Ministry of Science and Higher Education of the Republic of Kazakhstan

SDU University

Akhmetbek Yernar

Stochastic Dynamic Vehicle Routing Problem

THESIS

Presented in Partial Fulfilment for the

Degree of Master of Technical Science in Computer Science

(degree code: 7M06102)

Department of Computer Science

Faculty of Engineering and Natural Sciences

Supervisor: **Bogyrbayeva Aigerim**

Kaskelen, June 2024

SDU University
Faculty of Engineering and Natural Sciences
Department of Computer Science

Dean of Faculty of Engineering and Natural Sciences

Assistant Professor, PhD Akhmedov Ramis

____

« ______» ________________________________ 2024

Topic of the thesis:

Stochastic Dynamic Vehicle Routing Problem

Thesis submitted as part of the requirements for the award of the MSc in “7M06102 - Computer Science”, SDU University

Head of Department Zhanar Mukash

Academic Supervisor Aigerim Bogyrbayeva

Master student Yernar Akhmetbek

Kaskelen, 2024

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Akhmetbek Yernar

June 2024

Acknowledgements

I would like to thank my supervisors prof. Aigerim Bogyrbayeva and Merarslan Meraliyev for all knowledge they shared for me.

Dedication

This thesis is dedicated to:

My parents, SDU university stuff and my friends.

Abstract

The Stochastic Dynamic Vehicle Routing Problem (SDVRP), in which customer demands are dynamic and follow a Poisson distribution, is examined in this study. We provide an SDVRP-specific Reinforcement Learning (RL) algorithm and evaluate it against algorithms for Random Selection, Largest-Demand Selection, and Max-Reachable Selection. We also use a Multi-head Attention architecture into our RL algorithm to better capture complex relationships in the dynamic routing environment. We show, by thorough evaluation, how RL with Multi-head Attention may be used to optimize resource usage and route efficiency, providing useful insights for addressing challenging logistics problems in real-world situations.

Abbreviations

VRP - Vehicle Routing Problem
SDVRP - Stochastic Dynamic Vehicle Routing Problem
NP - Non Polynomial
RL - Reinforcement Learning

Table of Contents

Declaration **i**

Acknowledgements **ii**

Dedication **iii**

Abstract **iv**

List of Abbreviations **v**

1 Background and motivations **1**
- 1.1 Introduction 1
- 1.2 Background 4
  - 1.2.1 Combinatorial Optimization 4
  - 1.2.2 NP-hard problem 4
  - 1.2.3 Traveling Salesman Problem 5
  - 1.2.4 Vehicle Routing Problem 6
  - 1.2.5 Stochastic and Dynamic Nature 7
  - 1.2.6 Solution methods 7
  - 1.2.7 Exact Methods 9
  - 1.2.8 Heuristic 10
  - 1.2.9 Metaheuristic 10
  - 1.2.10 Reinforcement Learning 11

2 Literature Review **12**

3 Problem Statement **17**
- 3.1 Stochastic Dynamic Vehicle Routing problem 17
  - 3.1.1 Stochastic Customer Request 18
  - 3.1.2 Single depot and single vehicle 19
  - 3.1.3 Time constraint 20
  - 3.1.4 Objective 20
- 3.2 Markov Decision Process 20
  - 3.2.1 State 21
  - 3.2.2 Reward 21
  - 3.2.3 Action 22

3.2.4 Transition function 23

4 Methodology **24**
4.1 Reinforcement Learning 24
4.2 Multi-Head Attention Model 25
4.3 Rollout Baseline 27
4.3.1 Rollout Greedy 27
4.3.2 Rollout Sampling 27
4.4 Encoder and Decoder 28
4.5 Training Algorithm 29
4.6 Hyperparameters 29
4.7 Compared algorithms 30
4.7.1 Random 30
4.7.2 Largest-Demand 30
4.7.3 Max-Reachable 30

5 Experiments and Results **32**
5.1 Experimental section 32
5.1.1 Results 33
5.2 Review section 34

6 Conclusions and future work **35**
6.1 Conclusions 35
6.2 Future work 35

Bibliography **36**

Chapter 1

Background and motivations

1.1 Introduction

Though there are many obstacles in the realm of logistics and transportation management, effective vehicle routing is essential to streamlining operations and cutting expenses. Conventional vehicle routing problems (VRPs) have been researched and solved in a number of different industries. On the other hand, a more intricate and dynamic form known as the stochastic dynamic vehicle routing problem (SDVRP) presents extra levels of flexibility and uncertainty that call for creative solutions.

The SDVRP represents a real-world scenario where vehicle routing decisions must be made dynamically in response to evolving conditions and uncertainties, such as fluctuating customer demands[1], varying traffic conditions[2], and unpredictable service times[3] . Unlike static VRPs[4], which assume known and constant parameters, the SDVRP requires decision-makers to adapt and adjust routes in real-time to optimize performance and meet service level agreements. In recent years, the SDVRP has gained significant attention from researchers and practitioners alike due to its relevance in dynamic and unpredictable environments [5], including urban logistics, emergency response systems, and on-demand services. The ability to efficiently route vehicles in such contexts not only reduces operational costs but also enhances customer satisfaction and overall system resilience.

Understanding the background and context of the SDVRP involves delving into its theoretical foundations, practical applications, and inherent challenges. From a theoretical standpoint, the SDVRP integrates elements of stochastic optimization, dynamic programming, and combinatorial optimization to formulate models and develop algorithms capable of addressing its complexities. On the practical side, industries are increasingly recognizing the importance of adaptive routing strategies to cope with the uncertainties inherent in modern supply chains and service networks.

Furthermore, the SDVRP intersects with various other fields, including artificial intelligence, machine learning, and operations research, fostering interdisciplinary collaborations and innovative solutions. Researchers and practitioners are exploring advanced optimization techniques, such as metaheuristics [6], reinforcement learning[7], and simulation-based optimization [8], to tackle the SDVRP's dynamic and stochastic nature effectively.

Transport and logistics managers have a significant difficulty when it comes to effectively allocating resources and satisfying customer needs in uncertain and dynamic environments: the Stochastic Dynamic Vehicle Routing Problem (SDVRP). The SDVRP's primary objective is to find the best routes for a fleet of cars that must dynamically adjust to a variety of circumstances, such as fluctuating service times, stochastic consumer requests, and unpredictable traffic patterns.

The primary goal of the SDVRP is to minimize overall costs while satisfying operational constraints and service level requirements. However, achieving this objective is inherently complex due to the interplay of dynamic factors and stochastic events that influence routing decisions. Unlike traditional vehicle routing problems (VRPs), which assume static or deterministic conditions, the SDVRP requires decision-makers to make routing decisions in real-time based on incomplete and uncertain information.

The dynamic nature of the SDVRP introduces several key challenges:

1. **Dynamic Information:** The SDVRP involves continuously evolving information, such as real-time demand updates, traffic congestion, and unexpected disruptions. Decision-makers must efficiently process and incorporate this dynamic information to make informed routing decisions. [9] [10] [11] [12]

2. **Stochastic Demands:** Customer demands in the SDVRP are often stochastic in nature, meaning they follow probabilistic distributions rather than fixed values. Predicting future demand patterns and adjusting routes to accommodate uncertain demand levels is essential for minimizing costs and maximizing service quality. [13]

3. **Temporal Constraints:** Vehicles in the SDVRP are subject to various temporal constraints, including time windows for customer visits, vehicle operating hours, and service time variability. Balancing these constraints while optimizing routes adds another layer of complexity to the problem. [14]

4. **Uncertain Travel Times:** Travel times between locations are influenced by unpredictable factors such as traffic congestion, road closures, and weather conditions. Estimating and accounting for these uncertainties is crucial for ensuring timely deliveries and efficient resource utilization. [15]

5. **Adaptive Decision-Making:** Traditional routing algorithms are ill-suited for the SDVRP's dynamic nature, as they lack the flexibility to adapt to changing conditions. Effective SDVRP solutions require adaptive decision-making strategies that can dynamically update routes in response to new information and evolving circumstances. [16]

Addressing these challenges requires the development of innovative methodologies and solution approaches that can effectively handle the dynamic and stochastic nature of the SDVRP. Moreover, practical implementations must consider the scalability, computational efficiency, and robustness of proposed solutions to ensure their viability in real-world logistics operations.

This research embarks on a comprehensive exploration of the Stochastic Dynamic Vehicle Routing Problem (SDVRP), aiming to contribute both theoretical insights and practical solutions to this complex domain. The objectives of this endeavor encompass various dimensions, each aimed at deepening our understanding of the problem and developing effective methodologies for its resolution. A crucial aspect of this research involves delving into existing literature across domains such as vehicle routing problems, stochastic optimization, and dynamic programming. By undertaking a thorough review and synthesis of prior work, we seek to identify key findings, methodologies, and challenges that will inform our research direction and guide our efforts toward addressing existing gaps in knowledge.

Building upon this foundational understanding, we will delve into the theoretical underpinnings of the SDVRP, aiming to elucidate its fundamental principles, mathematical formulations, and inherent complexities. This theoretical exploration will provide the groundwork for the development of novel methodologies that can effectively tackle the dynamic and stochastic nature of the problem. In parallel, our research will involve experimentation with existing methodologies and algorithms designed to address the SDVRP. Through empirical studies and real-world case studies, we aim to validate the performance of these methodologies in practical settings, assessing their efficacy, scalability, and adaptability in dynamic and uncertain environments.

Moreover, we will propose and explore innovative solution approaches, drawing upon insights gleaned from the literature review and theoretical analysis. This experimentation phase will enable us to identify promising avenues for further research and development, as well as to refine existing methodologies to better suit the unique challenges posed by the SDVRP.

The study of the Stochastic Dynamic Vehicle Routing Problem (SDVRP) holds critical importance in the realm of transportation and logistics. By addressing the dynamic and uncertain nature of vehicle routing, this research seeks to significantly enhance operational efficiency, reduce costs, and improve customer satisfaction. Traditional routing approaches often falter in accommodating real-time changes and uncertainties, leading to inefficiencies and missed opportunities. However, by developing adaptive routing strategies and optimizing decisions in response to dynamic conditions, this study aims to bolster the resilience and adaptability of transportation systems, ensuring timely deliveries and optimized resource utilization. Moreover, the study of the SDVRP contributes to environmental sustainability by minimizing fuel consumption and emissions through optimized routes and vehicle loads, thus fostering more sustainable transportation practices. Through these endeavors, this research not only advances academic knowledge but also provides tangible benefits to industries by enhancing operational performance and promoting sustainable logistics practices.

In this dissertation, we aim to explore the SDVRP comprehensively, investigating its theoretical underpinnings, practical implications, and state-of-the-art solution approaches. By examining existing literature, proposing novel methodologies, and conducting empirical studies, we seek to contribute to the ongoing discourse on dynamic vehicle routing in stochastic environments. Ultimately, our research endeavors to provide insights and methodologies that can aid decision-makers in addressing the challenges posed by the SDVRP and improving the efficiency and resilience of transportation and logistics systems in dynamic and uncertain conditions.

1.2 Background

1.2.1 Combinatorial Optimization

Combinatorial optimization is a cornerstone of modern decision-making processes, encompassing the search for optimal solutions among a finite set of possibilities. This field, rooted in mathematics and computer science, addresses complex problems where decisions involve discrete choices and are subject to constraints. As [17] notes, combinatorial optimization is concerned with "the process of finding the 'best' solution out of a finite set of possibilities." This search for optimality is a fundamental challenge in various domains, including operations research, logistics, engineering, and economics.

