DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DESIGNING FEEDER BUS ROUTES BY USING SMART CARD DATA

by

Hassan Shuaibu ABDULRAHMAN

July, 2021

İZMİR

DESIGNING FEEDER BUS ROUTES BY USING SMART CARD DATA

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Civil Engineering, Transportation Program

by
Hassan Shuaibu ABDULRAHMAN

July, 2021
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **“DESIGNING FEEDER BUS ROUTES BY USING SMART CARD DATA”** completed by **HASSAN SHUAIBU ABDULRAHMAN** under supervision of **ASSOC.PROF.DR. MUSTAFA ÖZUYSAL** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

Assoc. Prof. Dr. Mustafa ÖZUYSAL

Supervisor

Prof. Dr. Serhan TANYEL

Thesis Committee Member

Assoc. Prof. Dr. Yalçın ALVER

Thesis Committee Member

Prof. Dr. Hüseyin CEYLAN

Examining Committee Member

Assoc. Prof. Dr. Görkem GÜLHAN

Examining Committee Member

Prof. Dr. Özgür ÖZÇELİK

Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENT

First and foremost, I am extremely grateful to my supervisor, Assoc. Prof. Dr. Mustafa ÖZUYSAL, and the supervısory team: Prof. Dr. Serhan TANYEL, and Assoc. Prof. Dr. Yalçın ALVER for their invaluable guidance, continuous support, and patience during my study. Their immense knowledge and plentiful experience have encouraged me in all the time of my academic pursuit. Besides my supervisor, I would like to thank the rest of my thesis committee: Prof. Dr. Hüseyin CEYLAN, Assoc. Prof. Dr. Görkem GÜLHAN for their constructive comments and suggestions. Also worthy of note are; Prof. Dr. Burak ŞENGÖZ and Prof. Dr. Ali TOPAL for their help during my thesis competency examination. I would like to give a special thanks to all my colleagues and friends for their motivation and sincere help during my studies.

I dedicate this thesis to the glory of Allah and my family. A special gratitude to my loving parents, Dr. Shuaibu ABDULRAHMAN and Hajiya Hadiza ABDULRAHMAN whose words of encouragement, sacrifice, incessant prayers, and push for tenacity has been a source of my strength. To my wife; Aisha T. IDRIS and my children; Ali and Abdallah, without their tremendous understanding, sacrifice and encouragement in these past years, it would have been impossible for me to complete my study. And to my siblings, friends, and everyone who believed in me and have supported me in any way, I say thank you.

This PhD studies is a scholarship scheme sponsored by the office of Presidency for Turks Abroad and Related Communities, Ministry of Culture and Tourism, Turkey for which I am grateful for the opportunity. This research was also supported by the Scientific and Technological Research Council of Turkey with project name; interpretation of settlement pattern changes in Turkey via Izmir case (project number:17k818) for which I am indebted. I am also grateful to the Graduate School of Natural and Applied Sciences of Dokuz Eylül University for coordinating the entire program.

Hassan Shuaibu ABDULRAHMAN

DESIGNING FEEDER BUS ROUTES BY USING SMART CARD DATA

ABSTRACT

Attempts have been made of over decades to design a more efficient and yet sustainable transportation system by using different mass transportation systems. A simple form of this, is the intermodal transportation system which can be described by feeder bus route design problem (FBRNDP). One of the fundamental components of transit planning is the estimation of origin-destination (O-D) matrix. Collecting such data, however, is extremely difficult and expensive using traditional paper-survey-based approaches which are rarely available. Therefore, a new data source was explored in the form of transit smart cards. This spatio-temporal data is most important for any meaningful transit smart card studies, and even so, if they are used alongside other data sources such as data base management system and geographic information systems, their potential can be harnessed. Multiple Traveling Salesman Problem (MTSP) is similar to FBRNDP in which there are many origins and one destination. The main purpose of this research is to develop a methodology for the planning of public feeder bus routes by using smart card data, transit network with existing bus stop coordinates and a distance optimization algorithm. In this study, important aspects of feeder bus route network design are highlighted, a method for the determining of potential feeder bus stop location developed, and a modified genetic algorithm was used to solve FBRNDP. The methodology was used to design a new feeder bus routes, and the best solution was selected via pareto efficiency and minimized total system cost.

Keywords: Feeder bus routes, smart card data, multiple travelling salesman problem, genetic algorithm

OTOBÜS BESLEME GÜZERGAHLARININ AKILLI KART VERİLERİ KULLANILARAK TASARLANMASI

ÖZ

Farklı toplu taşıma sistemleri kullanarak daha verimli ve yine de sürdürülebilir bir ulaşım sistemi tasarlamak için son yıllarda birçok girişimlerde bulunulmuştur. Bunun basit bir formu, besleyici otobüs güzergahı tasarım problemi (FBRNDP) ile tanımlanabilen intermodal ulaşım sistemidir. FBRNDP toplu ulaşım planlama sürecindeki farklı adımlar arasında rota tasarımı ilki ve belki de en önemlisidir. Toplu ulaşım planlamasında kullanılan diğer bir temel bileşen de başlangıç-varış (OD) matrisinin tahmin edilmesidir. Bununla birlikte, bu tür OD verilerini toplamak, geleneksel yüzyüze anket tabanlı yaklaşımlar kullanılarak oldukça zor ve yüksek maliyetlidir. Bu nedenle çalışmada, toplu ulaşım akıllı kart verileri gibi yeni bir veri kaynağının kullanılması incelenmiştir. Bu veriler, birçok toplu ulaşım çalışması için özel öneme sahiptir. Veri tabanı yönetim sistemi ve coğrafi bilgi sistemleri gibi diğer veri kaynakları ile birlikte kullanılırlarsa, tam potansiyellerinden yararlanılabilir. Çoklu gezgin satıcı problemi (MTSP), birçok kaynağın ve tek bir hedefin olduğu gerçek hayat problemi olan FBRNDP'ye oldukça benzer özelliklere sahiptir. Bu araştırmanın temel amacı, akıllı kart verilerini, toplu ulaşım ağı ile mevcut otobüs durağı koordinatlarını ve bir mesafe optimizasyon algoritmasını kullanarak genel besleyici otobüs güzergahlarının planlanması için bir metodoloji geliştirmektir. Bu çalışmada, besleyici otobüs güzergah ağı tasarımının önemli yönleri vurgulanmış, potansiyel besleyici otobüs durak konumunun belirlenmesi için bir yöntem geliştirilmiş ve FBRNDP'yi çözmek için değiştirilmiş bir genetik algoritma kullanılmıştır. Sonuç olarak, yeni bir besleyici otobüs güzergahı tasarlamak için metodoloji kullanılmış ve pareto verimliliği ve minimum toplam sistem maliyeti ile en iyi çözüm seçilmiştir.

Anahtar kelimeler: Besleyici otobüs güzergahları, akıllı kart verileri, çoklu gezgin satıcı problemi, genetik algoritma

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE

INTRODUCTION

In urban areas, transportation experts consider public transportation (PT) as a viable option in the sphere of sustainable transportation. Broadly speaking, they offer positives such as mobility enhancement, reduction in traffic congestion, energy, pollution, and to some extent provide social equity. However, in the past decades, factors such as socioeconomic growth has led to the need for a personalized mobility, and consequently a decrease in PT's share in daily commuting. Transportation has become a complex problem because it has different dimensions such as policymaking, planning, designing, infrastructure development and management. To simplify real life problems, models are used. Transportation modelling in its basic form are mathematical equations, which represents an actual transport system to predict travel behaviours/flows between origins and destinations of different modes of transportation in a geographic space (Improvements, 2011). Often, complications evolve from stakeholders' conflicting objectives, integration of different operational performances of modes of transport, data sources, mathematical techniques and computational power of the computers, etc. The aim these models are to guide in decision-making process especially in terms cost of effectiveness, appropriateness of the intended projects, and deliverance of an efficient system. If PT mode is to compete with private modes, then, adequate attention in planning, designing and management are necessary. The efforts for encouraging PT focuses on improving aspects like; capacity of the line, frequency of the services provided, service coverage, comfort and convenience of the service. Most of these aspects are important for the efficient and effective running of the PT system. The PT operation planning process has the following fundamental components often in a sequence.

1. Transit route network route design,
2. Service frequency determination,
3. Timetable determination,
4. Scheduling of vehicles and
5. Crew

Usually each of these phases are assumed to be separate problems, although, the result of one component in the previous phase becomes an input for the subsequent phase. Figure 1.1 below represent a transit network process by Ceder (2001).

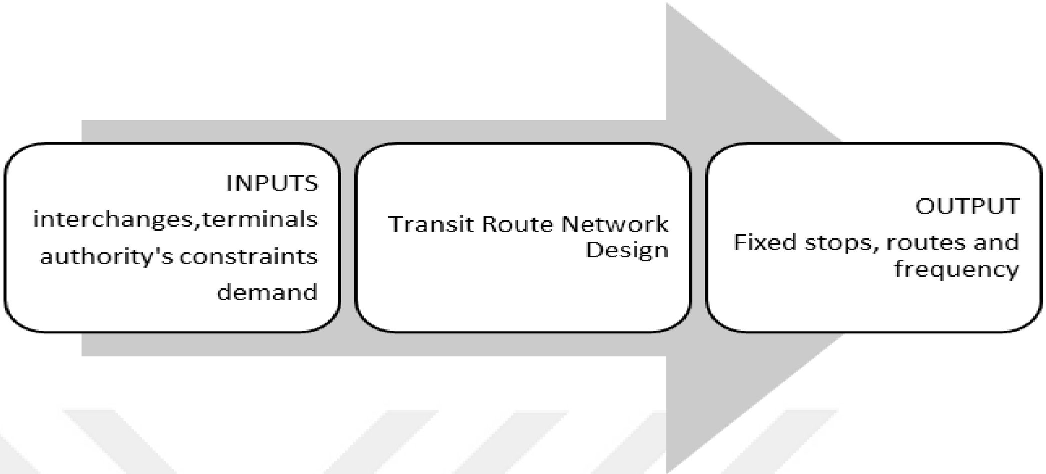

Figure 1.1 Schematic for TRNDP

The problem that formally describes the design of such a PT network is the Transit Route Network Design Problem (TRNDP). It involves the designing of PT networks based on a number of objectives that represents its efficiency under limiting factors such as length of transit routes, service frequencies, available resources (number of buses).

The basic functions of network route design are as follows:

1. To the locate the positions of the origins and destinations of the users
2. To define a relationship between where trips originate and terminate
3. To establish service schedules,location of vehicles and transfer points, and
4. To determine the optimal route paths.

As listed above TRNDP is primarily concerned with transit layout configuration which is beneficial to both user and operatos alike. Once that is done over the street network, the optimal frequency setting on these routes can be carried out and subsequently, the other aspects can be estimated. The optimality of transit network design problem solution will depend how these sub problems are tackled especially transit route network problem and frequency problems with others largely dependent on the first two. It is a well-documented norm to take the approach of sequential modelling that is to take the frequency setting stage after route generation (Baaj & Mahmassani, 1995; Mandl, 1980; Pattnaik et al., 1998). The argument for this is hinged upon strategic nature of route design, which usually has a validity of up to 15 years while frequency setting may change according to time (peak or offpeak), day (weekday or weekend) and so on. Therefore, the frequency setting should not have such impact on a strategic stage.

Specific to conventional bus routes is Bus Route Network Design Problem (BRNDP), which probably is the first and most important step in the planning of bus transit process as clearly highlighted by (Ceder & Wilson, 1986). Route structure once set, will form the part of input to be used in the subsequent planning processes and also bus operators will prefer not alter the route structure because new designs are usually capital intensive. Another type exists for multimodal/intermodal transit system which is the Feeder Bus Route Network Design Problem (FBRNDP).

A look at Figure 1.2 below consist of stops or demand locations for passengers in a transit network. BRNDP by implication is way of connecting these demand locations in such a way that, most of the passengers if not all can connect at least from one demand location to another, while ensuring that the objective function is optimized taking into consideration the perspectives imposed by the planners and operators. Modern metropolitan areas usually have a high transit demand, which is widely spread, across the entire city. This requires harnessing the advantages of different types of mass transportation systems or different bus services. With multiple modes of transportation or varying bus services, comes the challenge of integrating the various operations to develop a sustainable transit system that is more cost-effective and efficient. Mainline (rapid transit) often has a considerably larger capacity and relatively higher speeds, thus, it can function as a major transport corridor but may create problems of accessibility especially to residential areas where the demand normally originates. In the same way, a feeder bus system with lower capacity and speed can provide access services closer to the residential demand. Therefore, a simple form of intermodal transit system may consist of an integrated mainline movement which may be a rapid transit line and a lesser transit system called a feeder (bus) connecting the service areas (residential) to the transfer stations where their journey will continue on the mainline. The service coverage of the rail systems is expected to expand and an improvement in the utilization of different public transportation modes overall. Some of the main advantages of an integrated transit system include; reducing costs and increasing revenues, eliminating duplication of services, reduced travel times and access costs, and consequently a better overall quality of service of the system (Kuan et al., 2004). Therefore, the planning and designing of bus routes that will provide access between residential areas to a train station represents a FBRNDP. See the Figure 1.2 below for apt description of both cases.

