Data Aggregation, Gathering and Gossiping in Duty-Cycled Wireless Sensor Networks subject to Physical Interference

Lixin Wang[*,1], Jianhua Yang[1], Sean Gill[1], Xiaohua Xu[2]

[1]*TSYS School of Computer Science, Columbus State University, Columbus, GA 31907, USA*

[2]*College of Computing and Software Engineering, Kennesaw State University, Marietta, GA 30060, USA*

A B S T R A C T

Data aggregation, gathering and gossiping are all vital communication tasks in wireless sensor networks (WSNs). When all networking devices are always active, scheduling algorithms for these communication tasks have been extensively investigated under both the protocol and physical interference models. However, wireless devices usually switch between the sleep state and the active state for the purpose of energy saving. A networking device with duty-cycled scenarios having sleep/awake cycles may need to transmit the message to all neighbors more than once. Taking the duty-cycled scenarios into consideration, communication scheduling algorithms for these tasks have been extensively investigated under the protocol interference model. As far as we know, scheduling algorithms for these communication tasks have not yet been investigated in WSNs with duty-cycled scenarios under the physical interference model. In this paper, we propose minimum latency scheduling algorithms for these communication tasks in duty-cycled WSNs under the physical interference model. Our innovative scheduling algorithms for both data gathering and gossiping achieve approximation ratios at most a constant time of $|T|$, where $|T|$ is the length of a scheduling period. The approximation ratio of our proposed data aggregation scheduling algorithm is less than or equal to a constant times $|T|$ with bounded maximum degree of the network.

1   Introduction

This paper is an extension of our work originally presented on the 15th IEEE International Conference on Mobile Ad-hoc and Sensor Networks (IEEE MSN 2019) [1].

Wireless sensor networks (WSNs) have been widely employed in automated battlefields, emergency disaster relief, environmental monitoring, real-time pollution monitoring, military surveillance, etc. Most of these WSN applications are time critical. The definitions of data gossiping, data gathering, and data aggregation can be found in [2]. The concepts of a link schedule, data aggregation schedule, data gathering schedule and data gossiping schedule can also be found in [2].

The latency of a data aggregation schedule is the total number of time-slots needed for the sink node to receive the aggregated message from all other networking nodes. Minimum-Latency Aggregation Schedule (**MLAS**) is the research problem for developing a scheduling algorithm for data aggregation in a multi-hop WSNs with the smallest latency. Similarly, the latency of a data gathering

schedule is the total number of time-slots required for the sink node to receive a packet from every other networking node; the latency of a gossiping schedule is the total number of time-slots needed for every device to receive the packets from all other networking devices. Minimum-Latency Gathering Schedule (**MLGS**) is the research problem for developing a scheduling algorithm for data gathering in a multi-hop WSNs with the smallest latency. Minimum-Latency Gossiping Schedule (**MLGoS**) is the research problem for developing a scheduling algorithm for data gossiping in a multi-hop WSNs with the smallest latency. A great number of communication scheduling algorithms have been developed for MLAS, MLGS and MLGoS in multi-hop WSNs under either the physical interference model [2], or the protocol interference model [3]–[13] and the references therein.

However, wireless devices usually switch between the sleep state and the active state for the purpose of energy saving. A networking device with duty-cycled scenarios having sleep/active cycles may need several times to transmit the message to all neighbors. The duty-cycled scenarios is a popular energy-saving method for multi-

Corresponding Author: Lixin Wang, TSYS School of Computer Science, Columbus State University, Columbus, GA 31907, USA. Email: wang_lixin@columbusstate.edu

hop WSNs. The GreenOrbs projects [14] and the VigilNet projects [15] are two actual applications of duty-cycled WSNs (DC-WSNs). Thus, all the existing approximation solutions for MLAS, MLGS, or MLGoS without duty-cycled scenarios are not working any more on DC-WSNs.

We use the same duty cycle model in this paper as the one we used in our short conference version [1]. The problem MLAS on duty-cycled multi-hop WSNs is represented by **MLAS-DC**. Similarly, MLGS and MLGoS on duty-cycled multi-hop WSNs are represented by **MLGS-DC** and **MLGoS-DC**, respectively. With duty-cycled scenarios, all three problems MLAS-DC, MLGS-DC and MLGoS-DC have been extensively investigated subject to the protocol interference and its simple variations [5], [16]–[24]. To the best of our knowledge, all three duty-cycled scheduling problems have not yet been investigated on duty-cycled multi-hop WSNs subject to the physical interference.

In this paper, we develop short communication schedules for the above three duty-cycled scheduling problems on multi-hop WSNs under the physical interference model which is a realistic model that handles wireless interferences more accurately. Under such a realistic model, a receiver can successfully receive the message if and only if the *Signal-to-Interference-plus-Noise-Ratio* (*SINR*) of the receiver is greater than a threshold value. According to the way the SINR ratio is computed, the SINR value is based on those transmissions are that will be scheduled concurrently in every time slot. Therefore, constructing a conflict graph to model the wireless interference is challenging. Such a nature of the physical interference makes it very hard to analyze the proposed communication scheduling algorithms subject to physical interference. Let $\|uv\|$ represent the Euclidean distance between the two devices u and v. Let σ represent the threshold value for a receiving device can successfully receive the wanted signal subject to physical interference, and ξ power strength of the background noise. Suppose that all the devices use a uniform and fixed power P for transmission. The path loss is represented by a positive reference loss parameter η, and a constant path-loss exponent κ (a value between 2 and 6). Assume that the communication radius of each device is normalized to one. If a device u sends a signal using a transmission power P, the power strength of this signal received by the receiving device v is $\eta P \|uv\|^{-\kappa}$.

By far, [25] is the only work that studied **MLAS-DC** for multi-hop WSNs subject to physical interference. When the SINR ratio of a node u is calculated in EQ(1) of [25], the received power $P_r(u)$ at u is incorrectly calculated in EQ(2) of [25] as it will be greater than the transmission power when the distance $\|uv\| < 1$. When $\|uv\| \geq 1$, the received power $P_r(u)$ is simply set to be the transmission power in [25]. Similarly, the total interference power at u is also incorrectly calculated in EQ(3) of [25]. Therefore, radio signal attenuation is not taken into consideration and the algorithms proposed in [25] for **MLAS-DC** are invalid under the physical interference model. As far as we know, there are no existing algorithms proposed for **MLGS-DC** or **MLGoS-DC** for multi-hop WSNs subject to the physical interference constraint.

