2025 Alexandria Quantum Haqathon

Smart Traffic Optimization in the New Capital

Final Report

Team Members:
Bassel Ahmed
Mohamed Abo-Zeid
Omar Ayoub
Hassan Khalifa
Youssef Rezk

Mentors:
Ahmed Saad El Fiky
James Austin Myer

Why Smart Traffic Optimization Matters

Traffic optimization in the New Administrative Capital is not simply a matter of convenience; it is a matter of saving lives. Emergency response time is one of the most critical determinants of survival in urgent medical and accident cases. Research shows that:

- Survival to 30 days is **19.5%** when emergency medical services (EMS) arrive within 0–6 minutes, but drops to only **9.4%** when arrival is delayed to 10 minutes or later.
- Each minute of delay in CPR reduces the likelihood of neurologically favorable 1-month survival by approximately **6.4%**.
- Similar time-critical dependencies exist in trauma care, heart attack response, and stroke management, where faster arrival directly translates into higher chances of survival and reduced long-term complications.

In rapidly growing urban environments like the New Administrative Capital, traffic congestion poses a major challenge for timely EMS access. By applying intelligent traffic optimization strategies, including real-time route management, AI-based prediction, and smart signaling, we can drastically reduce travel time for ambulances and other emergency vehicles. This ensures that medical teams reach patients faster, increasing survival rates and improving overall public safety.

Problem Statement

Emergency Patient Transportation Challenge Scenario:
Five patients need urgent transport to a hospital. You have one ambulance.

Constraints:

- Each ambulance can make multiple trips.
- Maximum 3 stops per trip (3 patients per trip).
- All patients must reach the hospital.

Objective:
Find the optimal routes that minimize total travel distance.

Mathematical Formulation

Mathematically, the problem reduces to the discovery of a **shortest path** on a directed weighted graph, where:

- Each patient location is a vertex.
- The first vertex is reserved for the location of the hospital.
- Edge weights correspond to travel distances.

The cost matrix is obtained by transforming map API data into an adjacency matrix, which facilitates computation and optimization.

Classical Solution Approaches

Several classical optimization methods can be applied to this problem, including:

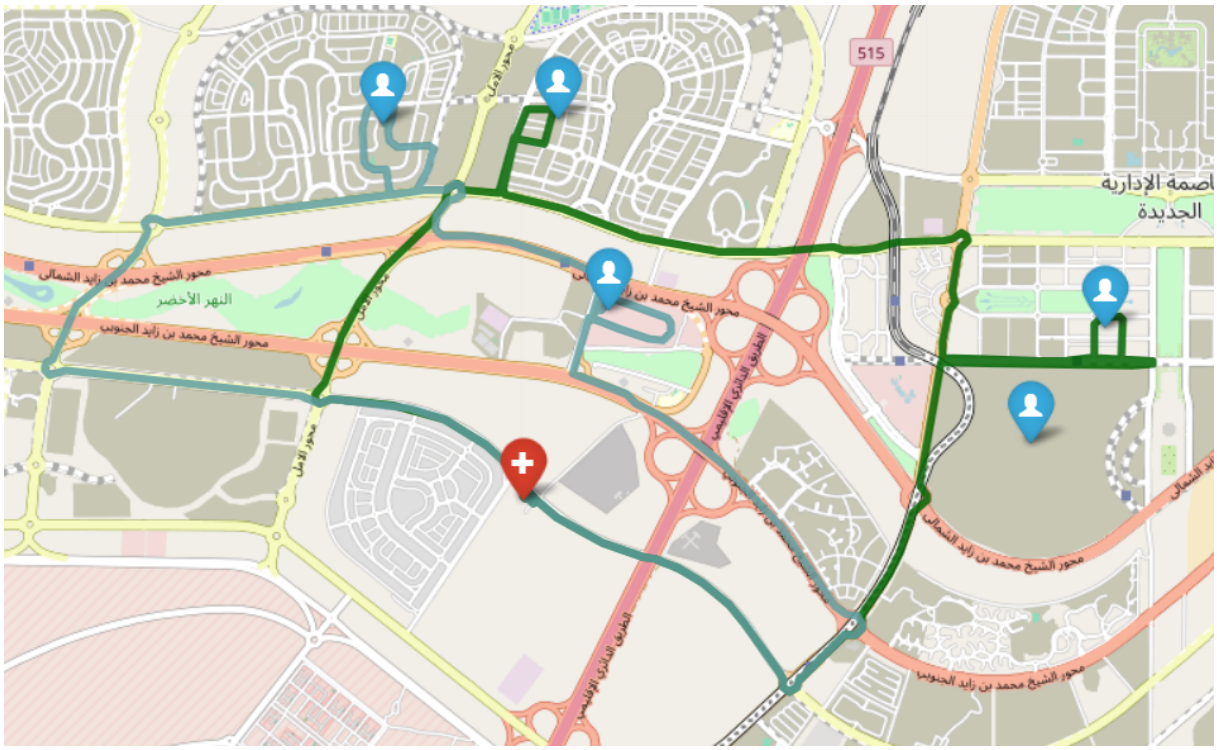
- **Brute Force:** Enumerates all possible routes. Guarantees optimality but becomes computationally infeasible as the number of patients grows.
- **A* Algorithm:** A graph-based heuristic search that uses a cost function combining path cost and an admissible heuristic to efficiently find near-optimal routes.
- **Heuristic Methods:** Approaches such as greedy algorithms or local search techniques that provide quick, approximate solutions for large-scale problems.
- **MILP (Mixed Integer Linear Programming):** A mathematical optimization approach that formulates the routing problem with linear constraints and objective, solvable by solvers like CPLEX or Gurobi.

Classical Results

The following table summarizes the performance of classical algorithms tested on the emergency routing problem:

Table 1: Classical algorithms: total distance and execution time

Algorithm	Total Distance (km)	Execution Time (s)
Brute Force	57.311100	0.0006
A* Search	57.311100	0.0008
OR-Tools	57.311100	0.0211
Heuristic	57.311100	0.0001



Example of classical solution results.

Quantum Approach for Emergency Vehicle Routing

The advent of quantum computing presents unprecedented opportunities to solve complex optimization problems that are intractable for classical computers. Our quantum approach transforms the emergency vehicle routing problem into a Quadratic Unconstrained Binary Optimization (QUBO) formulation, enabling solution on quantum annealers and gate-based quantum computers.

General Quantum Approach: Initial Problem Identification

Drawing inspiration from research on capacitated vehicle routing problems with column generation and reinforcement learning techniques, we identified our emergency patient transportation problem as a **Capacitated Vehicle Routing Problem (CVRP)** with specific constraints.

Mathematical Formulation:

The objective function minimizes total travel distance: [1] [3]

$$\min \sum_{k \in K} \sum_{i \in V} \sum_{j \in V} c_{ij} x_{ijk} \tag{1}$$

Where:

- K indexes the trip number
- V indexes the vertices/patients
- c_{ij} represents the travel cost between vertices i and j
- x_{ijk} is a binary decision variable

Key Constraints:[1]

1. Vehicle Leaves Node that it Enters

Ensure that the number of times a vehicle enters a node is equal to the number of times it leaves that node:

$$\sum_{i=1}^{n} x_{ijk} = \sum_{i=1}^{n} x_{jik} \quad \forall j \in \{1, ..., n\}, k \in \{1, ..., p\}$$

2. Ensure that Every Node is Entered Once

$$\sum_{k=1}^{p} \sum_{i=1}^{n} x_{ijk} = 1 \quad \forall j \in \{2, ..., n\}$$

Together with the first constraint, it ensures that every node is entered only once, and it is left by the same vehicle.

3. Every Vehicle Leaves the Depot

$$\sum_{j=2}^{n} x_{1jk} = 1 \quad \forall k \in \{1, ..., p\}$$

Together with constraint 1, we know that every vehicle arrives again at the depot.

4. Capacity Constraint

Respect the capacity of the vehicles. Note that all vehicles have the same capacity.

$$\sum_{i=1}^{n}\sum_{j=2}^{n} q_j x_{ijk} \leq Q \quad \forall k \in \{1, ..., p\}$$

General Quantum Approach: Problem Refinement

The initial formulation required a prohibitive number of qubits:

$$\text{Total qubits} = \text{trips} \times \text{initial vertices} \times \text{final vertices} = 5 \times 6 \times 5 = 180 \tag{2}$$

To make the problem tractable for current quantum hardware, we refined our approach by assuming the optimal solution requires exactly 2 trips, reducing the qubit count to 60.

Cost Function Optimization

We further reduced the qubit requirement from 60 to 30 by encoding the "2-trip hypothesis" as joint constraints within the objective function:

$$\min \left\{ \sum_{i=1}^{6}\sum_{j=1}^{6} d_{i,j} \cdot x_{i,j} \mid i \neq j, x_{i,j} \in \{0, 1\} \right\} \tag{3}$$

Where $d_{i,j}$ represents the distance matrix between locations, and $x_{i,j}$ indicates whether the ambulance travels directly from location i to location j.