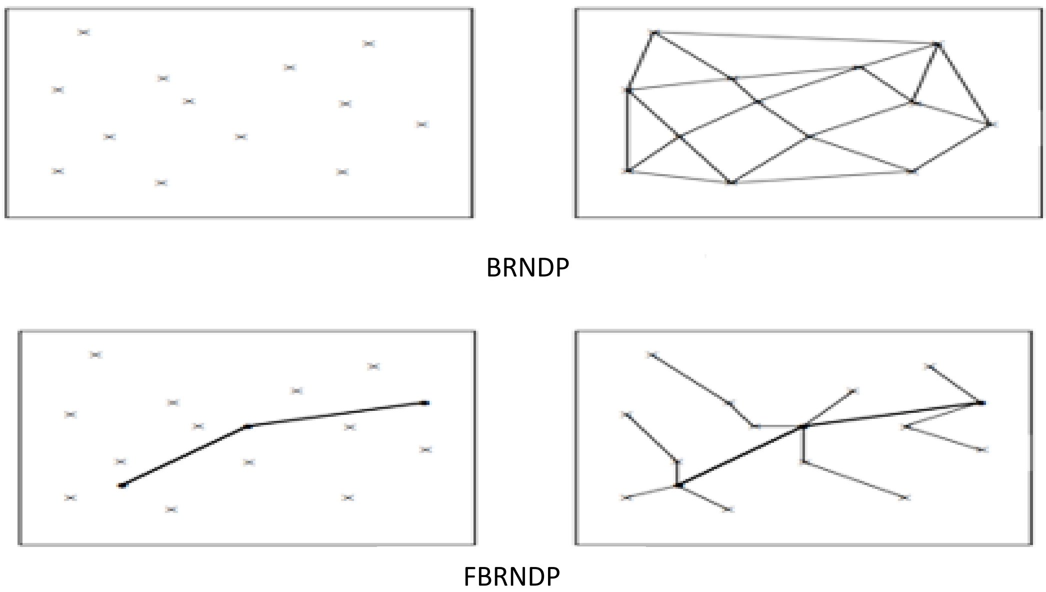

Figure 1.2 FBRNDP vs. BRNDP

1.1 Problem Statement

When planning a small bus route network, a planner's knowledge and expertise may be sufficient. This he/she can do by following simple guidelines, conducting surveys and analysis. However, for cosmopolitan areas with large bus route networks, huge data sources, the simple guidelines, and planner's expertise may be inadequate in providing a near optimal solution. Transit route design is dependent of demand at the stops which are generated from residential areas (origin) and attracted to activity related centres (destination) thus representing the demand for transit travel. One of the fundamental components of transit planning is the estimation of origin-destination (O-D) matrix. Collecting such OD data, however, is extremely difficult and tedious using traditional paper-survey-based approaches and rarely available. Generally, most transit smart cards (TSC) systems are designed for fare collection and management purposes and not OD data collection; hence, they require more analysis and data mining techniques to be carried out on them before meaningful data can be extracted for transportation planning purposes. They have been used to estimate potential travel origins and destinations and planning purposes, but the data collected shows only the present passenger demand and it is also dependent on existing routes and stops. Spatio temporal data are most important for any meaningful transit smart card studies, even so, they must be used alongside other data sources such data base management system and geographic information systems for its potential to be fully utilized. Because of the inherent disadvantage found in entry-only systems, researchers devoted most of their work in the prediction of destination of alighting passengers.

Feeder bus service is a secondary transportation services (e.g. tram lines, bus lines) connecting a primary transportation service (e.g. a rail line or metro line). It is used to solve the last mile problem and to increase the service area of the primary network. It is a useful way of serving train stations with high recurrent commuting traffic pattern. In designing this type of system will require reliable existing and potential passenger demand in the service area or at the stops within service area. Another aspect that is interesting is the resemblance FBRNDP to a known routing problem which is the multiple travelling salesman problem (MTSP), because they both belong to a group of NP-hard combinatorial optimization and they both involve designing a set of minimum cost routes and connecting them to a given location. FBRNDP is a large routing problem that has been solved satisfactorily by a heuristic, metaheuristic and sometimes a combination of solution approaches (Kuah & Perl, 1989; Kuan et al., 2006). Therefore, it will be interesting to come up with a strategy that combines TSC,GIS, database management system and using MTSP formulation and metaheuristics to simplify the ease of designing feeder bus routes, thus, the title of the research "Designing feeder bus routes by using smart card data".

Depending on the purpose, transit systems are classified into main lines and feeder lines. These multi modal systems or intermodal systems are often designed independently and can lead to increase in vehicle travel times, duplicated lines, and complicated fare structures etc. making the PTs less attractive especially from dissatisfied users. Improving this kind of systems can be very complex. It may be of benefit especially to passenger's comfort and convenience and reduction of both user and operating costs if these systems are well designed. Although there are many researches in this regards, there is a need to provide a background information on the systematic and integrated approaches for designing an efficient FBRNDP with the aid of big data analytics, new metaheuristics formulations, and a route selection process for multi perspective stakeholders. Because of the ensuing challenges and continuous developments in the fields of mathematics and computers analysis, these areas create opportunities for further research.

1.2 Aims and Objectives

The main purpose of this research is to develop a methodology for the planning of public feeder bus routes for predefined destination by using smart card data source for origin destination matrix. These routes should not be overtly biased towards any stakeholder, thereby minimizing the total cost of the feeder bus routes. To meet up with the aforementioned aim, the objectives are as follows;

1. To provide an understanding on current state of practice of feeder bus route design taking into cognisance the available data sources, solution methodologies. Hence as an output is a conceptual framework will be developed for feeder bus route network design problem.
2. To highlight the potential of using smart card data for feeder bus route configuration. The output of this objective is to carry exploratory data analysis on smart card data for use in feeder bus route configuration and mining of data from smartcard i.e. to obtain OD matrix. To develop a methodology for identifying potential feeder bus stops and their demands from smart card data and existing bus stop coordinates.

3. To develop a strategy for optimal feeder bus route construction that minimizes the total system cost. Investigation of tour construction technique; Fixed start Multiple travelling salesman genetic algorithm by Joseph, 2020 in the design of feeder bus routes.
4. To apply the methodologies so developed on some practical examples;
    a. Application on benchmark study
    b. Application on existing feeder bus route network
    c. Design of feeder bus routes using smart card data

1.3 Significance of Study

The solution to FBRNDP is driven by different and mostly contradicting objectives. This is largely due to the different stakeholders coexisting in the transportation planning space. Worthy of note is the fact that transportation requires a multi-dimensional and multi perspective and a typical example of objective is the social welfare which appears to be the most commonly used, usually interpreted as the minimization of total cost comprising mainly of user and operator’s costs. Operators in general, in maintaining an operationally efficient transportation system try to reduce the overall cost of operation in the form of route length, fleet size and consequently minimizing the size of the crew. Also rather than using conventional surveys that are unreliable, biased, capital and labour intensive, the resurgence of big data in the form of mobile phone records, transit smart card data will offer a relief and access to huge volumes of data often required for continuous re organization, evaluation and planning of feeder bus routes as time changes.

This study presents a new methodology for designing an efficient feeder bus services taking into cognizance smart card data, GIS data utilization for OD matrix estimation, and using conflicting objectives of the user and operator’s perspectives in the selection of best feeder route structures. Also it will provide a state of the art literature review on application of smart card data in the estimation origin and destination matrices for feeder bus services as well framework for feeder bus route network planning. It will also serve as an efficient and easy to understand way of planning feeder bus routes especially for planners.

1.4 Thesis Breakdown

Attempts have been made of over decades to design a more efficient and yet sustainable transportation system by integrating the independent operations of multi modal transportation through the coordination between main trunk lines and feeder services. The FBRNDP has many potentials and researches are still ongoing due to many reasons such as; new policies by operators, level of details (data collection and modelling) and the development of more sophisticated methods and computing power. These reasons create new requirements, challenges and potentials for planners and researchers. The FBRNDP is a large and complex routing problem that can be only solved satisfactorily using heuristic, metaheuristic and sometimes a combination of hybrid approaches. This work has three parts, see the Figure 1.3 below:

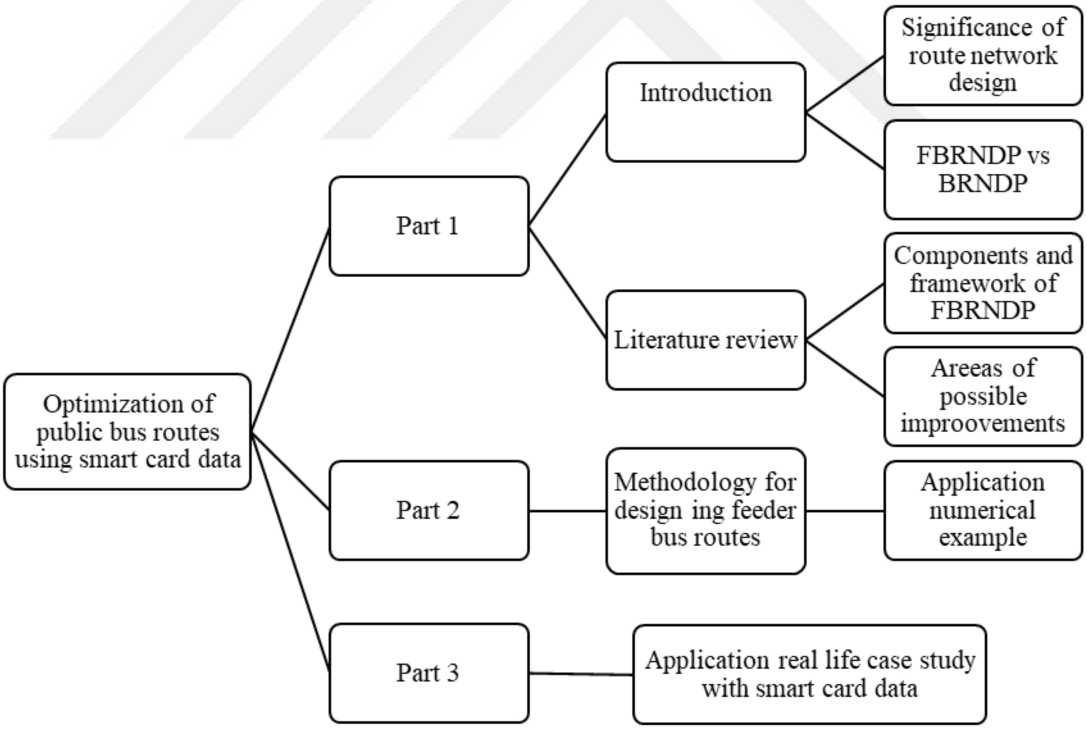

Figure 1.3 Thesis break down

CHAPTER TWO
FEEDER BUS ROUTE NETWORK DESIGN PROBLEM

2.1 Introduction

According to Lopez-Carreiro & Monzon (2018), “Smart Urban Mobility” can be defined as “connectivity in towns and cities that is affordable, effective, attractive and sustainable”. Therefore, developing a sustainable transport system is an integral part of modern day living. To avoid detrimental effects of urban transit systems, public transportation systems should be prioritized at early stages of planning to make it sustainable in a rapidly developing infrastructures of cities. Mass transportation systems like rail transit are always considered as an alternative to road transit. These transit systems always operate on main corridors which are less accessible to residential locations. The provision of feeders for main transit corridor is a sustainable way for solving problems associated with accessibility, limiting the problem of parking problems at the train station (park and ride system), and expansion of service area of the main corridor. With multiple transportation systems, comes the challenge of integrating the various operations to develop a sustainable transit system that is more cost-effective and efficient. Mainline (rapid transit) often has a considerably larger capacity and relatively higher speeds, thus, it can function as a major transport corridor but may create problems of accessibility especially to residential areas where the demand normally originates. In the same way, a feeder bus system with lower capacity and speed can provide access services closer to the residential demand. Therefore, a simple form of intermodal transit system may consist of an integrated mainline movement which may be a rapid transit line and a lesser transit system called a feeder (bus) connecting the service areas (residential or stops closer to residential areas) to the transfer stations where their journey will continue on the mainline. The service coverage of the rail systems is expected to expand and an improvement in the utilization of different public transportation modes overall. Some of the main advantages of an integrated transit system include; reducing costs and increasing revenues, eliminating duplication of services, reduced travel times and access costs, and consequently a better overall quality of service of the system (Kuan et.al.,2004).

Therefore, the planning and design of a set of connecting bus routes for the provision of access between residential areas to a train station can be defined as FBRNDP, in other words, it is the determination of feeder-bus routes consisting of stations, route structures and the operating frequency (Kuah & Perl, 1989). An example of this scenario is common with rush hour to work trips to the city centres in the mornings. In modelling terms, most passengers can be assumed to go to a common place (central city station or central business district). So, passengers aggregated at bus stops in the service area who wish to connect to CBD will do so by using a lesser mode say a bus to connect to any of the train stations going to the city centre.

Conventionally, in transit operation planning involves; designing a network of routes and stops, definition of frequency and headways, timetable development, scheduling of vehicles and drivers . These processes can be viewed from long, medium and short term perspectives. Strategic perspectives are mainly long term and is related the transit infrastructures themselves, in this case, setting of route network falls under this category. While tactical perspectives will be based on the efficient utilization of the infrastructure which is certainly related to frequency and headway of the transit networks which is viewed as mid-term goals. Finally, the time table and scheduling practices are categorized under operational decisions since they can be reviewed and changed as frequently as necessary (Farahani et al., 2013; Guihaire & Hao, 2008). Transit route network design is an important problem in overall transit network design, because the cost of the PT system depends on it. Once a transit network is defined, a set of routes is identified over a street network and subsequently all decisions about timetable development, bus and driver scheduling are considered. Although bulk of the cost of operating transit systems is based upon drivers' wages and fringe benefits (Ceder & Wilson, 1986), making the last steps most important for the operators, its however based on route structure and frequency settings that all other procedures are designed, making the whole process sequential in nature.