Let $|T|$ denote the total number of time-slots in a scheduling period. Below are this paper's main contributions:

1. The approximation ratio of the algorithm we proposed for

MLGS-DC is less than or equal to a constant times $|T|$;

2. The approximation ratio of the algorithm proposed for MLGoS-DC is less than or equal to a constant times $|T|$;

3. The approximation ratio of the algorithm proposed for MLAS-DC is less than or equal to a constant times $|T|$ when the network has bounded maximum degree.

All the notations used in this paper are listed in Table 1 for easy referencing.

The remaining of this paper is organized as follows. The related work for existing approximation scheduling algorithms for data aggregation, gathering and gossiping are given in Section 2. Section 3 gives some basic concepts and preliminary knowledge that are needed for the design of the scheduling algorithms. In Section 4, we create a tree rooted at the sink that will be used for routing. In Section 5, we present an approximation solution to MLAS-DC. In Section 2, we propose an approximation solution to MLGS-DC. An approximation solution to MLGoS-DC is developed in Section 7. Finally in Section 8, a conclusion for this paper and some future research are presented.

2 Related Works

First, let us review the related work for known algorithms that were proposed for MLAS in multi-hop WSNs when all the networking nodes are implicitly assumed to be always active [2]–[4], [7], [9]. If all the networking devices have uniform transmission radii normalized to one on a multi-hop WSN, the network topology forms a UDG (unit-disk graph). For convenience, the number of networking nodes is denoted by n, the radius of the network graph w.r.t. the sink node of data aggregation or gathering is denoted by R. For arbitrary interference radius, any optimal solution to MLAS uses at least R time-slots. So is $\log n$. [3] and [7] proposed two scheduling algorithms for data aggregation with the total latency less than or equal to $(\Delta - 1)R$ and $23R + \Delta - 18$, respectively, where Δ is the maximum degree when the transmission radius is the same as the interference radius. The single-level aggregation was investigated in [4] that developed aggregation scheduling algorithm. When the transmission radius is the same as the interference radius, three approximation algorithms were proposed in [9] with total latency less than or equal to $15R + \Delta - 4$, $2R + \Delta + O(\log R)$ and $\left(1 + O\left(\log R / \sqrt[3]{R}\right)\right) R + \Delta$, respectively. These existing scheduling algorithms for MLAS mentioned above were proposed subject to the protocol interference, and the networking devices are assumed to be always awake.

Subject to the physical interference or the protocol interference, MLGS in multi-hop WSNs has also been extensively investigated when all the networking devices are always awake [2], [10], [11], [13], [26]. Let us review the approximatino solutions developed subject to the protocol interference first. The NP-hardness of MLGS was verified by in [10] that developed a scheduling solution for MLGS that has approximation ratio at most four with the assumption that the communication topology could be reduced to a UDG. In [11], the author proposed a greedy approach for MLGS and proved it to achieve approximation ratio at most 4 in a general setting and achieve approximation ratio at most 3 when

Table 1: All notations used in this paper:

P	the uniform transmission power of every node
η	the parameter of the reference loss
κ	the path-loss exponent
$\|uv\|$	the Euclidean distance of the two nodes u and v
σ	the threshold ratio of the SINR model
ξ	the background noise in the SINR model
T	a scheduling period
$|T|$	the count of the time-slots in a scheduling period T
V	the set of all networking nodes
Δ	the maximum degree of the communication topology of the network
G_r	r-disk graph on V
SPT	a shortest path tree over V
$SPAN$	a spanning tree over V
$\zeta(x)$	the Riemann zeta function
$A(v)$	the active time slot of a node v
s	the sink node of aggregation and gathering
M	the largest latency of the shortest paths from all other nodes to node s in SPT
U_j	the set of the awake nodes in time-slot j
L_i	the subset of the nodes in V from which the latency of the shortest path to node s is i
I_j	the maximum independent set in U_j
I_{ji}	is defined to be $I_j \cap L_i$
$Lat(u,v)$	the Latency of the link (u,v)

the communication topology could be reduced to a UDG. When the transmission radius is the same as the interference radius, a scheduling solution for MLGS was proposed in [26] that has approximation ratio at most $(1 + 1/(k + 1))$ with k being a whole number. In [13], the author developed a scheduling solution for MLGS with the shortest gathering schedule in which several channels were used to speed up communications and make the scheduling solu-tions more efficient. For MLGS, when the interference radii r_i are uniform and no less than the transmission radius, an efficient scheduling solution was proposed in [13] having approximation ratio less than or equal to $2 \beta_{r_i}/\lambda$ with λ being the count of the channels that can be used and β_{r_i} the largest number of points in a half-disk of radius $r_i + 1$ with the distance between any two of them is greater than 1.

For MLGoS on multi-hop WSNs, under the UDG model, the authors of [8] designed a gossiping scheme that achieved approximation ratio 27, improved previously known algorithms which have approximation ratio 1000+. In [6], the author considered the problem MLGoS and presented respectively two algorithms with approximation guarantees of 20 and 34, thereby improving the approximation ratio of 27 presented in [8].

When all the devices are always active, [2] developed the short-est communication schedules for MLAS, MLGS and MLGoS sub-ject to the physical interference. [2] developed scheduling algo-rithms for the four important group communication tasks under the physical interference model. Both the gathering schedule and gos-siping schedule developed in [2] achieved constant approximation ratios. For MLGS, [2] proposed an efficient scheduling algorithm with approximation bound less than or equal to $2\beta_\rho$, where ρ and β_ρ are defined in EQ(1) and EQ(3), respectively. For MLGoS, [2] developed a gossiping solution with approximation bound less than

or equal to $5\beta_\rho$. Under mild assumptions, the broadcast schedule and data aggregation schedule developed in [2] could also achieve constant approximation ratios.