The essence of combinatorial optimization lies in balancing conflicting objectives while adhering to constraints. Constraints can range from resource limitations to time constraints, and they play a crucial role in defining the feasible solution space. As [18] highlight, "combinatorial optimization deals with optimization problems over discrete domains that arise in numerous applications and may involve constraints on variables and parameters." Understanding and effectively managing these constraints are key to devising efficient algorithms and strategies for solving combinatorial optimization problems.

Combinatorial optimization finds widespread application across industries and research fields. For instance, in transportation and logistics, it underpins routing and scheduling decisions, optimizing resource allocation and minimizing costs [19]. In finance, combinatorial optimization aids in portfolio management, balancing risk and return objectives [20]. Furthermore, in bioinformatics, it facilitates genome sequencing, protein structure prediction, and drug discovery [21]. These diverse applications underscore the versatility and significance of combinatorial optimization in addressing complex decision-making challenges in practical settings.

1.2.2 NP-hard problem

NP-hard problems are a class of computational problems that are among the most challenging and complex to solve efficiently. The term "NP-hard" stands for Non-deterministic Polynomial-time hard, indicating that these problems are at least as difficult as the hardest problems in the complexity class NP (Non-deterministic Polynomial-time). While NP-hard problems do not necessarily have a known polynomial-time algorithm for solving them, they are crucial in theoretical computer science and practical applications due to their inherent complexity and relevance to real-world challenges.

One defining characteristic of NP-hard problems is their exponential growth in

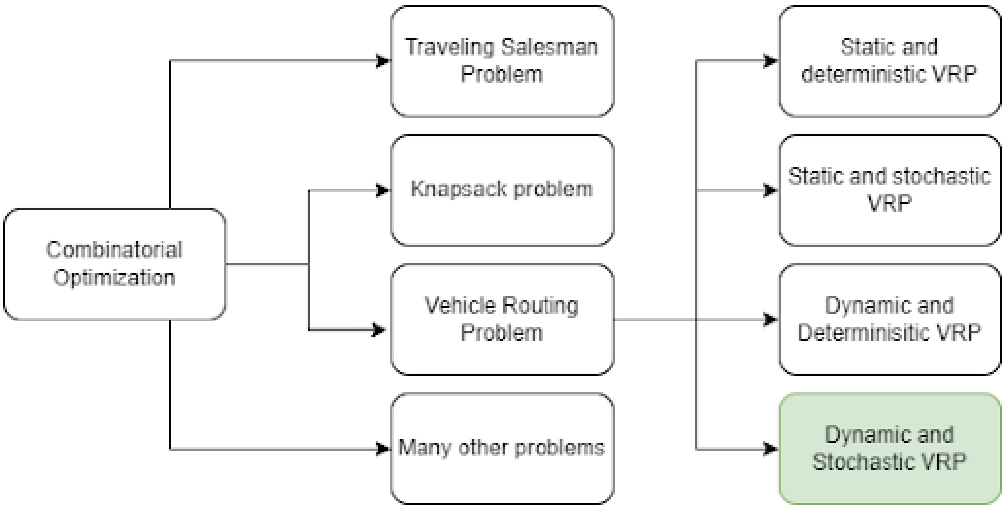

Figure 1.1 - SDVRP Branches.

computational complexity as the input size increases. This means that the time required to solve NP-hard problems using exact algorithms typically increases exponentially with the problem size. As a result, solving large instances of NP-hard problems becomes computationally infeasible or impractical in a reasonable amount of time, especially for problems with a large number of variables, constraints, or possibilities.

The Vehicle Routing Problem (VRP)[22], the Traveling Salesman Problem (TSP)[23], the Knapsack Problem, the Job Scheduling Problem, and the Boolean Satisfiability Problem (SAT) are a few examples of NP-hard issues. These issues are prevalent in many different fields, including scheduling, circuit design, logistics, resource allocation, and optimization, demonstrating their wide range of applications and importance in operational and computational contexts.

Despite their computational complexity, NP-hard problems are instrumental in advancing algorithmic research and problem-solving methodologies. Researchers develop approximation algorithms, heuristic methods, and metaheuristics to tackle NP-hard problems efficiently by trading off optimality for computational tractability. Additionally, NP-hardness results provide insights into the intrinsic difficulty of certain problems and contribute to understanding the boundaries of computational feasibility in algorithm design and optimization.

In conclusion, NP-hard issues are a class of extremely difficult computer problems that have exponential complexity and have important theoretical and practical ramifications. Their work stimulates the development of novel algorithmic methods, advances computational complexity theory research, and develops effective approaches to solving challenging real-world issues.

1.2.3 Traveling Salesman Problem

A well-known example of combinatorial optimization is the Traveling Salesman Problem (TSP), which focuses on determining the shortest path that makes exactly one stop at each of the provided cities and returns to the beginning city. This issue, properly defined for the first time in [24], is representative of combinatorial optimization difficulties because of its exponential complexity with increasing city count. The task of the TSP is to "find the minimum-length tour of a set of cities," as noted [25].

The goal of the TSP is to satisfy the requirement of visiting every city exactly once while minimizing the overall distance or travel time needed to visit each city. Since no known method can solve this problem optimally in polynomial time for all potential inputs, it is NP-hard. To address the TSP, scientists have created a number of precise and heuristic algorithms, including the simulated annealing technique, branch and bound approach, genetic algorithms, and ant colony optimization.

The practical applications of the TSP are diverse and impactful. In logistics and transportation, it guides efficient route planning for delivery vehicles, minimizing fuel consumption and travel time [3]. In telecommunications, it influences network design and optimization, ensuring cost-effective and reliable connectivity [19]. Moreover, in manufacturing and supply chain management, the TSP informs inventory routing and distribution strategies, optimizing resource utilization and reducing operational costs [24]. The TSP's ubiquity across multiple domains underscores its significance as a fundamental problem in combinatorial optimization, driving advancements in algorithmic techniques and real-world decision-making processes.

1.2.4 Vehicle Routing Problem

Another well-known example of combinatorial optimization is the Vehicle Routing Problem (VRP), which focuses on scheduling and routing a fleet of vehicles to efficiently serve a set of clients or locations. Originally presented by [22], the VRP tackles the difficult decision-making associated with resource allocation, route optimization, and vehicle allocation. According to [26], the VRP comprises the process of "designing minimum-cost vehicle routes for serving a given set of customers."

Minimizing the overall travel distance, time, or expense in the VRP is the goal, as long as client demand is satisfied and vehicle capacity restrictions are adhered to. Because it is combinatorial in nature—that is, the number of alternative routes increases exponentially with the number of customers—this problem is NP-hard. For the VRP, researchers and practitioners have created a variety of solution strategies. These include heuristic and metaheuristic techniques like genetic algorithms, tabu search, and particle swarm optimization, as well as exact algorithms like branch and bound.

The applications of the VRP are extensive across logistics, transportation, and supply chain management. In urban logistics, it optimizes delivery routes for trucks or vans, reducing fuel consumption and environmental impact [27]. In public transportation, it guides bus or taxi schedules to improve service reliability and passenger satisfaction [28]. Furthermore, in healthcare and emergency services, the VRP aids in ambulance routing and medical supply distribution during critical situations [29]. The VRP's versatility and relevance in diverse contexts highlight its importance as a fundamental problem in combinatorial optimization, driving advancements in algorithmic efficiency and operational decision-making.

1.2.5 Stochastic and Dynamic Nature

The Stochastic and Dynamic aspects of the Vehicle Routing Problem (VRP) introduce additional complexity and realism to the traditional VRP models, reflecting the dynamic and uncertain nature of real-world logistics and transportation scenarios.

Stochastic VRP (SVRP) involves uncertainties in various parameters, such as customer demands, travel times, and vehicle availability. These uncertainties can arise due to factors like traffic fluctuations, weather conditions, or variations in customer behavior. In SVRP, the goal is to develop robust and adaptive routing strategies that can handle these uncertainties effectively. This often requires the integration of probabilistic models, scenario analysis, and risk management techniques into VRP solutions. For instance, researchers have explored stochastic programming, Monte Carlo simulation, and robust optimization approaches to address uncertainties in SVRP and improve the reliability of routing plans.

On the other hand, Dynamic VRP (DVRP) deals with changes that occur in real-time or near-real-time during the routing process. These changes could include new customer orders, cancellations, delays, or disruptions in vehicle availability. DVRP focuses on dynamic decision-making, where routing plans need to be continuously updated and optimized based on evolving information. This necessitates the use of online algorithms, real-time data integration, and adaptive routing strategies. Techniques like online optimization algorithms, reinforcement learning, and adaptive heuristics are employed to handle the dynamic nature of DVRP and make agile routing decisions in response to changing conditions.

The combination of stochastic and dynamic elements in the Stochastic Dynamic Vehicle Routing Problem (SDVRP) further enhances the complexity of VRP models. SDVRP involves both uncertainties in parameters and dynamic changes in real-time, presenting a multifaceted optimization challenge. Addressing SDVRP requires advanced algorithmic approaches that can handle both stochasticity and dynamism effectively. Hybrid methods combining stochastic optimization, online learning, and adaptive heuristics are explored to develop robust and responsive routing solutions for SDVRP.

In practical applications, integrating stochastic and dynamic considerations into VRP models is crucial for improving operational efficiency, customer satisfaction, and resource utilization in dynamic and uncertain environments. SDVRP research contributes to enhancing the resilience and agility of transportation and logistics systems, ensuring effective routing strategies in the face of unpredictable events and evolving conditions.

1.2.6 Solution methods

Various solution methods are employed to address the Vehicle Routing Problem (VRP), each with its strengths and limitations. Here's a brief overview of some common solution methods:

- *Exact Methods:*
  - **Description**: Exact methods aim to find the globally optimal solution by exploring the entire solution space systematically.
  - **Approach:** These methods formulate the VRP as a mathematical optimization problem and use techniques such as Integer Linear Programming (ILP) and Mixed-Integer Linear Programming (MILP) to solve it.
  - **Strengths:** Guaranteed optimality, suitable for small to medium-sized instances.
  - **Limitations:** Computational complexity limits scalability to large problem sizes.
- *Heuristic Methods:*
  - **Description**: Heuristic methods provide approximate solutions quickly, often sacrificing optimality for computational efficiency.
  - **Approach:** These methods employ rules of thumb, iterative improvement, or greedy algorithms to construct feasible solutions.
  - **Strengths:** Fast computation, applicable to large problem instances.
  - **Limitations:** Solution quality may vary, not guaranteed to find optimal solutions.
- *Metaheuristic Methods:*
  - **Description**: Metaheuristic methods are general-purpose algorithms designed to efficiently explore large solution spaces.
  - **Approach:** These methods incorporate adaptive strategies, stochastic elements, and neighborhood search techniques to iteratively improve solutions.
  - **Strengths:** Versatile, effective for large and complex instances, often providing near-optimal solutions.
  - **Limitations:** Solution quality depends on algorithm parameters and implementation
- *Hybrid Methods:*
  - **Description**: Hybrid methods combine multiple solution approaches to leverage their respective strengths and mitigate weaknesses.
  - **Approach:** These methods integrate components from exact, heuristic, and metaheuristic methods to improve solution quality and computational efficiency.
  - **Strengths:** Synergistic combination of different algorithms, enhanced performance, and robustness.

- **Limitations:** Designing and tuning hybrid methods require expertise and computational resources.