Similarly, feeder bus route design is a simple form of transit route network design problem primarily concerned with transporting passengers from bus stops of a lesser transit mode to other stations on main corridor movement. To model this problem, requires nodes; variable nodes in the form of bus stops and fixed nodes in the form of train stations. To solve this problem variables such as train station and bus stop positions, hourly demand, cost parameters are predefined, thus, the problem will only require a connection between these nodes which optimizes the perspectives of some stakeholders depending who is planning the system. The operators is more concerned with revenue generation which is directly related to number of passengers and total route length of the network, these parameters might limited by other factors such as fleet size. On the other hand, the cost incurred by users are more towards travel time including the amount of time spent waiting and in vehicle.

Another important aspect to be considered when modelling feeder bus network is the way the system is represented mathematically. In general terms, network in transportation planning are represented as a well-knit structure comprising of roads, intersections used in modelling transportation problems (Guihaire & Hao, 2008). There are two ways of representing transit networks, the first type being zone type where zone demands are collected at its centroid and the second, is by considering the node itself as potential origin or destination. It is not wise to consider all these nodes as bus stops but rather a methodology should be constituted for the selection of the potential bus stops according some certain criteria such as distance and demand at the node.

Distance, travel time and demand matrices is the main data in transit route network design. It represents the demand for travel between the nodes and its interaction with overall network. For distance and travel time estimations, geographic information system (GIS) can be used along with numerous distance estimation methods (Euclidean distance). Demand is perhaps the main reason for transit network design, therefore, for a better solution, a more precise and accurate origin to destination matrix is pertinent. Fixed demand is usually considered for the simplicity of the system and ease of modelling, thus, most researches tow this line (Van Nes et al., 1988). For estimation of demand, traditional four step of transportation forecasting model may be used which is largely dependent on factors such as land use, socioeconomic factors, demography etc. This approach may be suitable when designing from scratch and most certainly difficult. However, a simpler approach is to count at existing demand at the bus stops which may be suitable for an already existing system which only needs improvements. Other aspects such as vehicle parameters and characteristics, constraints such as fleet size are also important especially in the determination frequency and operational performance of the system.

There are many factors that affects feeder bus route network design problem but depending on the nature of the problem, study area, expected target amidst of limitations will shape the way the problem is solved (Almasi et al., 2014). In this regards, we highlighted important aspects of feeder bus route network design as discussed by relevant literature. Therefore, the main components of FBRNDP will be discussed in the sub sections below.

2.2 Components of FBRNDP

This section describes the various components necessary for the design of FBRNDP. It is basically divided into 4 different components, each closely related to one another. These components are described as can be found in relevant literatures in the following sub sections. The Figure 2.1 below expresses how FBRNDP can be grouped into various parts for the design and improvement of feeder bus route network.

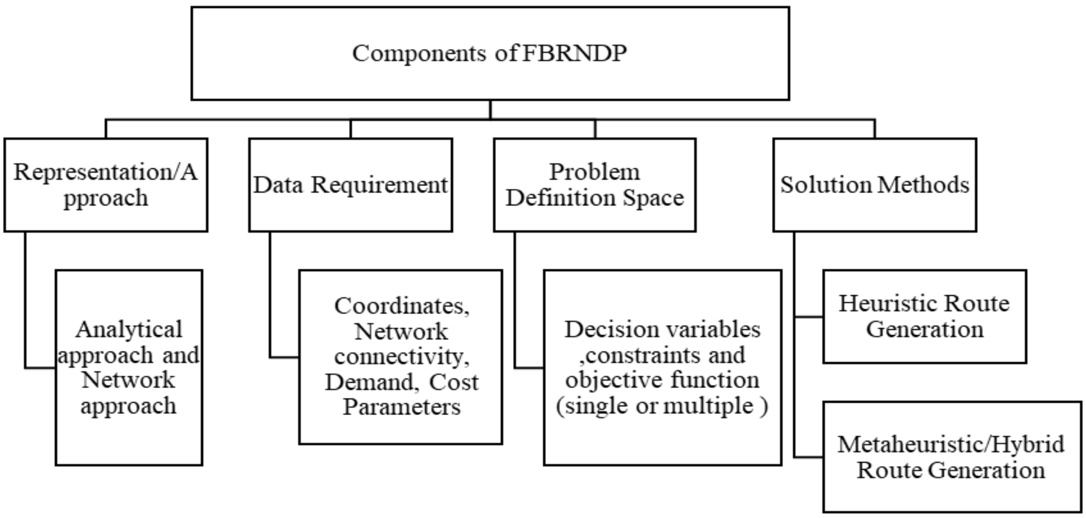

Figure 2.1 The components of FBRNDP

2.2.1 Representation of FBRNDP

In TNDP in general there is need for the representation of network problem in mathematical terms so as to solve them. A transit network is mainly comprised of routes and stops which is represented by a graph with interconnecting nodes and arcs. This representation normally varies from one planner to the other because of the data availability, level of details used, and method of analysis. According to (Owais, 2015), street network representation can be classified into two basic levels. The first one aggregates demand at the zonal level and the second one, identifies potential demand at each bus stop i.e. node level. Although, it may not be feasible to consider all bus stops, so consideration is given according to nodal demand and perhaps limits of walking distance from those nodes. It is worthy of note that, the representation of transit network differs from original road networks which they are derived from. As stated earlier, not all stops nor links can be fully represented and therefore, there should a distinction between road networks and transit networks. Street networks refers to all roads and intersections in the study area while transit networks include transit route and stops or stations. Considering the FBRNDP, the transportation network is built around a trunk line that forms the back bone of the network. This is usually a primary transportation service (e.g. a rail line or metro line or metro bus etc.) and to increase the service area, secondary transportation services (e.g. tram lines, bus lines, paratransit etc.) are added. Figure 2.2 shows the different feeder lines; evolving from linear minimal networks as a main line to addition of feeders and to addition detours to form Maeander networks. Others are demand responsive and flexible feeders. To increase the efficiency of a feeder bus system, there is need to consider the opposing perspectives of the user, operators, planners and even non users. While the users' perspective of a good feeder system is more of coverage and access in the service area, low access cost and so on, the operators' perspectives are to lower operation costs usually by keeping total route length within certain limits. Other perspectives are the non- users or the environment which these systems may affect. To put into perspective, FBRNDP as whole according to literatures (Byrne & Vuchic, 1972; Kuah & Perl, 1988; Kuah & Perl, 1989) is said to be composed mainly of feeder bus routes determination (stations and route structure) and operating frequency.

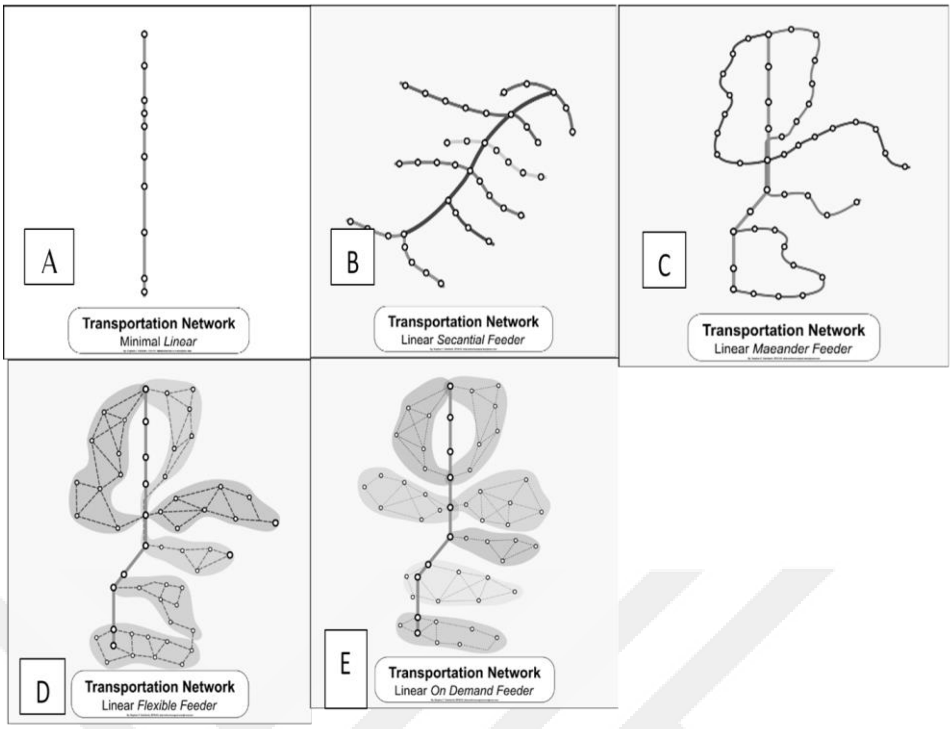

Figure 2.2 Feeder services around the main line using different feeder systems

2.2.2 Approaches to FBRNDP

According to Almasi et al.,(2014), there are two main approaches when trying to model FBRNDP as considered in literatures namely; analytical and network approaches. The analytical approaches use actual road networks therefore, the shape and geometry of the road becomes a very important factor. Also, the demand in the service area should be adequately represented in the demand function. The objective function is defined using a set of continuous design variables such as feeder bus route positions, main line station spacing and locations, and service frequencies. This implies that the optimal relationships between various components of feeder bus network problem can be found using extreme conditions on the objective function. So many researchers employ this analytical approaches and some of them are; (Chien et al., 2001; Chien & Schonfeld, 1998; Chowdhury & I-Jy Chien, 2002; Kuah & Perl, 1988).The disadvantage of analytic approaches is that they are only able to handle small and regular size networks. As the number of streets in the network increases so does the number of possible solutions increases which makes the search for optimal solution from all possible solution critical and thus, analytical approaches are mainly used for theoretical purposes (Kuan et al., 2004). Network approaches on the other hand, tends to avoid the complexities of analytical approaches by considering the stops or stations and links as the service area. Transit routes are represented by series of connected nodes while transit links can be represented by travel times or travel distance between these connected nodes. As expressed in the previous section, demand can be taken at the nodes or aggregated in zone centroids and the entire network can be represented as the number of trips between all pairs of nodes as origin – destination matrix. This approach can handle much larger size, irregular and real networks than the analytical approach. The network approaches have been used by many researchers (Kuah & Perl, 1989; Kuan et al., 2004, 2006; Martins & Pato, 1998; Mohaymany & Gholami, 2010; Shrivastav & Dhingra, 2001). See Figure 2.3 which shows the representation of transit networks for network approaches.

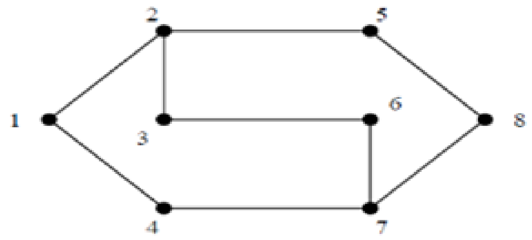

Figure 2.3 Representation of transit network

2.2.3 Data Requirement

Different objectives of FBRNDP require different data: be it data which provides the demography, land use properties, socio-economic, surveys/smart card data extractions etc. These together can be used to obtain details such as behavioural pattern of travel, and the distribution of trips amongst traffic zones in the study area can help in identifying potential zones to be served by regular feeder service. Kuah &Perl, (1989), used four basic data for his analysis which includes; coordinates of station and stop location, demand and stop density in the service area and also operating cost of both rail and bus transit which was derived from literatures. Kuan (2004), also uses the same bench mark study and same data but introduces a second set of random data consisting of problem size and structure which was used to compare performance of the different metaheuristics used. While on the other hand, other researchers like Shrivastava and O'Mahony (2007) uses traffic surveys with coded networks to create potential O-D matrix which was then used to design the feeder bus route. Tabassuma, (2016) uses three sources of data; to include demography of the study area, socio economic and land use characteristics so as to identify areas served by the feeder services and to obtain travel patterns. The researcher also used field surveys to obtain current travel patterns for main trunk users and assessment of current feeder modes used to and from the main trunk line. In order to design a feeder bus route network with a main line, since the main is fixed and stop locations are fixed, a list of nodes with their coordinates, a list of links connecting these nodes, demand matrix based on travel behaviour questionnaires or smart card data extractions and cost parameters which are based on ridership and financial reports. As stated earlier the type of data requirements depends on the level of details required and type of objective to be met.

2.3 Problem Definition Space

This section constitutes all the complexities associated with trying to model a seemingly real like situation in FBRNDP. It constitutes section that describes the demand pattern and characteristics, decision variables and constraints, stake holder's perspectives and objective functions.

2.3.1 Characteristics and Pattern of Demand

Demand can be classified based on its characteristics as "fixed-in elastic" and "variable-elastic"; this classification implies the effect of demand on the performance and services provided by a PT networks. By inelastic we mean demand is not changing with performance or quality of service and vice versa for elastic. The elastic demand is a closer model to the real world and more objective according literatures (Lee & Vuchic, 2005). However, for simplification purposes many researchers tend to use fixed demand (Fan & Machemehl, 2004). Moreover, for most developing cities, the issue of captive users is prevalent and the effect of variable demand is most likely minimized. That is to say that demand may be fixed for systems where passengers' preferences do not change with service quality or price of the service. However, variable demand can come to the fore when sharing or intense competition amongst the different modes of public transport, and an increasing demand for mobility. These factors are important in the modelling of urban transportation especially where different modes of transportation are present (Jakimavičius & Burinskiene, 2009).

