Supplemental Material: Capacity of Data Collection in Arbitrary Wireless Sensor Networks

Siyuan Chen* Minsu Huang* Shaojie Tang† Yu Wang*

I. Tighter Capacity Bound of Branch Scheduling on BFS Tree in General Graph Model

In this section, we provide a tighter bound of achievable collection capacity of the BFS-based branch scheduling method (Algorithm 1 in the paper). It is easy to show that the capacity is still $\frac{W}{\tilde{\Delta}}$. Here, $\tilde{\Delta}$ is the maximum path interference among all paths. However, in general graph model $\tilde{\Delta}$ is not bounded by a constant any more, and it could be $O(1)$ or $O(n)$. Thus, there is a gap between our lower bound of data collection $\frac{W}{\tilde{\Delta}}$ and the natural upper bound W. Considering both examples shown in Figure 4 of the paper, BFS-based method matches their tight upper bounds $\Theta(\frac{W}{n})$ and W. For the star topology, even though the sink has the maximal interference $\Delta = n$, each individual path has the path interference $\Delta_i = 1$ which leads to capacity of $\frac{W}{1} = W$. For the straight-line topology, the path interference of the single path $\Delta_i = n$, thus the capacity is $\frac{W}{n}$. In both cases, $\frac{W}{\tilde{\Delta}}$ matches the optimal capacity. However, similar to $\frac{W}{\Delta}$, $\frac{W}{\tilde{\Delta}}$ is still not a tight bound too. We will show such an example in Figure 1.

Now we modify the basic *Path Scheduling* of the BFS-method to achieve better collection capacity. Recall that in Section IV.B we claim that the path scheduling for a path P_i can be done in $\Delta_i \cdot |P_i|$ time slots. However, we can perform path scheduling in the following way to save more slots. Assume that path $P_i = v_0, v_1, v_2, \cdots, v_{|P_i|}$. Let $\delta_k^{P_i} = \max\{\delta^{P_i}(v_1), \cdots, \delta^{P_i}(v_k)\}$, *i.e.*, $\delta_k^{P_i}$ is the maximum interference number among *first* k nodes v_1 to v_k in path P_i. Clearly, $\delta_k^{P_i} \leq \delta_{k+1}^{P_i}$. In the first step, using $\delta_{|P_i|}^{P_i}$ slots, every node on the path transfers its data to its parent in the BFS tree. After the first step, the leaf $v_{|P_i|}$ already finishes its task in this round and has no data from current snapshot. In the second step, using $\delta_{|P_i|-1}^{P_i}$ slots, the current snapshot data will move one more level up along the path in BFS tree. Repeat these steps until all data along this path reach the sink. It is easy to show that the total number of time slots used by the above procedure is $\sum_{k=1}^{|P_i|} \delta_k^{P_i}$. Since $\delta_k^{P_i} \leq \Delta_i$, $\sum_{k=1}^{|P_i|} \delta_k^{P_i} \leq \Delta_i \cdot |P_i|$.

Figure 1 shows an example where $\sum_{k=1}^{|P_i|} \delta_k^{P_i}$ is much smaller than $\Delta_i \cdot |P_i|$. Again we have n sensors and the sink distributed on a line P as shown in the figure. Assume that $R = r$. On the left side, there are $\log n$ nodes close to each other, thus their $\delta(v_i) = \log n$ except for $\delta(v_{n-\log n+1}) = \log n + 1$. On the right side, every node has $\delta(v_i) = 3$. Thus, $\Delta = \log n + 1$ and $\Delta \cdot |P| = \Theta(n \log n)$. In addition,

S. Chen, M. Huang and Y. Wang are with Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC, USA. S. Tang is with Department of Computer Science, Illinois Institute of Technology, Chicago, Illinois, USA.

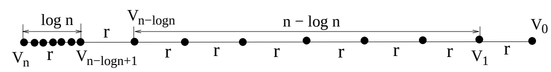


Fig. 1. Illustration of the advantage of a new path scheduling. Here, $R = r$.

