



# **Performance Analysis on Hybrid and Exact Methods for Solving Clustered VRP**

**A Comparative Study on VRP Algorithms**

**Tejaswi Nunna**

This thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Computer Science. The thesis is equivalent to 20 weeks of full time studies.

**Contact Information:**

Author(s):

Tejaswi Nunna

E-mail: [tenu15@student.bth.se](mailto:tenu15@student.bth.se)

University advisor:

Lawrence Henesey

DIDD

Faculty of Computing  
Blekinge Institute of Technology  
SE-371 79 Karlskrona, Sweden

Internet : [www.bth.se](http://www.bth.se)  
Phone : +46 455 38 50 00  
Fax : +46 455 38 50 57

# ABSTRACT

**Context:** The Vehicle Routing Problem is an NP-hard problem with a combination of varieties of topics like logistics, optimization research and data mining. There is a vast need of vehicle routing solutions in day to day like with different constraints. According to the requirements, this problem has been a field of interest to a lot of researchers who incorporate scientific methods to combine and innovate new solutions to optimize the routing. Being an np-hard problem, it is almost impossible to compute the solutions to optimality but years of research on this area has paid off quite significantly and the solutions are optimized little by little and better than before. Some applications may or may not find slight difference in the performance as a considerable affect but some applications or scenarios heavily depend on the performance of the solution where it is very vital that the solution is optimized to the fullest. As a data mining technique clustering has been used very prominently in case of portioning scenarios and similarly it has also began to surface in implementing VRP solutions. Although it has recently emerged into the Vehicle Routing era and shown some significant results, it has not yet come into an open state or awareness. The awareness regarding clustering matters in a huge extent to be considered by most of the recent researchers who formulate new algorithms to solve VRP and help them further optimize their solution.

**Objectives:** In this study the significance of clustering has been considered to find out how the usage of clustering techniques can alter the performance of VRP based solutions favorably. Then to test the results of two recently proposed cluster based algorithms, a comparison has been made to other types of algorithms which prove how the algorithms stand with various methods.

**Methods:** A literature review is performed using various articles that have been gathered from Google Scholar and then an empirical experiment was conducted on the results available in the papers. This experiment was done by performing a comparative analysis.

**Results:** For the literature review the results were gathered from all the articles based on their research, experience, use of clustering and how their result was improved by using clustering methods in their formulations. Considering the experiment, the results of both the algorithm were compared with the results of five other papers who aim to solve the VRP using exactly the same instances that were used in the two algorithms in order to compare valid results on the same variables. Then the results were analyzed for the purpose of comparison and conclusions were drawn accordingly.

**Conclusions:** From the research performed in this paper we can conclude the vast significance of clustering techniques that were drawn based on practical test results of various authors. From the experiment performed it is clear that the Hybrid algorithm has a much higher performance than any other algorithm it has been compared to. This algorithm has also been proven to enhance its performance due to the implementation of clustering techniques in their formulation. Since the results were only based on performance that is, in this case the total distance of the final route, future study indicates the implementation of algorithms to compare them on basis of time complexity and space complexity as well.

**Keywords:** Vehicle routing problem, logistics, clustering, genetic algorithm, exact algorithm.

## **ACKNOWLEDGEMENT**

With sincere gratitude I would like to thank my supervisor Mr. Lawrence Henesey for supporting my work on this research. Furthermore, I would also like to thank the authors of the literature that has aided me on a great extent that guided me to write this thesis. Finally, I would like to thank my friends and family who encouraged me throughout my research.

## **LIST OF ABBREVIATIONS**

VRP	Vehicle Routing Problem
TSP	Travelling Salesman Problem
VRPD	Vehicle Routing with Pickup and Delivery
CVRP	Capacitated Vehicle Routing Problem
GVRP	Green Vehicle Routing Problem
OR	Operations Research
GA	Genetic Algorithm
BC	Branch and Cut
VRPTW	Vehicle Routing Problem with Time Windows
PSO	Particle swarm Optimization
UCFL	Unsupervised Fuzzy Competitive Learning
CWS	Clarke and Wright Saving
ACO	Ant Colony Optimization
GWO	Grey Wolf Optimization

## LIST OF FIGURES

<b>Figure 1:</b> Basic structure of Vehicle Routing .....	4
<b>Figure 2:</b> Vehicle routing problem (waste collection) .....	5
<b>Figure 3:</b> A drone ready to take off for delivery .....	6
<b>Figure 4:</b> A VRP route before and after drone implementation.....	6
<b>Figure 5:</b> An example of benchmark instances of Augerat et al. set A.....	8
<b>Figure 6:</b> Implementation of Algorithm A.....	16
<b>Figure 7:</b> Sample data set for <i>Algorithm A</i> .....	17
<b>Figure 8:</b> Results comparison with Mark et al. on instance type A .....	24
<b>Figure 9:</b> Result table with Ganesh et al. Algorithm with instance type A.....	24
<b>Figure 10:</b> Result compare table with Feng et al. Algorithm instance type A .....	25
<b>Figure 11:</b> Result compare table with Feng et al. Algorithm instance type B .....	25
<b>Figure 12:</b> Result compare table with Bektás et al. Algorithm instance type A .....	26
<b>Figure 13:</b> Result compare table with Bektás et al. instance type B.....	26
<b>Figure 14:</b> Result compare table with Bektás et al. instance type P .....	26
<b>Figure 15:</b> Result compare table with Korayem et al. instance type A.....	27
<b>Figure 16:</b> Result compare table with Korayem et al. instance type B.....	27
<b>Figure 17:</b> Result compare table with Korayem et al. instance type P .....	27

## LIST OF TABLES

<b>Table 1:</b> Proposed Solution methods to solve VRP .....	6
<b>Table 2:</b> Comparison of literature analysis with other research methods for RQ1 .....	11
<b>Table 3:</b> Results of search string in Google Scholar .....	12
<b>Table 4:</b> Result of literature analysis from the selected literature .....	15
<b>Table 5:</b> Comparison of results with algorithm by Mark et al. ....	21
<b>Table 6:</b> Comparison of results with algorithm by Ganesh et al .....	21
<b>Table 7:</b> Comparison of results with algorithm by Feng et al .....	22
<b>Table 8:</b> Comparison of results with algorithm by Bektás et al. ....	22
<b>Table 9:</b> Comparison of results with algorithm by Korayem .....	22
<b>Table 10:</b> Analysis of the comparison of Algorithm A and B with Mark et al. ....	28

# CONTENTS

<b>ABSTRACT .....</b>	<b>I</b>
<b>ACKNOWLEDGEMENT .....</b>	<b>I</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>II</b>
<b>LIST OF FIGURES .....</b>	<b>III</b>
<b>LIST OF TABLES .....</b>	<b>IV</b>
<b>CONTENTS .....</b>	<b>V</b>
<b>1 INTRODUCTION .....</b>	<b>1</b>
1.1 DOCUMENT STRUCTURE .....	2
<b>2 BACKGROUND AND RELATED WORK .....</b>	<b>3</b>
2.1 OPERATIONS RESEARCH .....	3
2.1.1 OPTIMIZATION .....	3
2.2 VEHICLE ROUTING PROBLEM .....	3
2.2.1 APPLICATIONS .....	4
2.2.2 METHODS .....	6
2.2.3 BENCHMARK INSTANCES .....	7
<b>3 PROBLEM STATEMENT AND CONTRIBUTION .....</b>	<b>9</b>
3.1 AIM .....	9
3.2 OBJECTIVE .....	9
3.3 RESEARCH GAP AND CONTRIBUTION .....	9
3.4 RESEARCH QUESTIONS .....	10
<b>4 METHODOLOGY .....</b>	<b>11</b>
4.1 LITERATURE ANALYSIS .....	11
4.1.1 RISKS .....	12
4.2 EXPERIMENTATION .....	12
4.2.1 EXPERIMENT TYPE .....	13
4.2.2 VARIABLES CONSIDERED IN THE EXPERIMENT .....	13
4.2.3 EXPERIMENT PROCEDURE .....	13
4.2.4 RISKS .....	13
<b>5 RESULTS .....</b>	<b>14</b>
5.1 RESULT OF LITERATURE ANALYSIS .....	14
5.2 RESULT OF EXPERIMENT .....	16
5.2.1 IMPLEMENTATION OF Algorithm A .....	16
5.2.2 IMPLEMENTATION OF Algorithm B .....	18
5.2.3 Combining all the obtained results .....	21
<b>6 ANALYSIS AND DISCUSSION .....</b>	<b>23</b>
6.1 ANALYSIS BASED ON LITERATURE ANALYSIS .....	23
6.2 ANALYSIS BASED ON COMPARISON OF ALGORITHMS .....	23
6.3 STATISTICAL ANALYSIS .....	28
<b>7 CONCLUSION AND FUTURE WORK .....</b>	<b>30</b>
<b>REFERENCES .....</b>	<b>31</b>



# 1 INTRODUCTION

In a world dedicated to machines and artificial intelligence logistics play an important role that deal with minimizing the usage of valuable and scarce resources. As such, a primary goal of logistics is optimization[1]. Optimization is performed in cases of solving complex computing problems, routing, games, decision support, healthcare and many other instances related to problem solving and solutions are further formed to these problems as long as a perfectly optimized solution is obtained[2]. There are numerous ways of optimizing a problem which are classified based on the type of the problem such as tools, strategies, standard methods, tests etc. Coming to this paper, research is done regarding Vehicle Routing Problem, an np-hard problem, meaning that it is almost impossible to obtain a perfectly optimized solution. Many algorithms till date are being proposed to optimize the solution which is why it is known as an np-hard problem. This paper considers the significance of clustering in solving Vehicle Routing and how they can be helpful being implemented real time applications by comparing two recently proposed algorithms that aim to solve the vehicle routing problem which have been proven to be highly effective and efficient.

There has been extensive research on routing and the numerous classifications and constraints that fall under routing. Optimization plays a major role for the cause of research in the area of routing as a large number of day to day factors depend on routing like, traffic, fuel consumption, transportation etc[3]. Similarly, the research being conducted regarding this paper aims to uplift the significance of the area dealt with and the methods used, promoting and encouraging further research by aiding researchers to get an insight on the new methods that are proven to be more efficient.

The vehicle routing problem consists of a number of traditional and classical methods and approaches. Researchers extend the solutions of VRP by using these standard approaches and combining them with logic that they come up with or a different strategy intending to enhance the performance of the solution. As years passed by, research regarding this problem began to get more and more complicated due to the addition of multiple factors or constraints to the problem like capacity, time limit, non-uniform vehicles etc [4]. These factors added more significance to the problem and researchers worldwide found interest in this field as it always consisted research gaps regardless of the time and research conducted.

The aim of this thesis is to purely focus on clustering regarding vehicle routing problem where first a literature analysis will be performed to study various ways clustering techniques have helped improving the performance of vehicle routing problem solutions. The next step is to perform an experiment by implementing two algorithms that make use of clustering techniques. The results are then compared using their performances with five older state-of-the-art algorithms. This experimentation in a way is a practical testing of the theory to find the significance of clustering in solving the vehicle routing problem. The implementation has been done using 'MATLAB R2016b'. For the comparison of the results and conducting an analysis 'ANOVA' has been used. In ANOVA the two-factor without replication technique was used due to the presence of more number of factors in the data required to be analyzed. The analysis is made in ANOVA mainly by comparing the variances of the performances. This analysis has been conducted in 'Microsoft Excel' and the variances were provided in row wise and column wise which will be explained in detail in Section 6.3.