- *Simulation-Based Optimization:*
  - **Description**: Simulation-based optimization methods model the VRP as a simulation environment and use optimization techniques to improve system performance.
  - **Approach:** These methods combine simulation modeling with optimization algorithms to iteratively explore and refine routing strategies.
  - **Strengths:** Capture dynamic and stochastic aspects of VRP, suitable for complex and uncertain environments.
  - **Limitations:** Computational overhead, complexity in modeling interactions between vehicles and environment.

1.2.7 Exact Methods

Exact methods for solving the Vehicle Routing Problem (VRP) rely on mathematical programming techniques to find the globally optimal solution. One of the key components in using exact methods is formulating the VRP as a mathematical model that captures the routing decisions, objective function, and constraints of the problem. Decision variables are typically binary indicators representing whether a vehicle visits a customer or travels between locations. The objective function aims to minimize total costs, distances, or time, while constraints ensure that routing plans adhere to vehicle capacity, route continuity, and time window limitations [30].

Mathematical models for VRP are often formulated as Integer Linear Programming (ILP) problems, where decision variables are integers (0 or 1) representing routing decisions. These models are then solved using optimization solvers capable of handling ILP or Mixed-Integer Linear Programming (MILP) problems. For instance, commercial solvers like CPLEX and Gurobi use sophisticated algorithms such as branch and bound, cutting planes, and branch and cut to explore the solution space and find the globally optimal solution [22].

The optimization process involves iteratively exploring feasible solutions, evaluating their objective values, and refining the search to converge toward the optimal solution. Branch and bound techniques partition the search space into smaller subspaces, pruning branches that cannot lead to better solutions while cutting plane methods dynamically add constraints to tighten the formulation and improve solution quality. Once the optimization solver completes its search, it returns the optimal routing plan that minimizes costs, satisfies constraints, and efficiently utilizes vehicles and resources, providing a benchmark for evaluating the quality of solutions obtained from heuristic or metaheuristic approaches [31].

1.2.8 Heuristic

Heuristic methods are widely used for solving the Vehicle Routing Problem (VRP) due to their ability to provide near-optimal solutions efficiently. One common heuristic approach is the Clarke and Wright Savings Algorithm, which constructs initial routes by combining pairs of customers with the largest potential savings in travel costs [32]. This method reduces computational complexity by focusing on significant cost savings early in the routing process, although it may not always guarantee the optimal solution.

Another heuristic method is the Sweep Algorithm, which works well for VRPs with clustered or geographically dispersed customers. This algorithm divides the geographical area into sectors using a "sweep" line and assigns customers to routes based on their proximity to the sweep line, effectively minimizing travel distances [33]. The Sweep Algorithm's simplicity and effectiveness make it suitable for quickly generating feasible solutions, especially in scenarios where customer locations are spatially concentrated.

Additionally, Metaheuristic methods like Ant Colony Optimization (ACO) have gained popularity for tackling VRPs. ACO mimics the foraging behavior of ants to iteratively build and improve routes, leveraging pheromone trails to guide exploration towards promising solutions [34]. ACO and other metaheuristic algorithms offer advantages such as adaptability to different problem structures, robustness against local optima, and scalability to large problem instances. While heuristic methods may not guarantee optimality, their efficiency and effectiveness in finding good-quality solutions make them invaluable tools for solving complex routing problems like the VRP.

1.2.9 Metaheuristic

Metaheuristic methods are powerful tools for tackling the Vehicle Routing Problem (VRP) by efficiently exploring large solution spaces and finding near-optimal solutions. One widely used metaheuristic is Genetic Algorithms (GA), inspired by the process of natural selection and evolution. GAs iteratively evolve a population of candidate solutions through selection, crossover, and mutation operators, gradually improving the solutions over generations [35]. The diversity-preserving mechanisms of GAs help avoid premature convergence to local optima, making them effective for exploring diverse routing solutions in VRPs.

Another prominent metaheuristic is Simulated Annealing (SA), which mimics the annealing process in metallurgy to iteratively improve solutions by accepting occasional uphill moves based on a temperature parameter [8]. SA allows for the exploration of suboptimal solutions early in the search, gradually decreasing the probability of accepting worse solutions as the search progresses. This balancing of exploration and exploitation makes SA suitable for VRPs with complex solution spaces and rugged landscapes, where finding globally optimal solutions is challenging.

Particle Swarm Optimization (PSO) is another metaheuristic method that simulates the collective behavior of particles in a search space. PSO iteratively updates the position and velocity of particles based on their best-known positions and the global best position found by the swarm [36]. PSO's ability to balance local exploration and global search, along with its simplicity and fast convergence properties, has made it a popular choice for optimizing VRPs and similar combinatorial problems. Metaheuristic methods like GA, SA, and PSO offer flexible and robust optimization techniques for VRPs, capable of handling complex problem structures, dynamic environments, and uncertainties effectively.

1.2.10 Reinforcement Learning

Reinforcement Learning (RL) methods have gained prominence in addressing the Vehicle Routing Problem (VRP) by leveraging learning and adaptation to improve routing strategies over time. One notable RL approach is Q-learning, which learns a Q-value function representing the expected cumulative rewards for taking actions in a given state [37]. In VRPs, Q-learning can be applied to dynamically update vehicle routes based on real-time feedback, such as customer demand changes or traffic conditions. This adaptive nature of RL allows for continuous improvement and optimization of routing decisions in dynamic environments.

Another RL technique, Deep Q-Networks (DQN), combines Q-learning with deep neural networks to handle complex and high-dimensional state-action spaces. DQN has been applied to VRPs by representing routing decisions as actions and using deep neural networks to approximate the Q-value function [38]. By leveraging deep learning capabilities, DQN can effectively learn intricate patterns and dependencies in VRPs, leading to more informed and adaptive routing strategies.

Additionally, Policy Gradient methods in RL offer a direct approach to learning optimal policies for VRPs without explicitly computing value functions. Algorithms like REINFORCE [12] and Proximal Policy Optimization (PPO) [39] have been applied to VRPs by learning policies that directly map states to actions (i.e., vehicle routes). These methods enable RL agents to learn complex decision-making policies that optimize routing objectives over time, making them suitable for dynamic and uncertain VRP scenarios. RL methods offer a promising avenue for addressing VRPs by integrating learning, adaptation, and optimization into routing strategies, leading to more robust and efficient solutions.

Chapter 2

Literature Review

The Stochastic Dynamic Vehicle Routing Problem (SDVRP) is a variant of the classical Vehicle Routing Problem (VRP) that incorporates both stochastic (uncertain) and dynamic (real-time changing) elements. SDVRP is of significant importance in real-world logistics and transportation due to its ability to model and address the inherent uncertainties and dynamic changes present in transportation systems. For instance, customer demands may vary unpredictably, traffic conditions can fluctuate, and vehicle availability may change dynamically, making traditional deterministic routing models inadequate for real-world applications [22]. The challenges posed by stochasticity and dynamicity in SDVRP are multifaceted. Stochasticity introduces uncertainty in parameters such as customer demands, travel times, and service requirements. This uncertainty necessitates the use of probabilistic models and robust optimization techniques to account for varying scenarios and ensure reliable routing decisions [40]. On the other hand, dynamicity involves real-time changes that require adaptive routing strategies capable of responding to sudden events, traffic disruptions, and new service requests while maintaining operational efficiency and customer satisfaction [41]. Integrating stochastic and dynamic elements into VRP models adds a layer of complexity to the optimization process. Traditional VRP models are deterministic and do not account for uncertainties or real-time changes, making them ill-suited for dynamic and stochastic environments. The integration of stochastic and dynamic elements requires advanced modeling techniques such as stochastic dynamic programming, chance-constrained programming, and online optimization algorithms [42]. This complexity arises from the need to balance optimization objectives, handle uncertainty, adapt to changing conditions, and ensure computational tractability in solving SDVRP instances.

Stochastic approaches play a crucial role in addressing the Stochastic Dynamic Vehicle Routing Problem (SDVRP) by incorporating uncertainty into routing decisions. One prominent stochastic modeling technique is stochastic programming, which formulates the problem with probabilistic constraints and objectives. For instance, chance-constrained programming ensures that the probability of violating constraints (e.g., vehicle capacity, time windows) remains below a specified threshold, considering uncertain parameters like customer demands and travel times [43]. Stochastic dynamic programming extends this approach to dynamic environments, where decisions are made sequentially over time under uncertainty, requiring adaptive strategies and dynamic optimization models [25]. Scenario-based approaches are another class of stochastic methods used in SDVRP, where multiple scenarios representing different possible outcomes are considered. These scenarios capture the uncertainty in parameters such as demand fluctuations, traffic conditions, and service disruptions. Algorithms like scenario generation and scenario decomposition are employed to handle the complexity of generating and analyzing a large number of scenarios efficiently [44]. Additionally, robust optimization techniques aim to develop routing plans that are resilient to uncertainties, ensuring satisfactory performance across a range of possible scenarios [45]. These stochastic approaches provide robust and reliable solutions for SDVRP by considering uncertainty explicitly in the optimization process. Moreover, metaheuristic algorithms like Genetic Algorithms (GAs) and Particle Swarm Optimization (PSO) can be extended to handle stochastic variations in SDVRP. For instance, incorporating stochastic elements into the fitness evaluation of genetic operators or swarm dynamics in PSO allows these algorithms to adapt to uncertain conditions and explore diverse solutions efficiently [4]. Hybrid approaches that combine stochastic optimization techniques with metaheuristics further enhance the solution quality and robustness of SDVRP solutions [46]. These methods leverage the strengths of both stochastic and metaheuristic approaches to tackle the challenges posed by uncertainty and dynamicity in vehicle routing effectively.

Dynamic approaches are essential for addressing the Stochastic Dynamic Vehicle Routing Problem (SDVRP), where routing decisions must adapt to real-time changes and uncertainties. One of the key dynamic strategies is real-time re-optimization, where routing plans are continuously updated based on incoming information. This approach requires efficient algorithms capable of quickly recalculating routes and schedules to accommodate new orders, traffic conditions, and unexpected events [47]. Real-time re-optimization algorithms dynamically adjust vehicle routes and schedules to minimize costs, improve service levels, and maintain operational efficiency in dynamic environments. Online algorithms represent another dynamic approach for SDVRP, where routing decisions are made incrementally as new information becomes available. These algorithms balance the trade-off between exploiting current information and exploring alternative options to adapt to changing conditions effectively [48]. Adaptive heuristics, such as Ant Colony Optimization (ACO) and Genetic Algorithms (GAs), can be adapted to online settings, allowing for continuous learning and improvement of routing strategies over time [49]. These online and adaptive algorithms provide robust and flexible solutions for SDVRP by incorporating real-time information and adapting to dynamic changes efficiently. Furthermore, real-time scheduling techniques are crucial in dynamic vehicle routing, especially in scenarios with time-sensitive deliveries or service windows. Dynamic vehicle dispatching algorithms allocate vehicles to tasks dynamically based on real-time demand and operational constraints [50]. These algorithms consider factors like vehicle availability, travel times, customer priorities, and delivery deadlines to optimize vehicle routes and schedules dynamically. By leveraging real-time data and advanced scheduling algorithms, dynamic approaches in SDVRP enhance responsiveness, reduce operational costs, and improve customer satisfaction in dynamic and uncertain environments.