However, none of the known research papers described above have considered the situation that some networking devices may be in an inactive state. Taking duty-cycled scenarios into consideration under a graph-based interference model, the problem MLAS-DC has been extensively investigated [5], [17], [20]–[23], [27], [28].

Both [5] and [17] developed short data aggregation schedules for MLAS-DC. Both [20] and [21] studied the MLAS-DC prob-lem in order to minimize the communication delay and maximize sensor nodes' lifetimes, and propsed another distributed algorithm for MLAS-DC. All the above mentioned scheduling algorithms for MLAS-DC were proposed under the simple protocol interference model. A data aggregation scheduling algorithm was developed in [22] for MLAS-DC with minimum latency for the purpose of energy efficiency. [23] studied the MLAS-DC problem with no structures taken into consideration. A distributed aggregation algorithm was proposed for MLAS-DC under the protocol interference model, in which the routing tree used for aggregation and a data aggregation schedule without collision were produced at the same time by the algorithm. In [27], the author investigated the MLAS-DC and devel-oped two distributed aggregation algorithms for MLAS-DC, where the routing tree used for data aggregation and an aggregation sched-ule were produced at the same time. In [27], the author developed an approximation algorithm for MLAS-DC with better approxima-tion bound. When the networking nodes have changable radii, a aggregation schedule with less number of time slots was developed for MLAS-DC in [28] by constructing a nonlinear mathematical formula.

For MLGS-DC subject to protocol interference, [24] proposed

a data gathering scheduling algorithm that achieves an approximation bound of $10|T|$. For the problem MLGoS-DC, [18] proposed a solution called the UTB algorithm for the all-to-all broadcast in which every message must be sent separately without any combination or aggregation and a more efficient version UTB-P of this algorithm using the Prune method. The approximation bound of these two algorithms for MLGoS-DC are less than or equal to $17|T| + 20$. With the unit-size message model, [24] investigated MLGoS-DC and proposed a scheduling algorithm for gossiping that achieved approximation bound less than or equal to $20|T|$. The paper [19] studied the problem MLGoS-DC in uncoordinated duty-cycled multi-hop wireless networks with unbounded-size messages. The NP-hardness of MLGoS-DC was verified in [19]. This paper also developed two approximation algorithms for MLGoS-DC with approximation bounds less than or equal to $3\Phi^2(\Delta + 6)|T|$, with $\Phi = \lceil 2(\Psi + 2)/3 \rceil$, and Δ the maximum degree of the network. The value of Ψ is equal to the ratio obtained via dividing the interference radius by the transmission radius of each device.

3   Preliminaries

In this section, we introduce some notations, parameters and concepts that are needed for the design of our scheduling algorithms to be proposed in this paper. Set

$$R = \left(\frac{\eta P}{\sigma \xi} \right)^{1/\kappa}.$$

Given a pair of devices u and v. Clearly, there is a link between u and v in the communication topology when there is no interference from other nodes that are transmitting simultaneously if and only if $\|uv\| \leq R$. We call R the *maximum transmission radius*.

Suppose that r is a positive parameter. Denote by R' be the largest edge length of a Euclidean minimum spanning tree over V, the set of all the networking nodes. Clearly, R' equals the minimum value of r satisfying that the r-disk graph over V is connected. The relation $R' \leq R$ is needed for the network to be connected. An additional assumption we need is that $R' \leq (1 - \varepsilon)R$ which holds for certain small positive constant ε.

Denote by Γ the collection of all Euclidean distances of the networking nodes in V less than R, and greater than or equal to R'. That is, we have

$$\Gamma = \{ \|uv\| : u, v \in V \text{ and } R' \leq \|uv\| < R \} .$$

Through a straightforward calculation, we have $1 \leq |\Gamma| \leq n(n-1)/2$ with n being the number of nodes.

Under the physical interference model, a collect I of networking nodes is referred to as *r-independent* when the following two conditions hold

(1) if every node in I sends a message at the same time, each such a transmission could be correctly received by every node whose distance from the transmitter is no more than r;

(2) the distance between any two nodes in I is larger than r.

Now we introduce the well-known Riemann zeta function denoted by $\zeta(x)$:

$$\zeta(x) = \sum_{j=1}^{\infty} \frac{1}{j^x}.$$

For each $r \in [R', R)$, define the following parameter ρ that is needed for the design of our scheduling algorithms to be proposed in later sections and its value is clearly explained in [2].

$$\rho = 1 + \left(\frac{\sigma \left(16\zeta(\kappa - 1) + 8\zeta(\kappa) - 6 \right)}{1 - (r/R)^\kappa} \right)^{1/\kappa}. \tag{1}$$

Now we introduce the concept of independence for a given collection of links under the physical interference model. A collection C of links is called *independent* subject to the physical interference if the following two conditions hold:

(1) when the senders of all the communication links in C transmit at the same time, the receivers of every communication link in C can correctly receive the message; (2) all the communication links in C are disjoint.

4   Construction of A Routing Tree

A spanning tree $SPAN$ with the root node $s \in V$ is created in this section. This spanning tree will be employed in routing of the group communication tasks studied in this paper. Let $r > 0$. Denote by G_r a connected and undirected r-disk graph over V.

To begin with, a directed graph G'_r is created based on G_r as follows: replaced every edge (u, v) of G_r by two directed links (u, v) and (v, u) on G'_r, and then remove all links leaving s from G'_r. For a vertex $u \in V$, let $A(u) \in T$ be the awake time slot of u. For a link $(u, v) \in E(G'_r)$, it is assumed that u is the sender and v is the receiver. The latency of (u, v) is defined as follows:

$$Lat(u, v) = \begin{cases} A(v) - A(u), & if \ A(u) < A(v); \\ A(v) - A(u) + |T|, & if \ A(u) \geq A(v). \end{cases} \tag{2}$$

We associate each link (u, v) of G'_r with the weight equal to its latency $Lat(u, v)$, then G'_r is a directed weighted graph with all link weights non-negative. Therefore, a shortest path tree SPT of G'_r rooted at s can be computed by applying the Dijkstra's algorithm on G'_r. The tree SPT consists of all the minimum-weight paths from all nodes other than s to s. The largest latency among all the shortest paths in SPT from all other nodes to sink s is represented by M. Clearly, any optimal solution to either MLAS-DC or MLGS-DC is greater than or equal to M.