Quantum Constraint Formulation

1. Enter-Leave Constraint:

To ensure the ambulance returns to the hospital after each trip, we require that each vertex is entered as many times as it is left:

$$\sum_{i=1}^{6} x_{i,j} = \sum_{i=1}^{6} x_{j,i}, \quad \forall j \tag{4}$$

2. Two-Trip Constraint:

We limit the number of trips to exactly 2 by constraining the number of outgoing edges from the hospital (vertex 1):

$$\sum_{j=2}^{6} x_{1,j} = 2 \tag{5}$$

3. No Single-Patient Trips:

To prevent inefficient single-patient trips and ensure the constraint yields only the desired solution $\{2, 3\}$, we disallow the case $\{1, 4\}$:

$$x_{1,j} + x_{j,1} < 1, \quad \forall j \in \{2, 3, 4, 5, 6\} \tag{6}$$

4. Visit All Patients:

To ensure all patients are visited, every patient vertex must have exactly one incoming edge from a different vertex:

$$\sum_{i=1}^{6} x_{i,j} = 1, \quad \forall j \in \{2,3,4,5,6\} \tag{7}$$

QUBO Formulation

Following the methodology of Glover et al., we incorporate constraints into the objective function using penalty methods. The constraint-to-penalty conversion follows established patterns: [2]

Classical Constraint	Equivalent Penalty
$x + y \leq 1$	$P(xy)$
$x + y \geq 1$	$P(1 - x - y + xy)$
$x + y = 1$	$P(1 - x - y + 2xy)$
$x \leq y$	$P(x - xy)$
$x_1 + x_2 + x_3 \leq 1$	$P(x_1x_2 + x_1x_3 + x_2x_3)$
$x = y$	$P(x + y - 2xy)$

The final QUBO formulation incorporates all constraints as penalty terms:

$$P_1 := P \cdot \left(\sum_{i=1}^{6} x_{i,j} + \sum_{i=1}^{6} x_{j,i} - 2 \sum_{i=1}^{6} x_{i,j} \sum_{i=1}^{6} x_{j,i} \right) \tag{8}$$

$$P_2 := P \cdot \left(1 - \frac{1}{2} \cdot \sum_{j=2}^{6} x_{1,j} + 2 \cdot \frac{1}{2} \cdot \sum_{j=2}^{6} x_{1,j} \cdot x_{1,6} \right) \tag{9}$$

$$P_3 := P \cdot x_{1,j} \cdot x_{j,1} \tag{10}$$

$$P_4 := P \cdot \left(1 - \sum_{i=1}^{6} x_{i,j} + 2 \sum_{i=1}^{6} x_{i,j} \cdot x_{i,6} \right) \tag{11}$$

Quantum Formulations at Different Qubit Scales

We explored multiple formulations of the emergency patient transportation problem, each with a different qubit requirement depending on the encoding strategy, constraint embedding, or simplifying assumptions. This section documents each formulation separately.

180 Qubits Formulation

The initial full encoding of the emergency patient transportation problem required 180 qubits. This formulation represented the most comprehensive problem by explicitly modelling all trips and patient vertices without introducing any simplifying assumptions. This qubit count would be necessary when implementing the actual objective function that inherently considers all potential connections between vertices, prior to the application of specific constraints like disallowing self-loops ($i = j$) to prune the solution space. Due to its scale, this formulation was deemed theoretically correct but impractical for near-term quantum devices.

- **Underlying Objective Function Structure (General):** The objective function for this approach minimizes total travel distance and is broadly represented as:

$$\min \sum_{k \in K} \sum_{i \in V} \sum_{j \in V} c_{ij} x_{ijk}$$

Here, K indexes the trip number, V indexes the vertices/patients, c_{ij} represents the travel cost between vertices i and j, and x_{ijk} is a binary decision variable. This general form allows for all $i, j \in V$ potentially, including $i = j$.

150 Qubits Formulation

A subsequent refinement of the full model successfully reduced the qubit requirement to 150 qubits. This reduction was primarily achieved by eliminating redundant trip variables. This indicates a refinement where variables representing travel from a location to itself ($i = j$) were removed, as such trips are considered redundant in a routing context. Despite this optimization, the model's structure still mirrored the full Capacitated Vehicle Routing Problem (CVRP) formulation.

[label=]

- **Implicit Constraint (Eliminating Redundant Trips):** While no specific equation for this qubit count is provided in the sources, the elimination of redundant trip variables implicitly means that connections where $i = j$ are excluded from consideration. This principle is explicitly seen in later, more refined formulations, such as the 30-qubit model's objective function:

$$\min \sum_{i=1}^{6} \sum_{j=1}^{6} d_{i,j} \cdot x_{i,j} \Big| i \neq j, x_{i,j} \in 0, 1$$

This illustrates the exclusion of $i = j$ connections, thereby reducing the variable set.

72 Qubits Formulation

This formulation involved limiting the number of trips and partially embedding constraints, making it suitable as a mid-scale test case. The maximum number of trips was reduced, and constraints for trip balance were simplified. This represented an intermediate step towards making the problem tractable on current quantum hardware by reducing its overall complexity.

- **Formulation Details:** The provided sources describe this formulation as "Simplified Trips" where the "maximum number of trips" was reduced and "constraints for trip balance simplified". However, the sources do not provide a specific set of distinct mathematical equations solely associated with the 72-qubit formulation. The reduction of the maximum number of trips would imply a smaller set K in the initial general objective function and constraints.

60 Qubits Formulation

This formulation corresponds to the "two-trip hypothesis". The significant reduction in qubit count to 60 was achieved by assuming the optimal solution requires exactly two

trips per ambulance. This assumption greatly reduced the problem size while maintaining feasibility. Additionally, in such refined routing models, it is an inherent property that travel from a location to itself ($i = j$) is not permitted, thereby only considering valid travel segments where $i \neq j$. While the primary driver for this specific qubit count reduction was the "two-trip hypothesis," the $i \neq j$ constraint would be a standard part of such a refined routing formulation.

- **Two-Trip Constraint (Illustrative):** The "two-trip hypothesis" is a key simplification for this formulation. An explicit constraint that reflects this assumption, as seen in later detailed formulations (e.g., the 30-qubit model), is:

$$\sum_{j=2}^{6} x_{1,j} = 2$$

  This constraint enforces exactly two departures from the hospital (vertex 1), thus limiting the problem to two trips.

- **No Self-Loops ($i \neq j$):** Similar to other routing refinements, travel from a location to itself is excluded. This principle is embedded in the objective function structure, as exemplified in the 30-qubit formulation:

$$\min \left(\sum_{i=1}^{6} \sum_{j=1}^{6} d_{i,j} \cdot x_{i,j} \right) \quad \text{s.t.} \quad x_{i,j} \in \{0, 1\}, \quad i \neq j$$

  This ensures that only valid travel segments between distinct locations are considered.

35 and 45 Qubits Formulation

The 35- and 45-qubit formulations represent the most complete and theoretically correct versions of the emergency patient transportation problem. Unlike the 30-qubit formulation (which omitted capacity logic), these models incorporate all essential constraints directly into the objective function.

Variable Set (35 Qubits).

- **Routing variables** ($x_{i,j}$): 30 binary variables representing whether the ambulance travels from location i to location j ($i, j \in \{0, \ldots, 5\}, i \neq j$).
- **Assignment variables** (a_p): 5 binary variables assigning each patient p to either the 3-patient trip ($a_p = 1$) or the 2-patient trip ($a_p = 0$).

Core Objective.

$$H_{\text{distance}} = \sum_{i=0}^{5} \sum_{\substack{j=0 \\ j \neq i}}^{5} d_{ij}\, x_{i,j}$$

This term minimizes the total travel distance by summing all active edges.

Constraints as Penalties.

[label=**Constraint :**, leftmargin=*]

1. **3/2 patient split:**
$$\lambda_1 \Big(\sum_{p=1}^{5} a_p - 3 \Big)^2$$
Ensures exactly three patients are assigned to the 3-patient trip.

2. **Two distinct trips:**
$$\lambda_2 \Big(\sum_{j=1}^{5} x_{0,j} - 2 \Big)^2 + \lambda_3 \Big(\sum_{i=1}^{5} x_{i,0} - 2 \Big)^2$$
Enforces exactly two departures and two returns to the hospital.

3. **Service each patient once:**
$$\lambda_4 \sum_{p=1}^{5} \Big(\sum_{i=1}^{5} x_{i,p} - 1 \Big)^2 + \lambda_5 \sum_{p=1}^{5} \Big(\sum_{j=1}^{5} x_{p,j} - 1 \Big)^2$$
Guarantees exactly one arrival and one departure per patient.

4. **Link routes to assignments:**
$$\lambda_6 \sum_{p=1}^{5} \sum_{\substack{q=1 \\ q \neq p}}^{5} x_{p,q} \cdot (a_p - a_q)^2$$
Prevents routes between patients assigned to different trips.

5. **Prevent sub-tours:**
$$\lambda_7 \sum_{p=1}^{5} \sum_{\substack{q=1 \\ q \neq p}}^{5} (x_{p,q} + x_{q,p}) \cdot \Big[a_p a_q + (1 - a_p)(1 - a_q) \Big]$$
Eliminates 2-patient loops for patients on the same trip.