## **1.1 DOCUMENT STRUCTURE**

This document presents the research in various sections. The sections followed are the background and related work, problem statement and contribution, methodology, results, analysis and discussion, conclusion and future work and references. The background work describes all the basic necessary knowledge about the context and the thesis. In the problem statement and contribution section the main intention and purpose behind conducting the research is given including the contribution made by conducting the research. The methodology section contains description about the research methods used to accomplish the objectives and their implementation approach in the thesis. The methods are implemented according to the objectives and results are shown in the result section. In analysis and discussion the implemented results are analyzed using tools or other analysis methods to make it easier to extract data from the results. A summary of the complete research is written with potential future possible work which could lead to further development and aid to researchers, in conclusion and future work followed by the references used in the thesis.

## **2 BACKGROUND AND RELATED WORK**

This section will display the basic terminology and knowledge that covers the areas needed in this research. It will give a brief description about Operational Research, VRP, its methods and applications.

### **2.1 OPERATIONS RESEARCH**

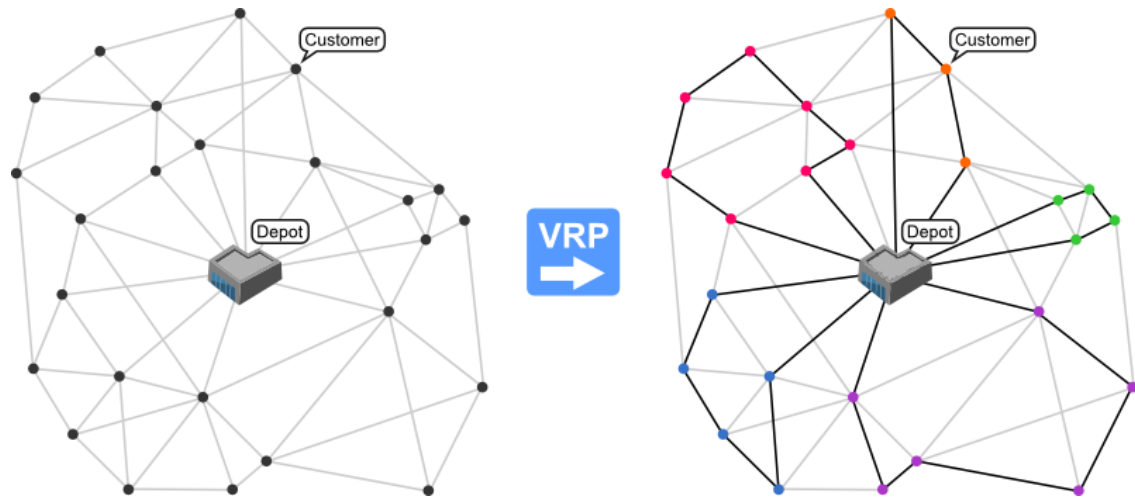
Operational research in the U.K, is a discipline that deals with the application of advanced analytical methods to help make better decisions. This study adapts methods from mathematical modelling, optimization, analysis techniques which finally aid in decision support systems[5]. Operations research is becoming more and more significant and[6]–[9] recognized due to the rise in decision support applications in real life scenarios such as container terminal operations, vehicle routing, biomass supply chain management etc. The need for operations research is growing as the real life problems are growing in sizes with more number of complex constraints and problem size and different logistic concepts, decision rules and optimization algorithms have to be compared by simulation before they are implemented into real systems [6], [7]. Vehicle routing problem is a similar complex scenario where the concept of operations research is adapted.

#### **2.1.1 OPTIMIZATION**

Optimization has been a source of research from a long time where its applicability is tested on endless number of applications and situations. Optimization aims in reducing the cost of performing a task in terms of time, resources used, space, labor etc[1], [2]. Optimization is performed while solving complex computing problems, routing, games, decision support, healthcare and many other instances related to problem solving. A solution to a problem is always further optimized until a perfectly desired solution is obtained. Regarding the research conducted in this thesis, optimization plays a very important role in affecting the performance of an algorithm. Many algorithms are being proposed so that every time a better optimized solution can be developed. When the results are significantly different and better from the other or older algorithm then they are optimized to some extent. Even the slightest gain in performance can show difference when it comes to a real life scenario for example in cases of disaster relief. This research aids optimization techniques to improve the performance of the algorithms.

### **2.2 VEHICLE ROUTING PROBLEM**

The vehicle routing problem is a classic routing problem which has come to recognition since 1959 [10]. The VRP is a logistics problem where a depot consists of more than one vehicle and scattered delivery points or customers. The problem lies behind the routing of each vehicle to cover all the delivery points such that time and cost (distance) is set to the minimal extent[11]. This problem is a generalization of the famous Travelling salesman problem (TSP). Obtaining an optimized solution to this problem contributes a great deal to its research due to its immense significance in real life and the nature of the problem always encourages further need to optimize the solution for feasible outcomes to serve logistics efficiently. A basic vehicle routing problem can be explained from Figure1.



**Figure 1:** Basic structure of Vehicle Routing

In Figure1, the center is the depot from where the vehicles dispatch. Considering the scenario in this image (the right side), there are 5 vehicles and 25 delivery points. Each color depicts a vehicle and the nodes are the delivery points. The solution (route) should be such that the vehicles should cover all the delivery points by travelling the least possible distance. From the figure it shows the routing before and after getting a solution, where in the left side all the nodes are connected to the possible neighboring nodes and the solution part shows that most of the connections are eliminated as a result of decreasing the total distance covered. It also concerns with the combination of delivery points a vehicle is appointed. No delivery point should be either left unattended or visited twice. The solution to this problem is formulated through mathematical form or (or mostly followed by) computational methods.

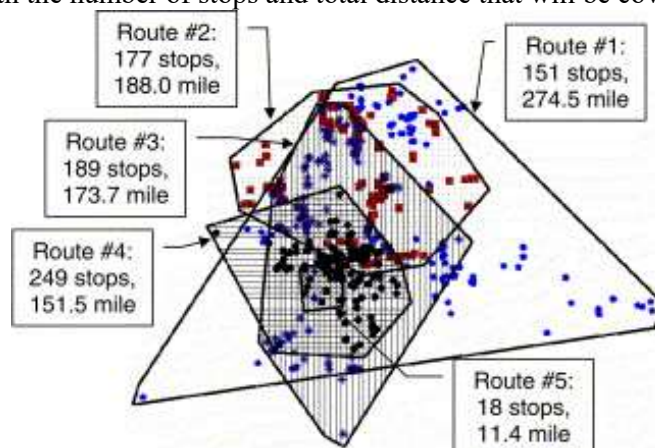
This problem has been taken to a higher level as mentioned above by considering various constraints like capacity, demand, time windows etc. For example, ‘*Capacity*’ constraints are considered when the freight carrying vehicles have equal or in some cases unequal capacities so the maximum number of houses or delivery points visited by the vehicle might differ considering this factor [4]. ‘*Demand*’, as the name says, it is considered when the customers need it earlier in which the delivery points are ranked according to demand and are attended to depending on the rank. ‘*Time windows*’ (VRPTW) is considered when the customers require the delivery during a specific period of time. For example a customer would prefer delivery on weekends or evenings as the person will not be available as he/she would go for work.

After certain amount of research the vehicle routing problem solutions which are optimized by many researchers can lead to decrement of resources directly or indirectly such as fuel consumption, time, and energy and can also reduce the number of vehicles assigned for the delivery.

## 2.2.1 APPLICATIONS

This section explains about the most widely used applications of the VRP solutions. Applications associated with vehicle routing are more than just delivering goods to customers on road. Olli Bäysy and Geir Hasle provided a vast research regarding the applications of Vehicle Routing and its uses and significance in various sectors and industries throughout the world. They concluded some day to day activities in which the application of Vehicle Routing solutions can help in making the job a little less complex[4].

**Waste collection:** Vehicle routing can be extremely useful in cases like these where it can be applied on an everyday basis. Waste collection is simply VRP with a lot of delivery points or nodes where garbage needs to be collected. Solutions for situations like these can come in a lot of handy giving them an organized route without leaving behind a node or by avoiding commotion[12], [13]. Figure 2 given below shows the routes of 5 vehicles that are to collect waste in a city with the number of stops and total distance that will be covered.



**Figure 2:** Vehicle routing problem (waste collection)

Real life applications usually incorporate many aspects and constraints which are not always sufficient in basic VRP which lead to addition of constraints to VRP like capacity, time windows etc. B.L. Golden et al. have also researched on some applications regarding VRP like disaster relief[4].

**Primary health care services:** In case of an unexpected disaster be it a natural disaster or a terrorist attack, it may lead to cause a lot of destruction and loss of life and property. In this case sometimes the health care preparations may not be sufficient. In these cases emergency logistics support can be provided by scheduled vehicle dispatch to help move people to safety[14]. More constraints can be added to this problem as availability of health care resources within the vehicles etc.

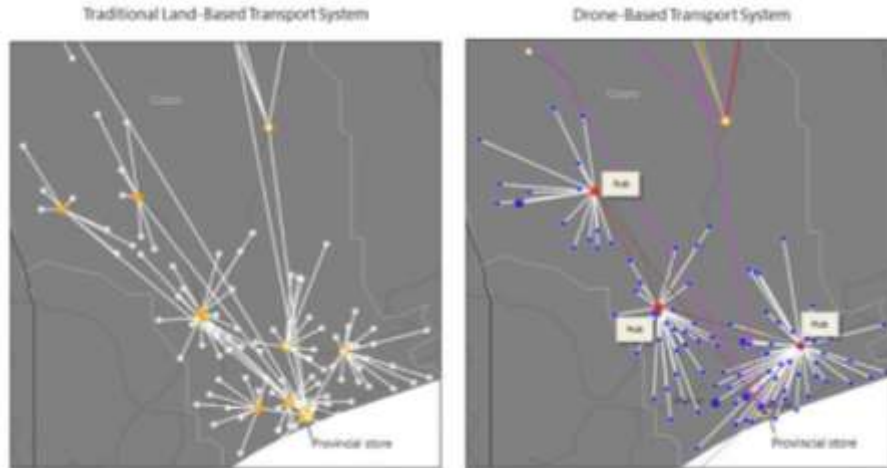
VRP can also be associated with other similar situations like school bus pickup and drop, considering other means of travel like, through sea (ships and ferries), aerial vehicles (airplanes and drones). The applications can vary depending on the situation and the constraints for the solutions vary depending on the problem statement as two problem statements may not require the same constraints.

Another recently emerging major application of VRP solutions are used by drones. These mobile bots are being fully utilized in ways that help in reducing time, fuel and also traffic.

**Drones:** Drones are entering a realm of popularity with a number of uses dealing with commercial and non-commercial uses, at the same time they are being used to act upon VRP solutions known as VRPD[15]. Drones are being used by few companies like Amazon, DHL, Federal express etc. where they aim to deliver goods through drones in a direct or indirect way. Xingyin Wang et al. suggested an algorithm using VRPD where delivery vehicles with a set of drones carrying goods are driven to certain points where the drones take over from there by delivering the goods to the nearby houses and returning to the vehicle. The vehicle then drives back to the depot once all the goods assigned have been delivered.



**Figure 3:** A drone ready to take off for delivery



**Figure 4:** A VRP route before and after drone implementation

Figure 3 shown above describes the structure of a drone built for delivery purposes which can withstand certain amount of weight it can carry. Below that, Figure 4 shows the comparative study of vehicle routing in an area between using a drone on the right side and not using a drone on the left side. It can be clearly seen from the figure that the number of clusters or substations are reduced due to the usage of drones. Apart from this research another research was conducted by Andrew Lotz et al. considering the possibilities that usage of drones in logistics might not be as feasible as they look. They consider factors such as limitations faced by drones and weather conditions or customer satisfaction[16]. Either way the study of drones under influence of logistics is another major source of research with an escalating trend attracting customers and industries around the globe for multiple applications.