$\delta_k^P = \log n + 1$ for $k = n - \log n + 1, \cdots, n$ and $\delta_k^P = 3$ for $k = 3, \cdots, n - \log n$, $\delta_2^P = 2$, and $\delta_1^P = 1$. Therefore, $\sum_{k=1}^{|P|} \delta_k^P = (\log n + 1)\log n + 3(n - \log n) - 3 = \Theta(n)$. It is obvious that $\sum_{k=1}^{|P|} \delta_k^P = \Theta(n)$ is smaller than $\Delta \cdot |P| = \Theta(n \log n)$ in order.

Using the above new path scheduling analysis, we now derive a tight lower bound for our BFS-based method. Recall that our method transfers data based on branches in BFS tree T. Given T, there are c paths P_i and c branches B_i as shown in Figure 3(a) and 3(b) in the paper. Then the total number of time slots used by Algorithm 1 with new path scheduling is at most

$$\sum_{i=1}^{c} \sum_{k=|P_i|-|B_i|+1}^{|P_i|} \delta_k^{P_i}.$$

It is clear that this number is much smaller than $\sum_{i=1}^{c} \Delta_i \cdot |B_i|$ from previous analysis. Notice that for path P_i our algorithm (Line 3-4 in Algorithm 1) will terminate the transmission until the branch B_i does not have data for current snapshot and switch to next path P_{i+1}. Thus, the index of k is only from $|P_i|$ to $|P_i| - |B_i| + 1$. Therefore, the capacity achieved by our algorithm is at least

$$\frac{W}{\frac{\sum_{i=1}^{c} \sum_{k=|P_i|-|B_i|+1}^{|P_i|} \delta_k^{P_i}}{n}}.$$

Let $\Delta^{**} = \frac{\sum_{i=1}^{c} \sum_{k=|P_i|-|B_i|+1}^{|P_i|} \delta_k^{P_i}}{n}$ which can be derived given the BFS tree. We now have a new lower bound of collection capacity as $\frac{W}{\Delta^{**}}$. Here Δ^{**} is a kind of weighted-average of the maximum interference among paths P_i and branches B_i in the BFS tree. We then have the following relationship:

$$n \geq \Delta \geq \tilde{\Delta} \geq \Delta^{**} \geq 1,$$

among the maximum interference number Δ in the whole graph, the maximum interference number $\tilde{\Delta}$ in the paths/braches of BFS tree, and the "average" maximum interference Δ^{**} in the paths/branches of BFS tree. These three interference numbers can be different from each other in order.

Compared with the lower bound $\frac{\lambda^*}{\lambda}\frac{W}{\Delta^*}$ which we derive from greedy scheduling on BFS tree, this new lower bound $\frac{W}{\Delta^{**}}$ may be smaller in some cases. Consider the example in Figure 1, $\frac{\lambda^*}{\lambda}\frac{W}{\Delta^*} = \Theta(\frac{W}{\log n})$, while $\frac{W}{\Delta^{**}} = \frac{W}{\Delta^*} = \Theta(W)$. However, the reason is mainly due to the rough relaxation in our capacity analysis of greedy scheduling. Actually, our simulations in Section V show that greedy scheduling always outperforms branch scheduling.

In summary, the bounds of collection capacity (Theorem 7 in the paper) could be revised as the following:

Theorem 7: Under protocol interference model and general graph model, data collection capacity for arbitrary sensor networks is at least $min(\frac{\lambda^*}{\lambda}\frac{W}{\Delta^*}, \frac{W}{\Delta^{**}})$ and at most $\frac{W}{\Delta^*}$.

II. Proof of $\lambda \geq \lambda^*$ in General Graph Model

In this section, we provide a proof of $\lambda \geq \lambda^*$, which is used in the proof of Lemma 6 in the paper. Let v_k be the node inside critical region with the largest level. We now consider two cases.

Case 1: If there is a node outside the critical region, as shown in Figure 5(a) in the paper, the transmission from v_s to v_k should interfere with the transmission from v_q to v_0. Thus, in the view of v_s, its $\lambda_s \geq l^* + 1 = \lambda^*$. Therefore $\lambda \geq \lambda^*$.

Case 2: If all nodes are inside the critical region, again consider the v_k with largest level. Then $\lambda = \lambda_k = l(v_k) + 1 > l(v_k) = \lambda^*$.

Consequently, we conclude $\lambda \geq \lambda^*$.