Let us take the undirected graph G_r with eight nodes in Fig. 1 as an example to explain the constructions of the directed graph G'_r and the shortest path tree SPT rooted at s. In this example, a scheduling period $T = \{0, 1, 2, 3\}$ with $|T| = 4$. Each node is represented by a circle with its active time slot contained inside it.

Based on our discussion above, the directed graph G'_r is constructed below

We compute a shortest path tree SPT of G'_r with the root s via employing the Dijkstra's shortest path algorithm on G'_r. This SPT is shown in Fig. 3.

This tree SPT has three leaf nodes v_3, v_6 and v_7. Let $\langle v_6, v_5, v_2, s \rangle$ denote the path from v_6 to s through v_5 and v_2.
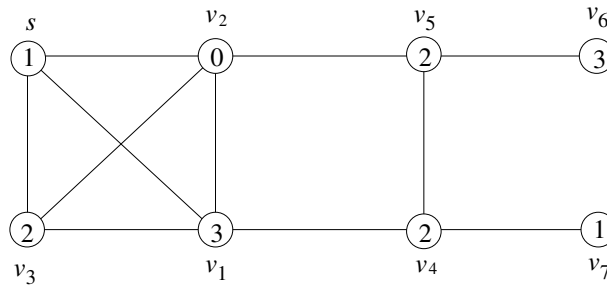
Figure 1: This network has eight vertices. We assume that a scheduling period T has four time slots, that is, $\{0, 1, 2, 3\}$. Each vertex contains a number that is the time-slot at which the vertex is awake.
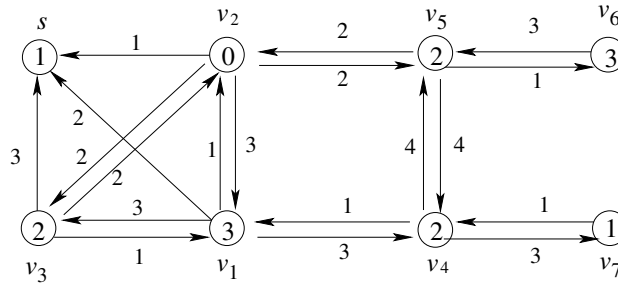

Figure 2: A directed graph with each link associated with the weight equal to its latency defined in EQ(2).
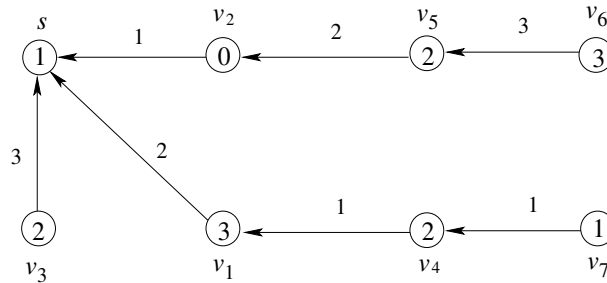

Figure 3: A shortest path tree T_{SPT} rooted at s of the above directed graph in Fig. 2.

Clearly, $Lat(\langle v_6, v_5, v_2, s\rangle) = 6$. Similarly, $Lat(\langle v_7, v_4, v_1, s\rangle) = 4$ and $Lat(v_3, s) = 3$. Therefore, the maximum latency $M = 6$ in this example.

Due to the likely interference with each other caused by the concurrent transmissions, however, the shortest path tree SPT constructed above cannot directly be used for routing of either data aggregation or data gathering. Therefore, a spanning tree $SPAN$ with root s must be created for the use of routing in data aggregation and gathering operations. Such a spanning tree $SPAN$ is constructed as follows:

The nodes in V are put into distinct layers L_0, L_1, L_2, $\cdots$, L_M according to the latency of the shortest path from each $v \in V$ with $v \neq s$ to s. Let's use the same example above to explain how to divide the nodes of V into different layers and construct the sets U_j with $j = 0, 1, 2, 3$. According to the shortest path tree SPT shown in Fig. 3, we have $L_0 = \{s\}$, $L_1 = \{v_2\}$, $L_2 = \{v_1\}$, $L_3 = \{v_3, v_4, v_5\}$, $L_4 = \{v_7\}$, $L_5 = \varnothing$, and $L_6 = \{v_6\}$. Based on EQ(??), it is easy to see that $U_0 = L_1 \cup L_5 = L_1$; $U_1 = L_0 \cup L_4$; $U_2 = L_3$; and $U_3 = L_2 \cup L_6$.

Note that likely interferences only happen at those active nodes that can receive messages at the time-slot that is under consideration.

For each time slot $0 \leq j \leq |T| - 1$, we construct an MIS $I_j \subset U_j$ layer by layer by following the steps below, based on the disjoint union of the nodes in distinct layers of U_j given in EQ(??):

- Initially, $I_j = \varnothing$.

- Sort all the nodes in the layer $L_{A(s)-j+K_0|T|}$ (with smallest index) in the ascending order of their identifications if $L_{A(s)-j+K_0|T|} \neq \varnothing$, say $L_{A(s)-j+K_0|T|} = \langle v_1, v_2, \cdots, v_z\rangle$ with z being the size of $L_{A(s)-j+K_0|T|}$, and add the first node v_1 to I_j.

- If $I_j \cup \{v_2\}$ is independent in G_r, then $I_j \leftarrow I_j \cup \{v_2\}$. If $I_j \cup \{v_2\}$ is dependent in G_r, then the node v_2 is skipped and then the next node in the list is taken into consideration.

- Do the same thing for v_3 until the last node v_z in $L_{A(s)-j+K_0|T|}$ is either added to I_j or skipped.

- Then do the same thing for all the nodes in $L_{A(s)-j+k|T|}$ with $k = 1 + K_0, \cdots, K_1$ until all the nodes in the last layer $L_{A(s)-j+K_1|T|}$ are processed.