## 2.2.2 METHODS

There are many methods to solve the VRP but the two most widely used methods are the exact and heuristic methods. Table 1 shows the most extensively used exact and heuristic procedures in solving the CVRP [4] which are solution methods. These methods are used with an additional personal logic that gives different results. These are the two main classifications of solutions methods. The vehicle routing solutions extensively use the study of Operations Research (OR) which helps them implement their solutions.

Type of Approach	Solution Procedure
Exact solutions	Branch and cut
	Branch and bound
	Pricing
Heuristics	Simple heuristics
	Meta heuristics
	Hybridization

**Table 1:** Proposed Solution methods to solve VRP

2.2.2.1 **Hybrid solutions:** The method combines genetic operators, selection, and crossover with an efficient local search which replaces the classical randomized mutation operator [4]. The difference between all the hybrid solutions is the selection of the type of local search procedure and known as GA (genetic algorithm) in short.

2.2.2.2 **Exact solutions:** Exact solutions were the very first solutions to VRP and they are known to solve the problem very close to an optimal extent[17]. A drawback regarding exact solutions is that they don't run efficiently on large scale problems[18]. But that was considering the classical exact solutions. The new exact solutions with proper cuts and pricings efficiently solve large scale problems.

## 2.2.3 BENCHMARK INSTANCES:

The benchmark instances are one of the major sources to aid this research with. Instances are nothing but some example problems that were generated and formulated for the researchers to test their algorithms on. For the vehicle routing problem, these instances act like an input during the implementation of the algorithm that provides a basic problem to run on. These instances contain the following basic information as input:

1. A graphical coordinates representing
  - The position of the depot
  - The position of the delivery points
2. The size of the problem (number of delivery points)
3. Number of Vehicles

Benchmark instances are the instances created by famous authors that made these problem samples to test their CVRP algorithms on them. These instances have been called as benchmark instances as many authors and researchers when come up with algorithms of their own to be able to compute the vehicle routing problem using these instances. There are a few benchmark instances that are widely used by most of the authors who proposed and implemented algorithms to solve the VRP and they can be available on various CVRP instances library such as they are given in [19].

The main benefit that comes from the benchmark instances are that since most of the authors and researchers use these instances to implement their algorithms, it becomes easier for them to compare and measure the performance of their algorithm to other recently proposed algorithms that used the same instances and also the known best solutions that have been obtained using those instances. A sample of an instance from the CVRP library is shown below:

Benchmark	Instance	$n$	$K$	$Q$	UB	Opt	Features
▼ Set A (Augerat, 1995)							 
	A-n32-k5	31	5	100	784	yes	   
	A-n33-k5	32	5	100	661	yes	   
	A-n33-k6	32	6	100	742	yes	   
	A-n34-k5	33	5	100	778	yes	   
	A-n36-k5	35	5	100	799	yes	   
	A-n37-k5	36	5	100	669	yes	   
	A-n37-k6	36	6	100	949	yes	   
	A-n38-k5	37	5	100	730	yes	   
	A-n39-k5	38	5	100	822	yes	   
	A-n39-k6	38	6	100	831	yes	   
	A-n44-k6	43	6	100	937	yes	   
	A-n45-k6	44	6	100	944	yes	   
	A-n45-k7	44	7	100	1146	yes	   
	A-n46-k7	45	7	100	914	yes	   
	A-n48-k7	47	7	100	1073	yes	   
	A-n53-k7	52	7	100	1010	yes	   

**Figure 5:** An example of benchmark instances of Augerat et al. set A



### 3 PROBLEM STATEMENT AND CONTRIBUTION

This section explains the cause behind the research and what contribution the research makes to science.

#### 3.1 AIM

The aim of this thesis is to determine the significance of clustering in VRP and compare recently proposed hybrid and exact algorithms solving the VRP that uses clustering to identify the best performing algorithm.

#### 3.2 OBJECTIVE

The objectives of the thesis are listed below:

- Conduct a literature analysis to determine the significance of clustering in implementing vehicle routing problems.
- Implement the two proposed methods under the suited environments and parameters.
- Conduct a performance analysis between the exact and hybrid algorithms and compare the results.

#### 3.3 RESEARCH GAP AND CONTRIBUTION

Two most recently proposed algorithms are chosen to solve the clustered VRP. The algorithm found for the hybrid method has been proposed by Pop et al. [20] This algorithm was implemented and tested and the results turned out favorable in terms of solution quality in comparison to existing approaches. Let us consider this algorithm as '*Algorithm A*' The algorithm found for the exact method has been proposed by Contardo et al. [21]. They have proposed an exact formulation based on specific set partitioning and pricing methods and their formulation has been tested on benchmarks instances that have proven to outperform other formulations in terms of performance. This formulation can be considered as '*Algorithm B*' for the purpose of using them in the research. This shows the productivity and significance of the both the algorithms and implies the opportunity to implement the algorithms on specific parameters such that an evaluation can be performed on both the algorithms.

Many algorithms are being proposed by researchers all over the world out of which most of them are sometimes left unattended or unevaluated. A comparison study should be more active among these algorithms that would benefit industries if the algorithms are found to be well performing and even the authors to put such effort on the research. In a comparison study, a strategy with lower performance is never implied to being ineffective or inefficient as their efficiency may lie elsewhere in other situations like size of the problem or the intention behind the problem etc. This research aims to consider the performances of two algorithms which are stated to perform well compared to the results of the previously proposed algorithms. This comparison is between two completely different methods in solving the VRP and will be tested over large benchmark instances to know which type of method will suit for solving large size problems.

Clustered VRP is chosen as the base of this research as it has been proven by Barreto, Ferreira, Paixão and Santos that clustering has resulted in faster problem solving and enabling alternative solutions in various scenarios [22]. They also suggested clustering to be a good source of investigation assuring promising results. Clustering has been proved in more ways than one that it can lead to minimizing the complexity of the problem and helps to yield a better result regarding the VRP [23]. This implies to assure an improvement to the optimization methods solving the VRP that could be practically implemented to deal with real life logistics. This will help in efficient delivery services by reduced travel distance and cost.

### 3.4 RESEARCH QUESTIONS

**RQ1.** What is the significance of clustering techniques when it comes to solve vehicle routing problem and how can it help the solution?

**Motivation:** When considering implementation of vehicle routing algorithms. A major step to initiate the solution is assigning nodes to a vehicle. The strategy applied in assigning the nodes to the vehicles makes a lot of difference in the result of the solution [18]. There has been limited research on the usage of clustering but the effect shown by the few researches conclude that clustering leads to an impressive performance [22]. These cases promote the need to share the significance of clustering that would lead to better awareness of this technique and encourage further researchers to implement VRP solutions using clustering.

**RQ2.** How can the performance and efficiency be evaluated between the hybrid and exact algorithms for solving the clustered VRP?

**Motivation:** Many algorithms are being proposed to solve the VRP from many years. The selected algorithms are the most recent which focus on solving the VRP using clustering. Both algorithms have been preliminarily proven to yield satisfactory results compared to older algorithms that show the productivity of these algorithms. Similarly clustering has also been proven to show better results in terms of performance and reducing problem complexity. A comparison of these algorithms will not only prove one algorithm as the better than the other but also open causes for researchers on the procedures used in the methods and encourage them to be able to further optimize the solutions. The evaluation could also catch the eye of various companies in the proposed algorithms by intending to implement the algorithms in real life, more specifically the algorithm that proves to be better in performance.

## 4 METHODOLOGY

This section gives a description about the research methods that are used to accomplish the objectives of the thesis and the motivation behind selecting these methods. There are two methodologies followed to acquire the aim of the thesis and they are described as follows.

### 4.1 LITERATURE ANALYSIS

This method has been selected in order to answer the first research question which is “What is the significance of clustering techniques when it comes to solve vehicle routing problem and how can it help the solution?” When considering a research question like this, a literature analysis can give the answer as needed. A literature review can also be used to justify the proposed study in order to find something new to which it contributes[24]. It also helps the reader to understand and differentiate between what has been already known and what new has been found out[25]. Comparing literature analysis with other research methods in this scenario the following table explains why other methods have been excluded to answer the research question and also why literature analysis will help in answering the research question.

Research Method	Cause of exclusion
<b>Survey</b>	A survey can be conducted when the participants of the survey are known or generalized. Considering a specific region or age group or a selected society, the survey can be very irrelevant to most of the participants. This problem specifies to selective researchers all around the world who have knowledge in the specific field which is not possible to list out or group.
<b>Case study</b>	A case study can be irrelevant as case studies deal with the product or result of a study and a result alone cannot conclude the significance of an approach used in a study. Thus, this research method has been excluded.
<b>Interview/ Questionnaire</b>	Similar to survey neither interview nor a questionnaire can answer this question as the participants cannot be easily listed or grouped. To answer this question, a study based on authors who have performed research in this field is required which can be found through research articles and such, which cannot be done using interviews or questionnaires. The results can also give biased answers or inaccurate results even though performed on a small range of selected participants.
<b>Literature analysis</b>	A literature analysis will help studying the experience, results and significance just by analyzing research papers and articles belonging to the required field written by researchers which can be easily listed out using databases. This proves the accuracy that is guaranteed in providing the right results and answers in an easy way which also saves time by using databases.

**Table 2:** Comparison of literature analysis with other research methods for RQ1

Literature analysis can deal with a wide range of databases to select the papers from and also a large number of papers in each database to choose the right papers. For this research the database ‘Google Scholar’ has been selected. Prior to the selection of this database other databases like ‘INSPEC’, ‘Science Direct’ and ‘Scopus’ were also considered. But for this field of research, higher accuracy rate and higher range of papers was found in

Google Scholar. A very less number of results was shown with relevant abstract, availability of full text and recent research on this topic in the other databases.

Considering the selection criteria of literature, I have used some key words to get various sets of results using different combinations and then refining the search result to a minimal extent when the papers found were relevant regarding the research. The search string and basic refinement is shown in the table below:

Search String	Number of results
“Clustering”	3 280 000
“Vehicle Routing Problem”	46 300
“Clustering” AND “Vehicle Routing Problem”	7512

**Table 3:** Results of search string in Google Scholar

After this search further refining has been done adjusting the dates and final result came down to 638 papers including books. These results were not all relevant. Most of the papers showed up by just mentioning the keywords somewhere in the papers or using similar scenarios like the ‘Travelling Salesmen Problem’ and some of the papers were not available with a full text. After performing forward and backward snowballing to the selected string, a total of 15 and 22 papers were gathered respectively. Finally, by refining the search once again and by removing the duplicate results a total of 10 papers were extracted.

The selected literature was then extensively and carefully analyzed. These papers as a source to answer the first research question in this paper were then categorized based on the objective of the research. The data was then extracted from the literature and documented in various sections of the paper that are ‘Introduction’, ‘Background Study’ and finally in the result of ‘Literature Analysis’.

#### 4.1.1 RISKS

The following show the possible risks or threats to validity regarding the literature analysis.

- In spite of following scientific techniques for literature extraction, there is a possibility that some literature with relevant description could be missed that were available on other databases. This could affect the literature analysis on a minor scale with at most partial possibility of biased conclusions.
- Another possible risk of selecting literature could possess an irrelevant description on the desired topic but the papers were carefully chosen based on the search string, title, year of publication, abstract and an overview of the complete document.

