq L$ as proposed in [11]. This condition implies that each sensor node has maximum two available time slots for data transmission to the sink.

Knowing that $|T| = L/l$, the problem of finding the optimal distance per time slot where the maximum network throughput is achieved (NTM-ODTS) can be formulated as MILP optimization model (10)-(13). In this optimization model, the variables a_{ij} and l are the binary and real decision variables, respectively. The value of parameters $v_m, r_{ij}, P_{ij}, B_i, I_i$ and $h_i(k)$

are known in advance. On the other side, with the optimal value of l , the value of variables d_{ij} are determined from equation (1), $|T| = \lfloor L/l \rfloor$, $\tau = l/v_m$, $\tau_{ij} = (d_{ij}/l) \times \tau$ and $c_i(k), b_i(k)$ depend on the binary values $a_{ij}, 2(k-1) \leq j \leq 2k$. Constraint (11) specifies the possible range for variable distance per time slot and constraints (12)-(13) have similar description with constraints (5)-(7) in NTM problem (2)-(9) except for here, the optimal value of l determines the time duration that is considered for each time slot as well as the number of time slots.

$$\underset{l, \{a_{ij}\}}{\text{Maximize}} \quad \frac{1}{L} \sum_{i=1}^{|V|} \sum_{j=1}^{\lfloor \frac{L}{l} \rfloor} a_{ij} \cdot r_{ij} \cdot d_{ij} \quad (10)$$

Subject to:

$$\text{Max}\{R_i\} \leq l \leq L, \quad \forall 1 \leq i \leq |V| \quad (11)$$

$$c_i(k) \leq b_i(k) \quad (12)$$

$$\forall 1 \leq i \leq |V|, 1 \leq k \leq |T|$$

$$c_i(k) = \sum_{2(k-1) < j \leq 2k} a_{ij} \cdot P_{ij} \cdot d_{ij} \cdot \left(\frac{1}{v_m}\right) \quad (13)$$

$$\forall 1 \leq i \leq |V|, 1 \leq k \leq |T|$$

with constraints (3), (4), (6), (8), (9).

4.2 Optimal Distance Allocation

As the second scenario for NTM problem, assuming that the mobile sink maintains a constant speed during whole of its trajectory on the path, the objective is to determine different optimal distances for time slots in order to further improve the network throughput. Note that under this scenario, the total number of time slots is fixed at $|T|$ and the time duration of time slots are not known in advance. With constant sink speed, different distances for time slots subsequently determine the time duration of time slots which their values vary depending on the optimal distance of each slot. Denoting $l_j, 1 \leq j \leq |T|$ as the distance considered for time slot t_j , the objective is to find the optimal distance vector $l_v = \{l_1, l_2, \dots, l_{|T|}\}$ for which the network throughput is maximized. Under this scenario, the problem of optimal distance allocation subject to maximizing the network throughput (NTM-ODA) can be formulated as the following MILP

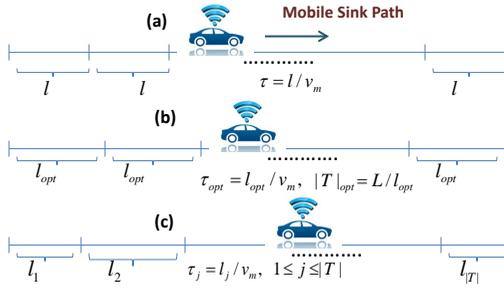


Fig. 3: The movement of sink under different scenarios. (a): NTM (b): NTM-ODTS (c): NTM-ODA

optimization model:

$$\text{Maximize}_{\{l_j\}, \{a_{ij}\}} \frac{1}{L} \sum_{i=1}^{|V|} \sum_{j=1}^{|T|} a_{ij} \cdot r_{ij} \cdot d_{ij} \quad (14)$$

Subject to:

$$\begin{aligned} l_{min} &\leq l_j \leq l_{max}, \quad \forall 1 \leq j \leq |T| \quad (15) \\ \sum_{j=1}^{|T|} l_j &= L, \quad \forall 1 \leq j \leq |T| \quad (16) \\ &\text{with constraints (3), (4), (6), (8),} \\ &\quad (9), (12), (13). \end{aligned}$$

In the above formulation, the variables a_{ij} and $l_j, 1 \leq j \leq |T|$ are respectively the binary and real decision variables. We should note that the value of parameters $|T|, v_m, r_{ij}, P_{ij}, B_i, I_i$ and $h_i(k)$ are known in advance while on the other side, the optimal value of $l_j, 1 \leq j \leq |T|$ subsequently determines $\tau_j = l_j/v_m, \tau_{ij} = (d_{ij}/l_j) \times \tau_j$ and $c_i(k), b_i(k)$ are determined based on binary values $a_{ij}, 2(k-1) < j \leq 2k$. Furthermore, d_{ij} quantities are computed using the following revised version of equation (1):

$$d_{ij} = \begin{cases} \text{Max} \left\{ \text{Min}(\sum_{k=1}^j l_k, x_i + \sqrt{R_i^2 - y_i^2}) - \right. \\ \left. \text{Max}(\sum_{k=1}^{j-1} l_k, x_i - \sqrt{R_i^2 - y_i^2}), 0 \right\} & R_i > y_i \\ 0, & R_i \leq y_i \end{cases} \quad (17)$$

Constraint (15) ensures that the allocated distance to each time slot falls within the minimum and maximum allowed distances for all time slots and constraint (16) guarantees that the sum of all allocated distances does not exceed the total length of path.

We have illustrated schematically in Fig. 3 the distinction between NTM and the proposed scenarios NTM-ODTS and NTM-ODA in term of travelled distance by sink at each time slot as well as the number of time slots when the speed of sink remains constant under each scenario.

4.3 Optimal Sink Speed Allocation

Motivated from delay-tolerant applications for which the total data delivery time must satisfy the application's deadline [1], [15], we define a scenario fro NTM problem where different speeds are allocated to the sink during different time slots.

For the third scenario, with total number of $|T|$ slots and assuming the fixed distance l considered for all time slots, the objective is to determine the optimal sink speed v_j for each time slot $t_j, 1 \leq j \leq |T|$ in order to achieve the maximum throughput. Since the distance is fixed for all time slots, the optimal speed at each time slot subsequently determines the time duration that sink must spend at that slot. With the aforementioned notations, the problem of optimal sink speed allocation with the objective of maximizing the network throughput (NTM-OSS) can be formulated as the following mixed integer nonlinear programming (MINLP) optimization model. Note that parameters $T_{deadline}, v_{min}$ and v_{max} are used to denote respectively the data delivery deadline, the minimum and the maximum speeds which can be allocated to sink at all time slots.

$$\text{Maximize}_{\{v_j\}, \{a_{ij}\}} \frac{\sum_{i=1}^{|V|} \sum_{j=1}^{|T|} a_{ij} \cdot r_{ij} \cdot d_{ij} \cdot (\frac{1}{v_j})}{\sum_{j=1}^{|T|} \frac{l}{v_j}} \quad (18)$$

Subject to:

$$\begin{aligned} v_{min} &\leq v_j \leq v_{max}, \quad \forall 1 \leq j \leq |T| \quad (19) \\ \sum_{j=1}^{|T|} \frac{l}{v_j} &\leq T_{deadline} \quad (20) \\ c_i(k) &\leq b_i(k) \quad (21) \\ \forall 1 \leq i \leq |V|, 1 \leq k \leq |TI| \\ c_i(k) &= \sum_{2(k-1) < j \leq 2k} a_{ij} \cdot P_{ij} \cdot d_{ij} \cdot (\frac{1}{v_j}) \quad (22) \\ \forall 1 \leq i \leq |V|, 1 \leq k \leq |TI| \\ &\text{with constraints (3), (4), (6), (8), (9).} \end{aligned}$$

In problem formulation (18)-(22), the variables a_{ij} and v_j are the binary and real decision variables, respectively. The value of parameters $|T|, l, r_{ij}, P_{ij}, B_i, I_i$ and $h_i(k)$ are known in advance and also d_{ij} values are computed from the relation (1). On the other hand, the optimal value of v_j determines the values of $\tau_j = l/v_j$ and $\tau_{ij} = (d_{ij}/l) \times \tau_j$. Constraint (19) specifies the possible range for the speed which can be allocated to the sink at each time slot and constraint (20) sates that the total time taken by sink for data collection must satisfy the data delivery deadline. The rest of the constraints have the similar description as their corresponding constraints in problem (2)-(9) except here, with fixed distance per time slot, the optimal value of speed at each time slot determines

TABLE 2: The NTM problems under different scenarios (optimization model and parameters).