Travel demand pattern is directly related to urban structure and spatial distribution of human activities and the decision between many-to-many (M-to-M) and many-to-one (M-to-1) is one based on modelling approaches and prevalent conditions. An example is a case of public transportation networks which are designed towards commuting trips to and from the centres of activities i.e. many trip origins going to a single destination. It is expected that M-to-1 may be more suitable in this case. However, M-to-M patterns may be most useful in situations where transit services are designed to serve study areas with many important activity centres and or passengers having with varying trip purposes. For FBRNDP both types of travel demand patterns, are relevant as mentioned above depending on the circumstance warranting their usage. The M-to-1 demand pattern is presented in many publications (Kuah & Perl, 1988, 1989; Chien & Schonfeld, 1998; Chien & Yang, 2000; Kuan, 2004; Kuan et al., 2004, 2006) etc. It is often than not more related to FBRNDP, which transports passengers to and fro a common destination e.g. central business district (CBD) or a transfer station). The M-to-M differs from the M-to-1 for FBNDP in that, the set of destinations includes the entire set of rail stations. Here, the demand at each bus stop is a multidimensional quantity and often, not simply a sum of M-to-1 FBNDPs, because under it, a single feeder-bus route usually serves demands to multiple destinations making it difficult. First, the design of the feeder-bus network should take into account not only the linking to alternative rail stations, but also alternative connections to rail lines. Depending on which rail line is chosen for connection, passengers may or may not have to transfer between rail lines. Second, the optimal feeder-bus network may include some bus stops on more than a single feeder-bus route. This results in a significantly more complex feeder-bus network.

2.3.2 Stakeholders' Perspectives and Objective Function

Public transportation systems often play a social role by attempting to lower operating cost as much as possible and also making mobility more accessibly for the community in general in an equitable and efficient manner. The purpose for designing operations of a PT system should capture aspects that are related to the stake holders. The existing literature focuses on both the effectiveness of the services provided and economic efficiency of the transit system being designed. These perspectives are opposite, since limiting the cost of operating the services for economic benefits may affect the quality of services being provided. Defining objectives, constraints, decision variables forms the basic structure of any TNDP complexity; this determines the problem space and consequently the type of solution method that can be employed. Determining objectives may come under the following factors such as, political, social, environmental and economic factors either as a single or combined (multiple) objectives. Transit agencies are normally responsible for the choice of factors to be considered based on the importance of these factors as it relates to their goals and taking into cognizance other relevant stake holders.

The design of FBRNDP is also driven by different and mostly contradicting objectives. This is largely due to the different stakeholders coexisting in the transportation planning space. Worthy of note is the fact that transportation requires a multi-dimensional and multi perspective and a typical example of objective is the social welfare which appears to be the most commonly used, usually interpreted as the minimization of total cost comprising mainly of user and operator's costs. Operators in general, in maintaining an operationally efficient transportation system try to reduce the overall cost of operation in the form of route length, fleet size and consequently minimizing the size of the crew.

A study by Kuah & Perl (1989), minimizes the total bus operating costs by optimizing routing structures and operating frequency. Baaj & Mahmassani (1991), in their publication stated that TNDP is inherently a multi objective problem. This allow planners, operators, users and even non users to interweave together to form a complex mix that is a semblance of reality (Baaj & Mahmassani, 1991). User costs for FBRNDP may include the in vehicle time, out of vehicle times (waiting time, transfer time), fare for both feeder buses and main trunk line vehicles. Figure 2.4 shows different stake holders perspectives taken from lin & Wong (2014). They further grouped these attributes into four distinct factors. Each of these factor contains attributes that are considered for different stakeholders. Since not all attributes can be used in a model, the route planners may support combined attributes of bus users and operators summarized in the form of service, cost, and revenue. It is always a neglected fact that new transit routes can have a negative impact on other road users (e.g., reduction in traffic safety and road capacity) which have seldom been addressed in the planning and design of transit routes.

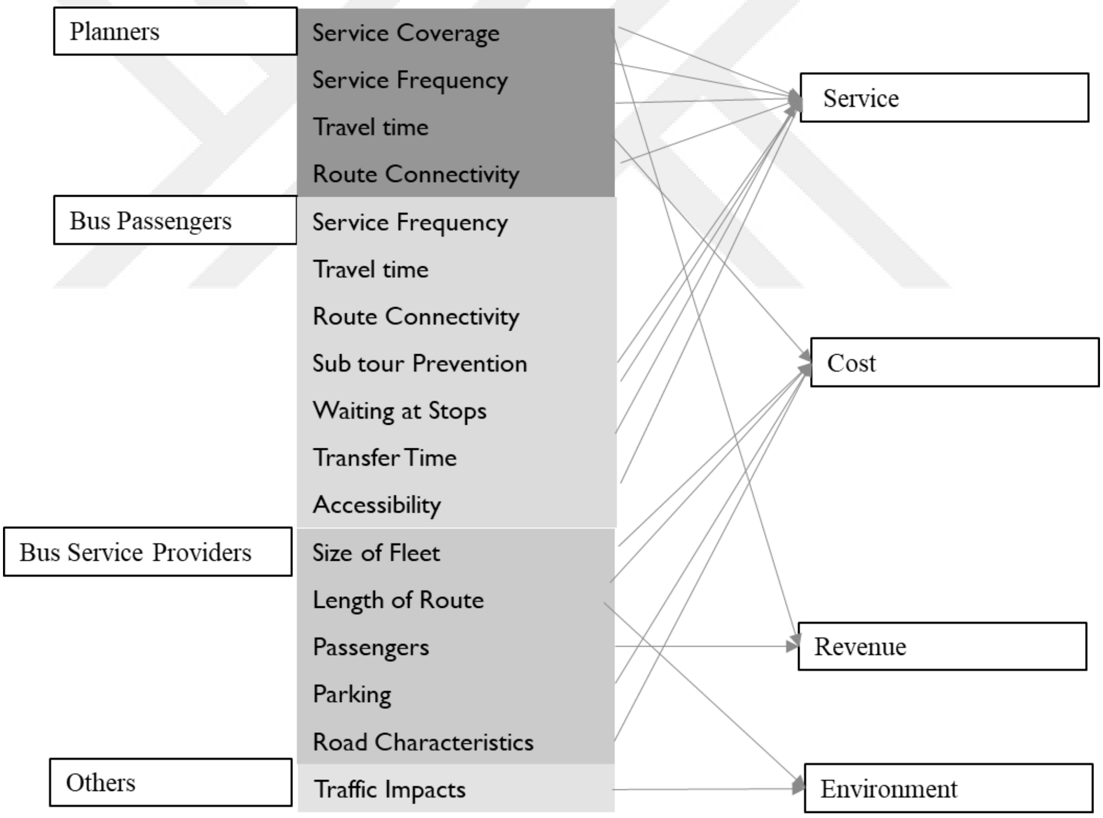

Figure 2.4 Factors considered for FBRNDP (Lin & Wong, 2014)

2.4 Solution Methods

Methodologies in solving transit route network design problem are mainly influenced by; level of details in design environment, quality of anticipated solution and available computational power.

TRNDP is usually partitioned in a sequence of procedures so as to be manageable. Two major approaches;

1. Route generation and configuration;
2. Route construction and improvement

2.4.1 Heuristics used in Initial Route Generation Methods

The first approach, are usually heuristic; they are based on experiences used in the past for solving similar problem which can be employed to reduce the size of the solution search space but in general do not guarantee an optimal solution but a practical solution. They are widely used in transit planning since they utilize planners' knowledge but may not be transferable to other systems. It involves finding a set of candidate routes considered as initial solution for route design stage i.e. shortest-path-based algorithms are used to generate some candidate routes under certain constraints. These constraints may include the maximum/minimum number of routes, length of route, travel time limit, etc. Because of the flexibility and practicability of heuristic methods, it is a common approach among literatures. Therefore, FBRNDP can be solved by applying heuristic algorithms to reduces the size of the large solution search space, by building initial routes followed by improvement of these initial routes using optimization technique. Kuah & Perl, 1989 used sequential building heuristic for building initial solutions which is an adoption from sequential saving approach for Multi-Depot Vehicle Routing Problem (MDVRP). In an expanded study (Martins & Pato, 1998) they generated the initial solution from two-phase building method by applying the sequential savings heuristics. The proposed Heuristic Feeder Route Generation Algorithm (Shrivastava & Dhingra, 2001) based on demand matrix of (Baaj & Mahmassani, 1995). Also metaheuristics like Genetic algorithm (GA) can be used to generate initial population or initial population can also be generated at random as suggested by (Chien et al., 2001) but it might not be a good selection for generating initial routes. Therefore, (Kuan et al., 2004) employed the concept of delimiter, proposed by (Van Breedam, 2001). Most of the studies follows this approach of generating initial routes using various kinds of heuristics and then try to improve the initial solution by using the initial solution as an input in some optimization techniques to obtain a better solution, improvements can be implemented on the routes. There are a lot of optimization methods used to improve the solutions.

2.4.2 Metaheuristics/Hybrid in Direct Route Construction and Improvement

An alternative to initial route generation and subsequent improvement is direct route construction by using some metaheuristic and hybrid methods. Meta-heuristic methods unlike heuristic approaches are able to generate local optimum solutions to combinatorial optimization problems where FBRNDP belongs. Most common examples used by researchers are; Ant colony optimization (ACO), Exhaustive search(ES), Genetic algorithm(GA), Simulated annealing (SA), Tabu search (TS), (Chien et al., 2001; Kuah & Perl, 1989; Kuan et al., 2004, 2006; Martins & Pato, 1998; Shrivastava & O'Mahony, 2007) For FBRNDP, the initial sets of routes or solutions are developed and a further improvement is required for a better solution by using nature inspired and non-nature inspired algorithms. These algorithms have their strengths and weaknesses , a comparism between the algorithms was carried out by several authors (Chien et al., 2001; Kuan et al., 2004, 2006). Some results of the performance comparison indicated that the local optimum solutions derived using ES and GA are identical even though ES had a higher computational time than GA, especially for large or complicated networks. Neighbourhood search methods like SA and TS are good methods but defining the neighbourhood seems to be very complex and difficult as solutions might keep changing. Also, difficulties arise from GAs in finding suitable cross overs every time. ACO does not depend on neighbourhoods but depends on continuous iteration of previous solution to generate better solutions thus making it dependent on number of iterations. Having discussed the pros and cons of some methods, a composite of some of them have been tried in the pasts. Some studies use heuristic method to generate the potential routes and then subsequently an optimization techniques scheduling problems (Shrivastava and O'Mahony, 2007). In another research (Shrivastava & O'Mahony, 2009) developed a hybrid algorithm using similar theme of first designing the bus routes using GA and using the developed heuristics for coordination problem. In a more recent work (Ciaffi et al., 2012), also used a hybrid metaheuristics comprising of GA and a heuristic See Table 2.4 for some generation methods found in literatures.

Table 2.4 Important route generation and improvement methods for FBRNDP

References	**Methods**
Kuah & Perl(1989)	Heuristics(sequential savings)
Martins &Pato(1998)	Heuristics(sequential savings and two phase)
Shrivastav & Dhingra (2001)	Heuristic Feeder Route generation algorithm and Dijkstra's algorithm
Chien *et al.* (2001)	Metaheuristics (Genetic algorithm)
Kuan *et al.* (2004)	Delimiter algorithm and Breedam(2000)
Kuan *et al.* (2006)	Delimiter algorithm and Breedam(2000)
	K-path algorithm, Eppstein (1994)
Shrivastava & O'Mahony (2007)	K-path algorithm, Eppstein (1994)
Shrivastava & O'Mahony (2009)	K-path algorithm, Eppstein (1994)
Shrivastava & O'Mahony (2009)	Heuristics(Dijkstra's algorithm)
Mohaymany & Gholami (2010)	Metaheuristics(Ant Colony optimization)
Gholami & Mohaymany (2011)	Metaheuristics (Ant Colony optimization)
Shrivastava & O'Mahony (2007)	Hybrid-Genetic algorithm and heuristics
Shrivastava & O'Mahony (2009a)	Hybrid-Genetic algorithm, heuristics, SOHFRGA
Shrivastava & O'Mahony (2009b)	Hybrid-Genetic algorithm and heuristics
Ciaffi *et.al.,* (2012)	Hybrid-Genetic algorithm and heuristics

Hybrid methods holds a lot of promises in terms of tackling problems that used to be intractable and can exploit the advantages of different methods thereby opening the possibility of integration of potential solution methods that can be combined. The figure 4 below shows a proposed methodology that can be used in designing FBRNDP. This flow of work takes into account the different components already discussed in the previous sections and presents the different approaches; heuristic route design and subsequent optimization of the initial route generated and direct route construction and improvement using metaheuristics.

From these works however, we delve further to identify the major components of feeder bus route design to include; network representation and approaches, data requirements, problem definition space, and solution methods. Hence a theoretical framework was developed to give a clear information on the different aspects of FBRNDP. From the literature review which was presented as components of FBRNDP, the following can be inferred from them.

