Continuous Data Collection Capacity of Wireless Sensor Networks Under Physical Interference Model

Shouling Ji, Yingshu Li
Georgia State University
Raheem Beyah
Georgia Institute of Technology

Outline

1. Introduction
   • Data gathering
     ◦ Data Aggregation, e.g. MAX, MIN, Average
     ◦ Data Collection
     • Snapshot Data Collection
     • Continuous Data Collection
   • Network Capacity
     ◦ Multicast/Unicast/Broadcast Capacity
     ◦ Snapshot Data Collection Capacity
     ◦ Continuous Data Collection Capacity

2. Interference Model
   • Protocol interference model
   \[ R_i = \rho \cdot r, \rho \geq 1 \]
   \( R_i \) is the interference radius, \( \rho \) communication radius
   • Physical interference model
   \[ \text{SNR} = \frac{P}{N_{e_i}} \sum_{j \in C_i} \frac{1}{d_{ij}} \geq \eta \]
   • Generalized interference model
   \[ R(i, s) = W \log(1 + \text{SNR}) \]
   • In this work, we take the physical interference model

Contribution

• A Cell-Based Path Scheduling (CBPS) for Snapshot Data Collection (SDC)
  • Partition the network into cells
  • Construct a cell based data collection tree
  • Achievable network capacity: \( \Omega(W) \), \( W \) is the bandwidth of a channel
  • Order-optimal

• A Segment-Based Pipeline Scheduling (SBPS) algorithm for Continuous Data Collection (CDC)
  • SBPS combines the Compressive Data Gathering (CDG) technique [20, Luo et al.] and the pipeline technique
  • Schedule continuous data transmission level by level
  • Achievable network capacity:
  \[ \left\{ \begin{array}{ll}
  \frac{\alpha}{\log N} & \text{if } N \leq \frac{1}{\alpha} \\
  \frac{N}{\log N} & \text{if } N > \frac{1}{\alpha}
  \end{array} \right. \]
  \( \alpha \) is the number of nodes in the WSN, \( N \) is the number of snapshots in a continuous data collection task
2. Network Partition

- **Network Model**
  - $n$ sensors, denoted by $s_1, s_2, \ldots, s_n$, and one sink deployed in a square area with size $A = c \cdot n$, $c$ is a constant.
  - The distributions of all the nodes are i.i.d.
  - Losing only a constant factor, the sink is located at the top-right corner of the square.
  - Communication radius: $r$, the size of a data packet: $b$
  - The network time is slotted with each slot of size $t = b/W$
  - Interference model: physical interference model
  - The achievable data collection capacity $C$ = the ratio between the amount of data successfully collected by the sink and the time used to collect these data

- **Network Partition**
  - Partition the network into square cells with side length $2 \log l\log n = \lambda$
  - The number of cells in each row/column is $A$
  - The cell with coordinates $(i, j)$ is denoted by $K_{ij}$

- For large $n$, the probability that a cell is empty is zero (Lemma 1). Thus, we assume each cell is not empty.
- It is almost surely that no cell contains more than $8 \log n$ sensors (Lemma 2). Hence, we use $8 \log n$ as the upper bound of sensors in a cell.

- **Interference zone**
  - Each cell forwards their data horizontally, vertically, or diagonally (upper-right)
  - To identify which cells can transmit data concurrently, further partition the network into large square zones with side length $R = \omega \cdot l$, called interference zones

- The next job is to determine $R$
  - If $R = \omega l$, where $\omega = \sqrt[2]{\frac{c}{2}}$, is a constant, the compatible cells can simultaneously and successfully transmit data without interference (Thm 1)
  - Proof idea of Thm 1: a layered method
3. Snapshot Data Collection

- Construct a data collection tree

- Cell-Based Path Scheduling
  - Schedule paths \( p_1, p_2, \ldots, p_r \) and \( p'_1, p'_2, \ldots, p'_s \) until all the data has been transmitted to \( p_1 \) and \( p_2 \), respectively.
  - Schedule \( p_1 \) and \( p_2 \) until all the data has been collected by the sink.

- Capacity Analysis
  - The number of time slots used by CBPS to collect a snapshot is bounded by \( O(n) \) (Thm 2).
  - The achievable network capacity of CBPS is \( \Omega(W) \), which is order-optimal (Thm 3).

- Discussion
  - Snapshot data collection algorithms + pipeline = continuous data collection algorithms?
  - Data accumulation effect

4. Continuous Data Collection

- Compressive Data Gathering (CDG) [30, Luo et al]
  - Traditional manner
  - CDG manner

- Segments
  - Consisting of zones

- Segment-Based Pipeline Scheduling (SBPS)
  - Scheduling at the segment-level
  - Scheduling at the row/column-level, i.e. within a segment
  - Scheduling at the cell-level, i.e. within each row/column
• Capacity analysis
  - The number of time slots used by SBPS to collect $N$ continuous snapshots is at most
    $$8\omega M \sqrt{2 \log n} + 16 \omega^2 MN \log n$$
    where $\omega$ is a constant, $M$ is a parameter in CDG (Thm 4)
  - The achievable capacity of SBPS is (Thm 5)
    $$\left\{ \begin{array}{ll}
    \frac{2 \omega \log^2 n}{\log n} & \text{if } N \leq \frac{2}{\log n} \\
    \frac{N \log^2 n}{\log n} & \text{otherwise}
    \end{array} \right.$$  

5. Conclusion
• The snapshot data collection under physical interference model is studied for WSNs.
  - A Cell-Based Path Scheduling (CBPS) algorithm is proposed
  - The achievable network capacity is order optimal
• The continuous data collection under physical interference model is studied for WSNs.
  - A Segment-Based Pipeline Scheduling (SBPS) algorithm is proposed
  - By using CDG and pipeline, the network capacity is improved significantly
• Future work
  - Continuous data collection capacity of arbitrary wireless networks
  - Distributed data collection algorithms and their achievable capacity for wireless networks