Clearly, the MIS I_j constructed above is an independent subset of U_j. For simplicity, let $I_{ji} = I_j \cap L_i$. For each layer $0 < i \leq M$,

every node $v \in L_i \backslash I_{ji}$ was skipped by the above algorithm for the construction of the MIS I_j. Thus, the union $I_{j1} \cup I_{j2} \cup \cdots \cup I_{ji}$ contains one or more neighbors of $v \in L_i \backslash I_{ji}$.

The algorithm for constructing the spanning tree *S PAN* over V rooted at s was described in the conference version [1] and omitted here due to space limitation. Let's use the same example above to explain how to construct the MIS $I_j \subset U_j$ for each time slot $0 \le j \le |T| - 1$, and how to construct the tree *S PAN* rooted at s. For $U_0 = L_1 = \{v_2\}$, $I_0 = \{v_2\}$. For $U_1 = L_0 \cup L_4 = \{s, v_7\}$, $I_1 = \{s, v_7\}$. For $U_2 = L_3 = \{v_3, v_4, v_5\}$, $I_2 = \{v_3, v_4\}$. For $U_3 = L_2 \cup L_6 = \{v_1, v_6\}$, $I_3 = \{v_1, v_6\}$.

Now the tree *S PAN* with root s for this example is created. For $L_1 = \{v_2\}$, v_2's parent in *S PAN* is s (the same parent as in *S PT*) as $v_2 \in I_0 \cap L_1$. For $L_2 = \{v_1\}$, v_1's parent in *S PAN* is also s as $v_1 \in I_3 \cap L_2$. For $L_3 = \{v_3, v_4, v_5\}$, v_3's parent in *S PAN* is s as $v_3 \in I_2 \cap L_3$; v_4's parent in *S PAN* is v_1 as $v_4 \in I_2 \cap L_3$; v_5's parent in *S PAN* is v_4 (a different parent from *S PT*) as $v_5 \notin I_2 \cap L_3$. For $L_4 = \{v_7\}$, v_7's parent in *S PAN* is v_4 as $v_7 \in I_1 \cap L_4$. For $L_6 = \{v_6\}$, v_6's parent in *S PAN* is v_5 as $v_6 \in I_3 \cap L_6$. The constructed tree *S PAN* rooted at s is shown in Fig. 4 below:

Finally in this section, the concept of the first-fit distance-d coloring for a subset of V in the lexicographic order is presented and adopted from our prior work [2]. Denote by U a subset of V. We consider the lexicographic order of the nodes in U, then we sort all nodes in U from the left to right, $u_1, u_2, \cdots, u_k$ with $k = |U|$. *The first-fit distance-d coloring of U* in the lexicographic order is described below:

- positive integers $1, 2, 3, \cdots$ represent colors;
- the first node u_1 in the sequence is assigned color 1;
- For $i = 2$ up to k, the node u_i is assigned the smallest possible number that is not utilized by any prior node u_j in the sequence with $j < i$ and $\|u_i u_j\| \le d$.

[2] proved that for a given independent set I of G_r, the first-fit distance-ρr coloring of I with the lexicographic ordering needs at most β_ρ colors. The value of β_ρ is given below (see Corollary 4 of [2]):

$$\beta_\rho = \left\lfloor \frac{\pi \rho^2}{\sqrt{3}} + \left(\frac{\pi}{2} + 1\right)\rho \right\rfloor + 1. \tag{3}$$

5 Aggregation Schedule under the Physical Interference Model Taking Duty Cycles into Consideration

In this section, we consider MLAS-DC subject to the physical interference and develop a short aggregation schedule for this problem. The sink node is represented by s. For a real number r satisfying $R' \le r < R$, we use EQ(1) to calculate ρ. Then follow the steps stated in Section 4, the spanning tree *S PAN* rooted at s is constructed. The transmission follows a bottom-up approach.

Next we present the first-fit scheduling algorithm for MLAS-DC subject to physical interference. The first layer to transmit is the bottom layer L_M, and then the nodes in layer L_{M-1} transmit. The scheduling is based on a bottom-up approach. For each layer L_i ($0 < i \le M$), we know that the nodes in L_i are active at time-slot j ($0 \le j \le |T| - 1$). At this layer, the nodes in $L_i \backslash I_j$ will be scheduled to transmit first, followed by the tranmissions of the nodes in $L_i \cap I_j$. Then the nodes in layer L_{M-1} will be scheduled to transmit. For each layer among $L_{M-1}, L_{M-2}, \cdots, L_0$, the nodes in these layers will receive the aggregated data from their children nodes in the decreasing order of the indices of the layers $M - 1, M - 2, \cdots$.

For the details of our first-fit data aggregation scheduling algorithm proposed for the problem MLAS-DC subject to the physical interference, please refer to the conference version [1]. Assume the graph in Figure 1 represents an r-disk graph G_r for some $r \in \Gamma$. We use this example in Figure 1 to illustrate the data aggregation scheduling algorithm we proposed for MLAS-DC.

The tranmission schedules for aggregation outputted by the above algorithm are listed in Table 2. Totally, 6 scheduling periods are required for s to receive eventually the aggregated message in this example.

The algorithm for this example works as follows: At $L_6 = \{v_6\}$, $v_6 \in I_3 \cap L_6$ and its parent in *S PAN* is v_5 which is active at time-slot 2. Thus, link (v_6, v_5) is scheduled at time-slot 2 of SP 1. We skip $L_5 = \varnothing$. For $L_4 = \{v_7\}$, $v_7 \in I_1 \cap L_4$ and its parent in *S PAN* is v_4 which is also active at time-slot 2. Thus, link (v_7, v_4) is scheduled at time-slot 2 of SP 2. For $L_3 = \{v_3, v_4, v_5\}$, node $v_5 \in L_3 \backslash I_2$ and its parent in *S PAN* is v_4 which is active at time-slot 2. Thus, link (v_5, v_4) is scheduled at time-slot 2 of SP 3. The nodes $v_3, v_4 \in I_2 \cap L_3$; v_3's parent in *S PAN* is s which is active at time-slot 1. Thus, link (v_3, s) is scheduled at time-slot 1 of SP 4. Node v_4's parent in *S PAN* is v_1 which is active at time-slot 3. Thus, link (v_4, v_1) is scheduled at time-slot 3 of SP 4. For $L_2 = \{v_1\}$, $v_1 \in I_3 \cap L_2$ and its parent in *S PAN* is s which is active at time-slot 1. Thus, link (v_1, s) is scheduled at time-slot 1 of SP 5. Finally for $L_1 = \{v_2\}$, $v_2 \in I_0 \cap L_1$ and its parent in *S PAN* is s which is active at time-slot 1. Thus, link (v_2, s) is scheduled at time-slot 1 of SP 6.