The accurate selection of route generation nodes will result in successful feeder routes and from that, demand data can be acquired to build a better demand matrix. A more detailed OD matrix at bus stop level might be possible with smartcard data. This is very important for a successful feeder route services where accurate selection of route generation nodes is pertinent. As discussed in earlier section, most studies rely on a M-to-1 pattern for simplification purposes even though most transit passengers in developed countries have varying origin and destination and most importantly this demand pattern depends on city planning.

Travel patterns, demand and capacity at the stations of main trunk line may change if a new feeder service is implemented. Aspects such as walking activities within transfer stations and capacity of the stations themselves can create serious concerns. Thus feasibility and post evaluation of the system might be necessary.

Considering the modelling aspects, the conflicting perspectives of stake holders and the fact that a sustainable transit system should be multimodal makes route generation problem a multi objective problem as suggested by researchers. The design of better metaheuristics can be explored as most of these solution methods have pros and cons. Hybrid methods also hold a lot of promises in terms of tackling problems that used to be intractable and can exploit the advantages of different methods thereby opening the possibility of integration of potential solution methods that can be combined.

The nature of problem in most researches, transit services are planned and designed for normal daily operations but aspects like disasters management, seasonal/mega events or even emergencies for hospitals, owl services etc. will change the nature of problem and will open a new frontier in urban transit planning. Thus making objectives a situation based and cannot be the same with the planning and design daily operating transit services. Also developing solution methods from previous models of daily operating transit systems and building new ones to fit the particular problem can be a good area of research.

Similarly, railways and feeder buses are the most researched even though different access modes or mode combinations exists. Modes like cycling, walking, can also be used to access main trunk lines station. Therefore, inclusion of different modes or mode combinations can be looked into. The consideration of these access modes can help model FBRNDP to be much closer to the real world.

Operators normally uses fare setting as a strategy for profit making while keeping in mind the total welfare in mind. This creates some sort of opposing objectives thereby leading to different fare strategies and products. Multi modal transportation also offers different variety of fare products depending on the characteristics of the mode and the intent of the planners. Thus integrating the different fare structures into one single one which is a necessary tool in modern transit system will be quite difficult to implement and model.

In this study, our focus is on the type of data used, that is utilization of smart card data to determine the feeder bus stops from existing bus stops and extract from the smart card system bus stops whose destination are towards a particular location in this case train station. This will further be used to plan feeder bus routes using genetic algorithm that has been formulated as a generalized travelling salesman problem. The following sections will discuss the proposed methodology of our study under subsections of smart card data for feeder bus stop location, genetic algorithm based MTSP.

CHAPTER THREE
MULTIPLE TRAVELING SALESMAN PROBLEM

3.1 Introduction

Transportation network can be described as a spatial structure which is used for vehicular movement through it, an example of which is rail and bus networks. To analyse the network, a mathematical graph theory is usually utilized which can be termed as a transportation network where the vertices are called nodes and the edges are called arcs. Routing problems are generally concerned with finding shortest path between two locations and constructing a complete tour among some locations in a network. Route construction problems can broadly be classified into two main categories; Travel Salesman Problem (TSP), Bus Routing Problem or Vehicle routing Problem (Eldrandy, et.al. 2008). TSP is a problem that entails a single salesman finding the shortest possible route to traverse through a known number of cities/nodes with the condition of visiting each node once and returning to the starting node. It finds application in wide range of areas but particularly in planning and logistics. Vehicle routing problem (VRP) is a problem of obtaining the maximum set of possible routes constrained by the number of available vehicles to deliver to a given set of customers to a particular destination, and it is composed of many variants such as capacitated VRP. But, when the vehicle capacity in this problem is assumed to be sufficiently large enough such that the vehicle capacity does not become a constraint, then the problem is the same as the MTSP. Similarly, MTSP is also a general form of the TSP in which multiple salesmen are allowed to visit a set of cities with a minimized cost, constrained by visitation of each city once, and by one salesman. A variation of this problem can be found in literature such as problem with multiple depots, fixed number of salesman, fixed cost, maximum or minimum distance of travel even though the main goal is to minimize the total travelling cost which is often formulated as integer programming (Jayasutha, R. and Zoraida, 2013).

Consider a graph Z= (D, F), where D is the set of n vertices, and F is the set of edges. Associated with each edge $(i, j) \in F$ is a cost (or distance) T_{ij}. Assuming the starting point is the first vertex and there are h salesmen at the vertex. The variable X_{ij} (takes the value 1 if edge (i,j) is included in a tour and X_{ij} takes the value 0 otherwise) can be defined for each edge (i,j) ∈ F. General formulation of MTSP is presented below (Arostegui Jr et al., 2006).

Minimize	$\sum_{ij\epsilon A} T_{ij}X_{ij}$	(3.1)
h salesmen leave vertex 1	$\sum_{j\in D:(1,j)\in F} X_{1,j} = h$	(3.2)
h salesmen return back to vertex 1	$\sum_{j\in D:(j,1)\in F} X_{j,1} = h$	(3.3)
one route enters each vertex	$\sum_{i\in D:(i,j)\in F} X_{1,j} = 1, \forall\, j \in D$	(3.4)
one route exits each vertex	$\sum_{j\in D:(i,j)\in F} X_{1,j} = 1, \forall\, j \in D$	(3.5)

Therefore, the mathematical formulations and solution approaches of the above-mentioned problems may be utilized for MTSP (Bektas, 2006). Various application of MTSP exist and the main application arises in real situation such as scheduling and routing problems (Matai et al., 2010).

The solutions to MTSP are broadly grouped into two; the heuristic/metaheuristic approaches and analytical approaches. The analytical approaches uses boundary conditions to ensure optimal solution. While this ensure best solutions found, its application is grossly dependent on the size of the problem (Laporte, 1992). Even though this problem appears simple conceptually, the optimal solution for larger size problems is often untenable and takes an un imaginable time to solve when using greedy search for solutions (Bektas, 2006). However, solutions deep rooted in experience or some metaheuristics algorithms can yield a near optimal solution in a reasonable amount of time regardless of the size of the problem. See the Figure 3.1 for a classification of route construction methods (Eldrandy, et.al, 2008).

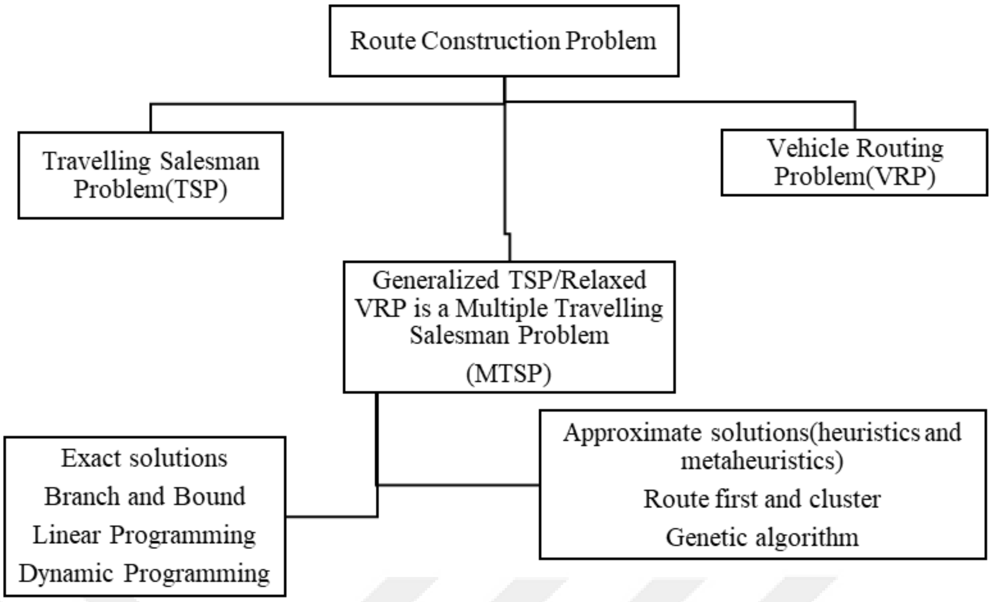

Figure 3.1 Route construction problem

3.2 Genetic Algorithm and MTSP Review

A simple heuristic method that solves this problem is the cluster based solutions which simply performs a single clustering of the nodes sets that are close geographically close. They are characterized by first ordering of customers based on their locations, ignoring the size of demand, and then grouping them into customers whose total demand does not exceed the capacity of the vehicle. One of its main advantage is that it obtains a practical solution especially when applied to real problems but its disadvantage is that it does generate the routes as it only processes the nodes . In 1974, Gillett and Miller have solved VRP Cluster-first, Route-second (Ghaziri, 1996).

In this study, we are particularly interested in using GA. They are search algorithms that copy the way populations evolve genetically through natural selection (Goldberg, 1989). This evolution is quite simple as it starts from a randomly generated sample space of strings, each string in the sample space is then evaluated through a fitness function, and consequently acted upon by GA operators which create a better population repeatedly through reproduction, crossovers, and mutation. The iteration of the algorithm may be stopped by using a termination criterion (e.g. number of iteration, crossover, mutation probability) because with a better population a better solution is expected. The research by authors (Arostegui Jr et al., 2006) states that, GA with limited computer power performs reasonably well when compared to other metaheuristics even though some of them yield better results. Some of its advantages are;

1. It directly searches from the potential solutions and the objective function themselves, not their derivatives which is used by exact approaches which makes it suitable for application in real-life problems like routing problems.
2. extensive algorithms have been developed using GA, even though they may not guarantee optimal solution but with manipulation of population, iterations, high crossover rate, low mutation rate, the probability of their solutions will tend towards an optimal solution.
3. Also, it is very popular especially in academic researches because of its implementation and its rigorous ability in solving practical engineering problems.

These reasons amongst others make GA a competitive algorithm in solving routing and NP-hard problems to which MTSP belongs. Because the computation time increases exponentially as the size of the problem is increased, it is therefore important to the heuristic, metaheuristic, or even hybrid optimization algorithms, such as GAs.

A research claimed that, their used in solving VRP has been on the increase considering publications written on the subject matter from 1993 to 2012 (Karakatič & Podgorelec, 2015). Perhaps the first work to deal with MTSP using GA is on team schedules (Zhang et al., 1999). A similar work used it for hot rolling scheduling and the algorithm was able to solve both TSP and MTSP (Tang et al., 2000). Many research done with regards to GA and MTSP mainly focus on the vehicle scheduling problem using the different variants such as vehicle capacities, time windows and fixed number salesman with no constraints on the length of the routes (Park, 2001).

Sometimes the solution search space becomes so large with redundant solutions (Carter & Ragsdale, 2006), therefore they developed an effective approach to cater to these problems using GA. They used a two part chromosomes for a combined crossover technique. Each part of the chromosomes functions independently and they processed by different crossover techniques. The first part uses the classic crossover technique (Goldberg, 1989), while the other part uses a different crossover operator (Chatterjee et al., 1996). With this introduction, the search space is reduced by limiting the diversity of the population thereby improving the computation time.

A proposal for a novel interpretable representation-based algorithm for MTSP using GA was presented by (Király & Abonyi, 2011). Their representation was specifically tailored to one depot one MTSP and the application of many chromosomes. Their results suggest that the algorithm allows for usage of heuristics and constraints. Their algorithm was highly sensitive to the number of iteration and the type of constraints used.

A hybrid algorithm termed GA2OPT was used on six benchmark studies . The algorithm comprises of two parts, the first part uses GA to solve the MTSP and using number of iteration as a stopping criteria and subsequently the 2-opt local algorithm was then implemented to improve the solution at the same stopping criteria Four out of the six problems' best known solutions were improved and the other two examples gave a solution close to the best known solutions (Sedighpour et al., 2012).

A research (Yuan et al., 2013) tried to improve the effectiveness and quality of solution of GA for MTSP by adopting a 2 part chromosome representation technique to minimize the size of the problem search space. This is because the redundancy in chromosome representation generates different alternative solutions, thereby increasing the search space. Their method was compared to methods with three distinct crossover techniques and it yielded a better solution.

Similarly, a modified GA was used to solve MTSP as presented by (Arya et al., 2014). They considered one chromosome technique and used order crossover. It involves choosing a random crossover point that divides the parent string into two substrings and it differs from conventional order cross over in that the right substrings are always chosen instead of randomizing several positions in the parent tour. This together with two mutations and a local search technique improves the solution at each iteration of the proposed algorithm especially for large sized problems.

Hussain et.al., 2017 like other researchers discussed above, suggested a crossover operator based on path representation to minimize the total distance and they found some improvement over the conventional crossover operators. A good review of the solving MTSP using GA can be found in (Bektas, 2006; Singh, 2016).

In summary, FBRNDP is a routing kind of problem which may be comparable to other routing problems like MTSP because they both belong to a group of NP-hard combinatorial optimization and they both involve designing a set of minimum cost routes and connecting them to a given location. While for small instances of these problems it may be solved exactly, a simple TSP solution with 30 cities can take an unimaginable time to solve for all possible routes to be evaluated. Therefore, a kind of intelligent algorithm which gives a good solution, but not necessarily optimal solution may be required. The FBRNDP falls under large routing formulations that can only be solved satisfactorily by a heuristic, metaheuristic (GA, SA, ACO) and sometimes a combination of solution approaches (Kuah & Perl, 1989; Kuan et al., 2006).Because of this similarity, it allows us to adopt an existing heuristics solution or a hybrid solutions used for the MTSP. It will be interesting to combine the clustering of nodes which are close together and subsequently use genetic algorithm to solve the FBRNDP. Most of these literature tries to improve the solution of MTSP by using a modified GA operators such as crossover and also these algorithms were applied on MTSP where the salesmen return to their origin . The FBRNDP problem that was considered did not consider return journey to the starting and 3 different cross overs will also be applied as stated in algorithm developed by Joseph, 2020. Details of this algorithm together with clustering technique used will be discussed in the following chapter: methodology.