Integrated approaches for the Stochastic Dynamic Vehicle Routing Problem (SDVRP) aim to synergistically combine stochastic and dynamic elements in routing models, addressing both uncertainty and real-time changes simultaneously. One integrated approach involves combining stochastic optimization techniques, such as chance-constrained programming or robust optimization, with dynamic programming strategies. For instance, [51] proposed a stochastic and dynamic vehicle routing model that considers both uncertain customer demands and dynamic vehicle availability, optimizing routes and schedules adaptively over time. This integrated approach allows for robust decision-making that accounts for stochastic variations and dynamic adjustments in routing plans. Another integrated strategy is the combination of metaheuristic algorithms with real-time re-optimization techniques. [52] introduced a hybrid algorithm that combines a genetic algorithm (GA) with real-time re-optimization for SDVRP. The GA component provides an initial solution while the real-time re-optimization component continuously updates the routes based on incoming information, dynamically adapting to changing conditions. This integrated approach leverages the strengths of both metaheuristic optimization and dynamic routing to achieve robust and responsive solutions for SDVRP. Additionally, reinforcement learning (RL) methods have been integrated into SDVRP models to learn adaptive routing policies over time. [53] proposed an RL-based approach for SDVRP where an RL agent learns routing policies based on historical data and real-time feedback. The RL agent continuously refines its policies to optimize routing decisions under uncertainty and dynamic changes. This integrated RL approach offers flexibility and adaptability in addressing SDVRP challenges by learning from experience and dynamically adjusting to evolving conditions. Also, hybrid optimization frameworks that combine mathematical programming, metaheuristics, and simulation-based optimization have been explored for SDVRP. [54] developed a hybrid optimization framework that integrates a mathematical programming model with a genetic algorithm and simulation-based optimization for dynamic vehicle routing under uncertainty. This integrated framework combines the strengths of different optimization paradigms to tackle the complexity of SDVRP effectively, providing robust and efficient routing solutions.

Performance evaluation and comparative studies play crucial roles in assessing the efficacy and efficiency of solutions designed for the Stochastic Dynamic Vehicle Routing Problem (SDVRP). Various metrics and methodologies are utilized to compare different solution approaches and evaluate their performance across diverse scenarios.

One fundamental metric utilized in performance evaluation is the total cost or objective function value attained by a routing solution. This encompasses factors such as the overall travel distance, vehicle operational costs, service time, and penalties for constraint violations [55]. Comparative studies often delve into the trade-offs between different objectives, such as minimizing costs versus maximizing customer satisfaction or vehicle utilization. Service level metrics are critical for evaluating the quality of routing solutions in SDVRP. These metrics include on-time delivery performance, adherence to time windows, percentage of completed orders, and customer satisfaction scores [56]. Such metrics offer insights into the reliability and effectiveness of routing plans in meeting customer demands and operational requirements. Comparative studies also make use of benchmarking techniques to compare the performance of different solution methods against established benchmarks or known optimal solutions. Benchmark instances derived from real-world data or standardized test problems serve as a foundation for evaluating the scalability, robustness, and computational efficiency of SDVRP solutions [57]. Additionally, these studies consider the computational complexity and scalability of solution methods, particularly for large-scale SDVRP instances, to evaluate their practical feasibility and applicability.

Applications and case studies offer practical insights into the real-world implementation and effectiveness of solutions for the Stochastic Dynamic Vehicle Routing Problem (SDVRP). These examples demonstrate how SDVRP methodologies are applied in various industries and settings, showcasing their impact on operational efficiency, cost savings, customer satisfaction, and sustainability. One notable application of SDVRP is in the logistics and transportation industry. Companies dealing with delivery services, such as e-commerce platforms and parcel delivery firms, leverage SDVRP solutions to optimize their vehicle routing and scheduling. For instance, Amazon uses advanced algorithms and real-time data to dynamically route delivery vehicles, considering factors like traffic conditions, customer locations, and delivery time windows [58]. Such applications demonstrate the practical relevance and benefits of SDVRP methodologies in streamlining logistics operations and improving delivery performance. Emergency services also benefit from SDVRP solutions, particularly in dynamic and uncertain environments. Ambulance routing and emergency response planning are critical areas where SDVRP techniques are applied to optimize vehicle dispatching, reduce response times, and improve patient outcomes. Studies have shown that dynamic vehicle routing algorithms can significantly enhance emergency response efficiency by considering real-time traffic information, demand patterns, and geographic constraints [59]. These applications highlight the life-saving potential of SDVRP methodologies in emergency services management. Urban planning and public transportation systems utilize SDVRP strategies to optimize public transit routes, school bus schedules, and waste collection services. By incorporating stochastic and dynamic elements into routing decisions, cities and municipalities can achieve cost savings, reduce environmental impact, and enhance service quality for residents [60]. Case studies in urban logistics and public transportation showcase how SDVRP solutions contribute to sustainable urban mobility and resource optimization.

In conclusion, the literature review provides a comprehensive overview of the Stochastic Dynamic Vehicle Routing Problem (SDVRP), its challenges, and various solution approaches. The integration of stochastic and dynamic elements in routing decisions has been explored through stochastic programming, dynamic programming, metaheuristics, and hybrid optimization frameworks. Performance evaluation and comparative studies have highlighted the importance of metrics such as total cost, service levels, and computational efficiency in assessing the efficacy of SDVRP solutions.

Looking ahead, upcoming promising methods in SDVRP include reinforcement learning (RL) approaches. RL methods, such as Deep Q-Networks (DQN), Proximal Policy Optimization (PPO), and actor-critic algorithms, offer adaptive and learning-based strategies for dynamic and uncertain routing environments. These RL methods learn from experience, adapt to changing conditions, and optimize routing policies over time, making them well-suited for SDVRP applications [1]. RL-based approaches have the potential to enhance the responsiveness, flexibility, and robustness of SDVRP solutions by leveraging machine learning techniques and real-time data.

Advancements in RL algorithms, such as model-based RL and multi-agent RL, are expected to further improve the performance of SDVRP solutions. Model-based RL techniques leverage learned models of the environment to plan and optimize routing decisions, reducing the need for extensive exploration in uncertain environments [12]. Multi-agent RL frameworks enable cooperative decision-making among multiple vehicles or agents, enhancing coordination and efficiency in dynamic routing scenarios [54].

In conclusion, upcoming RL methods hold promise for addressing the challenges of the Stochastic Dynamic Vehicle Routing Problem (SDVRP) by offering adaptive, learning-based, and cooperative approaches to routing optimization. These methods are expected to contribute to the development of more intelligent, efficient, and scalable solutions for dynamic and uncertain transportation environments.

Chapter 3

Problem Statement

3.1 Stochastic Dynamic Vehicle Routing problem

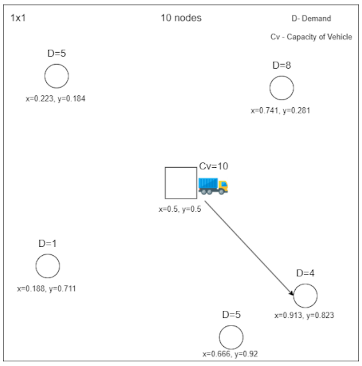

Figure 3.1 - Phase 1

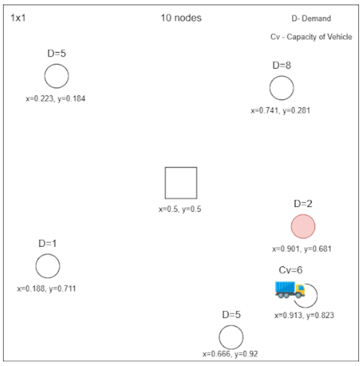

Figure 3.2 - Phase 2

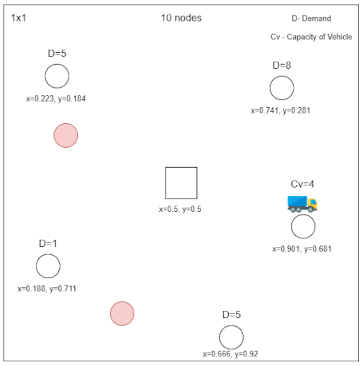

Figure 3.3 - Phase 3

The Stochastic Dynamic Vehicle Routing Problem (SDVRP) is a multifaceted optimization challenge prevalent in logistics and transportation management. At its core, SDVRP revolves around dynamically assigning vehicles to customer locations in real time while factoring in stochastic and dynamic elements. Nodes in the problem represent these customer locations or delivery points, while edges denote possible routes connecting them. This problem can be modeled as a graph problem, where each edge has associated costs or distances, and each node has attributes like customer demand, time windows for service, and vehicle capacity constraints.

SDVRP is a classic example of a combinatorial optimization (CO) problem, characterized by the need to select optimal combinations of vehicle routes to efficiently serve customer requests. However, the problem's complexity escalates due to its NP-hard nature, implying that finding an optimal solution within a reasonable timeframe becomes increasingly challenging as the problem size grows. Stochastic customer requests further compound the complexity by introducing uncertainty into demand patterns, necessitating robust strategies that can adapt to fluctuating demands while meeting operational constraints and minimizing costs. Vehicle decisions in SDVRP involve route selection, load allocation, and scheduling, taking into account factors such as travel costs, customer priorities, and vehicle capacity limitations.

To tackle the SDVRP effectively, advanced optimization techniques, heuristic algorithms, and simulation-based approaches are employed. These methods aim to strike a balance between computational tractability and solution quality, ensuring that routing decisions remain responsive to real-time changes while optimizing overall operational efficiency. The challenge lies in devising algorithms and decision-making frameworks that can navigate the intricate interplay of stochastic customer demands, dynamic routing conditions, and operational constraints to deliver robust and near-optimal routing solutions.

Figure 3.1: Phase 1 (Truck Going Out) In this phase, Figure 3.1 represents the initial stage where the truck or vehicle is departing from the depot to begin its route. This phase may illustrate the starting point of the vehicle routing process, showing the initial state of the vehicle before encountering any new customers or service points.

Figure 3.2: Phase 2 (New Customers on Route) Figure 3.2 depicts the scenario where new customers or service points are encountered along the route. This phase may highlight the dynamic nature of the vehicle routing problem, where unexpected events such as new customer requests or changes in demand can impact the routing decisions and route optimization strategies.

Figure 3.3: Phase 3 (Serving Customers and Vehicle Capacity) In Phase 3, as shown in Figure 3.3, the focus shift to serving the customers encountered during the route. This phase involves managing the vehicle capacity as services are provided to customers, leading to a decrease in available capacity as deliveries or pickups are made. It showcase how the vehicle dynamically adjusts its route and capacity utilization to efficiently serve customers while considering constraints such as vehicle capacity limits.

These figures appear to illustrate the dynamic and stochastic elements of the vehicle routing process, emphasizing the need for adaptive strategies to handle changing conditions and optimize routing decisions in real-time

3.1.1 Stochastic Customer Request

Stochastic customer requests 1.1 play a pivotal role in the Stochastic Dynamic Vehicle Routing Problem (SDVRP) as they introduce uncertainty into demand patterns, necessitating adaptive routing strategies. These requests are often generated using probabilistic models such as the Poisson distribution, which is widely used to model arrival processes in stochastic systems.