Data Collection Scenario	Optimization Model	Decision Variables	Independent Parameters	Dependent Parameters	Proposed Algorithm
Initial Distance per Time Slot [11]	ILP	$\{a_{ij}\}$	$ T , l, \tau, v_m, \tau_{ij}, d_{ij}, r_{ij}, P_{ij}, h_i(k), B_i$	$c_i(k), b_i(k)$	AdjustmentBased-Allocation
Optimal Distance per Time Slot [This Paper]	MILP	$l, \{a_{ij}\}$	$v_m, r_{ij}, P_{ij}, h_i(k), B_i$	$ T , \tau, \tau_{ij}, d_{ij}, c_i(k), b_i(k)$	ODSAA
Optimal Distance Allocation [This Paper]	MILP	$\{l_j\}, \{a_{ij}\}$	$ T , v_m, r_{ij}, P_{ij}, h_i(k), B_i$	$\tau_j, \tau_{ij}, d_{ij}, c_i(k), b_i(k)$	ODAA

the time duration that sink spends at each time slot.

We should note that MILP formulations (10)-(13) and (14)-(16) are not mathematically equivalent due to the difference in the number of their decision variables. This in turn implies that the proposed algorithm for each scenario is not applicable for the other one. It is also noted that although the problem formulations for the second and third scenarios are equivalent, however, they are not transformable to each other. The reason is that since the variable which has to be optimized is not identical under these two scenarios therefore, the time slots duration are not equal under these two scenarios. This in turn implies that the distance allocation in second scenario is treated as independent from the speed allocation in third scenario.

In Table 2, we have summarized the NTM problem and the first two proposed scenarios together with the optimization model, involving parameters and the proposed data collection algorithm for each scenario. We note that although third scenario of NTM problem has been investigated in this work, as we point out later in Section 6, designing an algorithm for this scenario is considered as an interesting future work.

5 PROBLEMS APPROACH

The proposed optimization models belong to the class of intractable problems and finding their optimal solutions is an NP-Hard problem [9]. The combination of branch and bound with linear programming relaxation can be applied to achieve the optimal solution of MILP models [26]. However, the computational complexity of the branch and bound significantly grows when the number of sensors or time slots increases. As another approach, one may apply the exhaustive method for searching one candidate for real decision variable from its domain for which the linear programming relaxation on its corresponding ILP subproblem results in the best upper bound. However, there is no guarantee on the optimality of the candidate value for real decision variable since there is no theoretically proven approximation factor for the relaxation technique.

Due to the computationally expensive of finding the optimal solutions of MILP problems, we rely on the well-used greedy heuristic [5] to find the approximate feasible solutions in a reasonable computation time.

Greedy Allocation Heuristic (GAH)

Input: $|T|$: Number of Time slots, $|V|$: number of sensors
Output: Allocation of $|T|$ time slots to $|V|$ sensor nodes

- 1: **for** each time slot $1 \leq t \leq |T|$ **do**
- 2: Neighbor(t) \leftarrow The set of nodes whose their transmission range covers the sink trajectory at t
- 3: **for** each sensor $s \in$ Neighbor(t) **do**
- 4: energyBudget(s) \leftarrow energyBudget(s)+harvestedEnergy
- 5: **end for**
- 6: eligibleNodes(t) \leftarrow The set of nodes from Neighbor(t) which satisfy constraint (5);
- 7: **for** each sensor $s' \in$ eligibleNodes(t) **do**
- 8: **Compute** Throughput(s')
- 9: **end for**
- 10: $s_{select} \leftarrow \arg \max_{s' \in \text{eligibleNodes}(t)} \{\text{Throughput}(s')\}$
- 11: **Allocate** timeslot t to s_{select}
- 12: **Update** energy budget of s_{select}
- 13: overallThroughput \leftarrow overallThroughput + Throughput(s_{select})
- 14: **end for**
- 15: **return** overallThroughput

Getting benefit of the greedy allocation, we develop two efficient algorithms for the first and second scenarios of NTM problem. As an advantage, the proposed algorithms not only result in significantly improvement in network throughput compared to the algorithm of [11] but most importantly, as we show later, they provide an approximation factor for the problems which is interestingly no larger than the approximation factor of the greedy algorithm for NTM problem. Before detailing the proposed algorithms, we need the following definitions:

Definition 1: Let ILP-ODTS($l', \{a_{ij}\}$) be the obtained integer linear programming subproblem when the fixed scalar distance $\text{Max}\{R_i, 1 \leq i \leq |V|\} \leq l' \leq L$ is substituted in NTM-ODTS problem (10)-(13).

Definition 2: Let ILP-ODA($l'_v, \{a_{ij}\}$) be the obtained integer linear programming subproblem when the fixed distance vector $l'_v = (l'_1, l'_2, \dots, l'_{|T|})$ satisfying conditions $l_{min} \leq l'_j \leq l_{max}, 1 \leq j \leq |T|$ and $\sum_{j=1}^{|T|} l'_j = L$ is substituted in NTM-ODA problem (14)-(16).

6 PROPOSED ALGORITHMS

According to the definitions in the previous section, we design two greedy-based efficient algorithms for the first and second scenarios of NTM problem. Having the pre-knowledge about the locations of

Algorithm 1 Optimal Distance per Slot Allocation (ODSAA)

Input: Initial distance per time slot
 $l = \text{Max}(R_i, 1 \leq i \leq |V|)$
Output: Optimal distance per time slot
 1: Throughput \leftarrow solution of GAH for subproblem
 $\text{ILP-ODTS}(l, \{a_{ij}\})$
 2: **while** Throughput is increasing **do**
 3: $l \leftarrow l + \Delta l$
 4: Throughput \leftarrow solution of GAH for subproblem
 $\text{ILP-ODTS}(l, \{a_{ij}\})$
 5: **end while**
 6: **return** l , Throughput;

nodes in the network using the intercommunication mechanism between them, the proposed algorithms run centrally by the sink and find the sub-optimal solutions within a reasonable computation time.

For the NTM-ODTS problem, as the authors provide a reasonable explanation in [11], by increasing the distance per time slot, l , from the initial value, the network throughput has an increasing behavior up to an optimal point. We get benefit of this observation to design the algorithm for NTM-ODTS problem. The body of the proposed algorithm namely as Optimal Distance per Slot Allocation Algorithm (ODSAA) has been shown in Algorithm 1. It starts with the initial distance and slightly increases it by small constant Δl until the throughput returned by greedy allocation heuristic (GAH), with the pseudocode shown in the previous page, for obtained ILP subproblem $\text{ILP-ODTS}(l, \{a_{ij}\})$ has increasing behavior. The algorithm stops and returns the last value of l where for the first time throughput decreases. Since the proposed algorithm is adopted from the justification of authors in [11] for the behavior of throughput, therefore, the convergence of ODTS algorithm to a near-optimal solution is guaranteed.

For the NTM-ODA problem, the algorithm initializes the distance vector $l_v = \{l_1, l_2, \dots, l_{|T|}\}$ with equal values for all time slots. Then, the greedy algorithm is run for the first round and the average of throughput at each time slot is obtained. Within a number of iterations equal to the number of time slots, the algorithm slightly updates the distance that sink must traverse at each time slot based on the amount of available data. More precisely, if the available data in the current time slot is below the average value, the algorithm decreases the distance at that time slot proportional to the difference between the average and available throughput at that time slot. Obviously, more available data to be collected less reduction in distance. The reverse operation is performed by the algorithm if the throughput at the current time slot is above the average. As a constraint, this updating process must satisfy the feasibility of distance vector. In other words, both constraints (15) and (16) must be satisfied during the updating process. The constraint (15) can

Algorithm 2 Optimal Distance Allocation (ODAA)

Input: Initial distance vector $l_v = \{l_1, l_2, \dots, l_{|T|}\}$ where
 $l_v(k) = \text{Max}(R_i, 1 \leq i \leq |V|), \forall 1 \leq k \leq |T|$
Output: Optimal distance vector and optimal throughput
 1: overallThroughput \leftarrow solution of GAH for
 subproblem $\text{ILP-ODA}(l_v, \{a_{ij}\})$
 2: optThroughput \leftarrow overallThroughput, optDVector $\leftarrow l_v$
 3: **for** each $1 \leq \text{iteration} \leq |T|$ **do**
 4: aveThr \leftarrow overallThroughput/ $|T|$
 5: **for** each $1 \leq t \leq \text{iteration}$ **do**
 6: $l_v(t) = l_v(t) +$
 $(\text{Throughput}(t) - \text{aveThr})/\text{aveThr}$
 7: **if** $l_v(t) < l_{\min}$ **OR** $l_v(t) > l_{\max}$ **then**
 8: $l_v(t) = l_v(t) +$
 $(\text{aveThr} - \text{Throughput}(t))/\text{aveThr}$
 9: **end if**
 10: **end for**
 11: sumofDistance $\leftarrow \sum_{t=1}^{\text{iteration}} l_v(t)$
 12: aveDistance $\leftarrow (L - \text{sumofDistance})/(|T| - \text{iteration})$
 13: **if** aveDistance $< l_{\min}$ **then**
 14: $l_v(t) = l_v(t) -$
 $((l_{\min} - \text{aveDistance}) * (|T| - \text{iteration}))/\text{iteration}$
 $\forall 1 \leq t \leq \text{iteration}$
 15: **end if**
 16: **if** aveDistance $> l_{\max}$ **then**
 17: $l_v(t) = l_v(t) +$
 $((\text{aveDistance} - l_{\max}) * (|T| - \text{iteration}))/\text{iteration}$
 $\forall 1 \leq t \leq \text{iteration}$
 18: **end if**
 19: sumofDistance $\leftarrow \sum_{t=1}^{\text{iteration}} l_v(t)$
 20: aveDistance $\leftarrow \text{sumofDistance}/(|T| - \text{iteration})$
 21: $l_v(t) \leftarrow \text{aveDistance}, \forall \text{iteration} + 1 \leq t \leq |T|$
 22: overallThroughput \leftarrow solution of GAH for
 subproblem $\text{ILP-ODA}(l_v, \{a_{ij}\})$
 23: **if** overallThroughput $>$ optThroughput **then**
 24: optThroughput \leftarrow overallThroughput
 25: optDVector $\leftarrow l_v$
 26: **end if**
 27: **end for**
 28: **return** optDVector, optThroughput