Final Objective (35 Qubits).

$$H = H_{\text{distance}} + \sum_{i=1}^{7} \lambda_i \cdot (\text{Constraint } i)$$

This is the complete formulation with 35 qubits. It is logically sufficient to model the problem but introduces higher-order (cubic) terms.

Extension to 45 Qubits. To convert cubic terms into quadratic form (required for QUBO), ancilla variables are introduced:

- **30 routing variables** ($x_{i,j}$).
- **5 assignment variables** (a_p).
- **10 ancilla variables** ($b_{p,q}$), each representing the product $a_p \cdot a_q$.

This yields a full quadratic 45×45 QUBO matrix that exactly encodes the problem.

Evaluation.

- The **30-qubit model** was incomplete and failed to enforce capacity.
- The **35-qubit model** is logically correct but involves higher-order interactions.
- The **45-qubit model** is the exact quadratic reduction but practically infeasible on current quantum hardware due to size and exponential complexity.

Hybrid Alternative. A practical approach is to avoid the full 45-qubit model by:

1. Solving small QUBO instances ($\sim$10 qubits) for trip subsets.
2. Using a classical outer loop to iterate over all $\binom{5}{3} = 10$ patient assignments.

This hybrid decomposition is computationally tractable and compatible with near-term quantum devices.

30 Qubits Formulation

The 30-qubit formulation represents the central practical model in this study. It balances between capturing realistic ambulance routing constraints and remaining feasible for execution on near-term quantum devices.

Problem Setup.

- Five patients and one hospital were modeled.
- Two ambulance trips were allowed ($num\_trips = 2$).
- Each trip could serve at most three patients ($max\_stops = 3$).
- A precomputed distance matrix was used as input cost coefficients.

QUBO Encoding.

- Binary edge variables $x_{i,j}$ indicate whether the route goes directly from node i to node j.
- Objective function:
$$\min \sum_{i \neq j} d_{ij}\, x_{i,j} + \text{penalty terms}$$
where d_{ij} are travel distances.
- Constraints were embedded as quadratic penalties:
  1. Each patient has exactly one incoming and one outgoing edge.
  2. Exactly two departures and two returns from the hospital.
  3. No single-patient loops (forbid $H \to p \to H$).
  4. Global trip capacity: total number of patient visits $\leq num\_trips \times max\_stops$.
- A penalty weight $P = 1000$ was chosen, sufficiently large relative to distances to enforce constraint satisfaction.

Solution Method.

- The problem was built as a `QuadraticProgram` and converted into QUBO.
- Optimization used **QAOA** with COBYLA (50 iterations) on the AerSimulator backend.
- QAOA depth was set to $p = 2$.

Results.

- First trial total distance: 57.5244 km,

  constraints satisfied: true,

  execution time: 234.2312946365421 sec.
- Second trial total distance: 63.107 km,

  constraints satisfied: true,

  execution time: 265.7258050441742 sec.

This 30-qubit formulation is the reference model of the project. It demonstrates how the emergency transportation problem can be realistically encoded within the constraints of near-term quantum devices, while enforcing both routing structure and capacity limits.

25 Qubits Formulation

This formulation adopts a compact set-partitioning strategy in which each feasible trip is treated as a single decision unit. All admissible subsets of requests, with cardinality not exceeding the allowed capacity, are enumerated, and for each subset the minimum round-trip distance from and back to the depot is precomputed. A binary variable is then associated with every such subset, indicating whether that trip is selected in the final plan. The optimization problem is expressed as a linear objective that minimizes the sum of the distances of the chosen trips, subject to coverage constraints requiring every request to appear in exactly one selected subset. Because the variables correspond to whole trips rather than individual arcs, the formulation drastically reduces the number of qubits needed while preserving exact feasibility of the routing structure. The resulting binary program is converted into an Ising-type Hamiltonian and solved using the Quantum Approximate Optimization Algorithm with depth two, executed on a simulator. This approach balances accuracy and hardware efficiency by encoding only high-level trip choices, producing a 25-qubit instance that remains tractable for near-term quantum devices while retaining a clear interpretation of the solution in terms of complete routes.

14 Qubits Formulation

The 14-qubit formulation was developed as a compact, testable version of the emergency patient transportation problem. The goal was to preserve core constraints (patient assignment and vehicle capacity) while minimizing qubit usage through the use of binary slack variables and simplified encoding.

Problem Setup. Let:

- $\mathcal{P}$: set of patients
- $\mathcal{T} = \{1, 2, \ldots, n_{\text{trips}}\}$: set of trips
- $x_{p,t} \in \{0, 1\}$: binary variable indicating if patient p is assigned to trip t
- $s_{t,i} \in \{0, 1\}$: binary slack variables for trip t (where $i = 0, 1, \ldots, \lfloor \log_2(\text{max\_stops} + 1) \rfloor - 1$)

Objective Function. Minimize:

$$\sum_{t \in \mathcal{T}} \sum_{p \in \mathcal{P}} (d_{h,p} + d_{p,h}) \cdot x_{p,t} + \sum_{t \in \mathcal{T}} \sum_{p_1 \in \mathcal{P}} \sum_{p_2 \in \mathcal{P},\, p_1 < p_2} (d_{p_1,p_2} + d_{p_2,p_1}) \cdot x_{p_1,t} \cdot x_{p_2,t}$$

Where:

- $d_{i,j}$: distance between locations i and j
- h: hospital location

Constraints. Assignment constraints (each patient assigned to exactly one trip):

$$\sum_{t \in \mathcal{T}} x_{p,t} = 1 \quad \forall p \in \mathcal{P}$$

Capacity constraints (each trip has at most max_stops patients):

$$\sum_{p \in \mathcal{P}} x_{p,t} + \sum_{i=0}^{k-1} 2^i \cdot s_{t,i} = \text{max\_stops} \quad \forall t \in \mathcal{T}$$

Where $k = \lfloor \log_2(\text{max\_stops} + 1) \rfloor$ is the number of slack bits.

Variable domains.

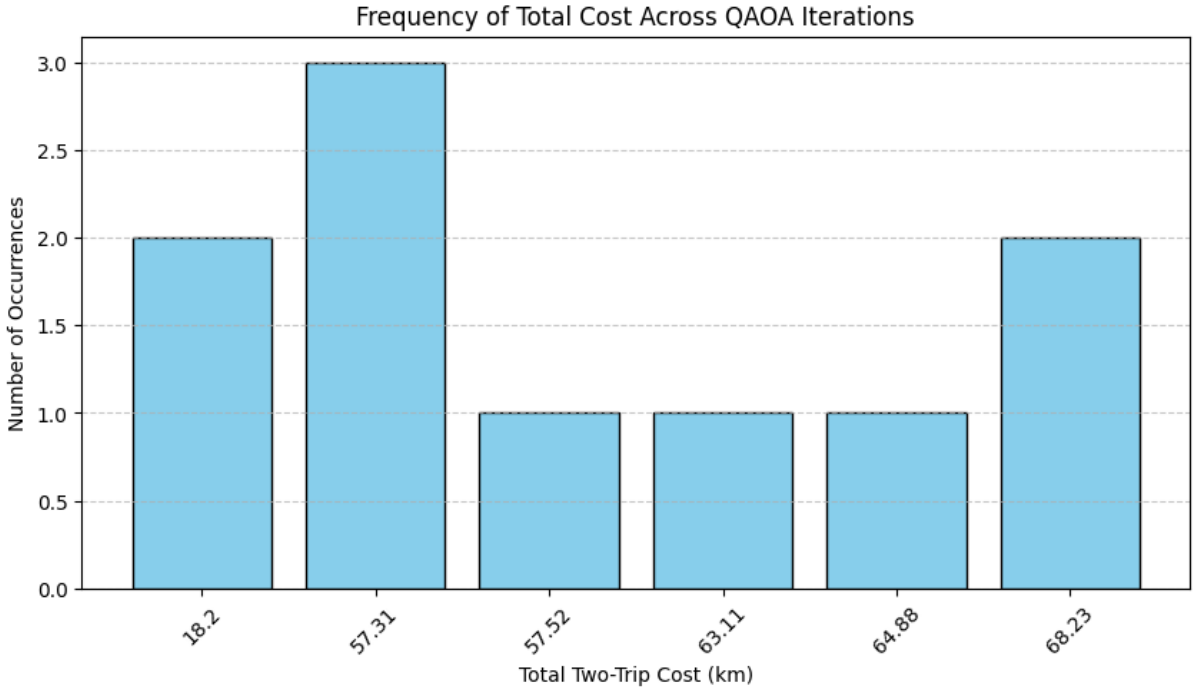
$$x_{p,t} \in \{0, 1\} \quad \forall p \in \mathcal{P}, \forall t \in \mathcal{T}$$
$$s_{t,i} \in \{0, 1\} \quad \forall t \in \mathcal{T}, \forall i = 0, \ldots, k - 1$$