## 4.2 EXPERIMENTATION

This research method has been selected to answer the second research question which is “How can the performance and efficiency be evaluated between the hybrid and exact algorithms for solving the clustered VRP?” With an experimentation and analysis performed on the results, an answer can be made to the second research question based on the final analysis. Experimentation has been selected to accomplish this objective as this research deals with algorithms.

### 4.2.1 EXPERIMENT TYPE

These algorithms are to be compared in order to evaluate their performances against each other that can provide a conclusion on which type of algorithm is better to solve the vehicle routing problem. Thus implementation and comparison is the type of experiment that is conducted to answer the second research question. The implementation is done for the two selected algorithms. The algorithms are also compared to earlier algorithms as it is important to know if both of the algorithms perform well compared to the previously proposed algorithms.

### 4.2.2 VARIABLES CONSIDERED IN THE EXPERIMENT

There are independent and dependent variables that are used in the experiment. The independent variables are those which can be changed with different environments or scenarios. Dependent variables are those which depend on the independent variables and change with the change in the independent variables[26]. The independent variable in this experiment is the size of the problem that is the number of delivery points. The dependent variable is the cost that is the total distance travelled. The cost changes with the change in the problem size and thus that is the dependent variable.

### 4.2.3 EXPERIMENT PROCEDURE

In this research, the two algorithms are implemented and their results are gathered based on the benchmark instances (the data sets). The comparison is done basing on the computational results provided by the algorithms. These computational results can be accurately compared as one advantage in VRP is that there exist some benchmark instances which most of the authors test their algorithms on. These instances are provided as an input file containing the size of the problem, the distance between two nodes and such, so that the computational results are easily comparable based on the instances used[27]. To evaluate the performance of these two algorithms, five other algorithms are considered that solve the CVRP using the same benchmark instances that the two algorithms have used. By comparing the two algorithms to the five algorithms, we will not only know which one of the two perform better but we can also determine how their performance stand compared to other algorithms using different methods and that have been implemented in different times.

All the five algorithms were chosen with the intention to consider various methods, for example exact, heuristics, hybrid etc. They were also chosen based on the type of VRP it solves and most importantly, all these five algorithms must be using the same benchmark instances in order to compare the performances on each instance credibly. Some of the papers in the five formulated multiple algorithms in which only the solutions of the best were considered to form the results.

### 4.2.4 RISKS

This section shows the possible risks that can take place in the validation of the experiment data which are given as follows:

- There could be a possibility that the results that were gathered from each of the papers were evaluated on different parameters which would make the comparison of just the results to be pointless. But to avoid this, the benchmark instances have been carefully considered in each of the papers such that the results obtained by each algorithm.
- There is also a risk based on the five algorithms that were selected to perform the comparison. There could be papers other than the selected ones that perform slightly different in comparison to others which have a chance of concluding the comparative analysis differently. To avoid strong conclusions which could be wrong, the conclusions were made carefully as favorably supporting a particular fact but not strongly confirming the conclusions based on a limited set of testing.

## 5 RESULTS

This section shows the results gathered from the literature and the experiment.

### 5.1 RESULT OF LITERATURE ANALYSIS

There has been extensive research about clustering for general partitioning purposes in the past but literature regarding the significance of clustering particularly regarding VRP consists of limited research. Various data mining applications state to improve the performance of analysis and applications from many years, but the idea of implementing clustering in the VRP has recently surfaced and the results are proven to be quite favorable to their solutions[18]. This literature is based on the experience and roles played by clustering techniques in various papers that have been implemented in combination with various VRP state-of-the-art methods and local search methods.

The following table shows the various ways clustering was implemented in the solutions of a few researchers and how it improved the efficiency of their solution.

Type of clustering in combination with VRP method	Metrics in which the use of clustering was favorable to the solution	Authors of the formulated algorithms
Heuristics with <ul style="list-style-type: none"><li>• Load balance clustering</li><li>• Clustering length approximation</li></ul>	<ul style="list-style-type: none"><li>• Minimized the variation of the distances travelled within each cluster</li><li>• Distributing equal load within every cluster</li></ul>	Nishanth Chandran et al.[28]
Clarke and Wright Saving Algorithm heuristics with <ul style="list-style-type: none"><li>• Cluster first, route second using K-means clustering</li></ul>	<ul style="list-style-type: none"><li>• Computational time</li><li>• Distributed orders within each cluster</li></ul>	S. Halim, and L. Yoanita2[29]
Ant colony optimization (ACO) with <ul style="list-style-type: none"><li>• Manual clustering</li></ul>	<ul style="list-style-type: none"><li>• Computational time</li></ul>	Marc Reimann and Marco Laumanns[30]
Meta heuristics with the use of <ul style="list-style-type: none"><li>• Unsupervised fuzzy clustering technique using unsupervised fuzzy competitive learning (UFCL)</li></ul>	<ul style="list-style-type: none"><li>• Performs well on large scale instances</li><li>• Computational time</li><li>• Distance travelled is reduced due to clustering</li></ul>	Henrique Ewbank et al.[31]
Meta heuristics with the use of <ul style="list-style-type: none"><li>• Hard cluster constraints</li><li>• Distance matrix clustering</li></ul>	<ul style="list-style-type: none"><li>• Reduced problem complexity</li><li>• Low calculation time</li></ul>	Christof Defryn and Kenneth Sörensen[32]
Memetic computation with <ul style="list-style-type: none"><li>• Pairwise distance comparison clustering</li></ul>	<ul style="list-style-type: none"><li>• Clustering caused over all enhancement of the performance.</li><li>• Time complexity</li></ul>	Liang Feng et al. [33]

	<ul style="list-style-type: none"> <li>• Total distance</li> </ul>	
Ant Colony Optimization and Particle Swarm Optimization using <ul style="list-style-type: none"> <li>• Sweep clustering</li> </ul>	The performance gave an <ul style="list-style-type: none"> <li>• Optimized solution</li> <li>• Reduced computational time</li> </ul>	M. A. H. Akhand et al. [34]
Heuristic algorithm with <ul style="list-style-type: none"> <li>• Ant-clustering technique</li> </ul>	The result by using Ant-clustering was <ul style="list-style-type: none"> <li>• Cost optimal</li> <li>• Improved the routing distance</li> </ul>	Jiashan Zhang et al.[35]
Particle swarm optimization using <ul style="list-style-type: none"> <li>• K-means clustering</li> </ul>	The performance was feasible in terms of <ul style="list-style-type: none"> <li>• Shortest route forming</li> <li>• Clustering in short time.</li> </ul>	Yong He et al.[36]
Use of Meta heuristics with <ul style="list-style-type: none"> <li>• Clustering using Grey Wolf Optimizer (C-GWO)</li> <li>• K-means clustering</li> </ul>	<ul style="list-style-type: none"> <li>• Improved performance</li> <li>• Reduced distance</li> </ul>	L Korayem et al[18]

**Table 4:** Result of literature analysis from the selected literature

These authors were significantly satisfied by the performance of their formulations due to the implementation of various clustering algorithms. Clustering alone does not improve the performance but also the right combination of methods used to solve the VRP is necessary. L Korayan et al. stated that K- means clustering is one of the most widely used clustering techniques for various applications but it possesses a certain drawback of falling into a local minima[18]. And to overcome the drawback they have used the GWO to enhance the clustering. They also stated that for clustering, in order to feasible to the most the total number of customer's demands should not exceed the capacity of the vehicle assigned to the cluster.

It has been stated that proximity or distance measures are often used as the basis for clustering the objects. The next basis is described as the demand where lies one of the biggest advantages of clustering[28]. For a homogenous fleet delivery, each vehicle will get equal distributed load within each of the formed cluster. Due to this from the gathered analysis shown in Table 4 it can be clearly seen that for most of the proposed algorithms the use of clustering has effected and reduced the computational time of the execution[36]. This can be useful as a little change in time can matter a lot in cases like disaster relief VRP applications, where the routing distance and computational time play a very major role as the scenario is very risky.

Traditionally, mathematical programming has been used to solve clustering issues within the ambit of vehicle routing problems. Chapleau et al.[37], for example, reexamined the school bus routing problem in an urban area and proposed building clusters to reduce the capacitated vehicle routing problem (CVRP) into smaller routing problems [31]. They especially state the significance of clustering in heavily populated city where distance and time play a major role in the structure of logistics.

It can be seen that the K-means clustering and Ant colony clustering have been used in more instances than the other methods and their results have shown to enhance the performance as found in the review in Table 4.

## 5.2 RESULT OF EXPERIMENT

This sections show the results that have been gathered from all the selected papers based on the benchmark instances that were in common in all these papers and through the result of the implementation of *Algorithm A* and *Algorithm B*. The comparison of performance is done in each case with the best solutions of one of the five papers. The performance that was calculated in the papers is the shortest distance covered through which all the delivery points are covered using  $k$  number of vehicles.

### 5.2.1 IMPLEMENTATION OF *Algorithm A*

The following figure shows the final result of the routing with the best cost and the reduction of route distance during the iteration.

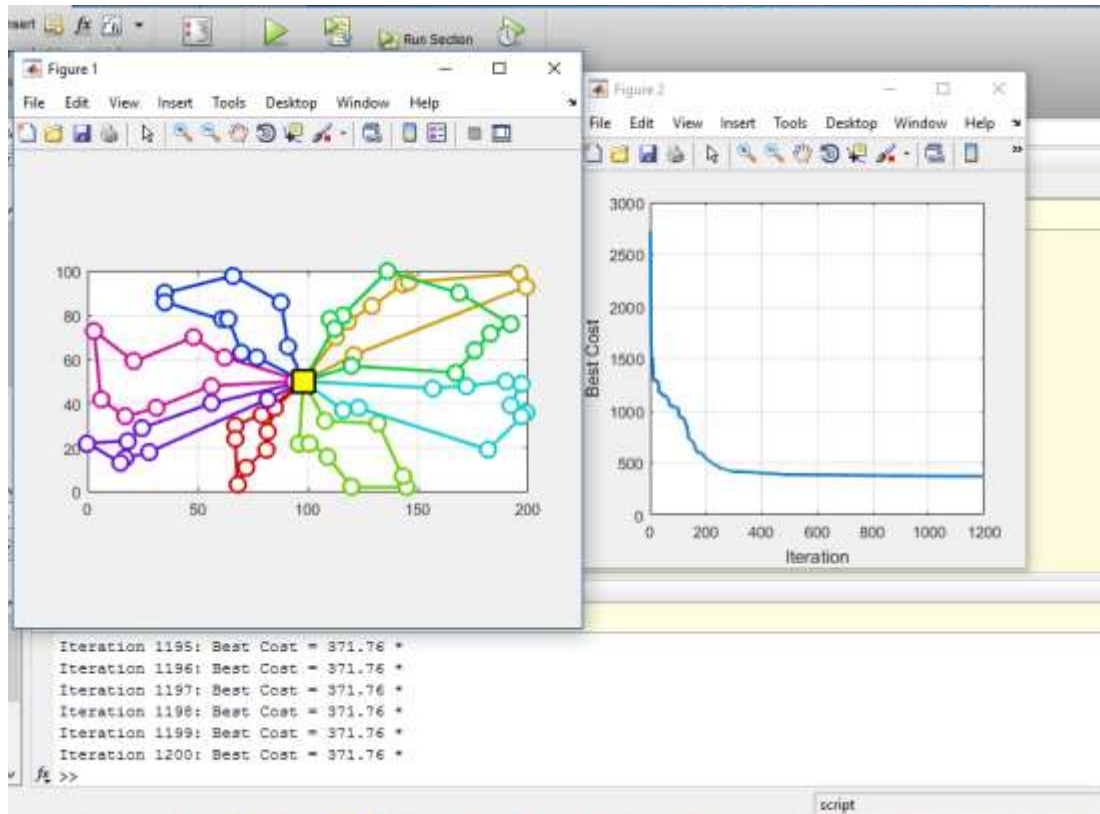
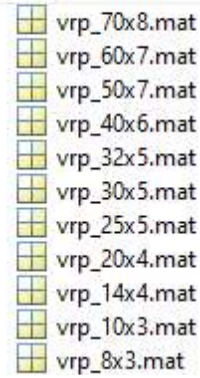


Figure 6: Implementation of Algorithm A

This algorithm has been tested on a wide range of datasets ranging from small scale to medium scale instances. MATLAB was chosen for the implementation of this algorithm as it is an efficient environment to perform the experiment and also easy to make it possible to visualize the output. From Figure 7 the implementation is done on the data set where the problem size was 69 and the number of vehicles is 9 which represent the benchmark instance A-n69-k9. These vehicles are represented in the simulation with a different color assigned to each. The clustering is done and the clustering doesn't depend on the nearest mean alone but also considers providing uniform capacity distribution among the vehicles so that the capacity of no vehicle exceeds its limit. This result in the intercrossing of clusters in some cases which can be seen in Figure 7 with vehicles represented using brown and dark green color.