be simply checked by performing two comparisons to ensure that the updated distance falls within the allowed range. To satisfy constraint (16) at each iteration, after updating the distance of the beginning time slots, the remaining slots are allocated with a distance which is the average of difference between the path length and residual distance. Two further conditions guarantee the validity of the allocated distance to the remaining time slots.

The greedy allocation heuristic (GAH) is run on the updated distance vector and the obtained throughput is compared with the best one found yet. The algorithm finally returns the optimal distance vector along with the maximum achievable throughput after the completion of all iterations. Since at each iteration, the distance of all time slots from the beginning up to the current iteration is updated based on the throughput achieved in the previous round, therefore, the convergence of the best solution returned by algorithm which as we show later provides the desired approximation factor is guaranteed. The body of the proposed algorithm called Optimal Distance

Allocation Algorithm (ODAA) has been summarized in Algorithm 2.

We note that an algorithm similar to ODAA can be designed for the third scenario with slightly different speed adjustment which we consider it as our interesting future work.

6.1 Approximation Factor

In this section, we analysis the approximation factor of the proposed suboptimal algorithms. Since the proposed algorithms employ the greedy allocation heuristic in each iteration, therefore, we need to find out the approximation factor of the greedy heuristic.

Definition 3: Let P be an NP-Hard maximization problem such that there is a polynomial algorithm A for it. We say A is an α -approximation algorithm ($0 < \alpha < 1$) for P if for every instance x of P , the solution returned by A , $s_A(x)$, is within the factor of α from the optimal solution $s_{opt}(x)$, namely, $s_A(x) \geq \alpha \cdot s_{opt}(x)$ [21].

As our methodology to prove the approximation factor of greedy heuristic, we first derive the approximation factor of AdjustmentBased-Allocation algorithm using the idea of interval repartitioning. Together with the approximate improvement ratio between the AdjustmentBased-Allocation and the greedy heuristic, we then derive the desired approximation factor.

Lemma 1: AdjustmentBased-Allocation algorithm proposed in [11] is a $\frac{6}{7}$ -approximation algorithm for the NTM problem (2)-(9).

Proof: Given an instance P of NTM problem (2)-(9), let assume S_A and S_{opt} be respectively the overall data throughput which is achieved by algorithm AdjustmentBased-Allocation and the optimal throughput for instance P . Under condition $l \geq \max(r_i, 1 \leq i \leq |V|)$, each sensor can send data to sink in at most two consecutive time slots and as authors show in [11], the difficulty of problem comes from the correlation between sensors when different interval partitioning is considered. To prove the approximation factor, we first find an upper bound on the optimal solution of instance P .

The AdjustmentBased-Allocation algorithm considers the interval partitioning $\{2, 2, \dots, 2\}$ for instance P and finds the optimal solution within each interval by doing four comparisons between the sensors with first and second maximum available data at two consecutive slots. In the following, we show how changing the interval partitioning leads to further improvement in throughput. Note that the sensor nodes harvest the same amount of energy within each time interval under different interval partitioning. Since the condition $l \geq \max\{R_i, 1 \leq i \leq |V|\}$ implies that each node

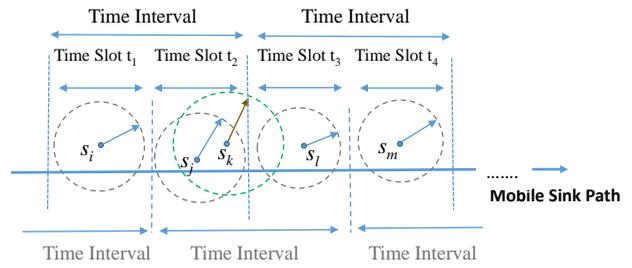


Fig. 4: Five sensor nodes at three consecutive time intervals.

can send its data to sink in at most two consecutive time slots, the following is the only scenario such that changing the interval partitioning leads to further energy harvesting and therefore further throughput obtained. Consider the scenario in Fig. 4 where five sensor nodes can send their data at four time slots within three consecutive time intervals. Assume sensor nodes s_i and s_m are the nodes with maximum available data at time slots respectively t_1 , the second time slot of the first interval, and t_4 , the first time slot of the third interval. In the second time interval, sensor node s_k is the node with maximum data at both time slots t_2 and t_3 but has not enough energy for data transmission to sink at both slots. Also, sensors s_j and s_l are the nodes with second maximum available data at time slots respectively t_2 and t_3 . Further assume that $D[s_j][t_2] + D[s_k][t_3] > D[s_k][t_2] + D[s_l][t_3]$ holds where $D[s][t]$, the volume of data that sensor node s can send to sink at time slot t , is given by $D[s][t] = r_{st} \times \tau_{st}$.

What AdjustmentBased-Allocation algorithm does is considering the interval partitioning $\{2, 2, 2\}$ (Shown in the bottom of Fig. 4) and allocating four time slots $\{t_1, t_2, t_3, t_4\}$ to sensor nodes $\{s_i, s_j, s_k, s_m\}$, respectively. Now to see how much further throughput can be achieved, we need to consider all possible interval partitioning. With interval partitioning $\{1, 2, 2, 1\}$ (shown in the top of Fig. 4), the sensor node s_k is located at two different time intervals and therefore finds the opportunity of two times energy harvesting. This in turn results in the selection of nodes $\{s_i, s_k, s_k, s_m\}$ for data transmission to the sink. It is easy to see that under the other possible interval partitioning $\{2, 1, 1, 2\}$, $\{1, 2, 1, 2\}$, $\{2, 1, 2, 1\}$ or $\{1, 1, 1, 1, 1, 1\}$, the same nodes $\{s_i, s_k, s_k, s_m\}$ are selected for data transmission to the sink. It can be seen that considering each of the above-mentioned interval partitioning, the further improvement $D[s_k][t_2] - D[s_j][t_2]$ is obtained. Depending on the deployed location of sensor node s_k at time slot t_2 , with fixed values τ and l , its transmission duration can change approximately from zero to maximum $\bar{R}\tau/l$, where \bar{R} is the average transmission range of nodes. Now, since the transmission duration of sensor node s_j at time slot t_2 can not be more than the one of node s_k at this

time slot, the average of improvement in throughput is approximately given by:

$$\begin{aligned} E[D[s_k][t_2] - D[s_j][t_2]] &= E[r_{kt_2}] \cdot E[\tau_{kt_2}] \\ &\quad - E[r_{jt_2}] \cdot E[\tau_{jt_2}] \\ &\approx \bar{r} \frac{\bar{R} \cdot \tau}{2l} - \bar{r} \frac{(\bar{R} \cdot \tau)/2l}{2} = \bar{r} \frac{\bar{R} \cdot \tau}{4l} \end{aligned} \quad (23)$$

Where \bar{r} is the average transmission rate of nodes. With $|T|/2$ time intervals, assume that for instance P of the problem, there are $k \leq |T|/2$ time intervals with available data (available node). Since the above improvement in throughput is achieved for three consecutive time intervals, we conclude that for instance P , the optimal throughput cannot be more than the throughput returned by AdjustmentBased-Allocation algorithm plus $k/3$ times of the above improvement. That means for the instance P of the problem:

$$S_{opt} \leq S_A + \left(\frac{k}{3}\right) \times \left(\frac{\bar{r} \cdot \bar{R} \cdot \tau}{4l}\right) \quad (24)$$

Note that here the equality holds when the improvement occurs in all three consecutive time intervals. Inequality (24) in turn implies that:

$$\frac{S_{opt}}{S_A} \leq 1 + \frac{k \cdot \bar{r} \cdot \bar{R} \cdot \tau}{12l \cdot S_A} \quad (25)$$

Since according to Fig. 4, at least the data from either nodes s_j and s_l in both or node s_k in one of the time slots t_2 and t_3 is collected, therefore, for instance P of the problem, at each time interval with available data, the AdjustmentBased-Allocation algorithm achieves an average data throughput of at least $\bar{r} \cdot (\bar{R} \cdot \tau / 2l)$. That means with k time intervals, the overall data throughput returned by AdjustmentBased-Allocation is at least $k \times (\bar{r} \cdot \bar{R} \cdot \tau) / 2l$ which yields the inequality $\frac{k \cdot \bar{r} \cdot \bar{R} \cdot \tau}{2l \cdot S_A} \leq 1$. This inequality in turn implies that $\frac{k \cdot \bar{r} \cdot \bar{R} \cdot \tau}{12l \cdot S_A} \leq 1/6$. Now, from (25), we get $\frac{S_{opt}}{S_A} \leq 1 + \delta_1$ where $0 < \delta_1 \leq 1/6$. This concludes that AdjustmentBased-Allocation is a $(\frac{6}{5})$ -approximation algorithm for NTM problem (2)-(9). \square

Corollary 1: Greedy allocation heuristic (GAH) is a $\frac{9}{14}$ -approximation heuristic for NTM problem (2)-(9).