We summarize our above discussion in the following theorem. The proof of this theorem can be found in the short conference version [1].

Theorem 1 *The data aggregation scheduling algorithm for MLAS-DC is technically correct and produces a data aggregation schedule without any collision. The latency of the transmission schedule for aggregation outputted via our proposed algorithm is less than or equal to $2\beta_\rho |T| (\Delta - 1)M$. The proposed algorithm for MLAS-DC has approximation ratio upper bounded by $2\beta_\rho |T| (\Delta - 1)$.*

Note that both ρ and β_ρ are constants defined in EQ(1) and EQ(3), respectively. Therefore, the approximation bound for our scheduling algorithm proposed for MLAS-DC is less than or equal to a constant times $|T|$ when the maximum degree Δ is bounded.
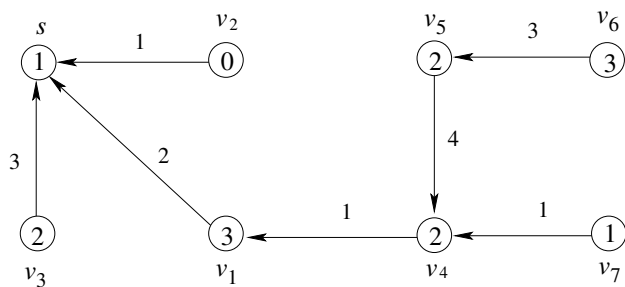
Figure 4: The construction of a spanning tree T_{span} rooted at s.

Table 2: The transmission schedule for the aggregation operation outputted by the algorithm proposed for MLAS-DC. In each row, "SP i" means the i-th scheduling period.

	timeslot0	timeslot1	timeslot2	timeslot3
$SP1$			(v_6, v_5)	
$SP2$			(v_7, v_4)	
$SP3$			(v_5, v_4)	
$SP4$		(v_3, s)		(v_4, v_1)
$SP5$		(v_1, s)		
$SP6$		(v_2, s)		

6 Gathering Schedule under the Physical Interference Model Taking Duty Cycles into Consideration

In this section, we develop a scheduling algorithm for MLGS-DC subject to physical interference. The sink node is represented by s. For a real number r satisfying $R' \leq r < R$ (see Section 3 for the definitions of R' and R), we use EQ(1) to calculate ρ. Then follow the steps stated in Section 4, the spanning tree $SPAN$ rooted at s is constructed.

For any time-slot $0 \leq j \leq |T| - 1$, each node in U_j is active and ready to receive packets from one of its children nodes in $SPAN$. Let

$$Children(U_j) =$$
$$\{u : u \text{ is a child in } SPAN \text{ of a node in } U_j\},$$

which is the collection of children devices of the devices in U_j. The devices in $Children(U_j)$ will send their packets at time-slot j.

For each $0 \leq j \leq |T| - 1$, we adopt the data gathering heuristic obtained in our previous work [2] for WSNs with all nodes always active for the time-slot j. Our propsed data gathering scheduling algorithm will have $2n - 3$ rounds in total. For each round k, represent by E_k the collection of tree links of $SPAN$ whose label is k. Suppose that A_k represents the collection of the links in the inward s-arborescence made by the tree edges in E_k. For each round k, a dominator connects no more than one link in A_k (see [2]). For each round k and time-slot j, represent by $A_k^{(j)}$ the sub-collection of links of A_k whose sender belongs to the set $Children(U_j)$.

Next the data gathering scheduling algorithm for MLGS-DC subject to physical interference is presented. In each time slot $0 \leq j \leq |T| - 1$, the algorithm is divided into $2n - 3$ rounds sequentially, each of which is dedicated to a subset of links in the list:

$$A_{2n-3}^{(j)}, A_{2n-4}^{(j)}, \cdots, A_2^{(j)}, A_1^{(j)},$$

respectively.

For each round k and time-slot j, the links in $A_k^{(j)}$ are scheduled for transmission: Let S denote the set of the dominator ends of the links in $A_k^{(j)}$. A first-fit distance-ρr coloring of the set S is first calculated. At the i-th time-slot of the k-th round, according to this coloring, the link with a dominator end that is assigned the i-th color will send a packet. By Corollary 4 of [2], the maximum number of time-slots each of these rounds uses is β_ρ. The total number of rounds is $2n - 3$. Therefore, for each time slot j, the maximum time-slots used by the solution outputted by our proposed gathering scheduling algorithm is $\beta_\rho(2n - 3)$. Each scheduling period has $|T|$ time-slots in total. Thus, the entire gathering operation uses no more than $\beta_\rho(2n - 3)|T|$ time-slots.

We summarize our above discussion in the following theorem for MLGS-DC under the physical interference model.

Theorem 2 *The gathering scheduling algorithm described above for MLGS-DC can correctly produce a interference-free schedule of data gathering. The latency of the produced data gathering schedule is less than or equal to $\beta_\rho(2n - 3)|T|$.*

Clearly, any optimal solution to the problem MLGS-DC uses at least $n - 1$ time-slots, the approximation bound for the gathering scheduling algorithm presented in this section is less than or equal to $2\beta_\rho|T|$, which is a multiple of $|T|$ as β_ρ is a constant.

At the end, a data gathering schedule is computed by applying the above scheduling algorithm for each $r \in \Gamma$. The one with the smallest latency among all these resulting data gathering schedules will be chosen. Note that the number of element in Γ is a finite and we have its upper bound $|\Gamma| \leq n(n - 1)/2$.