CHAPTER FOUR
TRANSIT SMART CARD DATA

4.1 Introduction

Transit route design is dependent of demand at the stops which are generated from residential areas (origin) and attracted to activity related centres (destination) thus representing the demand for transit travel. One of the fundamental components of transit planning is the estimation of origin-destination (O-D) matrix. The traditional way of gathering such data is both both capital and labour intensive, because it is based manual count at boarding and alighting locations The traditional way a transit agency would obtain an OD matrix is based on occasional on-board passenger surveys and using various techniques to expand the survey results based on manual boarding and alighting counts at the stops. This makes them rare and even when they are conducted it may be for other purposes.

Transit agencies are interested and are using a card technology that can store, identify various kinds of data (transportation fares, and other individual data) as a viable user validation and payment option (Blythe, 2004). These technologies are essentially automatic fare collection systems or sometimes known as transit smart card systems(TSCs). They are relatively a new source of data and are currently considered as a good substitute for traditional survey data. TSCs contains spatial and temporal information, unique IDs for users, and other information with regards fare which can be collected for traditional on board surveys but with greater details and less bias.

Table 4.1 Comparism between traditional surveys and TSC

Trip Characteristics	Traditional Travel Survey	Transit Smart Card Data
Traveller	Traveller	Smart card
Purpose	Declared by respondent	Not recorded
Origin(time and location)	Declared by respondent	First boarding location
Destination(time and location)	Declared by respondent	Can be estimated for specific trips
Transfer	Declared by respondent	Recorded

Generally, most TSC systems are designed for fare collection and management purposes and not OD data collection (Trepanier & Chapleau, 2006); hence, they require more analysis and data mining techniques to be carried out on them before meaningful data can be extracted for transportation planning purposes. They have been used in many fields of research such travel pattern inferences (Kieu et al., 2015; Ma, 2013; Zhao et al., 2017), transit policy and performance assessment (Eom et al., 2015; Kim et al., 2011; Pau, 2014; Smart et al., 2009), and for transit planning (Audouin et al., 2015; Gschwender et al., 2016; Utsunomiya et al., 2006; Yap et al., 2017). Spatio temporal data are most important for any meaningful transit smart card studies, even so, they must be used alongside other data sources from data base management system and geographic information systems for its potential to be fully utilized (Bagchi & White, 2005, 2004; D. Li et al., 2011; Nohl, 2007; F. Zhang et al., 2015). There are basically two types TSCs; the entry - only system (only boarding information is recorded) and entry-exit systems (both boarding and alighting information are recorded). The entry-only systems records only boarding information and are by far most commonly used (Kieu et al., 2015; Trepanier & Chapleau, 2006) and only few of these systems records both boarding and alighting information like those found in Australia and Seoul (Barry et al., 2002; Kieu et al., 2015). Because of the inherent disadvantage found in entry-only systems, researchers devoted most of their work in the prediction of destination of alighting passengers (Alsger et al., 2016; Barry et al., 2002; Jung & Sohn, 2017; Li et al., 2018; Trepanier & Chapleau, 2006). They used techniques such as trip chaining models (Barry et al., 2002), probability models (Huili et al., 2007), and deep learning (Jie & Yang, 2006; Jung & Sohn, 2017) to infer destination of passengers.

The advantages and disadvantages of smart card data are clearly identified in a write up by Pelletier, et.al. 2009. The following advantages over using manual data are not merely differences in degree but also differences in kind including:

1. The cost of obtaining an OD matrix is significantly reduced.
2. The resulting matrix is based on a significantly larger sample size.
3. The process is more suitable for automation which will make the process much
4. faster and therefore able to be updated more frequently.
5. This process can be combined with more targeted surveys to obtain a more cost effective and comprehensive picture of passenger travel behaviour.

4.2 Feeder Bus Route Related Data

Feeder bus service is a secondary transportation services (e.g. tram lines, bus lines) connecting a primary transportation service (e.g. a rail line or metro line). It is used to solve the last mile problem and to increase the connectivity of the primary network to local areas. It is a useful way of serving train stations with high recurrent commuting traffic pattern. In designing this type of system will require reliable existing and potential passenger demand in the service area or at the stops within service area. Most researches uses traditional surveys (Ceder, 2013; DiJoseph & Chien, 2013; Lund et al., 2004; Pan et al., 2015; Shrivastava & O'Mahony, 2005) for estimating commuting demand but they are expensive and time consuming to conduct. Some studies (Li et al., 2018; Munizaga Muñoz et al., 2014; Yu & Yang, 2006) use smart card data to estimate potential travel origins and destinations, with the restriction that the data depends on existing routes and stops and shows only the current passenger demand.

Different objectives requires different data and to achieve the object of FBRNDP, many data sources is required the required; such as city masterplan data which comprises of demographics, socio-economic details and land use characteristics of the city or surveys/smart card data extractions to obtain details like travel habits, important origin destination nodes. These data together can help in identifying the locations that has potential to be served by regular feederbus routes. A research (Kuah & Perl, 1989), used four basic data for his analysis which includes, service is identification station and stop location coordinates in the service area, demand density in the service area, stop density in the service area and also operating cost of both rail and bus transit which was derived from literatures. A similar work also uses the same bench mark study and same data but introduces a second sets of random data consisting of problem size and structure which was used to compare performance of the different metaheuristics used. In both cases cost demand was assumed at the stops (Kuan et al., 2004). While on the other hand, other researchers like (Shrivastava & O'Mahony, 2007) uses traffic surveys with coded networks to create potential O-D matrix which was then used to design the feeder bus route. Saadia (2016) uses 3 sources of data; to include demography of the study area, socio economic and land use characteristics so as to identify areas served by the feeder services and to obtain overall details overall travel pattern. Also field surveys were used to obtain current travel patterns for main trunk users and assessment of current feeder modes used by to and from the main trunk line. Transit smart cards portends a huge potential and the potentials of the entry only transit smart card used in this study is discussed in the sub section below.

4.3 Smart Card Data: Izmir Case Study

Izmir a city in Turkey has an urban population of about 4,061,074; (Stat, 2014). It has one of the most developed and well integrated public transit systems in the country because it can boast of quite a number of different and well integrated modes. In Izmir, the urban mass transportation network is comprised of the public bus system bus (Eshot and Izulas), three rail systems (Metro-rapid transit, the IzBan-commuter rail Service, Tramvay- light rail service), and a ferry service (Izdeniz) which operate in the inner bay.

The fare collection system in Izmir, Turkey was primarily designed to secure revenues and prevent fraud. Smart cards are swiped over validator located in buses or transit stations. The device then reads information from the smart card, such as card identity, validity, fare deductions, and even records the transaction data. Some of the purpose for developing this system are listed below;

1. To discard the conventional paper tickets along with its costs in terms of production, distribution and securities
2. To enhance inter modal and service integration of the various modes of transportation that exists
3. To reduce the cost of service operation and improve overall service quality
4. To allow for different fare structures to be implemented depending on the service being provided.
5. To allow for subsequent implementation information technology systems.

Many information, which can be used in transportation planning, are obtained using Smart card data based on fare collection system and GPS data based on vehicle location system used in İzmir Metropolitan Area (Deri & Kalpakci, 2014). The information obtained within the scope of the boarding data on a Card (ID) basis is summarized below. Some important data collected are;

1. Smart Card ID
2. Fare Type
3. Service Direction:
4. Time & Date of Boarding
5. Stop ID:
6. Route Number
7. State of Transfer

In the mass transportation network, trips are divided into five different categories: Student (1.64 Turkish Lira-TL), free (disabled, policemen, press card-holder professionals, the unpaid), 3–5 journey boarding pass (for passengers without a smart card: 5.75TL), discounted (teachers, municipal workers, the elderly over 65: price differs according to the type of pass 3.00 TL), and full fare boarding (all passengers not mentioned above 3.46TL).The fare system operated has various classes for example (Eshot/Metro/Izdeniz/Tramvay) has a time based fare system. The first for full fare boarding fee is 3.46 TL, the first transfer is 0.50 TL in 120 minutes, the second transfer is 0.50 TL and the next transfer is included in the price. While for IzBan has a distance based fare. For example, a full fare card type, the first 25km the fare is 3.46 TL and next 25km is 0.10 TL.

IzBan system is a rail line that has become the backbone of the city's transportation system, especially because of an effort to integrate with rubber wheel lines. It runs from the north to the southernmost of the city, and it has become a structure connects urban and rural areas. This represents the FBRNDP we are trying to solve and because aforementioned potentials in the previous sections, an preliminary exploration the data was done.

From 1,967,955 boarding records for 8.11.2018 was used an exploratory data analysis. The following can be deduced as depicted in the Figures 4.1, 4.2, 4.3, below. A modal split is the percentage of travellers using a particular type of transportation or number of trips using the mode type. It is an important component in developing sustainable transport within a city or region. In recent years, many cities have set modal share targets for balanced and sustainable transport modes, particularly equal percentages of non-motorized (cycling and walking) and of public transport. These goals reflect a desire for a modal shift, or a change between modes, and usually encompasses an increase in the proportion of trips made using sustainable modes. Izmir has a variety of public transit systems and even amongst them, the most widely used is the suburban train (IzBan) and an extensive bus network (Eshot) as can be seen in the Figure 4.1 considering a typical day for smart card users. Together they form more than 50% of daily boarding with IzBan representing high capacity and main transit corridor and Eshot representing a secondary corridor. Therefore, defining a composite system such as feeder transit network may improve the efficiency of the overall transit network.

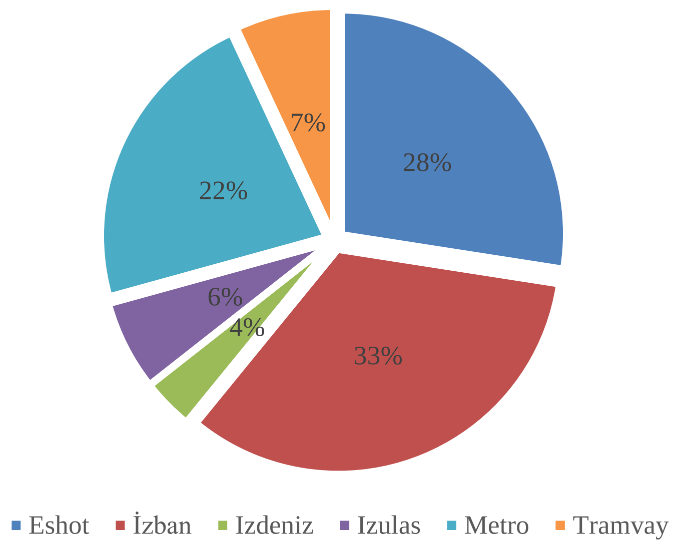

Figure 4.1 Modal split for PT (Smart card users)

In addition, the suburban train uses a distance based fare structure which require transit users to swipe their smart cards when boarding and alighting (distance-based fare). Figure 4 shows the number of passengers boarding normally, or transfer or a returned fee (Normal, Aktarmali and Iade respectively). These characteristics gives an insight into how many passengers use the train and transfer to other modes of transportation. The, the number of suburban train passengers with returned fees is 47% which suggests a high train to other modes transfer.

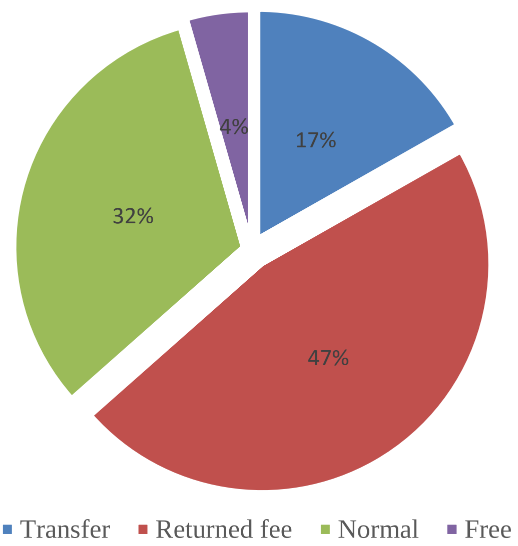

Figure 4.2 Type of boarding at IzBan stations

Similarly, Figure 4.3 presents the total number of passengers using IzBan stations for the particular day. Halkapinar and Sirinyer stations have a high activity which may be due transfer stations present at that location, particularly Halkapinar because it has Metro train and Eshot(Bus) transfer stations attached to it. Sirinyer station appears to be more suitable for designing an suburban train- feeder bus system because it is has only one public transit mode connected to it and it also has high passengers transferring to and leaving the station as shown in the Figure 4.4 below.