The following shows a sample data set that the Algorithm is run on:



**Figure 7:** Sample data set for *Algorithm A*

In these data sets are named after their specific parameters for example, vrp\_70x8.mat shows that the problem size is 70 and number of vehicles are 8.

#### 5.2.1.1 Pseudo Code for updating the best cost value after every iteration:

```

if xnew.Cost<=x.Cost

    x=xnew;

else

    delta=xnew.Cost-x.Cost;
    p=exp(-delta/T);

    if rand<=p
        x=xnew;
    end

end

if x.Cost<=BestSol.Cost
    BestSol=x;
end

end

BestCost(it)=BestSol.Cost;

if BestSol.Sol.IsFeasible
    FLAG=' *';
else
    FLAG='';
end
disp(['Iteration ' num2str(it) ': Best Cost = '
num2str(BestCost(it)) FLAG]);

T=alpha*T;
figure(1);
PlotSolution(BestSol.Sol,model);
pause(0.01);

end

```

This is the part of the implementation where the cost is updated every time it finds a lesser cost after every iteration. If the new cost is less than the current cost ( $x_{new}.Cost \leq x.Cost$ ) then the value of the new cost is updated to the current value. It is then compared to the best solution cost value and stores it in its place if it is better than that best solution cost. Then it displays the information after each iteration during the implementation.

The maximum number of iterations is set to a threshold value 1200 and a threshold value for maximum number of inner iterations to 80. The parser will keep testing other combinations of routes during these iterations to find the best cost till the threshold is reached. Once the threshold is reached the final best cost is displayed.

### 5.2.2 IMPLEMENTATION OF *Algorithm B*

This algorithm is the exact algorithm for solving the VRP using branch and cut procedure. This algorithm was implemented on Visual Studio using C++ since the source code was obtained with the same programming language. The functionality of the algorithms runs the same way Algorithm A runs although the background process is different. This algorithm was run on same data sets (benchmark instances) to gather the results.

The branch and cut algorithm works in the following way.

- Set bounds to contain the solution.
- This is done by forcing the variables within the existing bounds, and then tightens the bounds to fix the variables at the nearest integer value.
- If the branch created by object fixes the variables it returns a branching object which would give a new feasible point.
- Perform a branch by adjusting bounds (adding cuts).
- Each arm of the branch advances the object to the next arm by advancing the value of ‘\_way’.
- Comparison is done between the current and new values. It replaces them if feasible

#### 5.2.2.1 Sample code for branch and cut in *Algorithm B*

CbcBranchCut::~CbcBranchCut

()

```
{
}

double
CbcBranchCut::infeasibility(const
OsiBranchingInformation * /*info*/,
                           int &preferredWay) const
{
    throw CoinError("Use of base class",
"infeasibility", "CbcBranchCut");
    preferredWay = -1;
    return 0.0;
}

void
CbcBranchCut::feasibleRegion()
{
}
```

```

bool
CbcBranchCut::boundBranch() const
{
    return false;
}
CbcBranchingObject *
CbcBranchCut::createCbcBranch(OsiSolverInterface *
/*solver*/, const OsiBranchingInformation * /*info*/,
int /*way*/)
{
    throw CoinError("Use of base class",
"createCbcBranch", "CbcBranchCut");
    return new CbcCutBranchingObject();
}
CbcBranchingObject *
CbcBranchCut::preferredNewFeasible() const
{
    throw CoinError("Use of base class",
"preferredNewFeasible", "CbcBranchCut");
    return new CbcCutBranchingObject();
}
CbcBranchingObject *
CbcBranchCut::notPreferredNewFeasible() const
{
    throw CoinError("Use of base class",
"notPreferredNewFeasible", "CbcBranchCut");
    return new CbcCutBranchingObject();
}
void
CbcBranchCut::resetBounds()
// Default Constructor
CbcCutBranchingObject::CbcCutBranchingObject()
: CbcBranchingObject()
{
    down_ = OsiRowCut();
    up_ = OsiRowCut();
    canFix_ = false;
}
// Useful constructor
CbcCutBranchingObject::CbcCutBranchingObject
(CbcModel * model,
    OsiRowCut & down,
    OsiRowCut &up,
    bool canFix)
: CbcBranchingObject(model, 0, -1, 0.0)
{

```

```

        down_ = down;
        up_ = up;
        canFix_ = canFix;
    }
    CbcBranchingObject *
    CbcCutBranchingObject::clone() const
    {
        return (new CbcCutBranchingObject(*this));
    }
    // Destructor
    CbcCutBranchingObject::~CbcCutBranchingObject ()
    double
    CbcCutBranchingObject::branch()
    {
        decrementNumberBranchesLeft();
        OsiRowCut * cut;
        if (way_ < 0) {
            cut = &down_;
            way_ = 1;
        } else {
            cut = &up_;
            way_ = -1;           // Swap direction
        }
        printf("CUT %s ", (way_ == -1) ? "up" : "down");
        cut->print();
        double lb = cut->lb();
        double ub = cut->ub();
        int n = cut->row().getNumElements();
        const int * column = cut->row().getIndices();
        const double * element = cut->row().getElements();
        OsiSolverInterface * solver = model_->solver();
        const double * upper = solver->getColUpper();
        const double * lower = solver->getColLower();
        double low = 0.0;
        double high = 0.0;
        for (int i = 0; i < n; i++) {
            int iColumn = column[i];
            double value = element[i];
            if (value > 0.0) {
                high += upper[iColumn] * value;
                low += lower[iColumn] * value;
            } else {
                high += lower[iColumn] * value;
                low += upper[iColumn] * value;
            }
        }
    }

```

### 5.2.3 Combining all the obtained results

The first set of results were gathered from the algorithm proposed by Mark et al.[38] This algorithm is very much similar to the *Algorithm A* where they used a similar hybrid solution where they used k- means clustering with simulated annealing. There are still some specific differences that lead to difference in the results. The gathered results are shown in Table 5

Instance name	Algorithm A	Algorithm B	Mark et al.
A-n37-k5	437	671	573.41
A-n38-k5	476	734	510.31
A-n39-k5	557	826	620.30
A-n44-k6	608	937	731.17
A-n45-k6	613	930	715.51
A-n48-k7	667	1073	688.03

**Table 5:** Comparison of results with algorithm by Mark et al.

The Table 6 shows the results in case of the second paper proposed by Ganesh et al. [39] where they used GA to solve VRP on the benchmark instances with a combination of different methods including clustering.

Instance name	Algorithm A	Algorithm B	Ganesh et al.
A-n37-k5	430	671	669
A-n38-k5	464	734	730
A-n39-k5	557	826	822
A-n39-k6	544	833	831
A-n44-k6	610	937	937
A-n46-k7	593	915	914
A-n48-k7	672	1073	1073
A-n53-k7	633	1010	1017
A-n54-k7	687	1167	1172
A-n55-k9	699	1073	1073
A-n60-k9	770	1355	1358
A-n61-k9	635	1035	1038
A-n62-k8	738	1288	1288
A-n63-k10	800	1314	1322
A-n64-k9	759	1401	1410
A-n65-k9	687	1174	1177
A-n69-k9	690	1160	1163
A-n80-k10	1008	1761	1780

**Table 6:** Comparison of results with algorithm by Ganesh et al

This paper was proposed by Feng et al.[40], where they used a linear genetic model to formulate the algorithm for solving the CVRP. From this paper, two types of instances were considered that are the type A and type B. The gathered results regarding this paper in comparison to *Algorithm A* and *Algorithm B* are shown in Table 7

Instance name	Algorithm A	Algorithm B	Feng et al.
A-n54-k7	687	1167	1167
A-n60-k9	770	1355	1354
A-n69-k9	690	1160	1159
A-n80-k10	1008	1761	1763

B-n57-k7	552	1153	1140
B-n63-k10	588	1501	1496
B-n78-k10	601	1232	1221

**Table 7:** Comparison of results with algorithm by Feng et al.

The next set of results were extracted from the old and relatively successful formulation that have been proposed by Bektas et al[27]. They used an exact algorithm using BC as their procedure. Their results were known to be very effective but the computational time was rather higher than a normal heuristic or other type of algorithms.

The results that were gathered from this paper have been given below in the Table 8. They are almost close to the results obtained by Algorithm 1. From this paper three types of instances were compared. Type A, type B and type P are the instance type.

Instance name	Algorithm A	Algorithm B	Bektás et al.
A-n45-k6	613	930	480.83
A-n55-k9	699	1073	562.81
A-n63-k9	801	1314	654
A-n80-k10	1008	1761	806
B-n51-k7	513	1031	525.67
B-n63-k10	588	1501	713.51
B-n67-k10	562	1032	539.97
B-n78-k10	604	1232	659.81
P-n45-k5	245	510	272.03
P-n55-k8	268	570	298.75
P-n65-k10	375	792	415.20
P-n101-k4	379	681	375.30

**Table 8:** Comparison of results with algorithm by Bektás et al.

The final paper selected was proposed by Korayem et al. where they used a very unique method to formulate the results. This paper was also compared with the three types of instances as mentioned above. The results compared to both the algorithms are given below in the Table 9

Instance Name	Algorithm A	Algorithm B	Korayem
A-n60-k9	595	1355	1395
A-n62-k8	596	1288	1329
B-n39-k5	280	549	564
B-n50-k7	393	741	765
P-n45-k5	238	510	510
P-n76-k4	309	593	614
P-n76-k5	309	662	635

**Table 9:** Comparison of results with algorithm by Korayem

## 6 ANALYSIS AND DISCUSSION

This section then deal with the analysis that has been concluded from the results gathered in Section 4.

### 6.1 ANALYSIS BASED ON LITERATURE ANALYSIS

From the literature review conducted in Section 4 we can strongly recommend that the significance of various clustering techniques in the implementation of VRP algorithms benefit in a very satisfactory measure. It has proven to improve the performance of the algorithm by reducing the cost and computational time by being implemented on the right measures constraints. The analysis performed on the selected research can provide the answer to first research question (RQ1).