Proof: Let S_G be the optimal solution returned by greedy allocation heuristic GAH on instance P . Since the greedy heuristic collects data locally at each time slot without interval partitioning, therefore, finding the upper bound on ratio $\frac{S_{opt}}{S_G}$ is not straightforward following the repartitioning as in Lemma 1. To derive the upper bound on this ratio, we derive first an upper bound on the ratio between the data throughput returned by AdjustmentBased-Allocation and the greedy heuristic i.e. $\frac{S_A}{S_G}$ and then together with the result of Lemma 1, we get the desired approximation factor. In Fig. 4, since the greedy heuristic allocates time slots t_2 and t_3 to sensor nodes s_j, s_l , respectively, with respect to AdjustmentBased-Allocation

algorithm, a throughput improvement of $(D[s_j][t_2] + D[s_k][t_3]) - (D[s_j][t_2] + D[s_l][t_3])$ is achieved. Knowing that the transmission duration of node s_l at time slot t_3 can not exceed the transmission duration of node s_j at time slot t_2 , therefore, the average of improvement within one time interval is approximately given by:

$$\begin{aligned} E[(D[s_k][t_2] + D[s_j][t_3]) - (D[s_j][t_2] + D[s_l][t_3])] \\ \approx \left(\bar{r} \frac{\bar{R} \cdot \tau}{4l} + \bar{r} \frac{\bar{R} \cdot \tau}{2l}\right) - \left(\bar{r} \frac{\bar{R} \cdot \tau}{2l} + \bar{r} \frac{\bar{R} \cdot \tau}{8l}\right) = \bar{r} \frac{\bar{R} \cdot \tau}{8l} \end{aligned} \quad (26)$$

Having that $k \leq |T|/2$ time intervals with available data in Lemma 1, the following inequality holds:

$$S_A \leq S_G + k \times \left(\frac{\bar{r} \cdot \bar{R} \cdot \tau}{8l}\right) \quad (27)$$

And (27) in turn implies:

$$\frac{S_A}{S_G} \leq 1 + \frac{k \cdot \bar{r} \cdot \bar{R} \cdot \tau}{8l \cdot S_G} \quad (28)$$

Now, since according to Fig. 4, the greedy allocation collects at least the data from the nodes s_j and s_l at time slots t_2 and t_3 , therefore, for the instance P of the problem, the greedy heuristic achieves the average throughput of at least $(\frac{\bar{r} \cdot \bar{R} \cdot \tau}{4l} + \frac{\bar{r} \cdot \bar{R} \cdot \tau}{8l})$ in time interval with available data. That means the overall throughput returned by greedy allocation heuristic in k time intervals is at least $k \times (\frac{\bar{r} \cdot \bar{R} \cdot \tau}{4l} + \frac{\bar{r} \cdot \bar{R} \cdot \tau}{8l})$ which yields the inequality $\frac{3k \cdot \bar{r} \cdot \bar{R} \cdot \tau}{8l \cdot S_G} \leq 1$. Considering this later inequality in the right hand side of (28), we get $\frac{S_A}{S_G} \leq 1 + \delta_2$ where $0 < \delta_2 \leq 1/3$. Now from $\frac{S_{opt}}{S_A} \leq 7/6$ and $\frac{S_A}{S_G} \leq 4/3$, we conclude that greedy allocation GAH is a $\frac{9}{14}$ -approximation heuristic for NTM problem (2)-(9). \square

In the following, we show that the proposed algorithms achieve an approximation factor for the first and second scenarios of NTM problem which is same as the approximation factor of the greedy allocation for NTM problem. Lets bring first the following definition:

Definition 4: Let $opt(l')$ and $subopt(l')$ be respectively the optimal objective value of subproblem ILP-ODTS($l', \{a_{ij}\}$) and the corresponding suboptimal objective value when the greedy allocation heuristic is applied on it.

Theorem 1: Algorithm ODSAA achieves an approximation factor for NTM-ODTS problem which is same as the approximation factor of GAH for the NTM problem (2)-(9).

Proof: Let $Max\{R_i, 1 \leq i \leq |V|\} \leq l^* \leq L$ and $Max\{R_i, 1 \leq i \leq |V|\} \leq l^+ \leq L$ be respectively the optimal solution and the solution returned by algorithm ODSAA for NTM-ODTS problem. With $\alpha = 9/14$, the approximation factor of the greedy heuristic GAH, to prove that ODSAA is an α -approximation algorithm for NTM-ODTS problem,

we need to show the inequality $subopt(l^+) \geq \alpha \cdot opt(l^*)$.

Since the greedy heuristic GAH achieves the approximation factor of α for any instance of the NTM problem, therefore:

$$subopt(l^*) \geq \alpha \cdot opt(l^*) \quad (29)$$

Furthermore, since the ODSAA algorithm returns the distance scalar where the greedy allocation for its corresponding ILP subproblem achieves the maximum objective value among all possible subproblems, we have the following relation hold:

$$subopt(l^+) = Max \left\{ subopt(l'), \forall l' : Max \{ R_i, 1 \leq i \leq |V| \} \leq l' \leq L \right\} \quad (30)$$

This in turn implies that:

$$subopt(l^+) \geq subopt(l^*) \quad (31)$$

Now, combining (29) and (31), we get:

$$subopt(l^+) \geq \alpha \cdot opt(l^*) \quad \square \quad (32)$$

Corollary 2: Algorithm ODAA achieves a same approximation factor for the NTM-ODA problem as the approximation factor of GAH for NTM problem (2)-(9).

Proof: The proof is similar to the proof of Theorem 1 and is omitted here for the sake of simplicity. \square

6.2 Complexity Analysis

The two following theorems show that the proposed algorithms have the computational complexity of polynomial order in the worst case.

Theorem 2: Algorithm ODSAA has a computational complexity of order $O(|T| \cdot |V|)$ in the worst case where $|T|$ and $|V|$ are respectively the number of time slots and deployed nodes.

Proof: According to the ODSAA algorithm, in each time slot, the GAH constructs the neighborhood set and finds the *selectedNode* with $O(|V|)$ in the worst case. Since the mobile sink traverses the path within $|T|$ time slots, the overall time complexity of GAH is of order $O(|T| \cdot |V|)$. Assume that the algorithm finds the solution after N iterations in the worst case. Since the ODSAA algorithm runs the greedy allocation as its subroutine in each iteration, therefore, in the worst case, the overall time complexity of the algorithm ODSAA is of order $O(N \cdot |T| \cdot |V|)$.

With fixed value of $|T|$ and large $|V|$, the nodes lose their contribution in data transmission to sink with very high ratio when the distance per time slot increases from the initial value. This in turn implies that using ODSAA, the optimal distance per time slot is obtained after very small number of iterations. Therefore, from the definition of Big O notation:

$$f(|T|, |V|) \in O(g(|T|, |V|)) \iff \exists c, V_m > 0 \forall |V| \geq V_m \Rightarrow f(|T|, |V|) \leq c \cdot g(|T|, |V|)$$

with fixed $|T|$ and $|V| \rightarrow \infty$, the number of iterations, N , starts from a very small value and converges to 1. Considering $c = 1$ and large V_m , this implies that:

$$f(|T|, |V|) = T_{ODSAA}(|T|, |V|) \leq |T| \cdot |V| = g(|T|, |V|) \Rightarrow T_{ODSAA}(|T|, |V|) \in O(|T| \cdot |V|) \quad \square$$

Theorem 3: Algorithm ODAA has a computational complexity of order $O(|T|^2 \cdot |V|)$ in the worst case where $|T|$ and $|V|$ are respectively the number of time slots and deployed nodes.