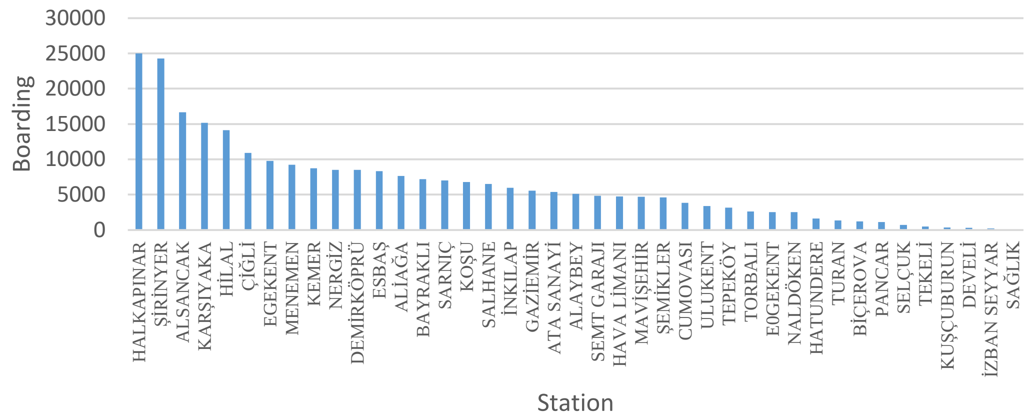

Figure 4.3 Boarding by IzBan stations

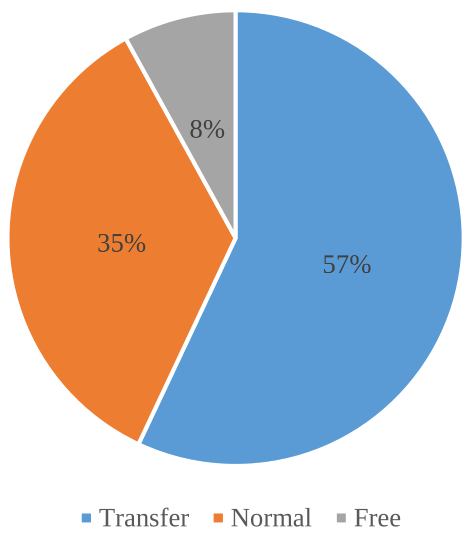

Figure 4.4 Type of Boarding at sirinyer station

Figure 4.5 shows a typical temporal count at a particular station. This can be useful if one is interested in designing a system for a particular time period. Clearly, the morning rush hour seems to have a high demand when people go to work while the return journey can be sparse because of varying return to home journeys.

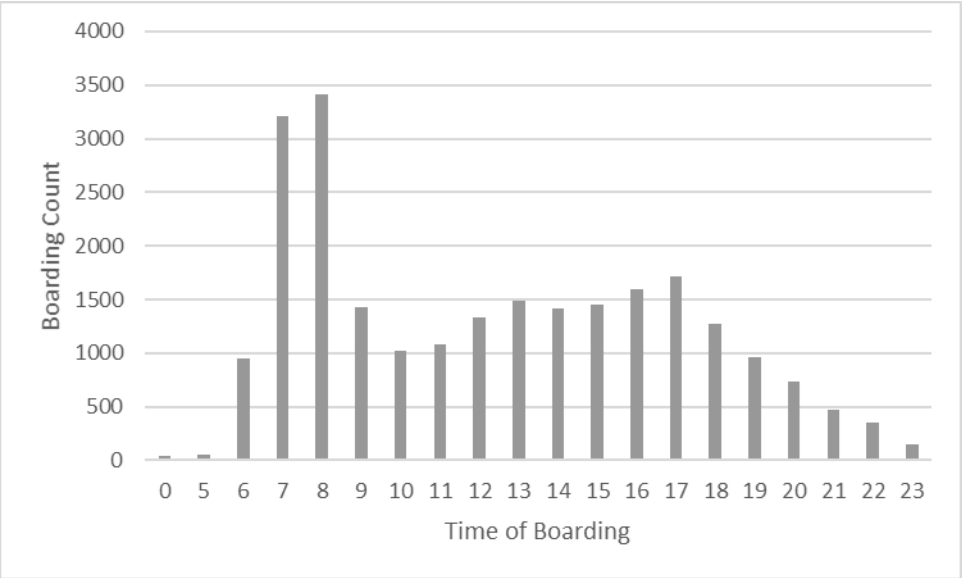

Figure 4.5 Temporal count at sirinyer station

In summary, this section discusses the literature concerning the use of transportation smart card data in transportation planning. It highlights a basic comparison between conventional survey data and an the transit smart card data and discusses the transit smart card system used in Izmir. A preliminary investigation was also carried out in this section.

CHAPTER FIVE
METHODOLOGY

5.1 Background

When planning feeder bus routes for a known destination especially for an intermodal transit system, the solution may be peculiar to that problem in terms of spatial distribution of the transportation system and temporal distribution of demand for the feeder bus routes. Methodologies used in solving this kind of problems are mainly influenced by; level of details in design environment, quality of anticipated solution and available computational power. Hence, they are often partitioned in a sequence of procedures so as to be manageable. Two major approaches are used namely; route generation and configuration and route construction and improvement (Kepaptsoglou & Karlaftis, 2009). Feeder bus route can be solved by applying heuristic algorithms to reduces the size of the large solution search space, by building initial routes followed by improvement of these initial routes using optimization technique. A typical example of this problem is well documented Kuah & Perl (1989). They used a sequential building heuristic for building initial solutions which is an adoption from sequential saving approach for MDVRP. Also metaheuristics like GA can be used generating initial population at random as suggested by (Chien et al., 2001) but it might not be a good selection for generating initial routes. Therefore, Kuan et al.(2004) employed the concept of delimiter, proposed by Van Breedam (2001). Most of the studies follows this approach of generating initial routes using various kinds of heuristics and then try to improve the initial solution by using the initial solution as an input in some optimization techniques to obtain a better solution, so that improvements can be implemented. In this study, we propose the design of a feeder bus route using smart card data as opposed to conventional data with the aid of genetic algorithm in order to allow the selection of best route that minimizes total cost. This work will allow the existing bus stops, its network, and the distribution of demand to be used rather than conducting new field study and obtain an optimized feeder bus route system.

Thus, this section describes the major steps in terms of methodology, algorithms, and adopted principles see the figure below for the procedure. In general, the first step when designing feeder bus routes requires certain data such as; the location of the bus stops and the train station, the demand at the bus stops going to the train stations, the network connectivity and operational characteristics of vehicle. These data are often gotten through conventional on board surveys which are usually costly and sometimes biased. In this study alternative data collection source was explored in the form transit smart card data. This module deals with smart card data preparation, in terms of data extraction, data cleaning and conversion into useable format as well as representation of the real network for the design of the feeder routes. The second step deals with adoption and application of genetic algorithm to MTSP, since it is similar to feeder bus services especially when considering a single destination been connected to multiple origins. This module allows for many solutions to be generated and compared. The last module is concerned with the selection of best route through the evaluation of the numerous solutions generated based on cost components and different perspective of user and the operator.

Over all Kuah & Perl (1989) network problem will be used to test the solution module and the evaluation module. Similarly, real bus route was also extracted from Eshot network but fixed demand was used and the algorithm was tested for its efficacy in design feeder bus routes. Finally, smart card data along with spatial information was used to identify key demand position to represent potential feeder bus stops and real network is superimposed to identify real network connectivity and consequently this was then used to design feeder bus routes, evaluated, and best solution was selected based on user and operators cost as well as total cost.

5.2 Major Assumptions

Modelling transportation systems usually requires some simplifications to allow for manipulation of the system. Similarly, in this study some fundamental assumptions were made listed below;

1. Intermodal transit system is consisting of only bus stops and fixed train stations i.e., no other transit modes is included in the system.

2. Service area was not used since the criteria for associating bus stops and train station is based on the smart card transfer system. Bus stops were related to a train station based on Smart card users that board a bus stop before transferring to the train station within threshold time of 90 minutes. Hence the demand pattern utilized was many origins to single destination. Also it is assumed that there is a good coordination between the operation of the railway service and bus service.

3. Each bus stop is served by one feeder bus route and all passengers have a certain railway station as their destination. The main purpose of transit assignment is the degree to which the network utilization. Thus, a passenger may even be proportionally assigned to several routes, in order to increase the overall accuracy of the assignment especially for conventional bus routes (Briem et.al 2017). For the case of feeder bus routes, all passengers are assigned to a particular route from their origin to specified destination (train station). The FBRNDP was modelled along MTSP, therefore ensuring that all demand nodes are served.

4. The coordinates of the bus stop and train station is given.

5. Parameters used in the evaluation and selection of the optimal feeder route are specified (such as vehicle operational characteristics, capacity of bus stops, and some cost parameters).

6. All Feeder routes are two ways and can be used in both directions.

7. The optimal route selection is based on route structure and estimated cost parameters.

Therefore, to plan a feeder bus routes for an intermodal system will require (see Figure 5.1)

1. Data Collection Module; In order to execute the feeder bus route problem, four important data sets must be made available, namely, the list of all nodes (bus stops and rail stations) coordination, the network available connectivity list, the transit demand matrix and cost parameters.
2. Solution Module: Route structure determination
3. Evaluation module: Analysis and selection of best alternative based user and operators cost.

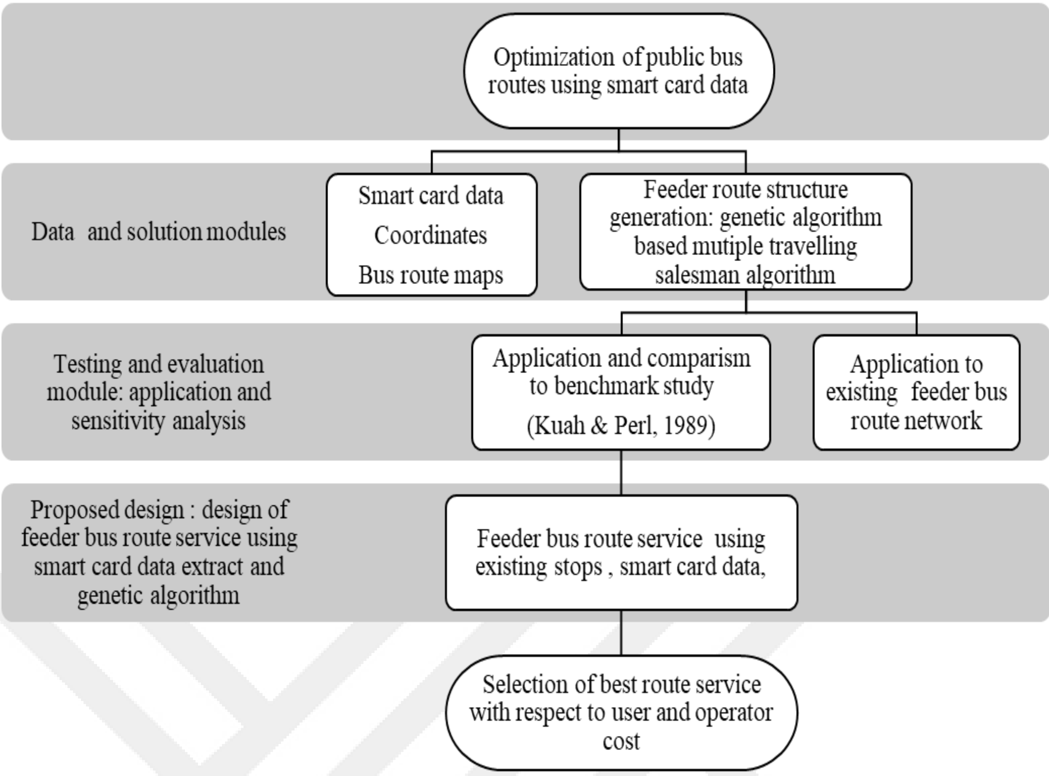

Figure 5.1 Methodology for feeder bus route planning

Three scenarios were generated for designing the feeder bus routes;

1. When all location of bus stops, train station as well as fixed demand are given and also network connectivity was assumed that all stops are connected. Here the Kuah & Perl,(1989) study was used as benchmark. The cost parameters were also presented in this benchmark study. For this scenario pairwise distance was used to cluster potential feeder bus stops around the train stations
2. The second scenario is the same as in one but the data is extracted from Eshot website. Also here real network connectivity was not used but an assumption that all locations are connected. The cost parameters here were estimated based on local condition.
3. The third scenario is that train station is given but potential feeder bus stops and demand are extracted from smart card data. Network connectivity was generated by superimposing real transit network on the potential bus stop locations. The cost parameters here were estimated based on local condition. To formulate an origin-destination matrix of the journeys using the public transport system which includes public buses and suburban train modes in the city by analyzing the smart card

information, an algorithm was formulated in the MATLAB programming language within the scope of the study and the flow diagram is presented below (Figure 5.2).