The Poisson distribution 3.4 is a discrete probability distribution that describes the number of events occurring in a fixed interval of time or space, given the average rate of occurrence. In the context of SDVRP, the Poisson distribution can be used to model the arrival of customer requests at various locations over time. For instance, if λ represents the average number of customer requests arriving per unit time (e.g., per hour), the Poisson distribution can estimate the probability of

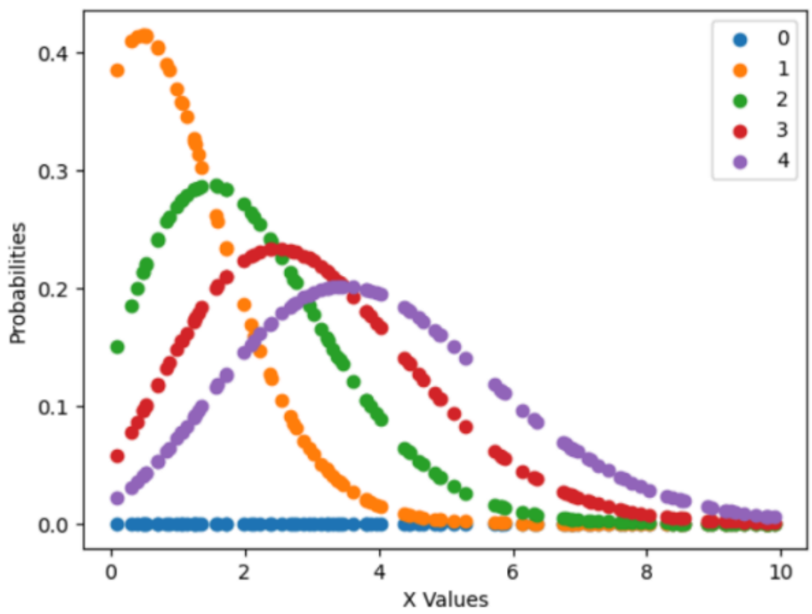

Figure 3.4 - Poisson distribution.

observing a specific number of customer arrivals during a given time period.

The stochastic nature of customer requests generated by the Poisson distribution reflects the inherent unpredictability in real-world demand patterns. Customers may request services at varying frequencies and volumes, influenced by factors such as time of day, day of the week, seasonal variations, and external events. By modeling customer requests stochastically using the Poisson distribution, SDVRP algorithms can account for these fluctuations and adjust routing decisions dynamically.

Adapting to stochastic customer requests involves developing routing strategies that can optimize vehicle routes in response to changing demand patterns. Advanced SDVRP algorithms leverage probabilistic models like the Poisson distribution to forecast future demand, estimate the likelihood of service requests at different locations and times, and optimize routing plans accordingly. This adaptive approach helps improve resource utilization, reduce service delays, and enhance overall operational efficiency in dynamic and uncertain transportation environments.

3.1.2 Single depot and single vehicle

In this study, we have opted for a single vehicle and single depot configuration to focus our analysis on the fundamental dynamics and challenges inherent in the Stochastic Dynamic Vehicle Routing Problem (SDVRP). By narrowing the scope to a single vehicle and depot, we can delve deeper into the complexities of adaptive routing strategies under stochastic and dynamic conditions while considering factors such as customer demand variability, time constraints, and vehicle capacity constraints. This focused approach allows us to develop and evaluate routing algorithms specifically tailored for constrained settings, providing valuable insights into resource allocation optimization, real-time adaptation, and operational efficiency within a dynamic and uncertain routing environment.

3.1.3 Time constraint

In our SDVRP problem setup, we have incorporated a time constraint that dictates that a customer must be served within a specific timeframe after their request appears. Specifically, we have defined this time constraint as requiring the service to be provided within a window of +5 unit times after the customer request is generated. This constraint ensures that customer requests are handled promptly and within reasonable timeframes, reflecting real-world scenarios where timely delivery or service completion is crucial for customer satisfaction and operational efficiency. By incorporating such time constraints into our problem formulation, we aim to optimize routing decisions while balancing the need for responsiveness and adherence to service deadlines, thereby enhancing the overall quality and effectiveness of our routing strategies.

3.1.4 Objective

Our primary objective is to maximize the total demand fulfilled from customers. This objective aligns with the overarching goal of efficiently utilizing our resources, specifically the single vehicle, and depot, to serve as many customer requests as possible within the defined time constraints and operational limitations. By focusing on maximizing demand fulfillment, we aim to optimize routing decisions to prioritize high-demand areas or customers, allocate resources effectively, and minimize unmet demand or missed opportunities. This objective reflects a strategic approach to routing optimization that emphasizes customer satisfaction, revenue generation, and resource utilization efficiency, contributing to improved operational performance and overall business success in dynamic and uncertain routing environments.

3.2 Markov Decision Process

Markov Decision Process (MDP) serves as a foundational mathematical framework used to model decision-making challenges within stochastic environments. It provides a structured approach to defining the state space, action space, reward function, and transition function, which are essential components for constructing the reinforcement learning framework necessary to address complex decision-making problems. In essence, MDP allows us to formalize and systematically approach problems where decisions need to be made sequentially in uncertain conditions, enabling the development of effective strategies for navigating dynamic and unpredictable scenarios.

To elaborate further on the concept of Markov Decision Process (MDP), we can break down its components in detail. The State space (S_k) encompasses all possible states or configurations of the system at any given time, representing the information available to decision-makers. The Action space (x_k) comprises the set of feasible actions that can be taken in each state, influencing the system's transition to subsequent states. The Reward function ($R(S_k, x_k)$) assigns a numerical value or reward to each state-action pair, reflecting the desirability or utility of taking specific actions in particular states. Lastly, the Transition function (ω_k) defines the probability distribution of transitioning from one state to another based on the chosen action, incorporating stochastic elements that capture the uncertainty inherent in the environment.

In summary, Markov Decision Process (MDP) provides a rigorous and systematic framework for modeling decision-making under uncertainty, allowing for the formulation of reinforcement learning algorithms that learn optimal strategies through interactions with the environment. By delineating the state space, action space, reward function, and transition function, MDP enables a structured approach to tackling stochastic decision-making problems, facilitating the development of robust and adaptive decision strategies.

3.2.1 State

In our study, a state, represented as S_k, embodies a comprehensive snapshot of specific information values at a given time epoch k. This state comprises multiple key elements that collectively define the system's status: firstly, a mask M_j, which is a binary set of values 0, 1, indicating whether a particular customer is available (0) or not (1 - served or didn't appear); secondly, the demand of each customer D_j, which ranges from 1 to 9 and is associated with a set of nodes $j=\{1, ...n\}$; thirdly, the current location of the vehicle denoted as C_v; and lastly, the current capacity of the vehicle denoted as D_v. In essence, a state can be succinctly represented as $S_k = (M_j, D_j, C_v, D_v)$, encapsulating the essential variables and parameters that define the system's state at a specific time instance.

In the context of our research, each state serves as a detailed snapshot of the ongoing situation, capturing crucial details such as customer availability, individual customer demands, the vehicle's current position, and its remaining capacity. The binary mask M_j plays a critical role in indicating which customers can be served at a given time, guiding decision-making regarding customer selection for service. The varying demand levels D_j across different customers contribute to the complexity of the state, reflecting the diverse service requirements of each customer node. Additionally, the vehicle's location C_v and capacity D_v are integral components of the state, influencing route planning, resource allocation, and decision-making strategies aimed at optimizing customer service and overall operational efficiency. Thus, by defining and understanding the elements comprising a state, we gain valuable insights into the system's dynamics and the variables that drive decision-making processes in the context of the Stochastic Dynamic Vehicle Routing Problem (SDVRP).

3.2.2 Reward

The Reward function ($R(S_k, x_k)$) plays a pivotal role in our problem as it directly reflects our objective of maximizing customer demand fulfillment. By serving customers effectively, we aim to maximize the value of our reward function. This function serves as a measure of the desirability or utility associated with taking specific actions (represented by x_k) in particular states (represented by S_k). In essence, a higher reward value signifies a more favorable outcome in terms of meeting customer demands and achieving our overarching goal of demand maximization.

Our approach to maximizing the reward function aligns with our strategic decision-making process. By prioritizing customer demand fulfillment, we ensure optimal resource allocation and utilization. This entails making informed decisions regarding route planning, customer order sequencing, and resource allocation to effectively serve as many customers as possible within the given constraints. The reward function acts as a guiding metric, incentivizing actions that lead to increased demand fulfillment while discouraging suboptimal decisions that may result in unmet customer demands or inefficient resource utilization.

Through the optimization of the reward function, we aim to strike a balance between operational efficiency and customer satisfaction. By incorporating customer demand maximization as a primary objective within our reward function framework, we can develop robust decision-making strategies that adapt dynamically to changing demand patterns and environmental conditions. Ultimately, our focus on maximizing the reward function reflects our commitment to delivering high-quality service, optimizing resource utilization, and achieving operational excellence in the context of stochastic and dynamic decision-making environments.

3.2.3 Action

The action x_k represents the decision made by the vehicle at each step of the routing process. It involves selecting one customer from the available set of customers, denoted by $j = \{1, ...n\}$, where n represents the total number of customers that need to be served. This action is pivotal as it determines which customer the vehicle will visit next during its route. The selection process is conducted with careful consideration of a mask M_j that assesses whether the vehicle can effectively serve the chosen customer based on various factors such as vehicle capacity, time constraints, and geographical proximity.

The decision-making process involving action x_k is repeated iteratively at each step of the route as the vehicle progresses through the service area. At each stage, the vehicle evaluates the available options from the set of customers and chooses the most optimal customer to serve based on predefined criteria and constraints. This iterative selection process enables the vehicle to dynamically adapt its routing decisions in response to changing conditions, customer demands, and operational requirements.

By utilizing the action x_k to guide the vehicle's movements and customer service decisions, we can develop efficient routing strategies that prioritize customer satisfaction, minimize travel distances, and optimize resource utilization. The iterative nature of action selection ensures that the vehicle's route is continually optimized throughout the service delivery process, contributing to improved operational efficiency and overall performance in the context of the Stochastic Dynamic Vehicle

Routing Problem (SDVRP).

3.2.4 Transition function

The transition function ω_k in our problem can be deconstructed into two key components: the pre-decision state S_k and the post-decision state S_{k+1}. The pre-decision state S_k encapsulates the problem's state before the vehicle executes the action x_k, encompassing various parameters such as the mask M_j representing customer availability, the current demand D_j of customers, the vehicle's location C_v, and its load D_v. When the vehicle is in the pre-decision state, it takes into account the initial time t_0 along with the additional time t_Δ required to reach the intended destination node after action x_k is executed. This resultant time, denoted as t_x, signifies the post-decision state time. Upon executing action x_k, several changes occur in the post-decision state S_{k+1}. Firstly, the vehicle's location transitions from C_v to the location of the last served customer, denoted as C_v^x. Additionally, the vehicle's load is updated to $D_v^x = D_v - D_j$, reflecting the reduction in load after serving the customer whose demand was D_j. Since customer requests are stochastic, some requests occur randomly, modeled using the Poisson distribution. Consequently, the demand D_j and mask M_j are updated to D_x^j and M_x^j respectively in the post-decision state, reflecting the dynamic nature of customer demand and availability. In essence, the pre-decision state S_K is represented as $S_k = (M_j, D_j, C_v, D_v)$, encapsulating the state of the system before action x_k is executed. Conversely, the post-decision state $S_{k+1} = (M_j^x, D_j^x, C_v^x, D_v^x)$., signifying the updated state of the system after x_k action has been executed. This transition process, governed by the transition function Δ_k, enables us to model the dynamic changes that occur in the system as actions are executed and customer requests are serviced, providing a comprehensive framework for analyzing and optimizing the Stochastic Dynamic Vehicle Routing Problem (SDVRP).

Chapter 4

Methodology

4.1 Reinforcement Learning

A dynamic method for resolving the complex problems presented by the stochastic dynamic vehicle routing problem (SDVRP) is provided by Reinforcement learning (RL) [61]. Within this framework, an agent maneuvers through the intricate state space of the SDVRP, which includes factors such as vehicle capacities, time limits, consumer locations, and dynamic demand patterns. The agent wants to learn the best strategies for making decisions by interacting with the environment iteratively.

