Chapter 6

Analytical Modeling of Strategies

While experimental results indicate that emergent optimization comes close to globally optimal solutions, it would be interesting to understand the individual strategies analytically. Unfortunately, for all but the simplest of strategies, analytical modeling involves modeling the behavior of closed-loop feedback systems, making it very complex. This chapter focuses on building analytical models explaining the behavior of some of the strategies. The focus of this chapter is in constructing an analytical framework forming the basis over which richer models can be built to explain more complex strategies. The model described here makes the following assumptions.
6.1 Assumptions

As described earlier in 1.1, the grid is modeled as $G = (X, d)$, where $X$ represents all the grid nodes and $d : X \times X \rightarrow \mathbb{R}^+$ is a distance function encapsulating latency between nodes. The grid is also assumed to have the following properties,

- $\forall x \in X, d(x, x) = 0$
- Triangle inequality holds for a multi-dimensional grid: $\forall x, y, z \in X, x \neq y \neq z, d(x, z) \leq d(x, y) + d(y, z)$
- The grid is divided into discrete regions such that the distance between consecutive regions is 1.

The following assumptions are expected to hold over the grid nodes which make up the grid,

- There is a single node present in each region of the grid
- The probability of nodes being a sensor node is equal
- Nodes receiving data from a source can behave as secondary sources.
- The effective number of sources $\mathcal{E}(q)$ for a query $q$ is defined as the sum of number of primary and secondary sources having the required data to answer query $q$.

The queries which arrive at the grid node have the following properties,

- Queries are either project or select queries
• Queries arrive on the nodes with equal probability

• Queries are selected from a query set $Q$ containing all the possible simple queries

• Queries are selected with uniform probability from $Q$. This would also mean that the $\mathcal{E}(q_i) = \mathcal{E}(q_j)$ for all $q_i, q_j \in Q$. Hence $\mathcal{E}(q_i)$ is the same for a given grid.

### 6.2 Single Dimensional Modeling

Given the above assumptions, a grid on a single dimension where all the nodes are placed along a straight line would be represented as shown in Figure 6.1. The grid shown in Figure 6.1 has four nodes in a straight line and each node in a discrete region $R_1$ to $R_4$ with the distance between any two consecutive regions being 1.

![Figure 6.1: Straight Line Grid](image)

If the effective number of sources for a grid having $N$ nodes is $\mathcal{E}$, the number of possible source locations is given by $\binom{N}{\mathcal{E}}$. In the grid shown in Figure 6.1 if $\mathcal{E} = 2$ then, the number of possible locations for sources is $\binom{4}{2}$ and the possible source locations are, $\{R1,R2\}, \{R1,R3\}, \{R1,R4\}, \{R2,R3\}, \{R2,R4\}, \{R3,R4\}$. 
A query arriving on the node in region selects a single node from the set of nodes in \( \binom{N}{2} \) possibilities based on the source selection strategy adopted by the node. For the grid in Figure 6.1, a query arriving on the node in region \( R_1 \), using strategy DO would select a node as given in Table 6.1 from the set of \( \binom{4}{2} \) source location possibilities.

<table>
<thead>
<tr>
<th>Source Location Possibilities</th>
<th>Selected Node (DO Strategy Used)</th>
</tr>
</thead>
<tbody>
<tr>
<td>{R1,R2}</td>
<td>R1</td>
</tr>
<tr>
<td>{R1,R3}</td>
<td>R1</td>
</tr>
<tr>
<td>{R1,R4}</td>
<td>R1</td>
</tr>
<tr>
<td>{R2,R3}</td>
<td>R2</td>
</tr>
<tr>
<td>{R2,R4}</td>
<td>R2</td>
</tr>
<tr>
<td>{R3,R4}</td>
<td>R3</td>
</tr>
</tbody>
</table>

Table 6.1: Source Selection Table

\( p(R_x, R_y) \) denotes probability of a node \( R_y \) being selected to answer a query incident on node \( R_x \) and is dependent on the source selection strategy. Assuming each of the source location possibilities to be equally probable, the selections described in Table 6.1 would lead to \( p(R_1, R_1) = \frac{1}{2} \), \( p(R_1, R_2) = \frac{1}{3} \) and \( p(R_1, R_3) = \frac{1}{6} \) at any given time.