III. Data Collection Capacity under Physical Interference Model

In this section, we study the data collection capacity under *physical interference model*. In *physical interference model*, node v_j can correctly receive signal from a sender v_i if and only if, given a constant $\eta > 0$, the SINR (Signal to Interference plus Noise Ratio)

$$\frac{P \cdot ||v_i - v_j||^{-\beta}}{B \cdot N_0 + \sum_{k \in I} P \cdot ||v_k - v_j||^{-\beta}} \geq \eta,$$

where B is the channel bandwidth, $N_0 > 0$ is the background Gaussian noise, I is the set of actively transmitting nodes when node v_i is transmitting, $\beta > 2$ is the pass loss exponent, and P is the fixed transmission power. Again, we assume that each node uses the same transmission power and the background noise N_0 is a fixed constant. Since the received signal has strength at most P, thus $\eta < \frac{P}{B \cdot N_0}$.

We now provide the formal proof of Theorem 8 in the paper.

Theorem 8: Under physical interference model and disk graph model, data collection capacity for arbitrary wireless sensor networks is $\Theta(W)$.

A. Upper Bound

To give an upper bound on the capacity of data collection, we will show that we can set an artificial transmission range r_0 and an artificial interference range R_0 such that (1) the receiving node v_j of a sender v_i is within distance r_0, and (2) a transmitting node v_k will cause interference at node v_j within distance R_0. I.e., if there is any interference among nodes in protocol interference model with these artificial ranges, there is also interference among them in physical interference model.

Given P, N_0, and η, we choose artificial ranges as follows:

$$r_0 \leq (\frac{P}{B \cdot N_0 \cdot \eta})^{1/\beta} \textit{ and } R_0 < (\frac{\eta \cdot P}{P - B \cdot N_0 \cdot \eta})^{1/\beta}.$$

Notice that the definition of R_0 is valid since $P - B \cdot N_0 \cdot \eta > 0$.

First, if the receiving node v_j within distance of d from the sender v_i can correctly decode the signal, $\frac{P \cdot d^{-\beta}}{BN_0} \geq \eta$. Obviously if $d \leq r_0 \leq (\frac{P}{B \cdot N_0 \cdot \eta})^{1/\beta}$, $\frac{P \cdot d^{-\beta}}{BN_0} \geq \eta$ holds. Thus r_0 is the maximum distance of a successful communication.

Second, if a receiving node v_j is within a distance R_0 of a transmitting node v_k and v_i is the legitimate sender of v_j, the SINR at node v_j is at most $\frac{P}{BN_0 + P||v_k - v_j||^{-\beta}} \leq \frac{P}{BN_0 + PR_0^{-\beta}}$, since the maximum strength of the signal from v_i received at node v_j is at most P. Given $R_0 < (\frac{\eta \cdot P}{P - B \cdot N_0 \cdot \eta})^{1/\beta}$, thus the SINR $< \eta$. Therefore, the node v_j cannot receive data from v_i correctly due to the interference from v_k.

By artificially setting r_0 and R_0 (which are both constants), we convert the physical interference model into a protocol interference model. Using previous proofs in protocol interference model, it is straightforward to show that the upper bound on the capacity under disk graph model is bounded by $\Theta(W)$ where the constant behind the $\Theta()$ is related to $\frac{R_0}{r_0}$.

B. Lower Bound

To give a lower bound on the capacity of data collection, we then set an artificial transmission range r_1 and an artificial interference range R_1 such that, when all simultaneously transmitting nodes are separated by a distance R_1, and the receiving nodes of a transmitting node is within r_1, the SINR of every receiving node is at least η. In other words, if there is no interference among nodes in the protocol interference model with artificial ranges r_1 and R_1, there is no interference among the nodes in the physical interference model as well.

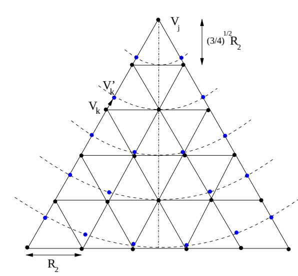


Fig. 2. Illustration of positions of active transmitting nodes causing the maximum interference at v_j. Here only nodes in one 60° direction from v_j are shown.