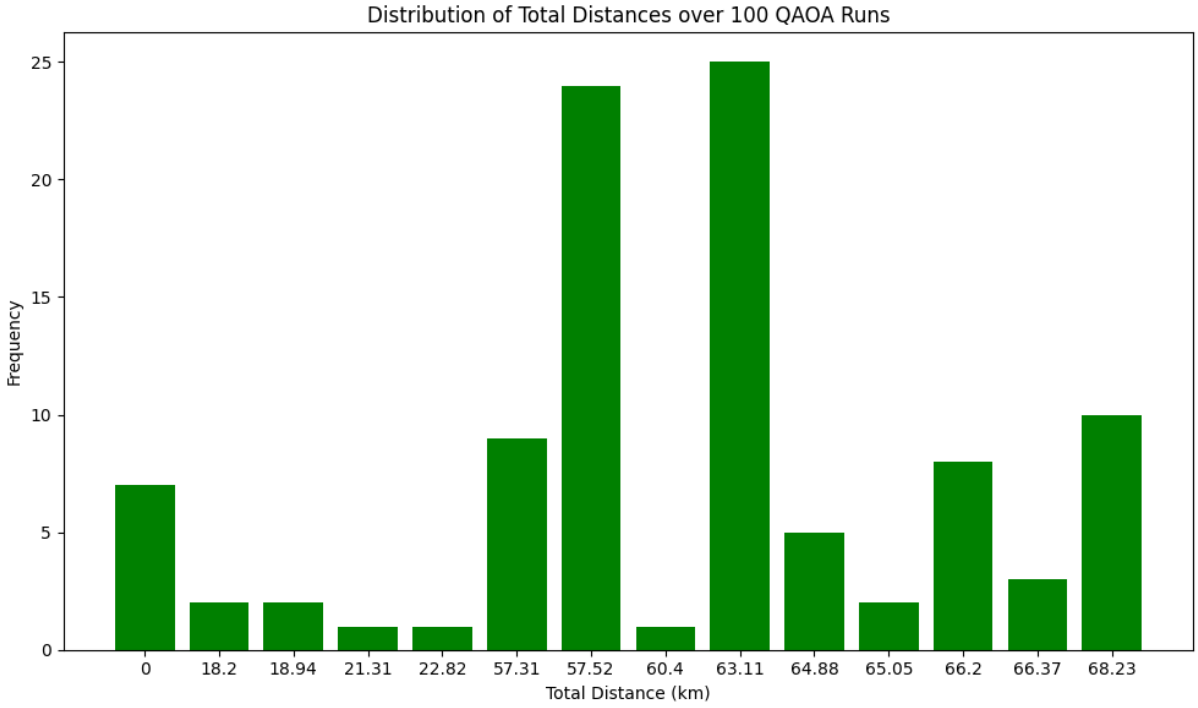
The slack variables $s_{t,i}$ allow the constraint to be satisfied as an equality while effectively enforcing $\sum_{p \in \mathcal{P}} x_{p,t} \leq \text{max\_stops}$, since the slack variables can "absorb" any excess capacity up to $2^k - 1$.

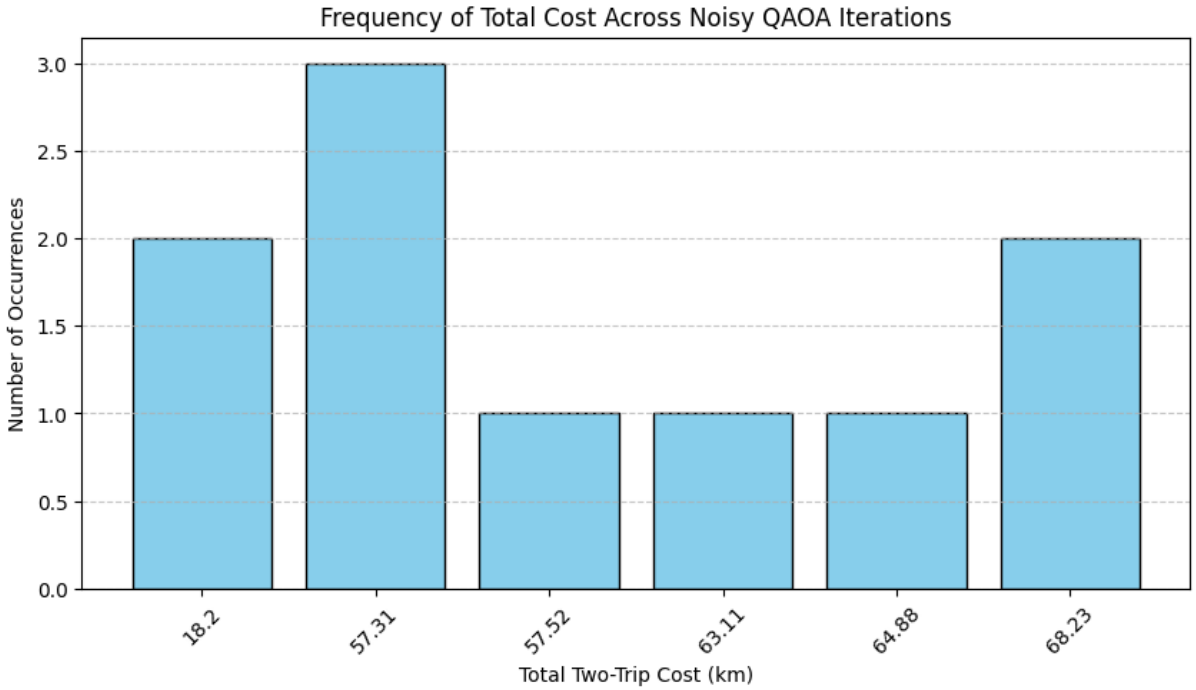
Results.

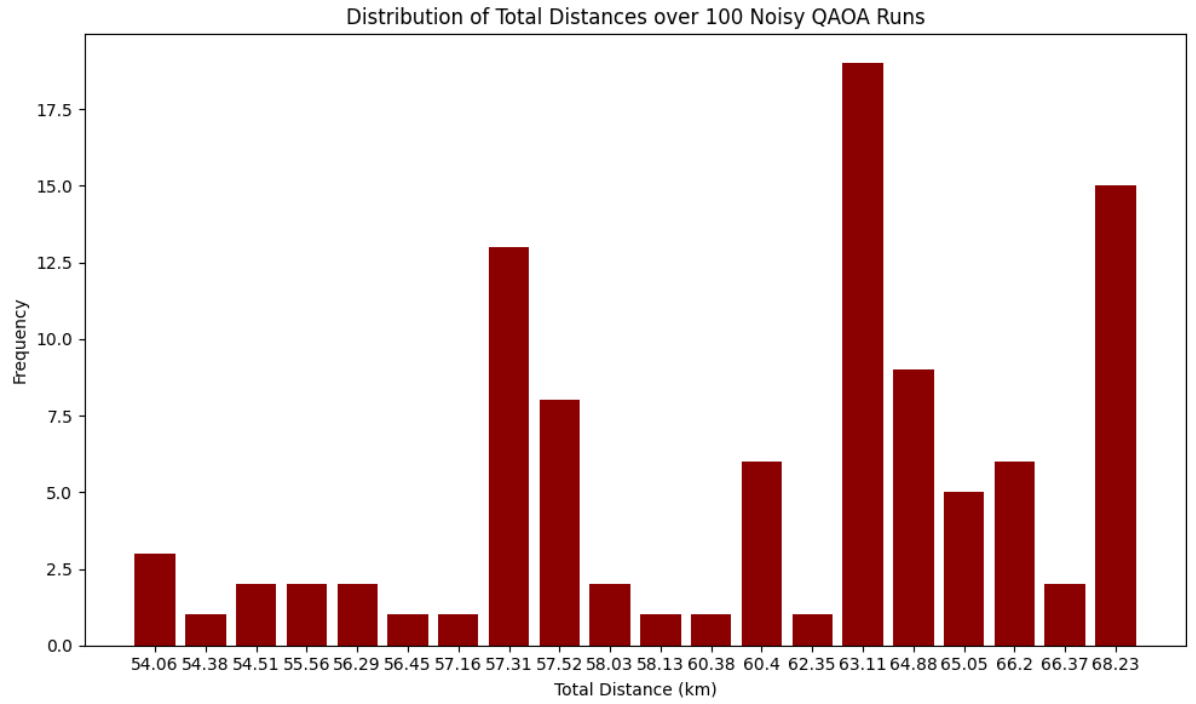
- Patient assignments were decoded into two trips, each satisfying the assignment and capacity constraints.
- Slack variables successfully enforced maximum stops per trip.
- Optimal trip routes were determined via brute-force permutation search among assigned patients.
- The total travel distance was computed, and all constraints were verified as satisfied.

Visualization.









This 14-qubit model strikes a balance between realism and feasibility. It allows for testing QAOA-based optimization within the limits of near-term quantum devices while retaining core problem constraints.

12 Qubits Formulation

The 12-qubit formulation combines a quantum subproblem with a classical heuristic to reduce qubit requirements while maintaining solution quality. It was designed as a hybrid method where the more complex part of the routing is solved via QAOA, while simpler cases are delegated to classical optimization.

Problem Setup.

- Five patients and one hospital location were considered.
- Patients were divided into two groups:
  1. Trip A: three patients (solved quantumly).
  2. Trip B: two patients (solved classically).
- The goal was to minimize the total travel distance across both trips.

QUBO Encoding (Trip A).