Proof: According to ODAA algorithm, the initial *optThroughput* is obtained using the greedy allocation heuristic with a complexity of order $O(|T| \cdot |V|)$. The total number of iterations in the algorithm is equal to the number of time slots. Within the iteration $i, 1 \leq i \leq |T|$, the first and second **For** loops are run with a complexity of order $O(i)$. To compute the complexity of the **If** part, we need to compute the probability that the event *AverageDistance* $< l_{min}$ happens. To do that, we need first to find the range of all possible values for the variable *AverageDistance*. It is obvious that the range of values for variable *AverageDistance* is equal to $\frac{L - il_{max}}{|T| - i} \leq AverageDistance \leq \frac{L - il_{min}}{|T| - i}$ where, L is the path length and l_{min} and l_{max} are respectively the minimum and maximum allowed distances. Considering variable *AverageDistance* as a random variable, its probability density function is stated as $f(l) = \frac{|T| - i}{i(l_{max} - l_{min})}$. Therefore, the probability of happening the event *AverageDistance* $< l_{min}$ is computed by the following integration:

$$Prob[AverageDistance < l_{min}] = Max \left\{ 0, \int_{\frac{L - il_{max}}{|T| - i}}^{l_{min}} \frac{|T| - i}{i(l_{max} - l_{min})} dl \right\} = Max \left\{ 0, \frac{|T|l_{min} + i(l_{max} - l_{min}) - L}{i(l_{max} - l_{min})} \right\} = \begin{cases} 0, & i < \lfloor \frac{L - |T|l_{min}}{l_{max} - l_{min}} \rfloor \\ \frac{|T|l_{min} + i(l_{max} - l_{min}) - L}{i(l_{max} - l_{min})}, & i \geq \lfloor \frac{L - |T|l_{min}}{l_{max} - l_{min}} \rfloor \end{cases}$$

With the above probability, we have an $O(i)$ complexity for the **For** loop in the **If** statement. Similarly, the probability that the average distance exceeds the maximum allowed range is given by:

$$Prob[AverageDistance > l_{max}] = Max \left\{ 0, \int_{l_{max}}^{\frac{L - il_{min}}{|T| - i}} \frac{|T| - i}{i(l_{max} - l_{min})} dl \right\} = Max \left\{ 0, \frac{L + i(l_{max} - l_{min}) - |T|l_{max}}{i(l_{max} - l_{min})} \right\} = \begin{cases} 0, & i < \lfloor \frac{|T|l_{max} - L}{l_{max} - l_{min}} \rfloor \\ \frac{L + i(l_{max} - l_{min}) - |T|l_{max}}{i(l_{max} - l_{min})}, & i \geq \lfloor \frac{|T|l_{max} - L}{l_{max} - l_{min}} \rfloor \end{cases}$$

With the above probability, there is $O(i)$ complexity for the **For** loop in the **If** statement. Finally, the last **For**

loop is performed with a complexity of order $O(|T|-i)$ and the solution of greedy allocation heuristic with the updated distance vector as the input is achieved with order $O(|T|\cdot|V|)$. Assuming that each **If-Else** statement takes $O(1)$ time, putting all the aforementioned complexities together, we have the following overall time complexity for the ODAA algorithm:

$$\begin{aligned}
 T_{ODAA}(|V|, |T|) &= |V|\cdot|T| \\
 &+ \sum_{i=1}^{\lfloor \frac{L-|T|l_{min}}{l_{max}-l_{min}} \rfloor} \left\{ 3i + |T| - i + |T|\cdot|V| \right\} \\
 &+ \sum_{i=\lceil \frac{|T|l_{max}-L}{l_{max}-l_{min}} \rceil}^{\lfloor \frac{|T|l_{max}-L}{l_{max}-l_{min}} \rfloor} \left\{ 3i + \left(\frac{|T|l_{min} + i(l_{max} - l_{min}) - L}{i(l_{max} - l_{min})} \right)(i) \right. \\
 &\quad \left. + |T| - i + |T|\cdot|V| \right\} \\
 &+ \sum_{i=\lceil \frac{|T|l_{max}-L}{l_{max}-l_{min}} \rceil}^{|T|} \left\{ 3i + \frac{i(2i - |T|)(l_{max} - l_{min})}{i(l_{max} - l_{min})} \right. \\
 &\quad \left. + |T| - i + |T|\cdot|V| \right\} \in O(|T|^2\cdot|V|) \quad \square
 \end{aligned}$$

7 ENERGY HARVESTING THRESHOLD

One important factor which affects the network throughput is the amount of harvested energy by sensor nodes at each time interval. Although increasing the amount of harvested energy causes the increase in throughput, there is a threshold on the harvested energy such that the network throughput becomes saturated when the sensors harvest energy more than the threshold. The reason is due to the selection of at most one eligible sensor at each time slot by the mobile sink. The increase in harvested energy yields the increase in throughput until in all time slots, the energy consumption of the sensor node with maximum available data falls below its energy budget. Obviously, after this point, by increasing the harvested energy, no more data can be collected by the sink. Having the system parameters, we are interested here to find the above-mentioned threshold. We note that since from the data collection model, the sensors send data with fixed transmission rate and for whole of the period that cover the sink's trajectory on the path and also sink has large data buffer size, therefore, the parameters such as sink's buffer size or the data size in sensor's queue have no effect on the threshold on energy harvesting. However, the battery capacity of nodes as a realistic parameter is considered in the derivation of this threshold.

In the following derivations, we assume the uniform distributions for the amount of harvested energy by nodes, the transmission range, the transmission rate and initial energy level of nodes denoted by random variables h , R , r and I ,

respectively. For each of the above-mentioned variables, the corresponding minimum and maximum values are considered for the uniform distribution and, therefore, the middle point of min and max is the average value of that random variable. Furthermore, the variable C is defined as a random variable to indicate the energy consumption of an eligible node for data transmission to sink within a time interval. We note that since the amount of harvested energy by nodes changes within a uniform interval, therefore, we are motivated to find out the analytical threshold on the energy harvesting mean.

Theorem 4: *With identical battery capacity B for all sensor nodes, the threshold on energy harvesting mean is approximately given by:*

$$\text{Threshold}_{\bar{h}} \approx \begin{cases} \frac{3B+pK-3\bar{I}}{3+p}, & B \leq K \\ \frac{(3+p)K-3\bar{I}}{3+p}, & B > K \end{cases} \quad (33)$$

Where $\bar{I} = \frac{1}{2}(I_{min} + I_{max})$ is the average initial energy of nodes, constant $K = \frac{1}{2}\tau\cdot r_{max}\cdot R_{max}^\alpha$ and p is the probability that a sensor node can send its data to sink in two consecutive time slots.

Proof: We know that at the threshold point, the energy consumption of the eligible node falls below its energy budget at the corresponding time interval. To find the average energy budget of an eligible node within the time interval, we need to consider two different cases.

Case 1: the eligible node can send data to sink in only one time slot. In this case, its average energy budget before data transmission to sink is equal to $\bar{I} + \bar{h}$, the average of its initial energy plus mean of the harvested energy.

Case 2: the eligible node can send its data to sink at both time slots of time interval. Here, three sub-cases can happen assuming with the same probability $\frac{1}{3}$. **Case 2-1:** the node can send data in both first and second time slots of time interval. Now, the node is selected for data transmission in either one time slot (first or second) or in both of them assuming with the same probability of $\frac{1}{2}$. If selected at one slot, its average energy budget is same as Case 1. Otherwise, if selected at both time slots, its average energy budget is equal to average of initial energy plus the amount of harvested energy minus the consumption in the first slot which is $\bar{I} + \bar{h} - \bar{C}$. **Case 2-2:** the eligible node can send data to sink in the second slot of the previous interval and the first time slot of the current interval. In this case, since the node is eligible in the current interval, therefore, either it is selected in the previous interval and the first slot of the current interval or is only selected in the first slot of the current interval assuming with same probability $\frac{1}{2}$. If first case happens its average energy budget is equal to $\bar{I} + 2\bar{h} - \bar{C}$ and if the second one happens its energy budget is equal to $\bar{I} + 2\bar{h}$. **Case 2-3:** The node can send

its data in the second time slot of the current interval and the first slot of the next interval. In this case, since the node is eligible in the current interval, therefore, its average energy budget is equal to $\bar{I} + \bar{h}$.