Consider v_i transmits signal to v_j and there are other active transmitting nodes v_k. Thus, the interference at the receiver v_j by all other transmitting nodes $v_k \in I$ is $\sum_{v_k \in I} P \cdot ||v_k - v_j||^{-\beta}$. Since every pair of active transmitting nodes need to be separated by at least $R_2 = R_1 - r_1$, the maximum number of active transmitting nodes are distributed in a regular triangulation pattern as shown in Figure 2. To calculate the upper bound of the maximum interference at v_j, we move every active transmitting node v_k to the position of v'_k on a circle centered at v_j, as shown in Figure 2. On the i-th circle from v_j (whose radius is $i\frac{\sqrt{3}}{2}R_2$), there are $6i$ active nodes. Therefore, the maximum interference at v_j is at most

$$\sum_{i=1}^{\infty} 6i \cdot P(i\frac{\sqrt{3}}{2}R_2)^{-\beta} = \frac{6P}{(\frac{\sqrt{3}}{2}R_2)^{\beta}} \sum_{i=1}^{\infty} \frac{1}{i^{\beta-1}} \le \frac{6P}{(\frac{\sqrt{3}}{2}R_2)^{\beta}} \zeta.$$

Here ζ is a constant bound of $\sum_{i=1}^{\infty} \frac{1}{i^{\beta-1}}$ for $\beta > 2$. Thus, the SINR at v_j is at least

$$\frac{P \cdot ||v_i - v_j||^{-\beta}}{B \cdot N_0 + \frac{6P}{(\frac{\sqrt{3}}{2}R_2)^{\beta}}\zeta} \ge \frac{P{r_1}^{-\beta}}{B \cdot N_0 + \frac{6P}{(\frac{\sqrt{3}}{2}R_2)^{\beta}}\zeta}.$$

The last inequality is due to $||v_i - v_j|| \le r_1$. To make the SINR value $\ge \eta$, we need

$$r_1 \le (\frac{P}{(B \cdot N_0 + \frac{6P}{(\frac{\sqrt{3}}{2}R_2)^{\beta}}\zeta) \cdot \eta})^{1/\beta}.$$

Then we can carefully choose r_1 and R_2 such that the above inequation holds. Notice that r_1 need to be larger than 1 to satisfy $P \cdot {r_1}^{-\beta} < P$. Since $\eta < \frac{P}{B \cdot N_0}$, we can choose sufficiently large constant R_2 such that $\frac{P}{B \cdot N_0 + \frac{6P}{(\frac{\sqrt{3}}{2}R_2)^{\beta}}\zeta} > \eta$. Then we can set $R_1 = r_1 + R_2$.

By artificially setting r_1 and R_1, we can convert the physical interference model into a protocol interference model. Using previous collection algorithms for protocol interference model, it can be shown that the lower bound $\Theta(W)$ on the capacity of data collection under disk graph model is achievable.

IV. Proof of Theorem 9: An Upper Bound for Gaussian Channel Model

We now provide a formal proof of Theorem 9 in the paper.

Theorem 9: An upper bound for data collection capacity under Gaussian channel model is at most

$$\max_i(W_{i0}) + W \cdot \log_2(n).$$

Proof: We first order all the incoming links of v_0 according to their length as follows: $l_{10} \le l_{20} \le \cdots \le l_{n'0}$. Here n' is the number of incoming links at sink v_0 and $n' \le n$. Next, we try to bound the SINR of the sink node v_0. For any link $v_i v_0$ ($i \ne 1$), its SINR

$$\begin{aligned} SINR_{i0} &\le \frac{P \cdot l_{i0}^{-\beta}}{N_0 + \sum_{k=1}^{i-1} P \cdot l_{k0}^{-\beta}} \\ &\le \frac{P \cdot l_{i0}^{-\beta}}{N_0 + \sum_{k=1}^{i-1} P \cdot l_{i0}^{-\beta}} < \frac{1}{i-1} \end{aligned}$$