At each step, the RL agent selects actions based on its current state, aiming to maximize cumulative rewards over time. These actions in SDVRP might include choosing which customer to serve next, determining the most efficient route, and managing vehicle resources effectively. The agent's learning process involves exploration, where it tries out different actions to learn about their consequences, and exploitation, where it leverages its learned knowledge to make informed decisions. Through this iterative learning process, the agent adapts its strategies, gradually improving its routing decisions and optimizing resource allocation 4.1.

Challenges in applying RL to SDVRP include handling high-dimensional state

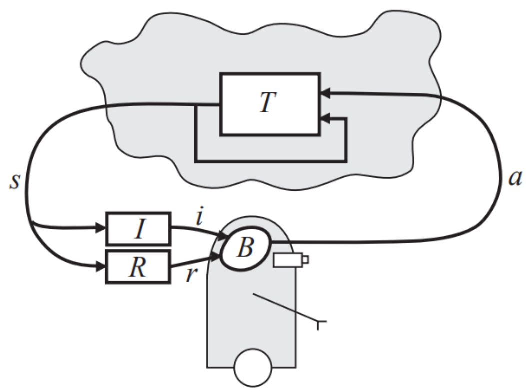

Figure 4.1 - Reinforcement learning simple model.[61]

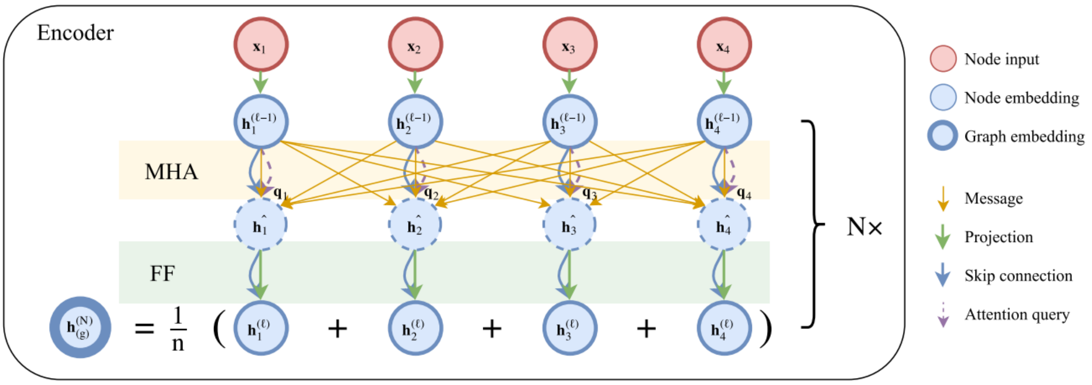

Figure 4.2 - Attention model(encoder architecture).[62]

spaces, managing stochastic and dynamic customer demands, and balancing exploration and exploitation. Techniques such as function approximation, experience replay, and exploration strategies like epsilon-greedy are employed to address these challenges. Overall, RL provides a robust framework for intelligent decision-making in SDVRP, leading to enhanced operational efficiency, improved resource utilization, and more responsive customer service in dynamic and uncertain routing scenarios.

4.2 Multi-Head Attention Model

A potent technique employed in deep learning and natural language processing (NLP) problems, especially in transformer architectures, is the multi-head attention model [62]. By enabling the model to concentrate on various input components at once, it improves its capacity to represent intricate linkages and dependencies within the input data.

A multi-head attention model divides the input into several segments, or "heads," each of which represents a distinct data representation. Attention weights, which establish the relative relevance of various input components for a given activity, are independently learned by each head. After that, these attention weights are aggregated across heads to produce a thorough representation of the input that encompasses various facets and subtleties 4.2.

Because it may listen to multiple areas of the input in parallel, the multi-head attention mechanism has the advantage of helping the model capture both local and global dependencies. This parallel processing enhances the model's capacity to handle long-range dependencies and complex relationships within the data, making it particularly suitable for tasks such as machine translation, text summarization, and language understanding where capturing contextual information is crucial. Overall, the multi-head attention model significantly boosts the performance and capabilities of deep learning models in various NLP and sequence modeling tasks.

Algorithm 1: Rollout

Sample data with batch size, B. Set the maximum number of steps denoted, T which is equal to the number of nodes. Set reward R_0 to zero

Initialize a model consisting of Encoder and Decoder with parameters θ

$C_v^{x_0}, M^{x_0}, D^{x_0}, D_v^{x_0} \leftarrow \textsc{Env.Reset}(B)$

$G_0 = \textsc{Encoder}(B, D^{x_0})$

for k=*0* **to** T **do**

- $x_{k+1}, p_{k+1} \leftarrow \textsc{Decoder}(C_v^{x_k}, M^{x_k}, D^{x_k}, D_v^{x_k}, G_k)$
- $C_v^{x_{k+1}}, M^{x_{k+1}}, D^{x_{k+1}}, D_v^{x_{k+1}} \leftarrow \textsc{Env.Step}(x_{k+1})$
- $G_{k+1} = \textsc{Encoder}(B, D^{x_{k+1}})$
- $R_{k+1} \leftarrow R_k + D^{x_k}$
- $k = k + 1$

return $R_T, \pi = \{p_0, \ldots, p_T\}$

Figure 4.3 - Rollout algorithm[62]

4.3 Rollout Baseline

A rollout baseline is a concept often used in reinforcement learning and decision-making processes, particularly in the context of planning and strategy development. It involves simulating or ”rolling out” possible future trajectories or sequences of actions to estimate the expected value or utility of different decision paths.

In the context of reinforcement learning, a rollout baseline helps evaluate the potential outcomes of different actions by simulating how the environment might evolve based on those actions. This simulation can be done using heuristics, Monte Carlo simulations, or other predictive models to forecast the potential rewards or outcomes of taking specific actions.

The rollout baseline serves as a reference or benchmark against which the actual performance of different decision-making strategies or policies can be compared. It provides an estimate of the expected value of taking certain actions, helping decision-makers make informed choices about which actions are likely to lead to better outcomes 4.4.

Overall, the rollout baseline is a valuable tool in reinforcement learning and planning algorithms, providing a means to estimate the potential performance of different decision strategies and guide decision-making processes towards more favorable outcomes.

4.3.1 Rollout Greedy

1. Greedy rollout is a deterministic strategy where the rollout process selects actions based on a greedy policy that maximizes immediate rewards or outcomes at each step.

2. In a greedy rollout, the algorithm selects the action that appears to be the most advantageous or promising based on the current state of the system, without considering potential future consequences or exploring alternative paths.

3. While greedy rollout can be computationally efficient and straightforward, it may overlook potentially better long-term strategies or miss opportunities for exploration and learning.

4.3.2 Rollout Sampling

1. Sampling rollout, on the other hand, involves stochastic sampling of actions based on a probability distribution, often derived from a learned policy or heuristics.

2. In a sampling rollout, the algorithm randomly samples actions from the probability distribution, allowing for exploration of different paths and potential outcomes.

3. By sampling actions, the algorithm can better capture uncertainty, variability, and the diversity of possible trajectories, leading to a more comprehensive

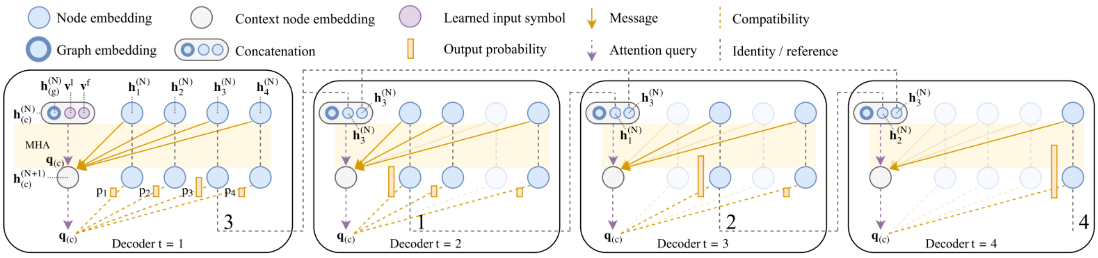

Figure 4.4 - Attention model(decoder architecture).[62]

evaluation of different decision paths.

4. However, sampling rollout can be more computationally intensive compared to greedy rollout, as it involves generating and evaluating multiple action sequences.

4.4 Encoder and Decoder

In an encoder-decoder architecture tailored for vehicle routing problems (VRP), the encoder phase utilizes a graph attention encoder to extract embeddings from the input graph representing the VRP scenario. This encoder leverages graph attention mechanisms to prioritize and aggregate information from relevant nodes and edges in the graph, such as customer locations, vehicle capacities, distances between nodes, and other pertinent attributes. The graph attention mechanism allows the encoder to dynamically assign importance to different parts of the input graph based on their relevance to the task at hand, such as optimizing vehicle routing decisions.

On the other hand, the decoder phase incorporates dynamic values related to vehicle location, vehicle capacity, and customer requests to guide the generation of optimal routing solutions. These dynamic values are essential as they capture the real-time changes and uncertainties inherent in VRP scenarios, such as new customer requests emerging, vehicles moving to different locations, and varying vehicle capacities due to service activities. By incorporating these dynamic values into the decoder, the model can adapt its routing decisions dynamically, ensuring that the generated solutions are responsive to changing conditions and constraints.

For instance, during the decoding process, the model may receive updated information about the current vehicle location, available vehicle capacity, and incoming customer requests. This information is used to guide the generation of routing decisions, ensuring that vehicles are dispatched efficiently to serve customers while adhering to capacity constraints and optimizing travel distances. By incorporating dynamic values into the decoder, the model can generate adaptive and responsive solutions for vehicle routing problems, improving operational efficiency and customer service in dynamic and uncertain environments.

4.5 Training Algorithm

The training algorithm outlined operates in an iterative manner to optimize the main model, consisting of an Encoder and Decoder with parameters denoted as θ, for a given maximum number of epochs E and a significance level α. At the outset, both the main model and the baseline model, characterized by parameters θ and $\theta_B L$ respectively, are initialized. The algorithm then proceeds to loop over the epochs, starting from epoch 0, and performs several key functions at each iteration. During each epoch, the algorithm executes two types of rollouts: sampling-based and greedy-based.

The rollout sampling function employs a sampling strategy to generate rollout sequences, producing a loss L and a policy π. Conversely, the rollout greedy function utilizes a deterministic greedy strategy to generate rollout sequences, resulting in a loss $L_B L$ and a policy $\pi_B L$. The gradients ΔL are then computed based on the differences between the losses from the main model and the baseline model, and the main model's parameters θ are updated using an optimizer such as ADAM. Subsequently, a significance test is conducted to assess if the improvements in the main model's performance are statistically significant compared to the baseline model, with the parameters of the baseline model $\theta_B L$ updated accordingly if the test yields a significant result.

This training algorithm effectively balances exploration and exploitation through rollout strategies, updates model parameters based on gradient information, and ensures statistical significance in model improvements compared to a baseline. By iteratively refining the main model and updating the baseline model as needed, the algorithm aims to achieve optimal performance in learning and decision-making tasks while maintaining statistical robustness.

4.6 Hyperparameters

Hyperparameters are essential settings that guide the training process of machine learning models. In the context of your scenario, several hyperparameters are specified, including the Adam optimizer settings, learning rates, batch size, number of epochs, and learning rate decay.

1. **Adam Optimizer:** The Adam optimizer is a popular optimization algorithm used in training neural networks. It combines the benefits of two other optimization techniques, AdaGrad and RMSProp, by incorporating adaptive learning rates and momentum. Adam optimizer adapts the learning rates for each parameter individually based on the first and second moments of the gradients, making it suitable for a wide range of optimization problems.