With known value of the probability that a node can send its data to sink at two consecutive time slots, p , which has been analytically derived in [11], putting the above two cases together and considering that the energy budget of an eligible node can not be more than its battery capacity, the following inequality holds for an eligible node:

$$\begin{aligned} \bar{C} \leq & \min\{(1-p)(\bar{I} + \bar{h}) \\ & + p\left\{\frac{1}{3}\left\{\frac{1}{2}(\bar{I} + \bar{h}) + \frac{1}{2}(\bar{I} + \bar{h} - \bar{C})\right\} \right. \\ & + \frac{1}{3}\left\{\frac{1}{2}(\bar{I} + 2\bar{h}) + \frac{1}{2}(\bar{I} + 2\bar{h} - \bar{C})\right\} \\ & \left. + \frac{1}{3}(\bar{I} + \bar{h})\right\}, B\} \end{aligned} \quad (34)$$

To find the average energy consumption of an eligible node, we need to find out the probability density function (pdf) of transmission power which has a complicated term. Therefore, for the sake of simplicity, with minimum and maximum energy consumption respectively zero and $\tau.r_{max}.R_{max}^\alpha$, we use the approximate value $\frac{1}{2}\tau.r_{max}.R_{max}^\alpha$ for the average of energy consumption. Knowing that the energy consumption of nodes can not be more than the battery capacity, we get the following approximation for \bar{C} :

$$\bar{C} \approx \min\left\{\frac{1}{2}\tau.r_{max}.R_{max}^\alpha, B\right\} \quad (35)$$

Now, replacing \bar{C} in the both sides of inequality (34) by relation (35) and analyzing the four possibility conditional cases, it is seen that $\bar{h} \geq \frac{3B+pK-3\bar{I}}{3+p}$ for $B \leq K$ and $\bar{h} \geq \frac{(3+p)K-3\bar{I}}{3+p}$ for $B > K$ where constant $K = \frac{1}{2}\tau.r_{max}.R_{max}^\alpha$. Therefore, we get the following result for the threshold on energy harvesting mean:

$$Threshold_{\bar{h}} \approx \begin{cases} \frac{3B+pK-3\bar{I}}{3+p}, & B \leq K \\ \frac{(3+p)K-3\bar{I}}{3+p}, & B > K \end{cases} \quad (36)$$

7.1 Battery Capacity Threshold

Another important factor which affects the network throughput is the battery capacity of sensor nodes. The experiment shows that although the increase in battery capacity causes the increase in throughput but after a threshold point, the network throughput becomes saturated by increasing the capacity. The reason is that initially, the harvested energy by nodes is bounded by the low battery capacity and therefore increase in the capacity leads to the increase in throughput. But this increase is up to the point when the capacity of nodes has enough space to accommodate the initial and available energy for harvesting. Although after this threshold sensors have enough battery capacity, there is no more energy to

TABLE 3: The list of system parameters and their corresponding values used in the simulations.

System Parameter	Corresponding Value (ODSAA, ODAA, AdjustmentBased-Allocation)		
Number of Sensor Nodes	2000 ~ 8000		
Path Length	10 km		
Sink Speed	7.5 m/s		
Sensors Transmission Range	Uniform[10 m, 15 m]		
Sensors Transmission Rate	Uniform[60 $\frac{KB}{Sec}$, 80 $\frac{KB}{Sec}$]		
Energy Harvesting Distribution	Uniform[480 Joule, 520 Joule]		
Battery Capacity of Nodes	4500 Joule		
Probability of Battery Failure (P_f)	0.05		
	ODSAA	ODAA	AdjustmentBased-Allocation
Time Slot Period	Variable	Variable	2 Sec

be harvested and therefore the throughput becomes stable. we are here interested to find out theoretically the threshold on battery capacity when the mean of energy harvesting distribution is given.

In the following, we denote B as the identical battery capacity of all sensor nodes. As it is seen from comparing the average energy budget of nodes in both **Case 1** and **Case 2** in Theorem 4, for every eligible node $s_i \in S$, the battery capacity must be at least:

$$\begin{aligned} B \geq & \max\{I_i + h_i, I_i + h_i - C_i, I_i + 2h_i - C_i, \\ & I_i + 2h_i\} = I_i + 2h_i. \end{aligned} \quad (37)$$

Since the above inequality must be satisfied for all eligible sensor nodes, therefore, we get the following result for the threshold on battery capacity of nodes:

Theorem 5: Given the maximum initial energy of nodes, I_{max} , and the harvested energy, h_{max} , the threshold on the battery capacity of nodes is given by:

$$Threshold_B = I_{max} + 2h_{max} \quad (38)$$

8 EXPERIMENTAL RESULTS

In this section, we evaluate the performance of three data collection algorithms ODSAA, ODAA, proposed in this work, and AdjustmentBased-Allocation [11] in terms of network throughput, energy efficiency and computational complexity. The algorithms have been implemented with the programs written in Java.

8.1 Simulation Setup

In the simulations, for the system parameters, we assume a direct path with fixed length 10 km for the trajectory of the mobile sink. The whole duration of sink's trajectory on the path is divided into several consecutive time slots. From the network side, we assume that approximately 2000 to 8000 sensors are uniformly deployed along the both sides of the path. The nodes have the transmission range uniformly distributed within interval [10 m, 15 m] and a uniform transmission rate [60 KB/Sec, 80 KB/Sec].

The performance of three algorithms are evaluated

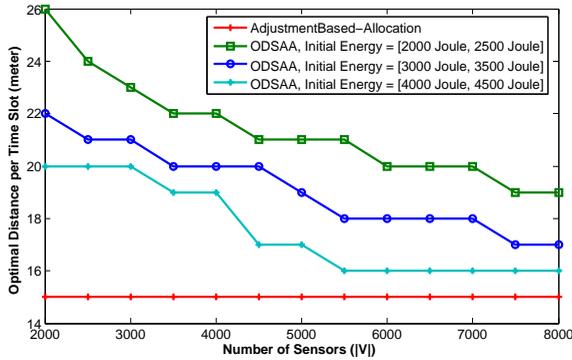


Fig. 5: The comparison between ODSAA and AdjustmentBased-Allocation algorithms in term of optimal distance per time slot.

considering the following model parameters: Since in the first and second scenarios of NTM problem the sink maintains a constant speed during whole of its trajectory on the path, therefore, the constant speed $v_m = 7.5m/s$ is considered for the mobile sink in the implementation of both ODSAA and ODA A algorithms. On the other side, the same sink speed, the constant time slot period 2 Seconds and fixed distance per time slot 15m are considered in the implementation of AdjustmentBased-Allocation algorithm. However, the time slot period is changed at different time slots depending on the optimal distance at each time slot under the both first and second scenarios of NTM problem.

For the energy harvesting parameters, the uniform distribution is used in the implementation of algorithms to express the stochastic amount of harvested energy by nodes at the beginning of each interval. The ideal battery capacity 4500 Joule and the probability of battery failure 0.05 are considered in the simulations. The unit of network throughput, the total collected data by the sink, is also KB. The list of system parameters and their corresponding values has been summarized in Table 3.

8.2 Performance Evaluation of Algorithms

As the first simulation, we are interested to find the optimal distance per time slot under different number of nodes using the ODSAA algorithm. Recall that the at the optimal distance, the maximum data throughput is achieved. With three initial energy levels and uniform energy harvesting [480Joule, 520Joule] with mean $\bar{h} = 500Joule$, Fig. 5 illustrates the distance per time slot achieved by the ODSAA algorithm when the number of sensors varies from 2000 to 8000 nodes. The following observations are seen: First, by increasing the number of nodes, the optimal distance per time slot returned by ODSAA decreases. The reason is that after some increase in throughput, with higher ratio the nodes loose their contribution in data transmission to sink

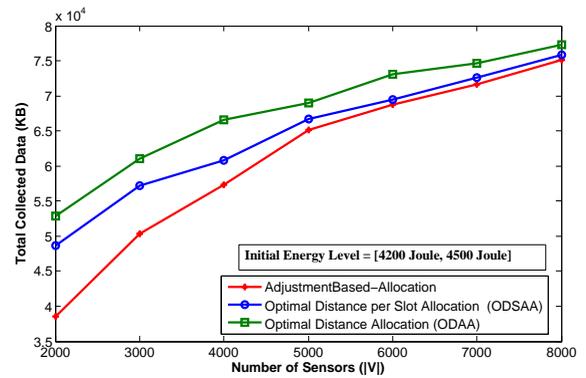


Fig. 6: The comparison between three algorithms in term of network throughput.

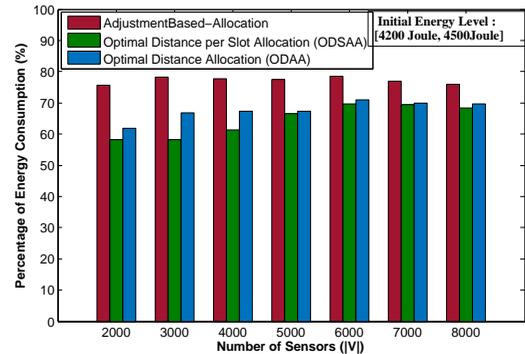


Fig. 7: The comparison between three algorithms in term of energy efficiency.

when they are densely deployed compared to the sparse deployment. Therefore, the optimal distance is reached earlier in the case of dense deployment.