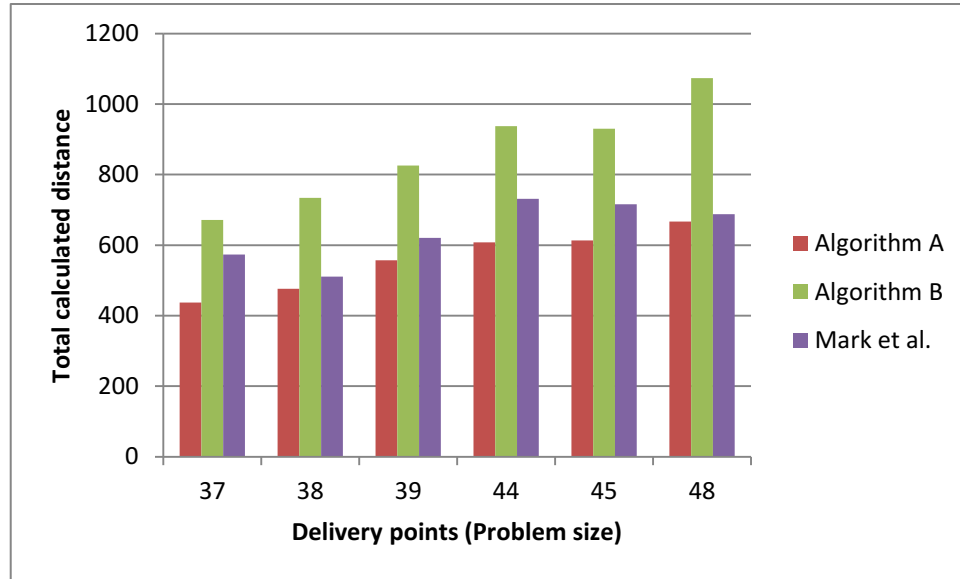
### 6.2 ANALYSIS BASED ON COMPARISON OF ALGORITHMS

The main objective regarding the second research question is to compare the performance of the algorithms, which is mainly focused in this section. The two algorithms are compared to the performances as well among which few algorithms were older than the selected 2 and few were more recently published solutions.

These algorithms are compared using the benchmark instances which provide the input problem to the algorithms, in which the main parameter is the size of the problem. The size of the problem is nothing but the number of vertices or delivery points that exist in the instances. The results show the performances using the both algorithms along with one previously proposed or newly developed algorithm which can help us compare the performances of not only the selected ones but also the other ones simultaneously. Comparing algorithms in this way will benefit to approximately show where the selected algorithms stand in the race of number of authors who come up with different methods of algorithms in order to optimize the VRP.

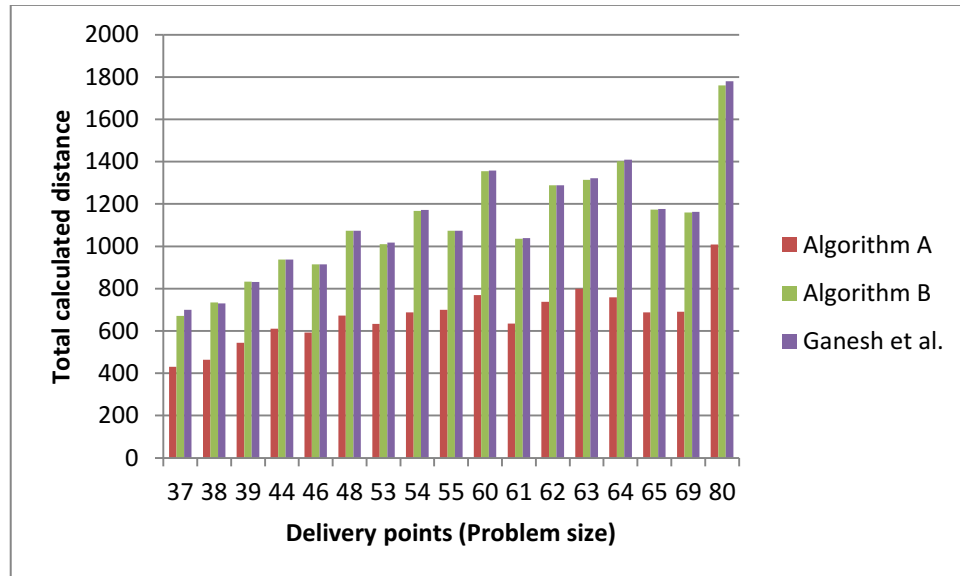
An analysis has been made using the results given in the previous section. In this research, the performance of the algorithms is solely monitored based on the problem size and the solution which was calculated in distance metrics. Three types of Benchmark instances have been selected on a total and each are denoted by the starting letter that are A, B and P. The instances with each letter increase with the problem size with every instance. Thus the analysis has been made based on the increment of problem size and the solution for each of the papers. The analysis and discussion related to each paper and the type of instance is given as follows:

For the first paper, the algorithm is similar to *Algorithm A* where they use a similar hybrid model with simulated annealing and k-means clustering. Looking at the graph below in Figure 6, we can see that there is a huge difference between the performances of all the three algorithms. We can see that as the problem size increases, the total distance also increases and the comparison shows that the hybrid algorithm by Mark et al. is almost as close to *Algorithm A*. Clearly, *Algorithm A* performs better than the other two.



**Figure 8:** Results comparison with Mark et al. on instance type A

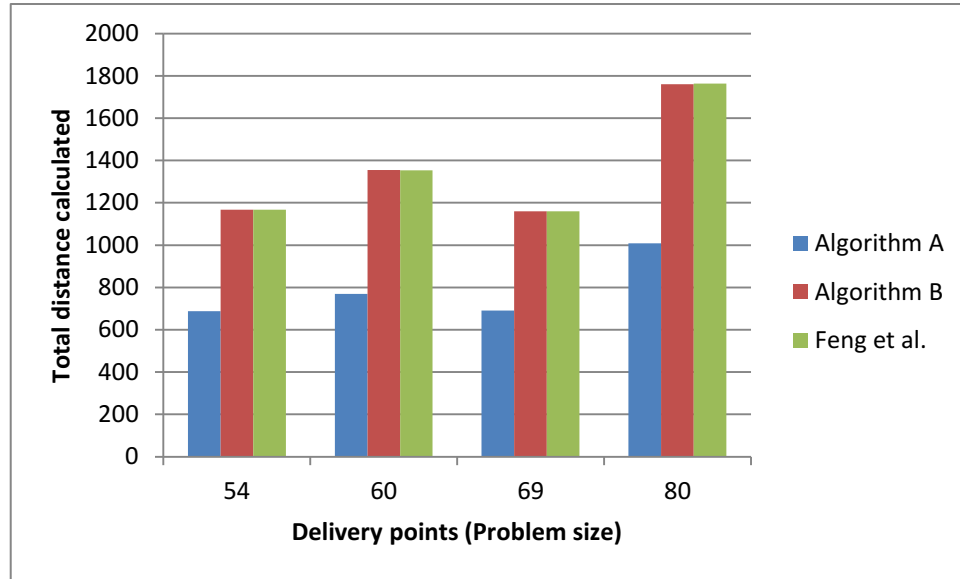
The next comparison in Figure 7 is done by comparing *Algorithm A* and *B* to Ganesh et al.[39]. The results belonging to this analysis are taken from the ‘A’ instances of Augerat [41]. The analysis shows that *Algorithm A* performs a lot better than both *Algorithm B* and the other paper. In this case, it can also be noticed that the performance of *Algorithm B* is relatively better than the performance of GA by Ganesh et al.



**Figure 9:** Result table with Ganesh et al. Algorithm with instance type A

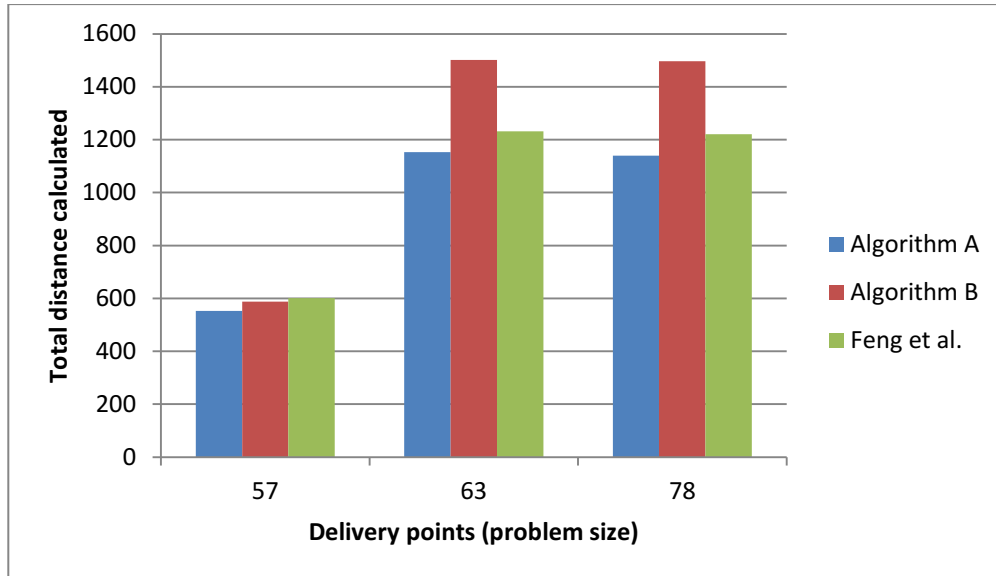
Coming to the third paper, as mentioned earlier this comparison is done to Feng et al.[40]. The performance analysis in this case is much similar to that of the previous case, where *Algorithm A* performs better than the other two and *Algorithm B* takes the second place. Also the performance of the algorithm proposed by Feng et al is much similar to the performance of *Algorithm B*.





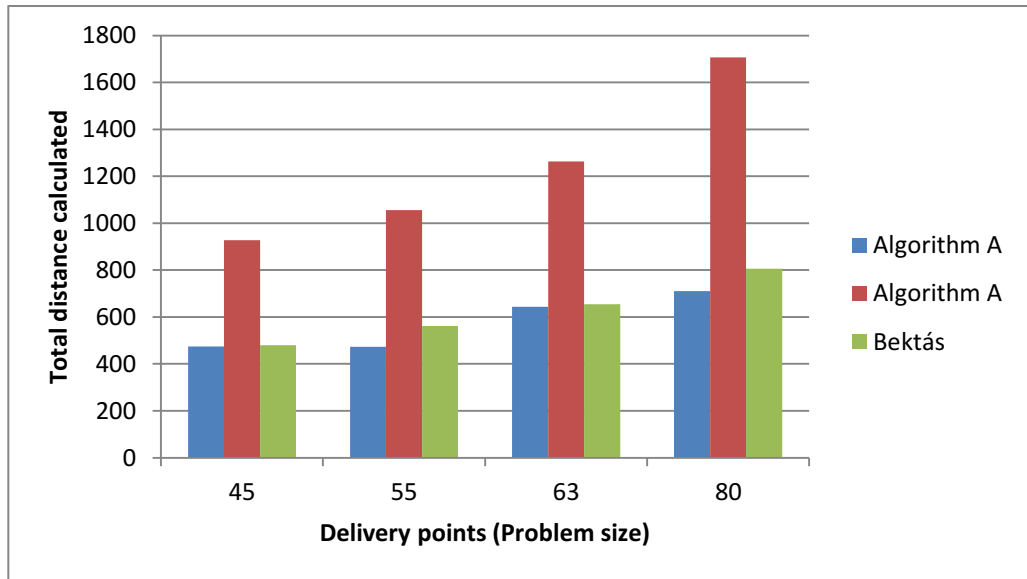
**Figure 10:** Result compare table with Feng et al. Algorithm instance type A

This algorithm has given the chance to compare the results on two instances where Figure 8 is compared with instance type A and Figure 9 is compared on instance type B. The performance is the same regarding *Algorithm A* versus the other two algorithms but *Algorithm B* performance has degraded in one case. But as the problem size increased the problem size increased it performed as same as it did in instance type A.

