INTRODUCTION AND MOTIVATION
- In many applications like military surveillance or habitat monitoring, sensors are required to send reports to a specific target (e.g. base station) periodically.
- The data gathering process is repetitive, hence it is necessary to reduce the number of packets and the transmission cost.
- A tree is used as the routing structure in many data aggregation algorithms, especially for the applications which monitor events continuously.
- In this paper, we have discussed the problem of finding the best data fusion points in a tree which maximizes the data accuracy and minimizes the transmission cost.

DATA AGGREGATION TREE
Let G = (V,E) be a graph with V vertices and E edges. Let c(E) be a cost function. Now, assume we have $S \subseteq V$, terminal vertices. The problem of finding a minimum steiner tree in graph is finding a subgraph, $G'=(V',E')$ such that:
1. $V' \subseteq S$.
2. $G'$ is connected.
3. $\text{Cost} = \sum c(E')$ is minimum.

ALL PAIR SEARCH
- All pair search is a dynamic approach for finding the combination of sensors in the network which has the maximum data quality and the minimum transmission cost.
- The algorithm uses the results of the previous step and checks for the following conditions to find the result of the current step:
  1. For the combination of sensors, S1 and S2:
     1. Check independent path of S2 to sink.
     2. Check path of S2 through S1 to sink.
     3. Check path of S2 through an intermediate node between S1 and sink.
- Selecting different data fusion points have no impact on data quality, but cost can be improved by selecting different data fusion points.

LEVEL ORDER SEARCH
- This algorithm uses level order traversal for searching the sensor nodes from the sink node.
- The optimal data aggregation tree for a group of sensors is formed using the below algorithm.
- The algorithm terminates when it satisfies the application requirements.

KALMAN FILTER
- Kalman filter is a computationally efficient way of estimating the variable of interest with the help of the set of equations.
- It is ideal for situations where the information about the linear dynamic system is uncertain.

\[ KG = \frac{E_{\text{EST}}}{E_{\text{EST}} + E_{\text{MEA}}} (1 - KG)E_{\text{EST}_{t-1}} \]
\[ E_{\text{ST}_{t}} = E_{\text{ST}_{t-1}} + KG(MEA - E_{\text{ST}_{t-1}}) \]

EXPERIMENTAL RESULTS

<table>
<thead>
<tr>
<th>Number of nodes</th>
<th>All pair search time (in ms)</th>
<th>Level order search time (in ms)</th>
<th>All pair search data accuracy</th>
<th>Level order search data accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2003</td>
<td>46</td>
<td>98.68%</td>
<td>93.15%</td>
</tr>
<tr>
<td>50</td>
<td>10019</td>
<td>123</td>
<td>97.89%</td>
<td>92.73%</td>
</tr>
<tr>
<td>100</td>
<td>-</td>
<td>268</td>
<td>-</td>
<td>94.67%</td>
</tr>
<tr>
<td>500</td>
<td>-</td>
<td>965</td>
<td>-</td>
<td>91.59%</td>
</tr>
</tbody>
</table>

CONCLUSION AND FUTURE WORK
- Level order search is faster than all pair search but at the cost of data quality.
- All pair search fails for a large number of nodes due to the memory limitation.
- Implement a heuristic approach and compare its performance for data quality and transmission cost.
- Compare different data fusion techniques for best results.

References