To model the benefits of data aggregation routing:

Address Centric Model (AC): Each source independently sends its data to the sink node using the shortest path between itself and the sink node.

Data Centric Method (DC): The source nodes send data to the sink, but the intermediate routing nodes look at the content of the data and perform data aggregation on the data originating at multiple source nodes.
Different types of data aggregation
- Duplicate suppression
- Aggregation functions
  - max
  - min
  - average
  - concatenation
  - sum

Comparative analysis between AC & DC

Assumptions: each node in the network generates only one packet of same size.

Notation: k sources S1, ..., Sk, one sink node: D
Network Graph: \( G(V, E) \) with \( \{\text{edges}\} \) and \( \{\text{vertices}\} \).

Performance Metric: total \( \# \) of transmissions
- same as message complexity

\( = \# \) of link in the data aggregation tree (convergecast tree)

The difference is that data is being collected from only a subset of the nodes in the network.

Multicast - when broadcast is restricted to a subset of nodes in the network.

We know that the problem of constructing optimal multicast tree (Steiner tree problem) is \( NP \)-hard.
P problem - polynomial time $O(n)$ $O(n^2)$

$NP^p$ problem - Non-deterministic polynomial time.

$NP^p$ - complete

$NP$ - hard

- Data aggregation tree is reverse of a multicast tree

⇒ optimal data aggregation tree construction problem is also $NP^p$-hard.

- So we will consider only heuristic or suboptimal solutions - which run in polynomial time.
Source node placement model:

- Random Source (RS) placement model
- Event-Radius (ER)

\[ A = \frac{\pi S^2 \cdot n}{A} \]

All nodes within a distance \( S \) (called sensing range) of the event that are not sinks are considered to be source nodes.

Normalize this and assume \( A = 1 \) sq. unit.
\( R \): Transmission radius of each node.

- each node has same transmission radius.

\( N_A \): optimal number of transmissions in the address-centric model

\( N_D \): optimal \# in the data-centric model to perform data aggregation

\[ S_1, S_2, \ldots, S_k \]

\[ d_1, d_2, \ldots, d_k \] — be their distances from the sink node.

\[ N_A = d_1 + d_2 + \cdots + d_k \]

\[ = \sum_{1 \leq i \leq k} (d_i) \] — in terms of hops.

\[ D \]

\[ S_1, S_2, S_k \]
$N_D > N_A$ : by combining we can only reduce the number of transmissions required because we are assuming that data size does not increase due to data aggregation.

$N_D \leq N_A$  

$X$ : diameter of the subgraph containing all the source nodes.

$\max$ of the pairwise shortest path between the source nodes.

$1 \leq X \leq $ diameter of network
Upper bound

\[ N_D \leq (k-1)X + \min(d_i) \]

Lower bound

\[ (k-1) + \min(d_i) \leq N_D \]

\[ N_D \leq N_A \]

Can we find condition under which \( N_D < N_A \)

\[ N_A = \sum d_i = d_1 + d_2 + \ldots + d_k \]

\[ d_i \geq \min(d_i) \]

\[ d_k \geq \min(d_i) \]

\[ d_i \geq \min(d_i) \]

\[ \sum(d_i) \geq k \min(c_i) \]
If \( x < \min(d_i) \) then \( N_D < N_A \)

**Fraction Saving (FS):**

\[
\frac{N_A - N_D}{N_A}
\]

\[
\frac{\text{sum} (d_i)}{-\left[ (k-1)x + \min(d_i) \right]} \leq FS \leq \frac{\text{sum} (d_i) - (\min(d_i) + k-1)}{\text{sum} (d_i)}
\]

\[
1 - \frac{(k-1)x + \min(d_i))}{\text{sum} (d_i)} \leq FS \leq 1 - \frac{(\min(d_i) + k-1)}{\text{sum}(d_i)}
\]

\[
\lim_{d \to \infty} FS = 1 - \frac{1}{k} \frac{N_D}{N_A}
\]

Assume all \( d_i \) are equal to \( d \).