Second, by increasing the initial energy, the optimal distance has an overall decreasing trend. The reason is that since the sensor nodes have enough energy for data transmission to sink in most of the cases when with high initial energy, therefore, with faster ratio they loose their contribution in data transmission to sink compared to the case when their initial energy level is low. As another observation, the distance for three initial energy levels remain constant for the number of nodes between 6000 and 7000. The reason is that in dense deployment, the ratio of losing contribution in data transmission to sink remains same with larger difference in the number of nodes compared to sparse deployment.

In the next simulation, we have compared the performance of three algorithms ODSAA, ODA A and AdjustmentBased-Allocation in term of network throughput in one round of data collection by sink. For different number of nodes varying from 2000 to 8000, the initial energy level [4200Joule, 4500Joule] and uniform distribution [480Joule, 520Joule], the result of throughput comparison has been shown in Fig. 6. As the results show, using ODSAA algorithm, the network throughput significantly increases

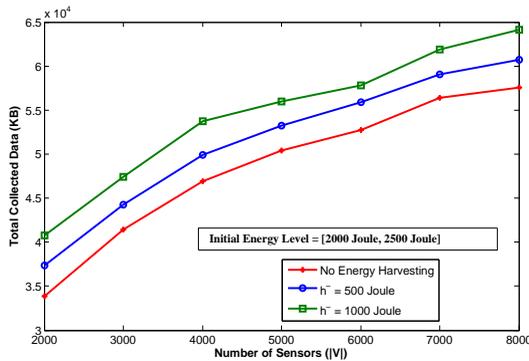


Fig. 8: The effect of increase in harvested energy on network throughput using ODSAA algorithm.

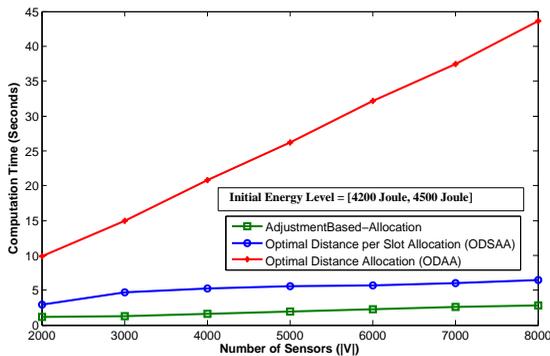


Fig. 9: The comparison between three algorithms in term of total computation time.

compared to AdjustmentBased-Allocation with an average improvement of approximately $10^3 KB$. As another observation, for each number of nodes, further improvement in network throughput is achieved by updating the distance at each time slot based on the available data at sink's vicinity using ODAA algorithm.

In the next simulation, we have evaluated the performance of three algorithms in term of energy efficiency. Denoted by E_c the total consumed energy by eligible nodes at all time slots in one round of data collection and E_b , their total energy budget, we define the energy efficiency criteria based on the total percentage of consumed energy obtained by $\frac{E_c}{E_b} \times 100\%$. As the result in Fig. 7 shows, the ODSAA algorithm has the best while AdjustmentBased-Allocation achieves the worst performance in term of energy efficiency. The reason is that AdjustmentBased-Allocation algorithm considers a fixed initial distance per time slot (fixed transmission period) for nodes which can lead to more energy consumption for nodes in some of the time slots.

In Fig. 8, we have shown the effect of increase in the amount of harvested energy by nodes on the network throughput using ODSAA algorithm. As we can see, the throughput increases when there are more energy for harvesting since as long as the nodes harvest more energy, they have more opportunities

TABLE 4: The comparison between experimental and theoretical thresholds on energy harvesting mean for different number of nodes ($B = 2 \times 10^4$ Joule).

Number of Sensor Nodes ($ V $)	Experimental Threshold (Joule)	Theoretical Threshold (Joule)
2000	17000	16293
3000	17500	16293
4000	17000	16293
5000	17500	16293
6000	17500	16293
7000	17500	16293
8000	17000	16293

TABLE 5: The comparison between experimental and theoretical thresholds on energy harvesting mean for different battery capacities ($|V| = 3000$).

Battery Capacity (Joule)	Experimental Threshold (Joule)	Theoretical Threshold (Joule)
5×10^3	5×10^3	6.4308×10^3
1×10^4	1×10^4	1.0224×10^4
1.5×10^4	1.5×10^4	1.4017×10^4
2×10^4	1.75×10^4	1.6293×10^4
2.5×10^4	1.7×10^4	1.6293×10^4
3×10^4	1.75×10^4	1.6293×10^4

for data transmission to the sink.

Since the energy harvesting means are considered for different rounds of sink path traversal, the observation from the result in Fig. 8 concludes that the nodes with more available data which have not enough energy in one round, find the opportunity for data transmission in subsequent round when they harvest more energy. This observation can in turn cover the fairness issue remarked in Section 3.

We have further compared the performance of ODSAA and ODAA with respect to AdjustmentBased-Allocation in term of computational complexity with the result illustrated in Fig. 9. The computational complexity of each algorithm is measured as the total time taken by the algorithm to find the suboptimal solutions. The unit of the time is considered as Seconds. As we can see, ODSAA and AdjustmentBased-Allocation algorithms have the same order of growth in computation time while the complexity of ODAA algorithm dramatically increases, therefore, confirming the statements of Theorems 2 and 3. Therefore, regarding the effectiveness of the proposed algorithms, with the same order of complexity, according to the result shown in Fig. 6, the significant improvement in network throughput is achieved using ODSAA algorithm. Also, further throughput can be obtained using ODAA with polynomial increase in computation time.

8.3 Energy Harvesting Thresholds

In this subsection, we evaluate the theoretical thresholds derived in Section 7 through the comparison with the experimental thresholds. With uniform

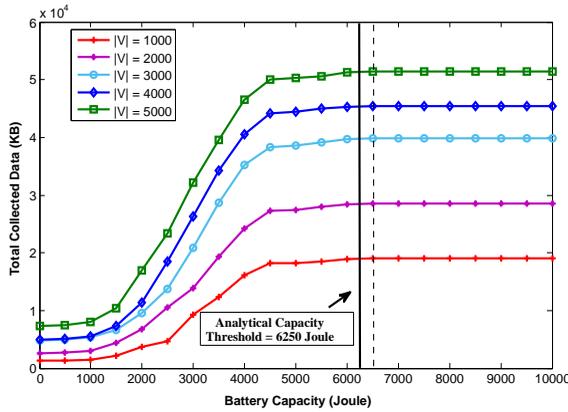


Fig. 10: The effect of increase in battery capacity of nodes on network throughput with fixed energy harvesting mean and different number of nodes.

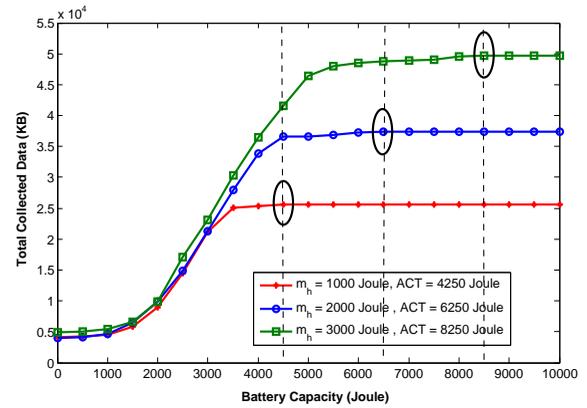


Fig. 11: The effect of increase in battery capacity of nodes on network throughput for different energy harvesting means and fixed number of nodes.

energy harvesting and initial energy from interval $[2000\text{Joule}, 2500\text{Joule}]$ for sensor nodes, the rest of the parameters have the same values as given in Table 3.

As the first simulation, we are interested to see the effect of increase in energy harvesting mean on network throughput in one round of data collection. With fixed variance of distribution and battery capacity $2 \times 10^4 \text{Joule}$, we have increased the mean of energy harvesting (\bar{h}) from 0 to $2 \times 10^4 \text{Joule}$. For different number of nodes, the experimental threshold has been obtained with 50 trails of simulation. As the results of comparison in Table 4 shows, with average energy consumption of approximately $\bar{C} \approx 1.8 \times 10^4 \text{Joule}$, the theoretical threshold derived from the relation (33) provides a confidence interval of at least 90% for the average experimental threshold for each number of nodes.

As the next simulation, with fixed number of sensors at 3000 nodes, the experimental threshold on energy harvesting mean has been derived considering different battery capacities ranging from $5 \times 10^3 \text{Joule}$ to $3 \times 10^4 \text{Joule}$. As we can see from the results illustrated in Table 5, with the same average energy consumption as the previous part, the theoretical threshold from equation (33) provides a confidence interval of in average 90% for the experimental threshold for each battery capacity.

As the next simulation, we have investigated the effect of increase in battery capacity of nodes on the network throughput in one round of data collection by sink. For the simulation, we have considered the mean of energy harvesting to justify the theoretical derivation in relation (38) in average sense. With fixed $\bar{h} = 2000 \text{Joule}$ and considering five different number of nodes, the battery capacity is increased from 0 to 10000 Joule. For each number of nodes, the threshold on battery capacity, where for the first time the throughput becomes saturated, has been shown with dashed line in Fig. 10. The theoretical threshold shown in relation (38) with average initial and harvested

energy has been also drawn with bold line. As the comparison shows, for different number of nodes, the theoretical threshold provides a confidence interval of almost 95%.