Therefore, for $i \ne 1$,

$$W_{i0} = W \log_2(1 + SINR_{i0}) < W \log_2(\frac{i}{i-1}).$$

So the maximum rate at sink v_0 is at most

$$\begin{aligned} & W_{10} + \sum_{i=2}^{n'} W \log_2(\frac{i}{i-1}) \\ = \; & W_{10} + W \log_2(\Pi_{i=2}^{n'} \frac{i}{i-1}) \\ \le \; & \max_i(W_{i0}) + W \cdot \log_2(n') \\ \le \; & \max_i(W_{i0}) + W \cdot \log_2(n) \end{aligned}$$

■

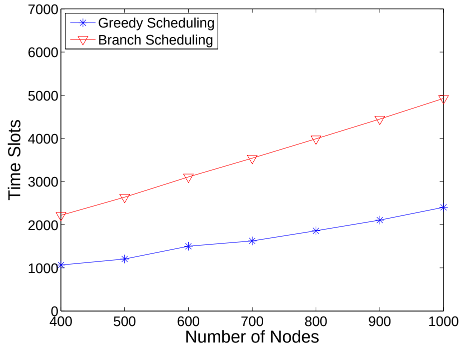
V. Simulation Results on Random Networks

In this section, we study the data collection performances of branch scheduling (Algorithm 1) and greedy scheduling (Algorithm 2) in random sensor networks. We implement both algorithms in a simulator developed by our group. The underlying wireless sensor networks are randomly generated in a square region of 1000×1000. We randomly pick one sink node and form the BFS tree rooted at the sink as the data collection tree. Both scheduling algorithms are performed on the same BFS tree. The total number of time slots is used for collecting one snapshot from the sensor network as the only performance measurement of our data collection algorithms. Clearly, larger total number of time slots leads to lower data collection capacity. For all the simulations, we repeat the experiment for multiple times and report the average values of total number of time slots used by these two methods.

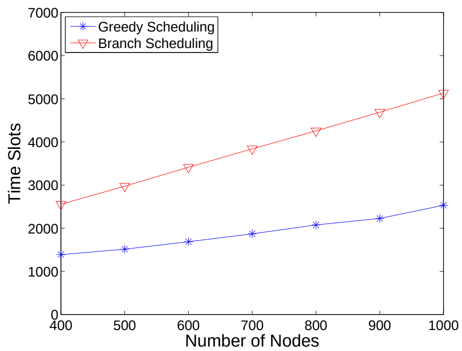
We first test our algorithms on disk graph model, where transmission range of r is set to 80 and $R = 2r$. We vary the number of sensors from 400 to 1000. Figure 3(a) shows the results. It is obvious that more time slots are needed to collect all data when the number of sensors increases. However, if consider the data collection capacity ($\frac{nb}{D}$), the capacity increases with the number of sensors. Compared the two scheduling algorithms, though they are both order-optimal for disk graph model, greedy scheduling algorithm uses much less time slots than the branch scheduling method, thus can achieve better collection capacity. This is mainly due to the advantage of parallel transmissions among multiple branches.

We then test both algorithms on general graph model, where links in the disk graph model ($r = 80$) are randomly removed with a fixed probability 0.5. The interference range R is still $2r$. Figure 3(b) illustrates the results. Compared with results in disk graph model, more time slots are needed since the communication graph is sparser, which usually leads longer branches in BFS tree. The advantage of greedy scheduling over branch scheduling is still obvious.

Finally, we enlarge the interference range R to $4r$ and repeat the above simulations. The results are summarized in Figure 4 for both disk graph model and general graph model. Due to the larger interference, more time slots are needed for both algorithms to perform collection task compared with previous results. Again the advantage of greedy scheduling is consistent. We also preform simulations on networks of different size, density, and link removing probability. The results and conclusions are similar, thus they are ignored here due to the space limit.