The expected network usage \( E[u(R_x)] \) of answering a query at node \( R_x \) in a grid of \( N \) nodes is given as,

\[
E[u(R_x)] = \sum_{i=1}^{N} p(R_x, R_i) \cdot d(R_x, R_i)
\]  

(6.1)

From Table 6.1, it can be seen that the probability of a query on \( R_1 \) being answered by \( R_1 \) is \( p(R_1, R_1) \) or \( \frac{1}{2} \). Similarly, when a query is on node \( R_2 \), the
probability that node $R_1$ answers the query is given by $p(R_2, R_1)$. If there are a total of $N$ queries in the system, given that all the queries are uniformly distributed on the grid nodes, the total load on node $R_1$ is $\sum_{i=1}^{N} p(R_i, R_1)$.

Generalizing, the expected load $E[L(R_x)]$, on node $R_x$ for answering queries is given as,

$$E(L(R_x)) = \sum_{i=1}^{N} p(R_i, R_x)$$  \hspace{1cm} (6.2)

For a set of $N$ queries incident on the grid in Figure 6.1, Figure 6.2 shows the expected network usage and Figure 6.3 shows the expected load distribution characteristics when the grid nodes adopt the DO strategy and the RO strategy.

It can be seen from Figure 6.2 that the expected network usage for the distance ordering strategy is lesser than that of the random ordering strategy. This is because when nodes use the random ordering strategy, the source selected for
answering a particular query may not be the nearest one available. However it is because of this random behavior that the expected load is evenly distributed between the nodes as seen in Figure 6.3. When the grid nodes adopt the distance ordering strategy, the expected load on the grid nodes towards the center of the grid are seen to be loaded more than the fringe nodes. This is because the average distance from any node to the centrally located nodes is less compared to the nodes located at the fringe of the grid.

While analytical modeling of RO strategy is simple, DO and LO strategies are history dependent. By assuming that the probability of source location possibilities as uniform, it is possible to model the DO strategy. However, assuming that all the nodes have uniform load to model the LO strategy does not provide any additional information. Hence, the LO strategy is not modeled in this work.
6.3 Two Dimensional Modeling

The single dimensional model explained in the previous section can be extended to two dimensions by considering a grid on a plane instead of a line. The equations for calculating the expected network usage and load remain the same. For a grid with 16 nodes in a two dimensional plane, arranged in a $4 \times 4$ matrix, the comparative expected network usage when the grid nodes adopt the DO and RO strategy is shown in Figure 6.4. As seen from Figure 6.4, the expected network usage is higher when the grid nodes use the RO strategy compared to when the grid nodes use the DO strategy.

![Expected Network Usage](image)

Figure 6.4: Expected Network Usage

The expected load distribution in the grid when the nodes adopt the DO and RO strategy is shown in Figure 6.5. As with a single dimensional grid, the expected load is even when the grid nodes use the RO strategy and skewed when the
nodes use the DO strategy. The reason for the nodes towards the center of the grid having a higher expected load compared to the fringe nodes is the same as in the single dimensional grid.

Modeling the distance ordering and random ordering strategy provides insights into the network usage and load distribution characteristics of these two strategies. Although the load ordering strategy is not modeled, it is expected to provide better load distribution compared to both distance and random ordering as neither of these two strategies consider the load on a grid node while selecting source node.

From the query optimization perspective it is imperative that a grid node is able to select the correct source selection strategy based on the query pattern and load to efficiently optimize on network usage or load distribution. In the experiments performed in the next section, all three strategies are used by the grid
nodes and it is shown that the grid nodes are able to automatically switch between source selection strategies to optimize in the presence of varying query patterns.