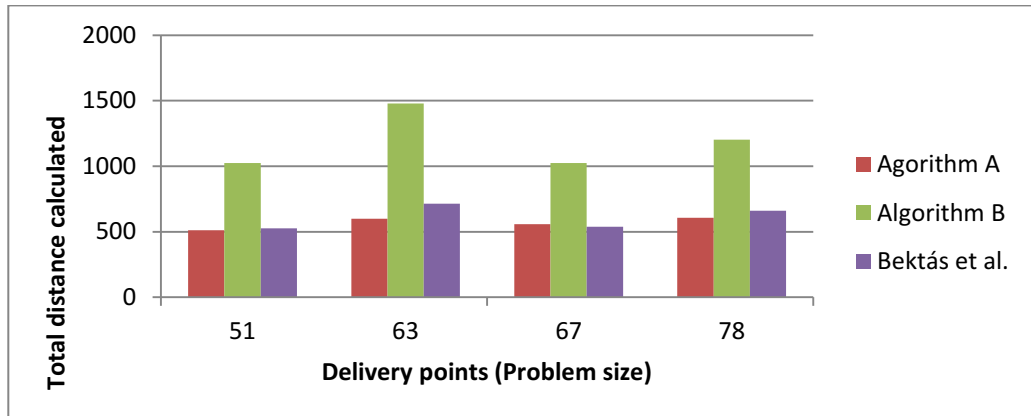


**Figure 11:** Result compare table with Feng et al. Algorithm instance type B

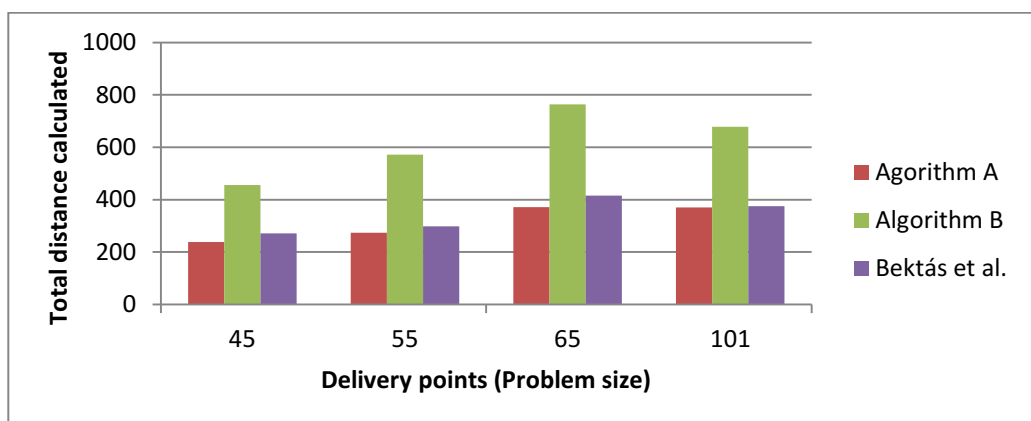
The next algorithm is an old yet fairly successful algorithm proposed by Bektás et al [27]. The results of this algorithm were specifically chosen in three types of instances to observe the performance of *Algorithm A* and *B* compared to this one. The graph below in Figure 10 clearly shows that the performance of the old algorithm is very close to the performance of *Algorithm A* and exactly in one instance the performance of the old Algorithm is a bit higher in Figure 11. The analyses of the algorithms are given below in Figure 10, Figure 11 and Figure 12 with instances A, B and P respectively.



**Figure 12:** Result compare table with Bektás et al. Algorithm instance type A



**Figure 13:** Result compare table with Bektás et al. instance type B



**Figure 14:** Result compare table with Bektás et al. instance type P

The final paper that was compared to the Algorithms was proposed by Korayem et al [18] which used a very unique methodology in their algorithm as mentioned in the previous section. The performance of this algorithm is similar to the ones in the first few cases. *Algorithm A* beats the performance of the remaining two and *Algorithm B* has a better performance compared to the Korayem's algorithm. This algorithm was also tested on three instances A, B and P which are shown in Figure 13, Figure14 and Figure 15 respectively.

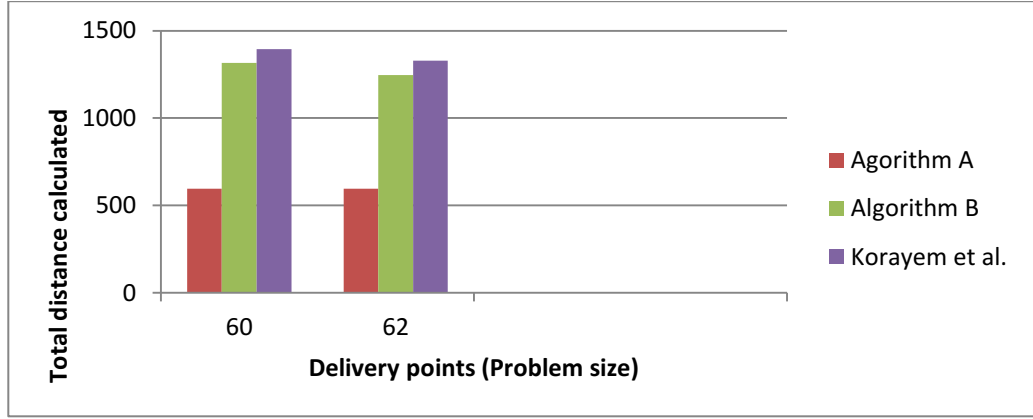


Figure 15: Result compare table with Korayem et al. instance type A

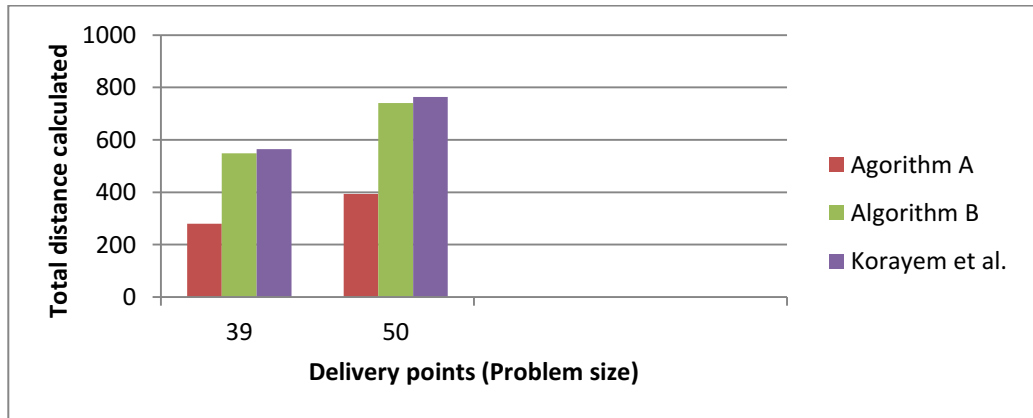


Figure 16: Result compare table with Korayem et al. instance type B

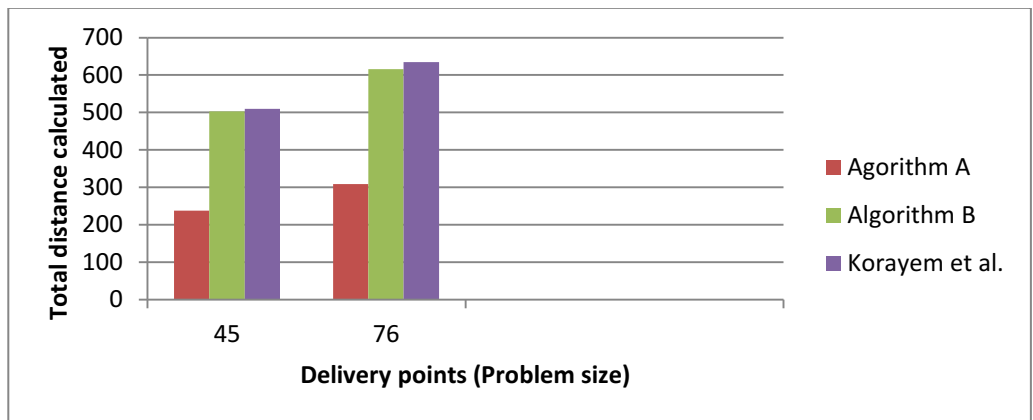


Figure 17: Result compare table with Korayem et al. instance type P

### 6.3 STATISTICAL ANALYSIS

ANOVA (Analysis of variance) two-factor without replication method has been chosen to perform the statistical analysis to the collected data as it was suitable for the kind of data used in this thesis. ANOVA is a widely used statistical hypothesis testing tool for analyzing experimental results. The type of comparison done is using data which has more number of factors that motivated to use the two-factor without replication method in ANOVA specifically. This analysis has been conducted in ‘Microsoft Excel’ using the data analysis tool and it provides the variance from which the standard deviation can be obtained. It provides the source of variance in rows and columns and the values obtained will allow us to reject or approve a hypothesis statistically. For example, if we were to compare the performance of each algorithm basing on only a single instance then that is the case where we should be using ANOVA two-factor with replication, since we compare the performances on more than one instances the two-factor without replication is used. To test the normality of the data it was tested using Shapiro-Wilk and also Kolmogorov’s test using SAS analysis tool. Both tests proved that the data was normally distributed. Although, according to [42] Shapiro’s results are only reliable on number of observations >2000, the current number of observations considered in the data sets are significantly lesser than 2000. This test was necessary as a process of selecting the right analysis test to perform on the data and ANOVA assumes the data to be normally distributed, which added the motivation to choose ANOVA.

A statistical analysis has been conducted to each of the cases individually. We only consider the column wise variance as we want to observe the difference in comparison of each of the algorithm in every case and not the row wise (difference in comparison of performance with each problem size). A sample of the complete analysis of the first comparison based on Table 5 and Figure 6 is given as follows:

<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
37	3	1681.41	560.47	13814.58		
38	3	1720.31	573.4367	19629.73		
39	3	2003.3	667.7667	19780.06		
44	3	2276.17	758.7233	27629.64		
45	3	2258.51	752.8367	26167.21		
48	3	2428.03	809.3433	52246.69		
Algorithm A	6	3358	559.6667	7751.067		
Algorithm B	6	5171	861.8333	21786.17		
Mark et al.	6	3838.73	639.7883	7592.431		
ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Rows	161168.8	5	32233.76	13.16764	0.000393	3.325835
Columns	294056.3	2	147028.2	60.06168	2.68E-06	4.102821
Error	24479.53	10	2447.953			
Total	479704.6	17				

**Table 10:** Analysis of the comparison of Algorithm A and B with Mark et al.

Let us consider two hypothesis to help answer the **RQ2**

- **Null Hypothesis:** The performance of all the algorithms are similar
- **Alternative Hypothesis:** There is significant difference in the performance of Algorithm A, B and the five algorithms and one of them performs better than the other.