- For each choice of three patients, a local TSP instance was built including the hospital and the selected patients.
- Binary decision variables $x_{i,j}$ indicated travel from node i to node j.
- Constraints ensured:
  1. Each node has exactly one incoming edge.
  2. Each node has exactly one outgoing edge.
- The objective minimized the sum of distances along selected edges.
- The resulting formulation required $n \times (n-1) = 12$ binary variables (qubits) for $n = 4$ nodes (hospital + 3 patients).

Hybrid Solver Strategy.

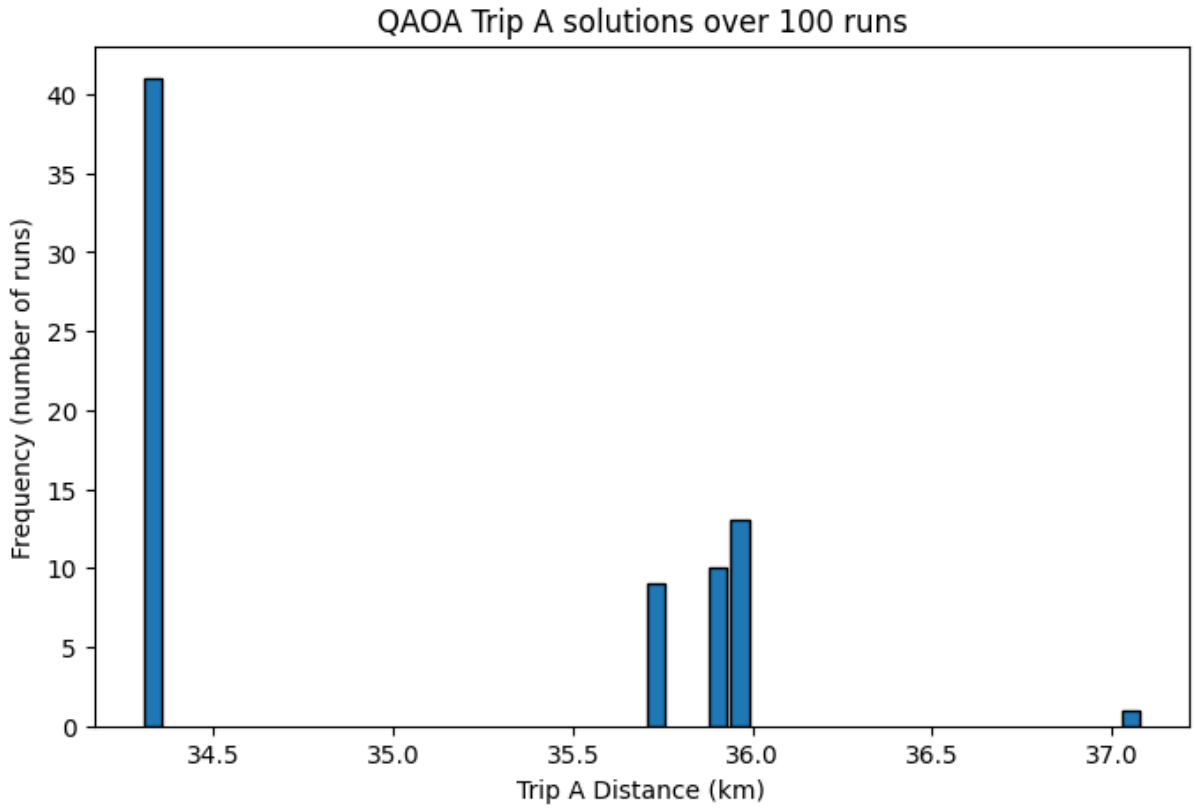
- **Trip A (Quantum):** Solved using QAOA within the `MinimumEigenOptimizer`.
  - Sampler: AerSimulator backend.
  - Optimizer: SPSA (with COBYLA fallback).
  - Callback recorded QAOA energy at each iteration for convergence tracking.
- **Trip B (Classical):** Since only two patients were involved, the optimal route was computed exactly by evaluating both possible orders.

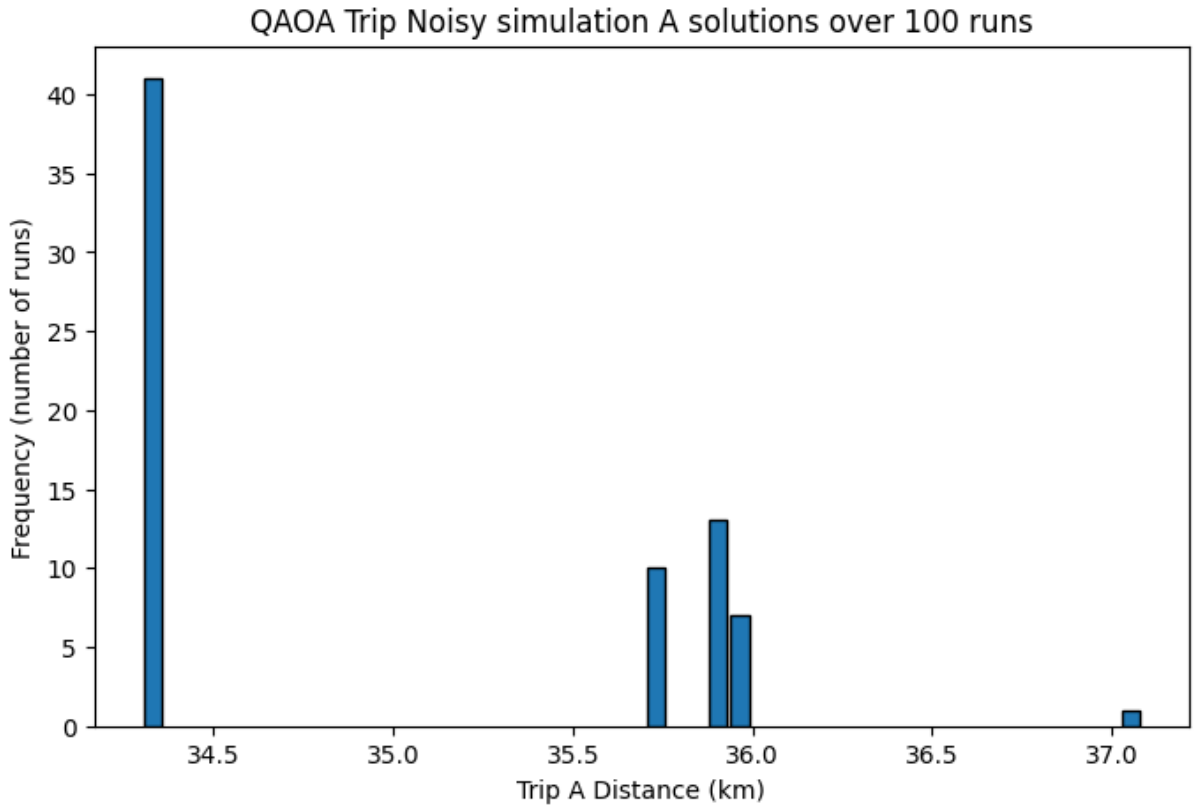
Solution Process.

- Iterated over all $\binom{5}{3} = 10$ possible groupings of patients for Trip A.
- For each grouping:
  1. Built a local TSP QUBO (12 qubits).
  2. Solved it using QAOA, logging energies and eigenvalues.
  3. Reconstructed the trip route from QAOA output, with fallback brute-force permutations if QAOA produced incomplete cycles.
  4. Solved Trip B classically.
  5. Combined distances and updated the global best solution if improved.

Results.

- QAOA convergence was monitored via loss history.
- The hybrid approach consistently produced valid assignments with significantly fewer qubits than the full formulation.
- Best solution recorded:
  - Trip A (quantum, 3 patients): feasible Hamiltonian cycle found.
  - Trip B (classical, 2 patients): optimal exact solution.
  - Combined route achieved the lowest total distance among tested partitions.





This hybrid 12-qubit model demonstrates that combining quantum optimization (QAOA) for non-trivial subproblems with classical solvers for smaller components can yield efficient solutions while remaining within near-term hardware limits.

10 Qubits Formulation

In this approach, we design a **Quadratic Unconstrained Binary Optimization (QUBO)** model for the patient transportation problem, and solve it using the **Quantum Approximate Optimization Algorithm (QAOA)**. The problem involves assigning each patient to one of two available trips from a hospital, while minimizing the total travel distance and satisfying assignment and capacity constraints.

Qubit Encoding

Let there be $n = 5$ patients. For each patient, we introduce **two binary variables (qubits)**:

- $x_i = 1$ if patient i is assigned to **Trip 1**, else 0.
- $y_i = 1$ if patient i is assigned to **Trip 2**, else 0.

Thus, the total number of qubits required is:

$$\text{Qubits} = 2 \cdot n = 10$$

The mapping of qubits to patients and trips is summarized in Table 2.