7 Gossiping Schedule under the Physical Interference Model Taking Duty Cycles into Consideration

In this section, we develop a scheduling algorithm that produces a short gossiping schedule for MLGoS-DC under the SINR model. Let $r = R'$ (see Section 3 for the definition of R'), we use EQ(1) to calculate ρ. The r-disk graph on V is denoted by G_r. Then we compute the graph center s for the graph G_r.

Our proposed gossiping scheduling algorithm for MLGoS-DC consists of two phases:

Phase 1: s collects all the packets from all other nodes.

Phase 2: s broadcasts all the n packets (including its own) to all other nodes.

For Phase 1, the data gathering scheduling algorithm for MLGS-DC presented in Section 6 will be adopted. After that, the sink node s sends all the n packets to all other $n - 1$ nodes in Phase 2. In this phase, for each time slot $0 \leq j \leq |T| - 1$, we adopt the algorithm that was used in the second phase of the gossiping algorithm from our prior work [2] we developed for MLGoS without duty-cycled scenarios. This algorithm works as follows: To begin with, a spanning tree $SPAN$ over V on the graph G_r is created according to the steps stated in Section 4. The routing used for the broadcast of Phase 2 is the outward spanning s-aborescence oriented from $SPAN$. Then we perform the following three steps:

- For all the dominating nodes, we calculate the first-fit distance-ρr coloring. Assume that this distance-ρr coloring uses Z colors. By Corollary 4 of [2], we have $Z \leq \beta_\rho$.

- For simplicity, suppose that the color 1 is assigned by the graph center s. We divide the time-slots into $2Z$ frames. In each frame of the time-slots, the first Z time-slots make a subframe of dominating nodes and the second Z time-slots make a subframe of connecting nodes. The center s sends one packet in each subframe.

- For each $1 \leq i \leq Z$, the connecting node with the color i receiving a packet in a subframe of dominating nodes sends the received packet in the i-th time-slot.

- For each $1 \leq i \leq Z$, the dominating node with the color i receiving a packet in a subframe of connecting nodes sends the received packet in the i-th time-slot.

According to Theorem 9 of [2] and the discussions thereafter, for each time slot $0 \leq j \leq |T| - 1$, the latency of both phase 1 and phase 2 is at most $\beta_\rho(5n - 5)$. Thus, our proposed gossiping scheduling algorithm generates a solution whose latency is less than or equal to $\beta_\rho(5n - 5)|T|$.

The theorem below summarizes our discussions and analyses in this section

Theorem 3 *The gossiping scheduling algorithm presented in this section works and it generates a gossiping schedule without collision. This gossiping scheduling algorithm produces a solution whose latency is less than or equal to $\beta_\rho(5n - 5)|T|$.*

Clearly, any optimal solution to the problem MLGoS-DC uses at least n time-slots, the approximation ratio of the gossiping scheduling algorithm presented in this section for MLGoS-DC is less than or equal to $5\beta_\rho|T|$, which is a multiple of $|T|$ as β_ρ is a constant.

8 Conclusion

In this paper, approximation algorithms for MLAS-DC, MLGS-DC and MLGoS-DC on duty-cycled multi-hop WSNs under the SINR model were proposed in order to minimized the communication latency. To the best of our knowledge, the paper is the first work that proposed efficient communication scheduling algorithms for these three problems subject to the physical interference. Denote by $|T|$ the count of time-slots contained in a scheduling period. Our proposed approximation algorithms for MLGS-DC has approx. ratios upper bounded by a constant time of $|T|$. So is the proposed approximation algorithms for MLGoS-DC. Our proposed approximation algorithms for MLAS-DC has an approximation bound no more than a constant times $|T|$ when the network maximum degree Δ is bounded. It has been well documented in the literature that analysis for scheduling algorithms developed for MLAS, MLGS and MLGoS under the SINR model is very challenging due to the impossibility of constructing related conflict graphs.

For future work in the design of scheduling algorithms for MLAS-DC, MLGS-DC and MLGoS-DC on duty-cycled multi-hop WSNs under the SINR model, the problems of proposing constant-approximation algorithms are still open, under either the SINR model or the protocol inference model. Any constant-approximation algorithms for MLAS-DC, MLGS-DC and MLGoS-DC will be significant.

9 Declarations

9.1 Availability of data and material

The data generated during the investigation of this research are included in this paper.

9.2 Conflicts of interest

The authors declare that there is no conflict of interest about the publication of this article.

9.3 Funding

References

[1] L. Wang, H. Liang, and P.-J. Wan, "Data Aggregation Scheduling in Duty-Cycled Multi-hop Wireless Networks Subject to Physical Interference, " the 15th IEEE International Conference on Mobile Ad-hoc and Sensor Networks (IEEE MSN), Dec. 11-13, 2019, DOI: 10.1109/MSN48538.2019.00041

[2] P.-J. Wan, L. Wang, O. Frieder, "Fast Group Communications in Multihop Wireless Networks Subject to Physical Interference," the 6th IEEE International Conference on Mobile Ad hoc and Sensor Systems (IEEE MASS), 526-533, 2009 DOI: 10.1109/MOBHOC.2009.5336958

[3] X.J. Chen, X.D. Hu, J.M. Zhu, "Minimum data aggregation time problem in wireless sensor networks," Lecture Notes in Computer Science 3794, 133-142, 2005.