The standard deviation from the given table is 3.74 and the critical value is 1.82. From the above table we can see that in the source of variation the F value and F critical value are unequal and  $F > F_{critical}$ . An alpha ( $\alpha$ ) value of 0.05 is given which is a standard value. If the P-value is less than the alpha value then it shows that there is significant difference in the performance of the algorithms. It is clearly seen that

- $F > F_{critical}$
- $\alpha > P\text{-value}$

Similarly, this is the case with every other comparison that is with the rest of the four algorithms. It shows us that we can reject the null hypothesis. The main objective of this research is to compare the performance of the two algorithms which is why only one single analysis on both the algorithm is shown. This proves us that there is significant difference in the performance of all the algorithms and one of them performs better than the other.

From the gathering of the results and conducting analysis on them it is clear that the results are highly favourable towards the conclusion where it is observable that the performance of both the algorithms are very different and *Algorithm A* performs much better than *Algorithm B*. The results also show that *Algorithm A* has given a better performance than all the five older and newer algorithms. This shows the dominance of the this algorithm, the combination of the right local search methods with k-means clustering has given a highly optimal solution. This cannot be proven as the best Algorithm but it can be used as a basis to further optimize the algorithm even more. This can also be tested on large scale problem instances and easily find way to provide real time services due to the efficient performance of the algorithm.

Coming to *Algorithm B* the performance is not as good compared to *Algorithm A* but compared to the five other algorithms, *Algorithm B* performs better in some of the cases. This shows that *Algorithm B* performs relatively acceptable compared to many algorithms that are being proposed but it can still be further optimized using the right parameters.

## 7 CONCLUSION AND FUTURE WORK

In this research two newly developed algorithms were considered for comparison that solves the vehicle routing problem. The significance of implementing clustering techniques in vehicle routing problems is determined in the first research question and the performance of the two algorithms that use clustering are compared with five other state-of-the-art algorithms and conduct a statistical analysis for determining the difference between the performances of the selected algorithms. From the research conducted in this paper we can say that the significance of clustering is very high from the literature analysis conducted as there have been several practical situations where researchers used clustering with the right combination of search and routing methods to increase the performance of their algorithm. The use of clustering has recently started emerging and has benefited the authors in a very favorable manner. This research aims to surface the significance of clustering and its methods to future researchers to consider the method for the better of their performance. It can be seen from the results and analysis regarding the comparison of multiple algorithms based on the total computed distance that genetic algorithms with the combination of clustering have a high scope of reaching further optimal solutions to solve the Vehicle Routing Problem.

Future work can indicate possibilities of implementing the algorithms and scaling further metrics of the solutions like computational time, space complexity and possible chances of reducing the usage of number of vehicles to indirectly help solve the problem using more constraints and finally test the algorithms using large scale instances. An algorithm can be developed using clustering with proper local search methods and can be compared with the best found solutions in this paper.

## REFERENCES

- [1] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, 2001.
- [2] S. Akpinar, “Hybrid large neighbourhood search algorithm for capacitated vehicle routing problem,” *Expert Syst. Appl.*, vol. 61, pp. 28–38, Nov. 2016.
- [3] T. Sooktip and N. Wattanapongsakorn, “Identifying preferred solutions for multi-objective optimization: application to capacitated vehicle routing problem,” *Clust. Comput.*, vol. 18, no. 4, pp. 1435–1448, Dec. 2015.
- [4] P. Toth and D. Vigo, *Vehicle Routing: Problems, Methods, and Applications, Second Edition*. SIAM, 2014.
- [5] “What is Operations Research? - INFORMS.” [Online]. Available: <https://www.informs.org/About-INFORMS/What-is-Operations-Research>. [Accessed: 09-Feb-2017].
- [6] D. Steenken, S. Voß, and R. Stahlbock, “Container terminal operation and operations research - a classification and literature review,” *Spectr.*, vol. 26, no. 1, pp. 3–49, Jan. 2004.
- [7] B. H. Ba, C. Prins, and C. Prodhon, “Models for optimization and performance evaluation of biomass supply chains: An Operations Research perspective,” *Renew. Energy*, vol. 87, Part 2, pp. 977–989, Mar. 2016.
- [8] R. Stahlbock and S. Voß, “Operations research at container terminals: a literature update,” *Spectr.*, vol. 30, no. 1, pp. 1–52, Jan. 2008.
- [9] G. Ghiani, D. Laganà, E. Manni, R. Musmanno, and D. Vigo, “Operations research in solid waste management: A survey of strategic and tactical issues,” *Comput. Oper. Res.*, vol. 44, pp. 22–32, Apr. 2014.
- [10] G. B. Dantzig and J. H. Ramser, “The Truck Dispatching Problem,” *Manag. Sci.*, vol. 6, no. 1, pp. 80–91, Oct. 1959.
- [11] G. Laporte, “The vehicle routing problem: An overview of exact and approximate algorithms,” *Eur. J. Oper. Res.*, vol. 59, no. 3, pp. 345–358, 1992.
- [12] B.-I. Kim, S. Kim, and S. Sahoo, “Waste collection vehicle routing problem with time windows,” *Comput. Oper. Res.*, vol. 33, no. 12, pp. 3624–3642, Dec. 2006.
- [13] M. Hoogeboom, M. Battarra, G. Erdoğan, and D. Vigo, “Erratum—Exact Algorithms for the Clustered Vehicle Routing Problem,” *Oper. Res.*, vol. 64, no. 2, pp. 456–457, 2016.
- [14] W. Yi and A. Kumar, “Ant colony optimization for disaster relief operations,” *Transp. Res. Part E Logist. Transp. Rev.*, vol. 43, no. 6, pp. 660–672, Nov. 2007.
- [15] X. Wang, S. Poikonen, and B. Golden, “The vehicle routing problem with drones: several worst-case results,” *Optim. Lett.*, pp. 1–19, 2016.
- [16] A. Lotz, “Drones in Logistics: A Feasible Future or a waste of effort.,” *Honors Proj.*, Dec. 2015.
- [17] D.-S. Chen, R. G. Batson, and Y. Dang, “Branch-and-Cut Approach,” in *Applied Integer Programming*, John Wiley & Sons, Inc., 2009, pp. 305–333.
- [18] L. Korayem, M. Khorsid, and S. S. Kassem, “Using Grey Wolf Algorithm to Solve the Capacitated Vehicle Routing Problem,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 83, no. 1, p. 12014, 2015.
- [19] “CVRPLIB - All Instances.” [Online]. Available: <http://vrp.atd-lab.inf.puc-rio.br/index.php/en/>. [Accessed: 10-Jan-2017].
- [20] P. C. Pop, O. Matei, and C. P. Sitar, “An improved hybrid algorithm for solving the generalized vehicle routing problem,” *Neurocomputing*, vol. 109, pp. 76–83, Jun. 2013.
- [21] C. Contardo and R. Martinelli, “A new exact algorithm for the multi-depot vehicle routing problem under capacity and route length constraints,” *Discrete Optim.*, vol. 12, pp. 129–146, May 2014.

- [22] S. Barreto, C. Ferreira, J. Paixão, and B. S. Santos, "Using clustering analysis in a capacitated location-routing problem," *Eur. J. Oper. Res.*, vol. 179, no. 3, pp. 968–977, Jun. 2007.
- [23] H.-S. Hwang, "Design of supply-chain logistics system considering service level," *Comput. Ind. Eng.*, vol. 43, no. 1–2, pp. 283–297, Jul. 2002.
- [24] Y. Levy and T. J. Ellis, "A systems approach to conduct an effective literature review in support of information systems research," *Informing Sci. Int. J. Emerg. Transdiscipl.*, vol. 9, no. 1, pp. 181–212, 2006.
- [25] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Q.*, pp. xiii–xxiii, 2002.
- [26] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Springer Science & Business Media, 2012.
- [27] T. Bektas, "The multiple traveling salesman problem: an overview of formulations and solution procedures," *Omega*, vol. 34, no. 3, pp. 209–219, Jun. 2006.
- [28] N. Chandran, T. t. Narendran, and K. Ganesh, "A clustering approach to solve the multiple travelling salesmen problem," *Int. J. Ind. Syst. Eng.*, vol. 1, no. 3, pp. 372–387, Jan. 2006.
- [29] S. Halim and L. Yoanita, "Adjusted clustering Clarke-Wright Saving Algorithm for two depots-N vehicles," in *2015 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2015, pp. 270–274.
- [30] "A hybrid ACO algorithm for the Capacitated Minimum Spanning Tree Problem." [Online]. Available: [https://www.researchgate.net/publication/221411193\\_A\\_hybrid\\_ACO\\_algorithm\\_for\\_the\\_Capacitated\\_Minimum\\_Spanning\\_Tree\\_Problem](https://www.researchgate.net/publication/221411193_A_hybrid_ACO_algorithm_for_the_Capacitated_Minimum_Spanning_Tree_Problem). [Accessed: 11-Jan-2017].
- [31] H. Ewbank, P. Wanke, and A. Hadi-Vencheh, "An unsupervised fuzzy clustering approach to the capacitated vehicle routing problem," *Neural Comput. Appl.*, vol. 27, no. 4, pp. 857–867, May 2016.
- [32] C. DEFERNY and K. SÖRENSEN, "A two-level variable neighbourhood search for the Euclidean clustered vehicle routing problem," University of Antwerp, Faculty of Applied Economics, Working Paper 2015002, 2015.
- [33] L. Feng, Y. S. Ong, M. H. Lim, and I. W. Tsang, "Memetic Search With Interdomain Learning: A Realization Between CVRP and CARP," *IEEE Trans. Evol. Comput.*, vol. 19, no. 5, pp. 644–658, Oct. 2015.
- [34] M. A. H. Akhand, Zahrul Jannat Peya, T. Sultana, and Al-Mahmud, "Solving Capacitated Vehicle Routing Problem with route optimization using Swarm Intelligence," 2015, pp. 112–117.
- [35] J. Zhang, X. Lin, J. Yi, and Q. Li, "Solving the Capacitated Vehicle Routing Problem Based on Improved Ant-clustering Algorithm," in *MATEC Web of Conferences*, Les Ulis, France, 2015, vol. 22.
- [36] Y. He, J. Wen, and M. Huang, "Study on Emergency Relief VRP Based on Clustering and PSO," in *2015 11th International Conference on Computational Intelligence and Security (CIS)*, 2015, pp. 43–47.
- [37] L. Chapleau, J.-A. Ferland, and J.-M. Rousseau, "Clustering for routing in densely populated areas," *Eur. J. Oper. Res.*, vol. 20, no. 1, pp. 48–57, 1985.
- [38] A. H. Marc, L. Fuksz, P. C. Pop, and D. Dănciulescu, "A Novel Hybrid Algorithm for Solving the Clustered Vehicle Routing Problem," in *Hybrid Artificial Intelligent Systems*, E. Onieva, I. Santos, E. Osaba, H. Quintián, and E. Corchado, Eds. Springer International Publishing, 2015, pp. 679–689.
- [39] K. Ganesh and T. T. Narendran, "CLOVES: A cluster-and-search heuristic to solve the vehicle routing problem with delivery and pick-up," *Eur. J. Oper. Res.*, vol. 178, no. 3, pp. 699–717, May 2007.
- [40] L. Feng, Y.-S. Ong, C. Chen, and X. Chen, "Conceptual modeling of evolvable local searches in memetic algorithms using linear genetic programming: a case study on capacitated vehicle routing problem," *Soft Comput.*, vol. 20, no. 9, pp. 3745–3769, Sep. 2016.



- [41]P. Augerat, J. M. Belenguer, E. Benavent, A. Corberan, D. Naddef, and G. Rinaldi,  
“Computational results with a branch and cut code for the capacitated vehicle routing  
problem,” *Rapp. Rech. - IMAG*, 1995.
- [42]H. M. Park, “Univariate analysis and normality test using SAS, Stata, and SPSS,” 2015.