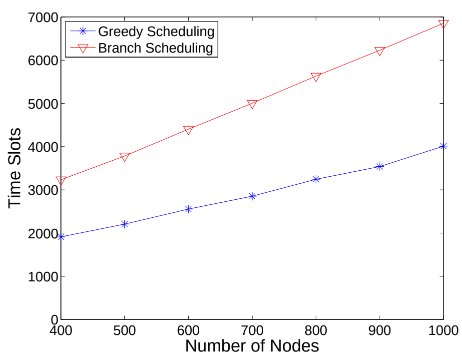


(a) Disk graph model

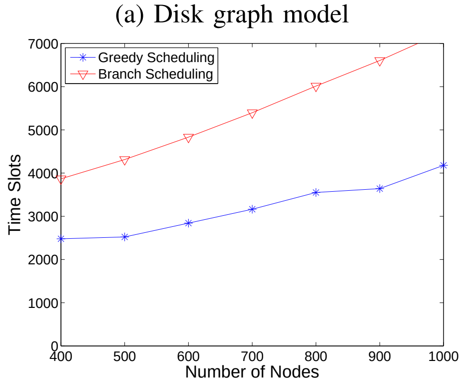


(b) General graph model

Fig. 3. Total time slots used by our data collection algorithms with different network settings. Here, $R = 2r$.



(a) Disk graph model



(b) General graph model

Fig. 4. Total time slots used by our data collection algorithms with different network settings. Here, $R = 4r$.

VI. All Notations

Table I summarizes all notations used in the paper and this supplemental material.

TABLE I

Summary of Notations

Term	Notation or Definition
W, t, b	fixed link rate, size of a time slot, unit packet size
G, V, E	communication graph, set of sensors, set of links
v_0, v_i	the sink, i-th sensor node
v_iv_j	edge/link between v_i and v_j
$\|\|v_i - v_j\|\|$	Euclidean distance between v_i and v_j
$h(v_i, v_j)$	shortest hop number from v_i to v_j in graph G
n	total number of sensors $\|V\|$
r, R, α	transmission range, interference range, and ratio between R and r
D, C	delay and capacity of data collection
T,	BFS tree
c	total number of paths or branches in T
v_i^l, P_i, B_i	i-th leaf, i-th path and i-th branch in T
$\|P_i\|$, $\|B_i\|$	number of links in a path or a branch
$par(v_i)$	parent of v_i in T
$l(v_i)$	level of v_i in T
$\delta(v_j)$, $\delta^{P_i}(v_j)$	interference number of v_j, interference number on path P_i of v_j, both include v_j
Δ_i, Δ	maximum interference number on path P_i of T ($\max\{\delta^{P_i}(v_j)\}$ for all $v_j \in P_i$), maximum interference number in G ($\max\{\delta(v_i)\}$)
$\tilde{\Delta}$	maximum path interference number of T, *i.e.*, $\max\{\Delta_1, \cdots, \Delta_c\}$
$\delta_k^{P_i}$	maximum path interference among first k nodes in path P_i, *i.e.*, $\max\{\delta^{P_i}(v_1), \cdots, \delta^{P_i}(v_k)\}$
Δ^{**}	$\frac{\sum_{i=1}^{c} \sum_{k=\|P_i\|-\|B_i\|+1}^{\|P_i\|} \delta_k^{P_i}}{n}$ for BFS T
$D(v_0, l)$	l-hop neighborhood around v_0
l^*	critical radius around v_0
$D(v_0, l^*)$	critical region around v_0
λ_i^*, λ^*	min number of hops to reach v_0 after entering critical region, $\lambda^* = \max_i\{\lambda_i^*\}$
Δ^*	$\frac{\sum_i \lambda_i^*}{n}$
λ_i, λ	minimal hops that a packet needs to be forwarded from v_i before a new packet at v_i can be safely forwarded along T, $\lambda = \max_i\{\lambda_i\}$
p_i	packet originated from node v_i
ρ_i	priority of p_i defined as $\frac{1}{l(v_i)}$
v_j^τ	the node of packet p_j in the end of time slot τ
D_j	delay of packet p_j, defined as $t \cdot \min\{\tau : v_j^\tau = v_0\}$
$p_k \prec p_j$	p_j blocked by a higher order packets p_k
T_D	blocking tree of all packets
p_r	root of T_D
$P(j)$,$h(j)$	path in T_D from p_r to p_j, hop count of $P(j)$
R_0, r_0	artificial interference and transmission ranges for upper bound in physical interference model
R_1, r_1	artificial interference and transmission ranges for lower bound in physical interference model
B, N_0, β, η	channel bandwidth, background Gaussian noise, pass loss exponent, SINR threshold
I	the set of actively transmitting nodes
P	transmission power of every node
W_{ij}	transmission rate of a link between v_i and v_j
$SINR_{i0}$	SINR of a link v_iv_0