[4] Y. P. Chen, A. L. Liestman, J. Liu, "A hierarchical energy-efficient framework for data aggregation in wireless sensor networks," IEEE Transactions on Vehicular Technology, **55**(3), 789-796, 2006, DOI: 10.1109/TVT.2006.873841

[5] Z. Chen, G. Yang, L. Chen, J. Xu, H. Wang, "A load-balanced data aggregation scheduling for duty-cycled wireless sensor networks," In Cloud Computing Technology and Science (CloudCom), 4th IEEE International Conference on, 888-893, December, 2012, DOI: 10.1109/CloudCom.2012.6427533

[6] R. Gandhi, Y.-A. Kim, S. Lee, J. Ryu, P.-J. Wan, "Approximation Algorithms for Data Broadcast in Wireless Networks," IEEE INFOCOM Mini-conference, 2009, DOI: 10.1109/TMC.2011.162

[7] S.C.-H. Huang, P.-J. Wan, C. T. Vu, Y. Li, F. Yao: "Nearly Constant Approximation for Data aggregation Scheduling in Wireless Sensor Networks," IEEE INFOCOM, 2007, DOI: 10.1109/INFCOM.2007.50

[8] S.C.-H. Huang, H. Du, E.-K. Park, "Minimum-latency gossiping in multi-hop wireless networks," ACM Mobihoc, 2008, DOI: 10.1145/1374618.1374662

[9] P.-J. Wan, C.-H. Huang, L. Wang, Z.-Y. Wan, X. Jia: "Minimum Latency aggregation Scheduling in Multihop Wireless Networks," ACM MOBIHOC, 2009, DOI: 10.1145/1530748.1530773

[10] J.-C. Bermond, J. Galtier, R. Klasing, N. Morales, S. Perennes, "Hardness and approx. of gathering in static radio networks," Proceedings FAWN06, 2006, DOI: DOI: 10.1109/PERCOMW.2006.62

[11] V. Bonifaci, P. Korteweg, A. Marchetti-Spaccamela, L. Stougie, "An approx. Algorithm for the Wireless Gathering Problem," Proceedings of SWAT, 328-338, 2006, DOI: 10.1007/11785293_31

[12] R. Gandhi, S. Parthasarathy, A. Mishra, "Minimizing broadcast latency and redundancy in ad hoc networks," in Proceedings of the 4th ACM international symposium on Mobile Ad hoc networking and computing (MobiHoc), 222-232, 2003, DOI: 10.1145/778415.778442

[13] P.-J. Wan, Z. Wang, Z. Wan, S. C.-H. Huang, H. Liu, "Minimum-Latency Schedulings for Group Communications in Multi-channel Multihop Wireless Networks," WASA, 2009, DOI: 10.1007/978-3-642-03417-6_46

[14] The GreenOrb Wireless Sensor Network Projects available at http://www.greenorbs.org/all/projects.htm, 2011

[15] P. Vicaire, T. He, Q. Cao, T. Yan, G. Zhou, L. Gu, T. F. Abdelzaher, "Achieving long-term surveillance in vigilnet," ACM Transactions on Sensor Networks (TOSN), **5**(1), 9, 2009, DOI: 10.1145/1464420.1464429

[16] K. Han, J. Luo, Y. Liu, A. Vasilakos, "Algorithm Design for Data Communications in Duty-Cycled Wireless Sensor Networks: A Survey," Communications Magazine, IEEE, **51**(7), 2013, DOI: 10.1109/MCOM.2013.6553686

[17] X. Jiao, W. Lou, X. Feng, X. Wang, L. Yang, G. Chen, "Delay efficient data aggregation scheduling in multi-channel duty-cycled WSNs," In 15th IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS), 326-334. IEEE, 2018, DOI: 10.1109/MASS.2018.00055

[18] X. Jiao, W. Lou, J. Ma, J. Cao, X. Wang, X. Zhou, "Minimum latency broadcast scheduling in duty-cycled multihop wireless networks," IEEE Transactions on Parallel and Distributed Systems, **23**(1), 110-117, 2011, DOI: 10.1109/TPDS.2011.106

[19] X. Jiao, W. Lou, X. Wang, J. Ma, J. Cao, X. Zhou, "On interference-aware gossiping in uncoordinated duty-cycled multi-hop wireless networks," Ad Hoc Networks, **11**(4), 1319-1330, 2013, DOI: 10.1007/978-3-642-14654-1_24

[20] B. Kang, P. K. H. Nguyen, V. Zalyubovskiy, H. Choo, "A distributed delay-efficient data aggregation scheduling for duty-cycled WSNs," IEEE Sensors Journal, **17**(11), 3422-3437, 2017, DOI: 10.1109/JSEN.2017.2692246

[21] Z. Li, Y. Peng, D. Qiao, W. Zhang, "LBA: Lifetime balanced data aggregation in low duty cycle sensor networks," In 2012 Proceedings IEEE INFOCOM, 1844-1852, IEEE, 2012, DOI: 10.1109/INFCOM.2012.6195559

[22] F. Y.-S. Lin, Y.-F. Wen, "Multi-sink data aggregation routing and schedul-ing with dynamic radii in WSNs," Communications Letters, IEEE, **10**(10), 692-694, 2006, DOI: 10.1109/LCOMM.2006.06058

[23] Q. Chen, H. Gao, S. Cheng, J. Li, Z. Cai, "Distributed non-structure based data aggregation for duty-cycle wireless sensor networks," In IEEE INFOCOM 2017-IEEE Conference on Computer Communications, 1-9, IEEE, 2017, DOI: 10.1109/INFOCOM.2017.8056960

[24] X. Xu, J. Cao, and P.-J. Wan, "Fast Group Communication Scheduling in Duty-Cycled Multihop Wireless Sensor Networks," WASA, 2012, DOI: 10.1007/978-3-642-31869-6_17

[25] J. Tang, J. Xiao, W. Xiao, "Minimum-latency data aggregation in duty-cycled wireless sensor networks under physical interference model," In the 22nd Wireless and Optical Communication Conference, 309-314, *IEEE,* 2013, DOI: 10.1109/WOCC.2013.6676328

[26] X. Zhu, B. Tang, and H. Gupta, "Delay efficient data gathering in sensor networks," International Conference on Mobile Ad-Hoc and Sensor Networks, Springer, Berlin, Heidelberg, 380-389, 2005, DOI: 10.1007/11599463_38

[27] Q. Chen, H. Gao, Z. Cai, L. Cheng, J. Li, "Distributed low-latency data aggregation for duty-cycle wireless sensor networks," IEEE/ACM Transactions on Networking, **26**(5), 2347-2360, 2018, DOI: 10.1109/TNET.2018.2868943

[28] Y.-F. Wen, F. Y.-S. Lin, "Cross-layer duty cycle scheduling with data aggrega-tion routing in wireless sensor networks," Embedded and Ubiquitous Comput-ing. Springer Berlin Heidelberg, 894-903, 2006, DOI: 10.1007/11802167_90