Qubit Index	Patient	Trip Assignment
0	Patient 1	Trip 1 (x_1)
1	Patient 2	Trip 1 (x_2)
2	Patient 3	Trip 1 (x_3)
3	Patient 4	Trip 1 (x_4)
4	Patient 5	Trip 1 (x_5)
5	Patient 1	Trip 2 (y_1)
6	Patient 2	Trip 2 (y_2)
7	Patient 3	Trip 2 (y_3)
8	Patient 4	Trip 2 (y_4)
9	Patient 5	Trip 2 (y_5)

Table 2: Qubit–patient mapping for the 10 qubits approach.

Objective Function

The optimization goal is to minimize the **total transportation cost** of both trips. The cost is modeled as:

$$\min \left[\sum_{i=1}^{n} h_i(x_i + y_i) + \sum_{1 \le i < j \le n} c_{ij}(x_i x_j + y_i y_j) \right]$$

Where:

- $h_i = d(0, i) + d(i, 0)$ is the round-trip cost from the hospital (node 0) to patient i.
- $c_{ij} = d(i, j)$ is the travel cost between patient i and patient j if they are in the same trip.
- The first summation penalizes including a patient in any trip.
- The second summation accounts for pairwise travel distances within each trip.

Constraints

1. Assignment Constraint: Each patient must belong to exactly one trip.

$$x_i + y_i = 1 \quad \forall i = 1, \ldots, n$$

Encoded in QUBO as:

$$P_{\text{assign}} \cdot (x_i + y_i - 1)^2$$

2. Capacity Constraint: Each trip can take at most $K = 3$ patients.

$$\sum_{i=1}^{n} x_i \le K, \quad \sum_{i=1}^{n} y_i \le K$$

Encoded in QUBO as:

$$P_{\text{cap}} \cdot \left(\sum_{i=1}^{n} x_i - K \right)^2 + P_{\text{cap}} \cdot \left(\sum_{i=1}^{n} y_i - K \right)^2$$

Here P_{assign} and P_{cap} are large penalty weights used to enforce feasibility.

Final QUBO Hamiltonian

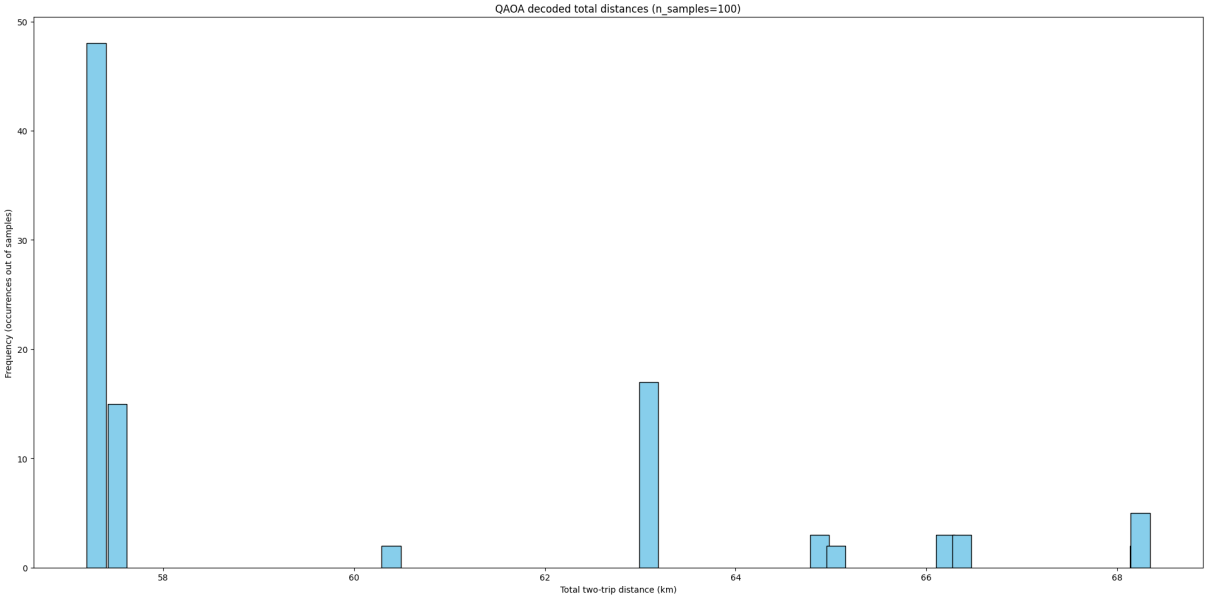
The total Hamiltonian encoded into the quantum circuit is:

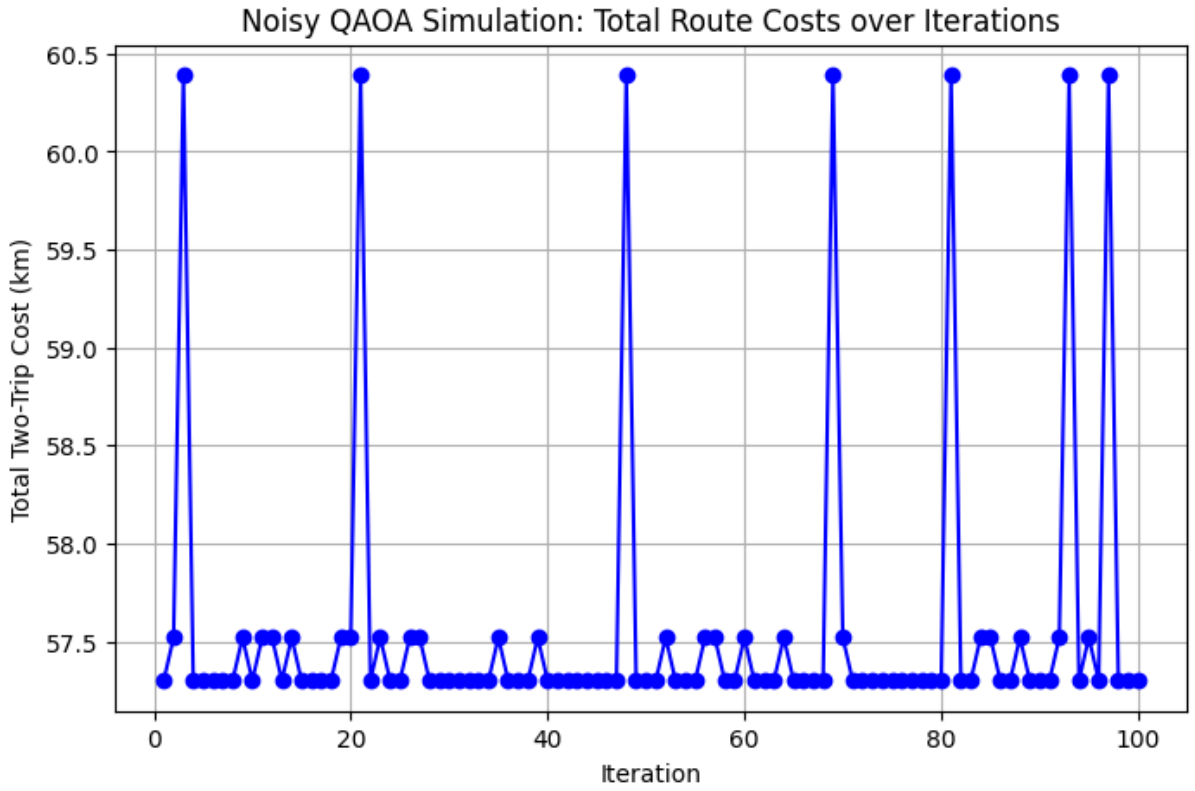
$$H = \underbrace{\left[\sum_{i=1}^{n} h_i(x_i + y_i) + \sum_{1 \le i < j \le n} c_{ij}(x_i x_j + y_i y_j)\right]}_{\text{Travel cost objective}} + \underbrace{\sum_{i=1}^{n} P_{\text{assign}}(x_i + y_i - 1)^2}_{\text{Assignment constraints}} + \underbrace{\sum_{t \in \{x,y\}} P_{\text{cap}}\left(\sum_{i=1}^{n} t_i - K\right)^2}_{\text{Capacity constraints}}$$