We have further simulated the effect of increase in battery capacity of nodes when with fixed number of sensors at 3000 nodes, three different values for \bar{h} are considered. As the results illustrated in Fig. 11 show, for each energy harvesting mean, the theoretical threshold (ACT) provides a confidence interval of 95% for the experimental threshold.

9 CONCLUSIVE REMARKS

Toward the network throughput improvement using a mobile sink in energy harvesting wireless sensor networks (EH-WSNs), we propose a general framework for network throughput maximization (NTM) problem. By optimizing feasible problem parameters, the NTM problem is investigated under different scenarios by formulating each scenario as a mixed integer linear programming (MILP) optimization model subject to some scenario-specific constraints. Due to the NP-Hardness of the MILP problems, we design two efficient algorithms namely as ODSAA and ODAAs for two practically feasible scenarios which run centrally by the sink to find the sub-optimal solutions in a reasonable computation time. Furthermore, under the uniform energy harvesting distribution, we establish two theoretical thresholds on energy harvesting mean and battery capacity of nodes such that by increasing either one after the threshold, the throughput becomes stable. The results of simulations conducted on different set of node deployments confirm the significant improvement in network throughput by optimizing the feasible parameters. Furthermore, the results of simulation is an evidence that theoretical thresholds provide a confidence interval of 90%. As an advantage, these thresholds can help the designers of sink based EH-WSNs to optimize efficiently the

cost of energy harvesting resources in practical data collection scenarios.

ACKNOWLEDGMENTS

This research was a part of the project titled Domestic products development of marine survey and ocean exploration equipments, funded by the Ministry of Oceans and Fisheries, and by GIST Research Institute (GRI) in 2016. The authors would like to thank the anonymous referees for their valuable comments which helped to improve the quality of the paper.

REFERENCES

- [1] M. D. Francesco, S. K. Das, and G. Anastasi, "Data collection in wireless sensor networks with mobile elements: a survey", *ACM Trans. on Sensor Netw.*, vol. 8, no. 1, pp. 1-34, Aug. 2011.
- [2] Yimin Chen, Long Cheng, Canfeng Chen, and Jian Ma, "Wireless sensor network for data sensing in intelligent transportation system", in *Proc. 69th IEEE Veh. Tech. Conf.*, pp. 1-5, Apr. 2009.
- [3] E. B. Hamida, and G. Chelius, "A line-based data dissemination protocol for wireless sensor networks with mobile sink", in *Proc. IEEE Int. Conf. on Commun. (ICC)*, pp. 2201-2205, May 2008.
- [4] S. Gao, H. Zhang, and S. K. Das, "Efficient data collection in wireless sensor networks with path-constrained mobile sinks", *IEEE Trans. on Mobile Comput.*, vol.10, no. 4, pp. 592-608, Apr. 2011.
- [5] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, "Introduction to Algorithms", *MIT Press, 3rd Edition*, 2009.
- [6] J. Chen, S. He, and Y. Sun, "Rechargeable Sensor Networks: Technology, Theory, and Application", *World Scientific Publishing*, 2014.
- [7] X. Ren, and W. Liang, "The use of a mobile sink for quality data collection in energy harvesting sensor networks," in *Proc. of IEEE Wireless Commun. and Netw. Conference (WCNC)*, pp. 1145-1150, Apr. 2013.
- [8] X. Ren, W. Liang, and W. Xu, "Use of a mobile sink for maximizing data collection in energy harvesting sensor networks," in *Proc. of IEEE 42nd Int. Conf. on Parallel Processing*, pp. 439-448, Oct. 2013.
- [9] M. Garey, and D. Johnson, "Computers and Intractability: A Guide to the Theory of NP-Completeness", *Bell Laboratories Incorporated*, 1979.
- [10] D. Gunduz, K. Stamatiou, N. Michelusi, and M. Zorzi, "Designing Intelligent Energy Harvesting Communication Systems", *IEEE Communications Magazine*, vol. 52, no. 1, pp. 210-216, Jan. 2014.
- [11] A. Mehrabi, and K. Kim, "Maximizing data collection throughput on a path in energy harvesting sensor networks using a mobile sink", *IEEE Trans. on Mobile Comput.*, vol. 15, no. 3, pp. 690-704, Mar. 2016.
- [12] R. S. Liu, K. W. Fan, Z. Zheng and P. Sinha. "Perpetual and fair data collection for environmental energy harvesting sensor networks", *IEEE/ACM Trans. on Netw.*, vol. 19, no. 4, Aug. 2011.
- [13] 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.
- [14] X. Ren and W. Liang, "Delay-tolerant data gathering in energy harvesting sensor networks with a mobile sink," in *Proc. of IEEE GLOBECOM Conference*, pp. 93-99, Dec. 2012.
- [15] A. Kinalis, S. Nikolettseas, D. Patroumpa, and J. Rolim, "Biased sink mobility with adaptive stop times for low latency data collection in sensor networks", *Elsevier Information Fusion*, vol 15, Jan. 2014.
- [16] X. Xu and W. Liang, "Monitoring quality optimization in wireless sensor networks with a mobile sink," in *Proc. 14th ACM Int. Conf. Modeling, Analysis Simulation of Wireless and Mobile Systems*, pp. 77-84, Oct. 2011.

- [17] Y. Yun and Y. Xia, "Maximizing the Lifetime of wireless sensor networks with mobile sink in delay-tolerant applications", *IEEE Trans. Mobile Comput.*, vol. 9, no. 9, pp. 1308-1318, Sep. 2010.
- [18] P. L. Zhi Ang Eu, M. Han and H. P. Tan, "Empirical modeling of a solar-powered energy harvesting wireless sensor node for time-slotted operation", in *Proc. of IEEE Wireless Commun. and Netw. Conference (WCNC)*, pp. 179-184, Mar. 2011.
- [19] S. Zhang and A. Seyedi, "Statistical models for harvested power from human motion", in *Proc. of IEEE International Workshop on Energy Harvesting for Communication*, pp. 5777-5781, Jun. 2012.
- [20] V. Raghunathan, A. Kansal, J. Hsu, J. K. Friedman, and M. B. Srivastava, "Design considerations for solar energy harvesting wireless embedded systems", in *Proc. IEEE International Conf. on Inf. Proc. in Sensor Netw. (IPSN)*, pp. 457-462, Apr. 2005.
- [21] D.P. Williamson, and D. B. Shmoys, "The Design of Approximation Algorithms", *Cambridge University Press*, 2010.
- [22] Y. Zhang, S. He, J. Chen, "Data gathering optimization by dynamic sensing and routing in rechargeable sensor networks", *IEEE/ACM Trans. on Netw.*, vol. 24, no. 3, pp. 1632-1646, Jun. 2016.
- [23] S. Chen, P. Sinha, N. B. Shroff, and C. Joo, "A simple asymptotically optimal joint energy allocation and routing schema in rechargeable sensor networks", *IEEE/ACM Trans. on Netw.*, vol 22, no. 4, pp. 1325-1336, Aug. 2014.
- [24] V. Schrijver, "Theory of linear and integer programming", *Wiley Publications*, 1998.
- [25] Y. T. Hou, Y. Shi, and H. D. Sherali, "Applied Optimization Methods for Wireless Sensor Networks", *Cambridge University*, 2014.
- [26] P. Blasco, D. Gunduz, M. Dohler, "A learning theoretic approach to energy harvesting communication system optimization", *IEEE Trans. on Wireless Commun.*, vol. 12, no. 4, pp. 1872-1882, Apr. 2013.



Abbas Mehrabi received the BSc and MSc degrees in computer engineering in 2007 and 2010, respectively, from Shahid Bahonar University of Kerman and Azad University, South Tehran, Iran, respectively. He has been currently working toward the PhD program since March 2013 in the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea. His main research interests include mobile computing,

scheduling and planning problems in smart grid and transportation systems and graph theory. He is a student member of the IEEE Computer Society and the IEEE.



Kiseon Kim received the BEng and MEng degrees in electronics engineering from Seoul National University, Korea, in 1978 and 1980, respectively, and the PhD degree in electrical engineering from the University of Southern California, Los Angeles in 1987. From 1988 to 1991, he was with Schlumberger, Houston, Texas. From 1991 to 1994, he was with the Superconducting Super Collider Lab, Texas. He joined Gwangju Institute of Science and Technology (GIST), Korea,

in 1994, where he is currently a full professor at the School of Information and Communications. His current research interests include the design and analysis of wideband digital communications systems, and the analysis and implementation of sensor networks at both physical and the resource management layers. He is a senior member of the IEEE Communications Society and the IEEE.