**Task 5 (30 marks).** Write a brief report (no more than 3 A4 pages) containing the following:

1. a) A short explanation of your choice of data structure and algorithm.
2. b) A run of your algorithm on a small benchmark example. This should include the supporting information as described in Task 4.
3. c) A performance analysis of your algorithmic design and implementation. This can be based either on an empirical study, e.g., doubling hypothesis, or on purely theoretical considerations, as discussed in the lectures and tutorials. It should include a suggested order-of-growth classification (Big-O notation).
4. ---------------------------------------------------------------------------------------------------------------------------------------

Algorithmic Approach

Dijkstra Algorithm was selected to be used in the implementation.

For similar problems where path finding algorithms can be used, some of the most used algorithms are A\* and Breadth-First Search (BFS). BFS is very much similar to Dijkstra.

The same puzzle can be solved using A\* and Breadth First Search (BFS).

**Why Dijkstra is better than A\* and Breadth First Search (BFS)?**

Both BFS and Dijkstra fall under SSSP (Single Source Shortest Path) algorithms where it solves the problem of finding the shortest path from a starting node (source) to all other nodes inside the graph. After the algorithm ends, there will be the shortest paths from the source node to all other nodes in the graph.

A\* algorithm is relatively faster than Dijkstra since it is an improvised version of Dijkstra, but if you have many target nodes and you don't know which one is closest to the main one, A\* is not very optimal. This is because it needs to be run several times (once per target node) to get to all of them.

Dijkstra is an uninformed algorithm. This means that it does not need to know the target node beforehand. For this reason, it's optimal in cases where you don't have any prior knowledge of the graph when you cannot estimate the distance between each node and the target. Since Dijkstra picks edges with the smallest cost at each step it usually covers a large area of the graph. This is especially useful when you have multiple target nodes, but you don't know which one is the closest.

**Why Priority Queue?**

The chosen data structure was Priority Queue.

In Dijkstra's algorithm, the important step is selecting an unexplored vertex *v*such that there is an edge *(u, v)*in the graph, where *u*is an already explored vertex, and *d'(v) = dist(u)*+*cost(u, v)* is minimum. Here, *dist(u)* is the length of the already found shortest path from the source vertex *s*to vertex *u*, and *cost(u, v)* is the weight of the edge from *u*to *v*.

If we store the unexplored vertices in a simple array/linked list, we would have to iterate over the whole list each time to find the desired vertex *v,*with minimum *d'(v)*(*d'(v) = dist(u) + cost(u, v)*).

If we store the vertices in a priority queue with *d'(v)* as the key for each vertex, we can get the vertex with minimum *d'(v)* by using the Extract-Min operation. If we use a binary (min-)heap, the asymptotic complexity of Extract-Min operation will be ***O(log n)***.

After selecting vertex *v*, and updating *dist(v)*, we may find that we have a shorter path to another unexplored vertex *w*through *v*, i.e. a path with cost *d'(w) = dist(v) + cost(v, w)*, which is less than the cost of the existing path to *w*.

This vertex will be in the priority queue, and its key value will need to be changed to the new value using the Decrease-Key operation, which will decrease the key of a certain element, and bubble it up if necessary to ensure the min-heap property. A binary (min-)heap will support doing Decrease-Key operation with ***O(log n)*** complexity.

Again, if we had used a simple list, this step would have required us to iterate over the whole list to update the key of a vertex.

Examples of running the Algorithm on Benchmark examples

Time taken for the largest puzzle with 2688 rows and 2688 columns with 98417 steps.

maze25\_1.txt

A screenshot of a computer

Description automatically generated with medium confidence

Empirical Analysis of Algorithm Performance

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **File name** | **No. of elements** | **Trial 1** | **Trial 2** | **Trial 3** | **Trial 4** | **Trial 5** | **Average** | **Change of Ratio** |
| puzzle\_21.txt | 441 | 0.071 | 0.069 | 0.062 | 0.063 | 0.059 | 0.0648 | 1.614198 |
| puzzle\_42.txt | 1764 | 0.165 | 0.112 | 0.08 | 0.086 | 0.08 | 0.1046 | 1.397706 |
| puzzle\_84.txt | 7056 | 0.11 | 0.134 | 0.187 | 0.177 | 0.123 | 0.1462 | 1.826265 |
| puzzle\_168.txt | 28224 | 0.248 | 0.299 | 0.323 | 0.233 | 0.232 | 0.267 | 1.719101 |
| puzzle\_336.txt | 112896 | 0.339 | 0.558 | 0.522 | 0.501 | 0.375 | 0.459 | 1.934205 |
| puzzle\_672.txt | 451584 | 0.86 | 0.856 | 0.865 | 1.06 | 0.798 | 0.8878 | 3.224825 |
| puzzle\_1344.txt | 1806336 | 2.857 | 2.78 | 2.946 | 2.8 | 2.932 | 2.863 | 4.098708 |
| puzzle\_2688.txt | 7225344 | 11.565 | 12.044 | 11.216 | 12.066 | 11.782 | 11.7346 | 0 |

steps for the shortest path

To accurately describe the algorithms efficiency Big O Notation is used