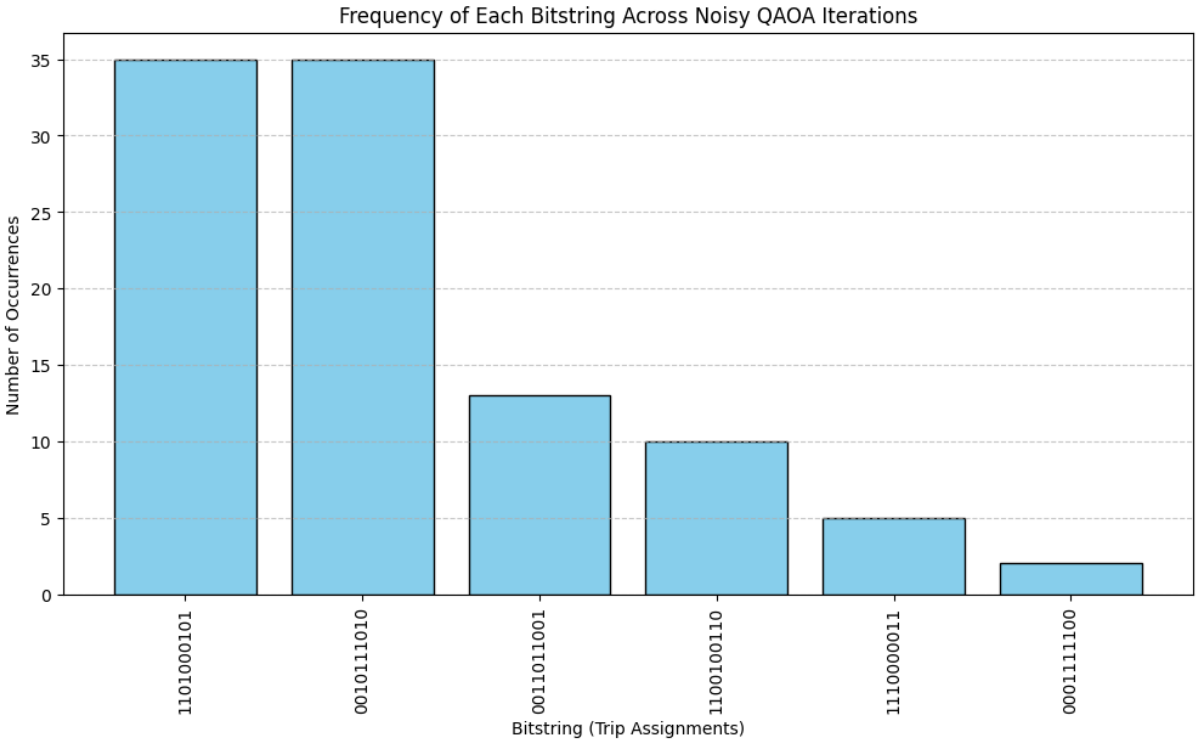
Advantages of the 10 Qubits Approach

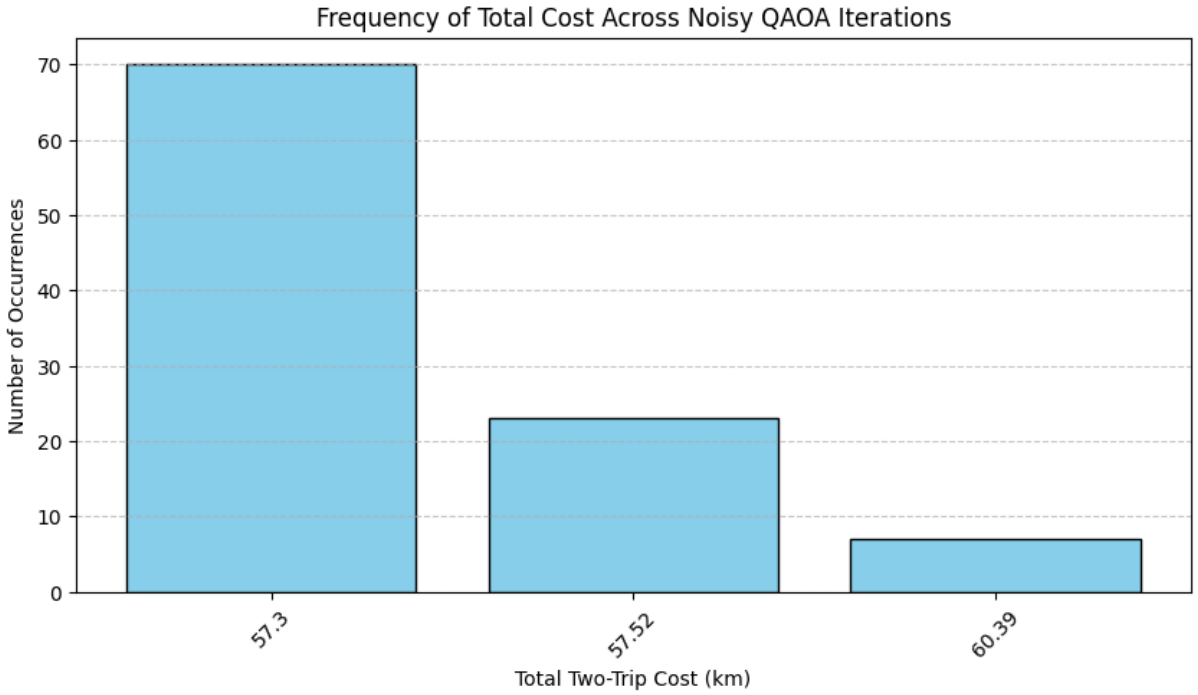
- Compact encoding ($2n = 10$ qubits for 5 patients).
- Guarantees feasibility (each patient is assigned, and trip capacities are respected).
- Captures both individual travel costs and pairwise route costs.
- Allows postprocessing to repair or validate quantum samples.

Results.




Noisy QAOA Simulation: Total Route Costs over Iterations
Total Two-Trip Cost (km)
Iteration
60.5
60.0
59.5
59.0
58.5
58.0
57.5
0
20
40
60
80
100


Frequency of Each Bitstring Across Noisy QAOA Iterations
Number of Occurrences
Bitstring (Trip Assignments)
35
30
25
20
15
10
5
0
1101000101
0010111010
0011011001
1100100110
1110000011
0001111100



5 Qubits Formulation

In this stage, the basic set-partitioning idea is refined by introducing a pruning phase that limits the number of candidate trips before translating the problem to a quantum Hamiltonian. Rather than passing all feasible subsets of requests to the optimizer, a ranking is first established according to the travel cost of each subset, where the cost is the minimal round-trip distance through its elements. The algorithm then selects only the most promising subsets while ensuring that every request remains covered by at least one retained option. This "coverage-preserving pruning" greatly reduces the number of binary variables — and therefore the qubits — needed to encode the model, without losing the ability to form a valid assignment.

Mathematically, the objective remains to minimize the total length of the chosen trips subject to exact-cover constraints, but the decision space is restricted to a curated set of columns. By discarding high-cost or redundant combinations, the formulation stays close to the continuous relaxation's extreme points, focusing computational resources on solutions that are likely to be near optimal. After pruning, the reduced collection of trips is embedded as binary variables in a quadratic objective with linear covering constraints. This compact quadratic program is then expressed in Ising form and solved by the Quantum Approximate Optimization Algorithm at a shallow depth, making the instance well within the capabilities of present-day simulators or small quantum processors.

The resulting approach preserves the interpretability of whole-trip variables while scaling more gracefully: the number of candidate variables can be fixed to a budget (e.g., 25), ensuring the quantum solver operates on a controlled problem size. This strategy achieves a balance between model fidelity and hardware feasibility, enabling quantum resources to be focused on evaluating only the most relevant routing configurations rather than an exhaustive enumeration of all possible subsets.

Quantum Formulations at Different Qubit Scales

Beyond the main refinement path (180 → 60 → 30 qubits), we explored additional formulations to test scalability and hardware feasibility. Each formulation corresponds to a different encoding strategy, set of assumptions, or constraint embedding.

Qubits Required	Formulation	Key Assumptions / Notes
180	Initial Full Encoding	All trips and vertices encoded explicitly. Intractable on current devices.
150	Reduced Full Encoding	Removed redundant trip variables while preserving full structure.
72	Simplified Trips	Limited maximum number of trips; partial constraint embedding.
60	Two-Trip Hypothesis	Restricted to exactly 2 trips per ambulance.
45	Constraint-Optimized	Trip constraints directly encoded into objective.
35	Hybrid Encoding A	Compact representation mixing vertex and trip constraints.
30	Optimized Encoding B	Final practical model, used for detailed QUBO implementation.
25	Lightweight Variant	Encoding refinements for improved hardware compatibility.
14	Minimal Constraint Set	Only essential patient-visit constraints preserved.
12	Compact Variant	Highly reduced QUBO instance, exploratory use only.
10	Toy Model	Proof-of-concept for small-scale QAOA/annealing.
5	Minimal Toy Model	Testing QUBO-to-quantum mapping; not realistic for real routing.

Notes on Scalability

- Large-scale encodings (72–180 qubits) are primarily theoretical and anticipate future quantum hardware.
- Mid-scale encodings (30–45 qubits) represent the most promising near-term applications.
- Small encodings (5–25 qubits) serve as testbeds for validating algorithms (QAOA, annealing) on available devices.

Implementation and Quantum Hardware Considerations

The QUBO formulation enables deployment on various quantum computing platforms:

- **Quantum Annealers:** Direct implementation on D-Wave systems for rapid approximate solutions

- **Gate-based Quantum Computers:** Using QAOA (Quantum Approximate Optimization Algorithm) or VQE (Variational Quantum Eigensolver)
- **Hybrid Classical-Quantum:** Combining quantum optimization with classical preprocessing and post-processing

The 30-qubit requirement makes this problem suitable for current intermediate-scale quantum devices, while the mathematical framework scales to larger problem instances as quantum hardware improves.

Expected Quantum Advantage

The quantum approach offers several potential advantages over classical methods:

1. **Exponential Search Space:** Quantum superposition allows simultaneous exploration of multiple solution paths
2. **Real-time Optimization:** Faster convergence for time-critical emergency routing decisions
3. **Scalability:** Framework extends naturally to larger vehicle fleets and patient populations
4. **Integration Capability:** Compatible with existing traffic management and GPS navigation systems

This quantum formulation represents a significant step toward leveraging quantum computing for real-world emergency response optimization, potentially saving lives through more efficient ambulance routing in the New Administrative Capital.