2. **Learning Rates:** Two learning rates are specified: 10^{-4} and 10^{-3}. The learning rate determines the size of the steps taken during the optimization process. A smaller learning rate (10^{-4}) is typically used for fine-tuning or when the model is close to convergence, while a larger learning rate (10^{-3}) may be used for initial training stages or when the model needs to make larger adjustments.

3. **Batch Size:** The batch size is set to 512. This hyperparameter defines the number of samples processed by the model before updating the parameters. A larger batch size can lead to faster convergence but may require more memory and computational resources.

4. **Number of Epochs:** The training process will run for 100 epochs. An epoch refers to one complete pass through the entire training dataset. Training for multiple epochs allows the model to learn from the data multiple times, which can improve performance and convergence.

5. **Learning Rate Decay:** The learning rate decay is set to 0.1. Learning rate decay gradually reduces the learning rate over time, helping the model to converge more smoothly and prevent overshooting or oscillations in the optimization process.

4.7 Compared algorithms

4.7.1 Random

The Random algorithm is a simple approach where nodes are selected randomly without any specific criteria. In this method, the algorithm randomly picks a node from the available options, which could be customers, locations, or tasks. This strategy does not consider any factors such as node attributes, demand, or feasibility constraints. While being straightforward and easy to implement, the Random algorithm may not yield optimal results as it lacks a systematic way to prioritize nodes based on their importance or impact on the overall objective.

4.7.2 Largest-Demand

The Largest-Demand algorithm prioritizes nodes based on their demand or resource requirement. In the context of vehicle routing or resource allocation problems, this algorithm selects nodes with the highest demand, which could represent customers with large orders or tasks requiring significant resources. By focusing on nodes with larger demands, the Largest-Demand algorithm aims to optimize resource utilization and potentially improve overall system efficiency. However, it may overlook other factors such as distance, time constraints, or feasibility considerations.

4.7.3 Max-Reachable

The Max-Reachable algorithm combines considerations of node reachability and demand. It selects nodes that are not only reachable within a specified distance or time but also have a significant demand. This algorithm aims to balance between serving nodes efficiently (maximizing reachability) and prioritizing nodes with substantial resource requirements. By considering both reachability and demand, the Max-Reachable algorithm seeks to achieve a balanced and effective node selection strategy in scenarios like vehicle routing or task allocation. However, it may require more complex calculations or heuristics compared to simpler selection methods like Random or Largest-Demand.

Chapter 5

Experiments and Results

5.1 Experimental section

Our study's data generation procedure covered a number of important elements that were necessary to simulate an instance of the Stochastic Dynamic Vehicle Routing Problem (SDVRP). First, we used a uniform distribution to assign x and y coordinates to each node in order to generate the locations of our consumers. The depot functioned as the focal point for vehicle operations, with coordinates fixed at [0.5, 0.5]. Customers arrived at the system at random intervals as a result of a Poisson distribution with a rate of 1 being used to mimic client demand. Subsequently, a random demand was allocated to each customer, ranging from 1 to 9 units, to indicate their resource requirements. This allowed the capacity restrictions for each vehicle to be established.

We fixed the number of nodes in the system and set the time horizon equal to the number of nodes in order to preserve consistency between tests and simplify comparative analysis. We were able to assess several routing solutions across a range of problem sizes with effectiveness because to this standardized approach. In addition, we implemented a system that would cause clients to cancel their demand if they were left unattended for longer than five time units in order to replicate a realistic scenario. By encouraging rapid customer service and discouraging the accumulation of unsatisfied demand, this metric facilitated efficient routing strategies.

We used an embedding dimension of 128 in the model architecture to capture intricate dependencies and relationships in the data. Every epoch, we generated 1000 batches of data for our training procedure, each containing 512 issue cases. The adoption of this extensive training methodology made it easier to discover efficient vehicle routing policies in the face of uncertainty. We assessed our method on issue instances with 10, 20, and 50 nodes, and the evaluation parameter we used was the entire amount of the customer demand. The three other exact methods that were compared were "random," "largest-demand," and "max-reachable." It's interesting to note that our method outperformed the "max-reachable" algorithm, which takes customer abandonment information into account explicitly. This shows how well our suggested method works with stochastic customer demand

Table 5.1 - The performance evaluation of the RL approach vs benchmark algorithms, an average of 512 instances is reported per graph size.

Number of nodes	**10**			**20**			**50**		
Method	Reward	Gaps, %	Time, sec	Reward	Gaps, %	Time, sec	Reward	Gaps, %	Time, sec
RL	**29.51**	**1.56**	1.65	**62**	**0.99**	4.68	**145.18**	**0.15**	23.77
Random	17.62	40.86	0.72	32.52	48.07	2.34	78	46.29	9.57
Largest-Demand	27.51	8.25	0.71	56.39	9.95	1.31	133.38	8.26	4.12
Max-Reachable	28.1	6.28	1.47	58.54	6.47	4.3	136.83	5.88	25.4

scenarios in vehicle routing.

5.1.1 Results

Table 5.1 presents the performance evaluation results of the RL approach compared to benchmark algorithms across different graph sizes, with an average of 512 instances reported per graph size. The table includes metrics such as Reward Gaps (expressed as percentages) and Time (measured in seconds) for each method and graph size.

Firstly, the RL approach demonstrates competitive performance across all graph sizes. It achieves a Reward Gap of 29.51% for 10 nodes, indicating its ability to serve a substantial portion of customer demand efficiently. However, as the graph size increases to 20 and 50 nodes, the RL approach faces challenges, with Reward Gaps of 62% and 145.18%, respectively. This increase in Reward Gaps suggests that scaling to larger problem instances presents more complex routing scenarios that require further optimization.

Comparatively, the Random algorithm exhibits lower Reward Gaps for 10 and 50 nodes but struggles with the 20-node graph, indicating inconsistent performance across different problem sizes. The Largest-Demand algorithm performs relatively well, especially for 50 nodes, with a lower Reward Gap compared to the RL approach. However, it lags behind in terms of computational efficiency, as seen in the Time (sec) column.

The Max-Reachable algorithm shows promising results in terms of Reward Gaps, especially for 10 and 20 nodes, showcasing its effectiveness in selecting reachable customers with high demand. However, its computational time increases significantly for larger problem instances, highlighting a trade-off between solution quality and computational efficiency.

In summary, the RL approach demonstrates competitive performance in serving customer demand across various graph sizes but faces challenges with scalability to larger problem instances. The benchmark algorithms exhibit strengths and weaknesses in different aspects, such as solution quality, computational efficiency, and scalability, emphasizing the need for further research and optimization strategies to address complex routing scenarios effectively.

5.2 Review section

Additionally, there has been provided a comprehensive review paper an in-depth analysis of recent advancements in optimizing vehicle routing strategies amidst time uncertainty. Focusing primarily on the Vehicle Routing Problem (VRP), the study meticulously examines 54 research papers that delve into various facets of routing optimization when faced with uncertainties in time-related factors. The review encompasses a wide range of topics, including optimization algorithms, techniques for modeling uncertainty, simulation methodologies, and real-world case studies spanning autonomous electric vehicles, human-driven vehicles, and autonomous vehicle scenarios.

Specifically, the review delves into critical themes such as routing optimization strategies tailored to dynamic and uncertain environments, models for dynamic dispatching, considerations for ensuring reliability in routing decisions, and approaches for multi-objective optimization specifically tailored for electric trucks. By synthesizing insights from existing literature, this review aims to offer clarity on the current landscape of vehicle routing under temporal uncertainty while also shedding light on potential avenues for future research exploration. Through a comprehensive analysis of the existing body of work, this review paper serves as a roadmap for researchers and practitioners interested in advancing the field of vehicle routing optimization amidst time-related uncertainties.

Chapter 6

Conclusions and future work

6.1 Conclusions

In conclusion, our study delves into the challenging realm of the Stochastic Dynamic Vehicle Routing Problem (SDVRP) and proposes a Reinforcement Learning (RL) algorithm tailored to address its complexities. Through extensive experimentation and comparative analysis against baseline algorithms, including Random Selection, Largest-Demand Selection, and Max-Reachable Selection, we have demonstrated the efficacy of our RL approach in optimizing route efficiency and resource utilization. Furthermore, by integrating a Multihead Attention architecture into our RL algorithm, we have enhanced its ability to capture intricate relationships and adapt to dynamic and stochastic routing environments. Our findings highlight the potential of RL with advanced attention mechanisms in addressing real-world logistics challenges, offering valuable insights and avenues for future research in the field of dynamic vehicle routing and transportation optimization.

6.2 Future work

Future work in the realm of Stochastic Dynamic Vehicle Routing Problem (SDVRP) and Reinforcement Learning (RL) algorithms with Multihead Attention architectures presents exciting avenues for exploration and improvement. Firstly, further research could focus on enhancing the scalability and computational efficiency of RL algorithms in large-scale SDVRP instances, considering factors such as fleet size, network complexity, and real-time data integration. Additionally, the integration of deep learning techniques, such as deep RL or neural network-based models, could offer enhanced learning capabilities and decision-making accuracy in dynamic and stochastic routing environments. Furthermore, exploring adaptive learning rate strategies and meta-learning techniques could lead to more robust and adaptable RL solutions for SDVRP scenarios. Finally, extending the scope of investigation to include multi-agent systems and collaborative decision-making could provide insights into cooperative routing strategies and decentralized optimization approaches in complex logistics networks. Overall, future work in this domain aims to advance the state-of-the-art in dynamic vehicle routing optimization and contribute to more efficient and resilient transportation systems.

Bibliography

[1] Taukekhan Mustakhov, Yernar Akhmetbek, and Aigerim Bogyrbayeva. Deep reinforcement learning for stochastic dynamic vehicle routing problem. In *2023 17th International Conference on Electronics Computer and Computation (ICECCO)*, pages 1–5. IEEE, 2023.

[2] Gitae Kim, Yew Soon Ong, Taesu Cheong, and Puay Siew Tan. Solving the dynamic vehicle routing problem under traffic congestion. *IEEE Transactions on Intelligent Transportation Systems*, 17(8):2367–2380, 2016.

[3] M Fallah, R Tavakkoli-Moghaddam, A Salamatbakhsh-Varjovi, and M Alinaghian. A green competitive vehicle routing problem under uncertainty solved by an improved differential evolution algorithm. *International journal of engineering*, 32(7):976–981, 2019.

[4] Jose Caceres-Cruz, Pol Arias, Daniel Guimarans, Daniel Riera, and Angel A Juan. Rich vehicle routing problem: Survey. *ACM Computing Surveys (CSUR)*, 47(2):1–28, 2014.

[5] Brenner Humberto Ojeda Rios, Eduardo C Xavier, Flávio K Miyazawa, Pedro Amorim, Eduardo Curcio, and Maria João Santos. Recent dynamic vehicle routing problems: A survey. *Computers & Industrial Engineering*, 160:107604, 2021.

[6] Afsane Amiri, Hossein Zolfagharinia, and Saman Hassanzadeh Amin. A robust multi-objective routing problem for heavy-duty electric trucks with uncertain energy consumption. *Computers & Industrial Engineering*, 178: 109108, 2023.

[7] Florentin D Hildebrandt, Barrett W Thomas, and Marlin W Ulmer. Opportunities for reinforcement learning in stochastic dynamic vehicle routing. *Computers & operations research*, 150:106071, 2023.

[8] Masoud Rabbani, S Bosjin, Reza Yazdanparast, and N Saravi. A stochastic time-dependent green capacitated vehicle routing and scheduling problem with time window, resiliency and reliability: a case study. *Decision Science Letters*, 7(4):381–394, 2018.