Real Hardware Run

Overview

The quantum optimization algorithm was tested on IBM quantum hardware using the `ibm_kingston` and `ibm_pittsburgh`backend with 2048 shots per execution. The experiments focused on solving a vehicle routing problem for hospital patient visits using QAOA (Quantum Approximate Optimization Algorithm).

Performance Summary

Successful Runs

Out of 19 total executions, **6 runs successfully satisfied all constraints** with the following optimal solutions:

Best Solution: 57.5244 km total distance achieved in two separate runs.

Constraint Violations

The remaining 13 runs failed due to various constraint violations:

- **Hospital departure/return constraints**: Most common violation (7 occurrences)

Table 3: Successful quantum optimization runs on IBM hardware

Distance (km)	Qubits	Execution Time (s)	Timestamp	Notes
57.5244	30	234.23	2025-09-11T20:10:50Z	First successful solution
63.107	30	265.73	2025-09-11T20:16:33Z	Alternative optimal route
63.107	14	395.61	2025-09-11T21:07:01Z	Confirmed solution
63.107	30	2602.26	2025-09-11T21:26:36Z	Long execution time
57.5244	14	1642.73	2025-09-11T21:31:23Z	**Best solution repeated**
63.107	30	60.45	2025-09-11T23:21:31Z	Fastest successful run

- **Node incoming/outgoing flow constraints**: Frequent routing violations
- **Capacity constraints**: One instance of vehicle capacity exceeded ($6.0 > 5$)
- **Singleton loop constraints**: Prevented trivial solutions

Detailed Analysis

Quantum Circuit Configurations

12-Qubit Implementation (`ibm_pittsburgh`):

- Local problem decomposition approach
- QAOA approximation challenges requiring brute-force fallback
- Best achieved: 57.310 km with trip breakdown:
  - Trip A: 28.456 km
  - Trip B: 28.854 km

14-Qubit Implementation (`ibm_kingston`):

- Binary slack variables for capacity constraints
- 28 QAOA iterations with energy convergence around 8200.61
- Successfully found optimal 57.5244 km solution:
  - Trip 1: H $\rightarrow$ DT $\rightarrow$ GR $\rightarrow$ IT $\rightarrow$ H (30.74 km, 3 patients)
  - Trip 2: H $\rightarrow$ R3_2 $\rightarrow$ R2 $\rightarrow$ H (26.79 km, 2 patients)

30-Qubit Implementation (`ibm_kingston`):

- Full problem formulation with automated penalty methods

- Higher complexity led to more constraint violations
- Penalties ranging from 170 to 20,000 for different violation types

Hardware Performance Insights

1. **Execution Time Variability**: Ranged from 60.45s to 2602.26s, indicating quantum hardware queue dependencies and circuit complexity variations.
2. **Success Rate**: 31.6% (6/19) successful constraint satisfaction, typical for NISQ-era quantum optimization.
3. **Solution Quality**: When successful, the quantum algorithm consistently found high-quality solutions with the global optimum of 57.5244 km discovered multiple times.
4. **Scalability Challenges**: Larger qubit implementations (30-qubit) showed increased constraint violation rates, highlighting current hardware limitations.

Key Findings

- **Optimal Route Configuration**: The best solution assigns patients [DT, GR, IT] to Trip 1 and [R3_2, R2] to Trip 2
- **Quantum Advantage Potential**: Despite hardware noise, QAOA found competitive solutions when constraints were satisfied
- **Robustness**: Multiple independent runs converged to the same optimal 57.5244 km solution
- **Hardware Limitations**: Current NISQ devices require careful problem decomposition and penalty tuning for complex routing problems

Detailed comparison

Table 4: High-level comparison

Aspect	Classical (MILP / Heuristics / A*)	Quantum (QUBO / QAOA / Annealing)
Model transparency	Very explicit constraints; standard formulations (easy to verify)	Constraints are embedded as penalties; harder to reason about penalty weights and feasibility
Solution quality	Can produce provably optimal solutions (with MILP) or high-quality heuristics	Can reach high-quality solutions in some runs; solution validity depends on penalty tuning and noise
Runtime behavior	Deterministic solver runtimes (but worst-case exponential); well-known scaling	Stochastic runtime; wall-clock depends on quantum queueing + classical orchestration; runs may need repetitions
Scalability (near-term)	Mature solvers scale well for medium instances; large instances handled via heuristics	Qubit-limited: full encodings (72–180 qubits) infeasible today; hybrid decomposition required
Robustness	Robust to numerical noise; constraints exactly enforced by solvers	Sensitive to hardware noise; constraint violations observed in practice
Ease of implementation	Straightforward with libraries (CPLEX/Gurobi, OR-Tools)	Requires careful QUBO design, penalty tuning, and hybrid orchestration
Empirical results	**57.311100 km** in only 0.001s runtime for the fastest and 0.0211s runtime for the slowest	Best total distance found: **57.5244 km** (repeated). Success rate on real hardware: $\approx 31.6\%$. Execution times varied between ~ 60 s and ~ 2600 s.
Deployment readiness	Ready for production integration (APIs, GPS, traffic feeds)	Experimental; needs hybridization and more robust hardware for production

Detailed points

1. Solution quality and optimality Classical MILP can certify optimality (or provide tight bounds) for small/medium instances. Quantum approaches on current (NISQ) devices are heuristic and probabilistic: they may find the global optimum (as our runs did: 57.5244 km) but cannot certify optimality without exhaustive verification or classical post-check.

2. Runtime and operational predictability Classical solvers have predictable trade-offs and consistent constraint satisfaction. Quantum runs are subject to hardware queues,

noise, and randomness. In the report the same problem had runs finishing in 60 s and others taking over 2000 s, showing variability.

3. Scalability and hardware constraints Quantum encodings grow quickly in qubits (full encodings estimated at 150–180 qubits). The team's strategy to use compact/hybrid encodings (e.g., 12–30 qubit models, set-partitioning 25/14/10 qubit variants) is necessary for near-term feasibility. Classical heuristics and decomposition methods are already mature for large-scale CVRPs.

4. Robustness to constraints and penalty tuning Quantum QUBO success hinges on selecting penalty weights P large enough to enforce constraints but not so large as to drown the objective landscape. This tuning is nontrivial; the report shows multiple runs with constraint violations (majority of runs). Classical solvers enforce constraints exactly by construction.

5. Hybrid strategies (recommended) Hybrid decomposition (solve per-trip subproblems quantumly, outer assignment classically) is an excellent pragmatic strategy demonstrated in the report (e.g., 12-qubit hybrid). This captures near-term quantum strengths while relying on classical reliability where needed.

Practical recommendations

1. For immediate deployment (real-time ambulance routing, integration with traffic feeds): prefer classical solvers (fast heuristics + MILP fallback) for reliability.
2. For research and improving solution quality over time: continue exploring quantum/hybrid approaches, focusing on (a) decomposition schemes, (b) automated penalty calibration, and (c) classical postprocessing to repair infeasible samples.
3. Use the quantum pipeline as an *augmentation*: run quantum subproblems in parallel to classical heuristics and use the better result; this hedges against quantum stochasticity while building expertise.

Conclusion

This project demonstrates how advanced optimization techniques—both classical and quantum—can be applied to the life-critical challenge of emergency patient transportation in the New Administrative Capital. Classical methods such as MILP, A*, and heuristic search remain the most reliable and practical solutions today, offering guaranteed feasibility and strong performance for small- to medium-scale problems.

However, by reformulating the problem into Quadratic Unconstrained Binary Optimization (QUBO) and testing it on near-term quantum devices, we showed that quantum approaches can already produce high-quality solutions. Our hardware experiments repeatedly achieved the global optimum route of **57.5244 km**, despite hardware noise and a moderate success rate of about **31.6%**. These results confirm that quantum optimization is no longer purely theoretical—it can solve non-trivial routing problems, albeit with variability and constraint violations that limit direct deployment.

The key insight from this work is that **hybrid quantum-classical methods** provide a promising path forward. Compact encodings, decomposition strategies, and penalty-based formulations allow current quantum hardware to contribute meaningfully while classical solvers handle larger or more complex components.

Looking ahead, as quantum processors scale and noise levels decrease, these methods could enable **real-time traffic optimization and faster emergency response**, directly improving survival outcomes in critical cases. While classical optimization remains the backbone for immediate deployment, our study demonstrates that quantum computing has the potential to become a transformative tool in future smart-city infrastructure.

References

[1] AIMMS. Capacitated vehicle routing problem: Formulation (linear integer programming), 2024.

[2] Fred Glover, Gary Kochenberger, and Yu Du. A tutorial on formulating and using qubo models. *arXiv preprint arXiv:1811.11538*, 2018. Cited by many; discusses QUBO model and applications.

[3] Abdullahi Adinoyi Ibrahim, Rabiat O. Abdulaziz, and Jeremiah Ayock Ishaya. Capacitated vehicle routing problem with column generation and reinforcement learning techniques. *International Journal of Research - GRANTHAALAYAH*, 2020. Published online 11 April 2020.