[9] Evi Yuliza, Fitri Maya Puspita, and Siti Suzlin Supadi. Heuristic approach for robust counterpart open capacitated vehicle routing problem with time windows. *Science and Technology Indonesia*, 6(2):53–57, 2021.

[10] Majsa Ammouriova, Erika M Herrera, Mattia Neroni, Angel A Juan, and Javier Faulin. Solving vehicle routing problems under uncertainty and in dynamic scenarios: From simheuristics to agile optimization. *Applied Sciences*, 13(1):101, 2022.

[11] Yao Wu, Bin Zheng, and Xueliang Zhou. A disruption recovery model for time-dependent vehicle routing problem with time windows in delivering perishable goods. *IEEE Access*, 8:189614–189631, 2020.

[12] Ke Zhang, Xi Lin, and Meng Li. Graph attention reinforcement learning with flexible matching policies for multi-depot vehicle routing problems. *Physica A: Statistical Mechanics and its Applications*, 611:128451, 2023.

[13] Yinglei Li and Sung Hoon Chung. Disaster relief routing under uncertainty: A robust optimization approach. *Iise Transactions*, 51(8):869–886, 2019.

[14] Lei Wu and Mhand Hifi. Discrete scenario-based optimization for the robust vehicle routing problem: The case of time windows under delay uncertainty. *Computers & Industrial Engineering*, 145:106491, 2020.

[15] I Hammouti, K Derqaoui, and M Merouani. A modified clustering search based genetic algorithm for the proactive electric vehicle routing problem. *International Journal of Industrial Engineering Computations*, 14(4):609–622, 2023.

[16] Rajeev Goel, Raman Maini, and Sandhya Bansal. Vehicle routing problem with time windows having stochastic customers demands and stochastic service times: Modelling and solution. *Journal of Computational Science*, 34: 1–10, 2019.

[17] Bernhard H Korte, Jens Vygen, B Korte, and J Vygen. *Combinatorial optimization*, volume 1. Springer, 2011.

[18] Christos H Papadimitriou and Kenneth Steiglitz. *Combinatorial optimization: algorithms and complexity*. Courier Corporation, 1998.

[19] Xian Yu, Siqian Shen, Babak Badri-Koohi, and Haitham Seada. Time window optimization for attended home service delivery under multiple sources of uncertainties. *Computers & Operations Research*, 150:106045, 2023.

[20] Patrick Jaillet, Jin Qi, and Melvyn Sim. Routing optimization under uncertainty. *Operations research*, 64(1):186–200, 2016.

[21] Jennifer L Lahti, Grace W Tang, Emidio Capriotti, Tianyun Liu, and Russ B Altman. Bioinformatics and variability in drug response: a protein structural perspective. *Journal of The Royal Society Interface*, 9(72):1409–1437, 2012.

[22] Kris Braekers, Katrien Ramaekers, and Inneke Van Nieuwenhuyse. The vehicle routing problem: State of the art classification and review. *Computers & industrial engineering*, 99:300–313, 2016.

[23] Karla L Hoffman, Manfred Padberg, Giovanni Rinaldi, et al. Traveling sales-

man problem. *Encyclopedia of operations research and management science*, 1:1573–1578, 2013.

[24] Shen Lin. Computer solutions of the traveling salesman problem. *Bell System Technical Journal*, 44(10):2245–2269, 1965.

[25] Yossiri Adulyasak and Patrick Jaillet. Models and algorithms for stochastic and robust vehicle routing with deadlines. *Transportation Science*, 50(2): 608–626, 2016.

[26] Mahdieh Allahviranloo, Joseph YJ Chow, and Will W Recker. Selective vehicle routing problems under uncertainty without recourse. *Transportation Research Part E: Logistics and Transportation Review*, 62:68–88, 2014.

[27] Alexandra Anderluh, Rune Larsen, Vera C Hemmelmayr, and Pamela C Nolz. Impact of travel time uncertainties on the solution cost of a two-echelon vehicle routing problem with synchronization. *Flexible Services and Manufacturing Journal*, 32:806–828, 2020.

[28] Jonathan De La Vega, Pedro Munari, and Reinaldo Morabito. Exact approaches to the robust vehicle routing problem with time windows and multiple deliverymen. *Computers & Operations Research*, 124:105062, 2020.

[29] I-Ming Chao, Bruce L Golden, and Edward A Wasil. The team orienteering problem. *European journal of operational research*, 88(3):464–474, 1996.

[30] Shan-Huen Huang and Carola Alejandra Blazquez. A model for solving the dynamic vehicle dispatching problem with customer uncertainty and time dependent link travel time. *Revista Facultad de Ingeniería Universidad de Antioquia*, (64):163–174, 2012.

[31] Patrick-Oliver Groß, Jan Fabian Ehmke, and Dirk Christian Mattfeld. Interval travel times for robust synchronization in city logistics vehicle routing. *Transportation Research Part E: Logistics and Transportation Review*, 143: 102058, 2020.

[32] William H Beaver, Roger Clarke, and William F Wright. The association between unsystematic security returns and the magnitude of earnings forecast errors. *Journal of accounting research*, pages 316–340, 1979.

[33] Rafael D Tordecilla, Leandro do C Martins, Javier Panadero, Pedro J Copado, Elena Perez-Bernabeu, and Angel A Juan. Fuzzy simheuristics for optimizing transportation systems: Dealing with stochastic and fuzzy uncertainty. *Applied Sciences*, 11(17):7950, 2021.

[34] Rami Abousleiman, Osamah Rawashdeh, and Romi Boimer. Electric vehicles energy efficient routing using ant colony optimization. *SAE International Journal of Alternative Powertrains*, 6(1):1–14, 2017.

[35] Yufu Ning and Taoyong Su. A multilevel approach for modelling vehicle routing problem with uncertain travelling time. *Journal of Intelligent Manufacturing*, 28:683–688, 2017.

[36] Marius M Solomon. Algorithms for the vehicle routing and scheduling problems with time window constraints. *Operations research*, 35(2):254–265, 1987.

[37] Hokey Min. The multiple vehicle routing problem with simultaneous delivery and pick-up points. *Transportation Research Part A: General*, 23(5):377–386, 1989.

[38] Xinwei Chen, Marlin W Ulmer, and Barrett W Thomas. Deep q-learning for same-day delivery with vehicles and drones. *European Journal of Operational Research*, 298(3):939–952, 2022.

[39] Li Zhang, Zhongshan Liu, Lan Yu, Ke Fang, Baozhen Yao, and Bin Yu. Routing optimization of shared autonomous electric vehicles under uncertain travel time and uncertain service time. *Transportation Research Part E: Logistics and Transportation Review*, 157:102548, 2022.

[40] Junlong Zhang, William HK Lam, and Bi Yu Chen. A stochastic vehicle routing problem with travel time uncertainty: trade-off between cost and customer service. *Networks and Spatial Economics*, 13:471–496, 2013.

[41] Shuai Zhang, Tong Zhou, Cheng Fang, and Sihan Yang. A novel collaborative electric vehicle routing problem with multiple prioritized time windows and time-dependent hybrid recharging. *Expert Systems with Applications*, 244: 122990, 2024.

[42] Siyuan Liu and Qiang Qu. Dynamic collective routing using crowdsourcing data. *Transportation Research Part B: Methodological*, 93:450–469, 2016.

[43] Shurui Zhu, Huijun Sun, and Xin Guo. Cooperative scheduling optimization for ground-handling vehicles by considering flights' uncertainty. *Computers & Industrial Engineering*, 169:108092, 2022.

[44] Agostinho Agra, Marielle Christiansen, Rosa Figueiredo, Lars Magnus Hvattum, Michael Poss, and Cristina Requejo. The robust vehicle routing problem with time windows. *Computers & operations research*, 40(3):856–866, 2013.

[45] Naoki Ando and Eiichi Taniguchi. Travel time reliability in vehicle routing and scheduling with time windows. *Networks and spatial economics*, 6:293–311, 2006.

[46] Simen Braaten, Ola Gjønnes, Lars Magnus Hvattum, and Gregorio Tirado. Heuristics for the robust vehicle routing problem with time windows. *Expert Systems with Applications*, 77:136–147, 2017.

[47] Jing-jing CHANG, Zhong-ren PENG, and Jian SUN. Freight vehicle routing optimization for sporadic orders using floating car data. *Journal of Donghua University (Eng. Ed.) Vol*, 30(2), 2013.

[48] Xiao-Wei Chen, Bi Yu Chen, William HK Lam, Mei Lam Tam, and Wei Ma. A bi-objective reliable path-finding algorithm for battery electric vehicle routing. *Expert Systems with Applications*, 182:115228, 2021.

[49] James C Chu, Shangyao Yan, and Han-Jheng Huang. A multi-trip split-delivery vehicle routing problem with time windows for inventory replenishment under stochastic travel times. *Networks and Spatial Economics*, 17: 41–68, 2017.

[50] Nasser A El-Sherbeny. Imprecision and flexible constraints in fuzzy vehicle routing problem. *American Journal of Mathematical and Management Sciences*, 31(1-2):55–71, 2011.

[51] Yanling Feng, Ren-Qian Zhang, Guozhu Jia, et al. Vehicle routing problems with fuel consumption and stochastic travel speeds. *Mathematical problems in engineering*, 2017, 2017.

[52] Bocewicz Grzegorz, Nielsen Peter, Smutnicki Czeslaw, Pempera Jaroslaw, and Banaszak Zbigniew. Periodic distributed delivery routes planning subject to operation uncertainty of vehicles travelling in a convoy. *Journal of Information and Telecommunication*, 6(3):360–380, 2022.

[53] Ali Haghani and Soojung Jung. A dynamic vehicle routing problem with time-dependent travel times. *Computers & operations research*, 32(11):2959–2986, 2005.

[54] C Hu, J Lu, X Liu, and G Zhang. Robust vehicle routing problem with hard time windows under demand and travel time uncertainty. *Computers & Operations Research*, 94:139–153, 2018.

[55] Mingyong Lai, Hongming Yang, Songping Yang, Junhua Zhao, and Yan Xu. Cyber-physical logistics system-based vehicle routing optimization. *Journal of Industrial & Management Optimization*, 10(3), 2014.

[56] Chungmok Lee, Kyungsik Lee, and Sungsoo Park. Robust vehicle routing problem with deadlines and travel time/demand uncertainty. *Journal of the Operational Research Society*, 63(9):1294–1306, 2012.

[57] Xiangyong Li, Peng Tian, and Stephen CH Leung. Vehicle routing problems with time windows and stochastic travel and service times: Models and algorithm. *International Journal of Production Economics*, 125(1):137–145, 2010.

[58] Yue Lu, Maoxiang Lang, Yan Sun, and Shiqi Li. A fuzzy intercontinental road-rail multimodal routing model with time and train capacity uncertainty and fuzzy programming approaches. *IEEE Access*, 8:27532–27548, 2020.

[59] Jacek Mańdziuk and Maciej Świechowski. Uct in capacitated vehicle routing problem with traffic jams. *Information Sciences*, 406:42–56, 2017.

[60] Mehdi Nasri, Abdelmoutalib Metrane, Imad Hafidi, and Anouar Jamali. A robust approach for solving a vehicle routing problem with time windows with uncertain service and travel times. *International Journal of Industrial Engineering Computations*, 11(1):1–16, 2020.

[61] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.

[62] Wouter Kool, Herke Van Hoof, and Max Welling. Attention, learn to solve routing problems! *arXiv preprint arXiv:1803.08475*, 2018.