5.2.1 Data Module: Algorithm for Feeder Bus Stop Location from Smart Card Data

This section is focused on the data preparation, and distance matrix extraction from smart card data and cost estimation parameters. These procedures are discussed in the sub sections below. Designing a feeder bus route will require the following important data ;

1. Location of all transit nodes both bus stops and train stations
2. Transit network connectivity
3. Transit demand matrix
4. Transit cost parameters

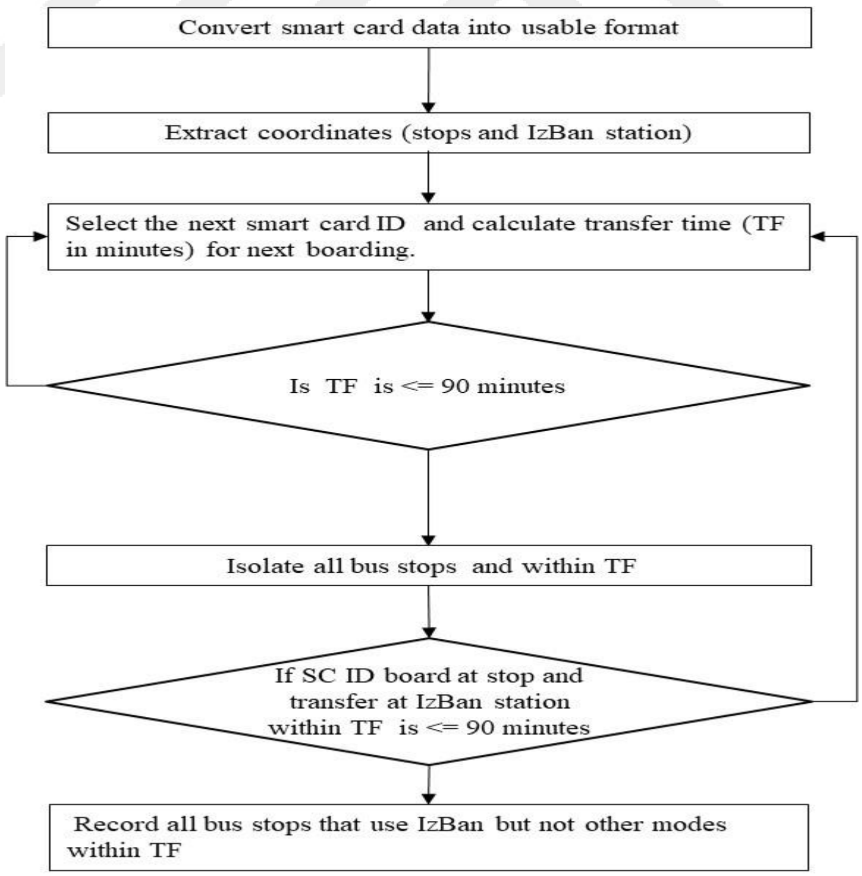

Figure 5.2 Algorithm for demand and potential origin estimation

This exercise is rigorous and requires a lot of computer software to prepare the smart data into useable data. It includes cleaning and coding of the smart card data since not all collected data are relevant and there are also missing data. Coordinates of bus stop and train station was extracted and transfer time was calculated for each smart card data. In Izmir the transfer time is less or equal 90 minutes. Therefore, if transfer time is less than 90 minutes, then we isolate all bus stops that uses that IzBan station and no any other modes.

Two popular data partition algorithms (K-means and K-medoids) are used . Both algorithms clusters data into groups by minimizing the distance between coordinate points and the centroid of the clusters. While K-medoids selects the centroid from one the coordinate points as a centroid, K-means selects the centroid from average between the points in the cluster. Also, K-medoid also uses dissimilarity measures which reduces the sum pairwise dissimilarity instead of euclidean distances used by K-means. Importantly, it is assumed that the “k” number of clusters is known or evaluated by other methods or certain criteria. Therefore, for the purpose of this study, the procedure for determining the feeder bus stop location through clustering restricted by the capacity of the ensuing cluster is proposed.

The procedure is stated below;

1. Calculate the number of clusters; It is calculated based on the demand (di) at the stops and expected capacity of cluster (C), k is the sum of all demands at the stops divided by expected capacity of the cluster.
2. Select initial centroids: the initial k centroids are selected by arranging the
3. stops based on their demand in their non-increasing order d1 > d2 > d3 > dn. Then the first k stops become k centroids.
4. select first k stops as initial centroids
5. Assign the stops to clusters

   a) The Euclidean distances between each requester to all the k centroids are calculated. Group all the stops to the closest centroid.
6. Centroid Calculation

a) If the capacity of the centroid is not exceeded, the initial k centroid is adopted.
b) If the capacity of a centroid is exceeded, that centroid is split into the required number to satisfy the cluster constraint

Estimating the demand, a total 321 stops used the train station for the period for this particular date as shown in the Figure 5.3(A). Not all stations have significant demand therefore the stops were chosen based on the card usage at the various stops. The selected stops were limited to those stops with at least 3 passengers per hour thereby reducing the stops to 87 stops with a total demand of 633 passengers per hour as shown Figure (5.3B). Feeder bus service differ from conventional bus service in that they are mostly servicing suburban areas with high commuter volumes. This implies most passengers gather in one location to proceed to somewhat common destination i.e. bus stops may be crammed. Therefore, the need locate suitable position for siting feeder bus stops and it is also important to take into consideration the capacity of the stops themselves in terms of passenger volumes.

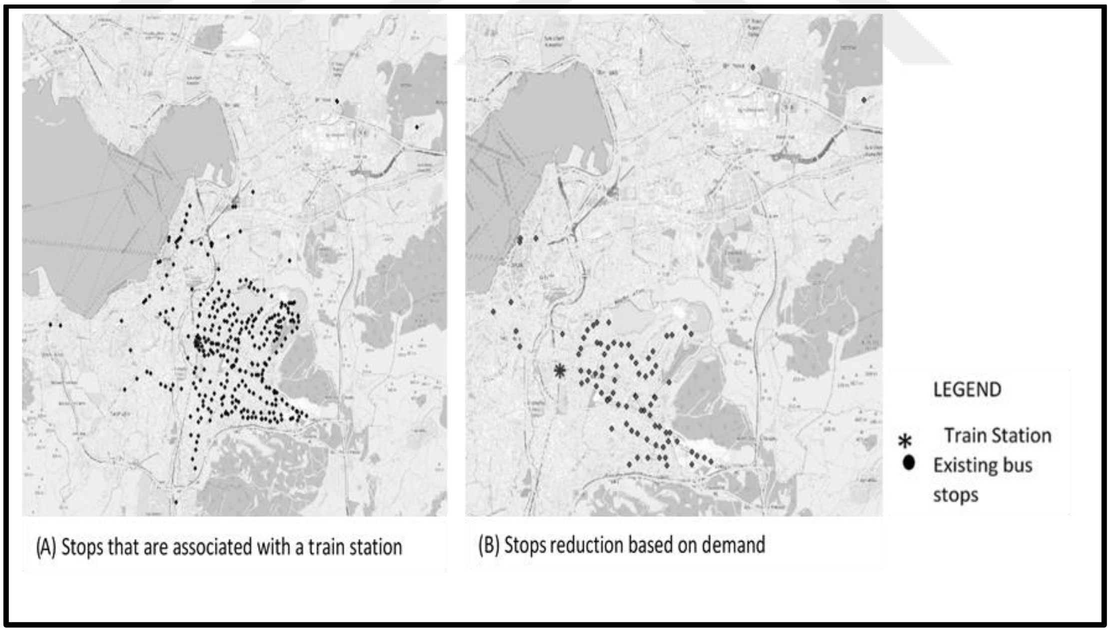

Figure 5.3 Stop reduction

To determine the feeder bus stop location, clustering of the bus stops are needed. After reduction, the area has 87 existing bus stops whose demands are known and are distributed in (x, y) coordinates. The K-medoid algorithm tries to establish the suitable cluster centroid by using existing nodes(bus stop) unlike the K-means algorithm which creates a new centroid. Using the K-medoid allows the use existing bus stop locations for the siting of the feeder bus stop location(see Figure 5.4), however because the demand at every bus stop exists, the new point might be cramped based on demand which might be un desirable for users . as shown in Table 5.1.

Table 5.1 Passenger demand/hour at the centroids (number of stops)

Cluster	K-means	K-medoids	Capacitated clustering
C1	152(12)	10(1)	67(6)
C2	10(1)	38(5)	73(8)
C3	19(3)	16(1)	42(3)
C4	38(5)	18(4)	63(11)
C5	48(11)	73(7)	47(9)
C6	127(15)	121(15)	48(7)
C7	8(2)	19(3)	26(2)
C8	121(15)	140(12)	21(3)
C9	42(9)	75(8)	44(9)
C10	17(4)	37(8)	34(4)
C11	23(6)	31(5)	38(9)
C12	16(1)	17(4)	32(4)
C13	12(3)	39(8)	97(18)

In our case, since the centroid is known, we find the distance for all remainder nodes to the selected initial centroids. After the shortest distance to all centroids are assigned, a group of clusters formed around the initial centroids and the Table5.1 ensued. But we have a constraint that no cluster demand shall exceed 50 passengers per hour. Hence the red colored cells exceed and must be split into more centroids to accommodate the excessive demand. Centroids 1,2,4 and 13 are split further to accommodate more demand and the final feeder bus nodes their respective demands.

Table 5.2 Average distance from stops to centroids

Cluster	K-means	K-medoids	Capacitated
C1	2931	0	649
C2	0	264	708
C3	3926	0	576
C4	3851	550	782
C5	1811	329	400
C6	1686	519	408
C7	4500	222	4077
C8	2384	447	841
C9	1877	328	1143
C10	3186	432	463
C11	1017	439	460
C12	0	350	283
C13	3146	404	1989

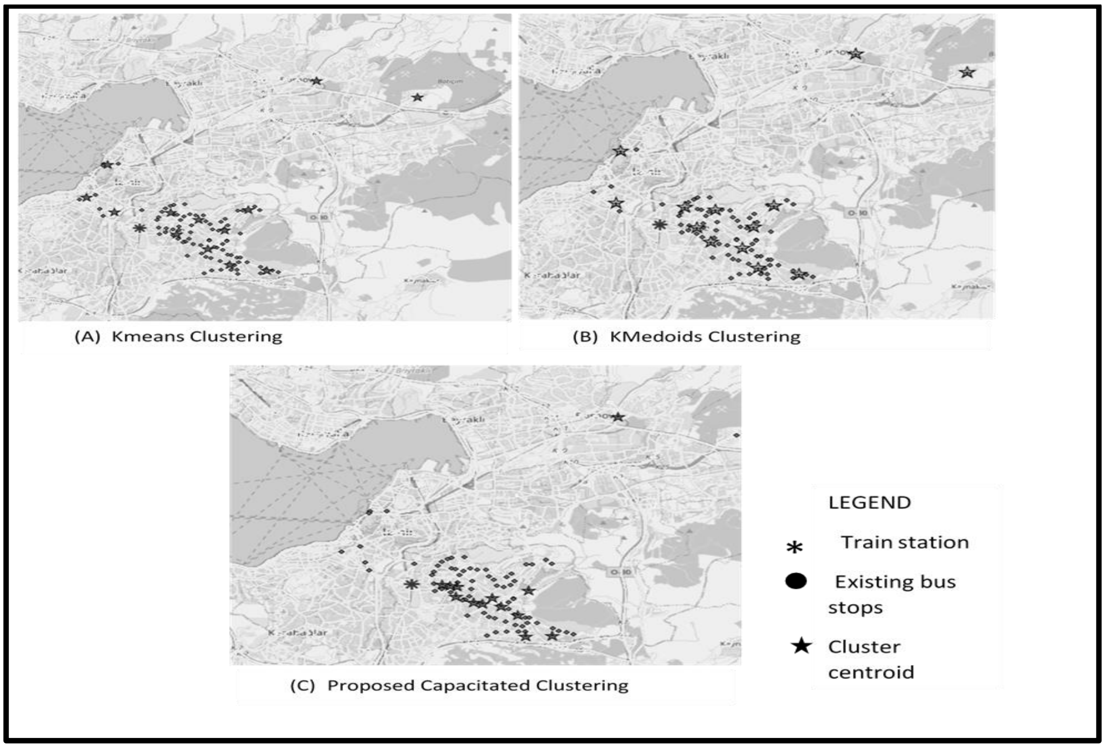

Figure 5.4 Clustering

To analyze the results, we evaluate the clustering results (Figure 5.4) by summarizing the clustering by a quality score and also based on the purpose of the clustering carried out. The quality scores used are calculated by understanding the cohesion of how near the data points in a cluster are to the cluster centroid (intra cluster distance) and also the distance between the centroids of different clusters(inter cluster distance). It is well desired a good cluster should maximize inter cluster distance and minimize intra distance cluster. These distances are then used to calculate the Dunn index which is basically is ratio of inter distance (separation) and intra distance (compactness), and the Davies-Bouldin index which is the ratio intra distance (compactness) to inter distance (separation). The clustering algorithm that produces a collection of clusters with the smallest Davies–Bouldin index is considered the best algorithm based on this criterion, while the algorithms that produce clusters with high Dunn index are more desirable. Figure 5.5 shows the cluster analysis and the proposed algorithm has the lowest inter cluster distance while the K-means has the largest distance while for the Intra distance K-medoids algorithm has the lowest value. This may be largely due to the fact that the proposed algorithm pays attention to locating the centroids based on demand and it is expected that passengers that use the train station are closer to the station. To go further, See Figure 5.6. Based on both indices, the K-medoids algorithms performed better than our proposed method, even though our proposed method performed better the K-means algorithm. These analysis do not really shows how useful these algorithms are especially in this specific problem of estimating the location of feeder bus stops based on demand. While The K-means algorithms locates the feeder bus stop generating a mean centroid for all the clusters , the K-medoids uses existing centroid and tries to find which centroid is more at the center of the cluster. Interesting enough, our main focus is to develop a feeder bus stops by locating important centers of demand and grouping them together. Therefore, for the purpose of achieving this with 13 clusters based on the expected capacity of each centroid. For all the algorithms some stops exceeded the estimated capacity. This is expected as K-means and K-medoids are mainly dependent on distance and even our proposed method. The number of stops has to be expanded by dividing any stops that exceeded the capacity into desired number of stops to accommodate the capacity. In all, our method can be very useful in planning bus stop location especially by considering how close they are but also how important does demand points are, see the and Figure 5.7. It is important to note that, when evaluating clustering algorithms, to emphasize the purpose to which the clustering is done (which is highly subjective) even though the statistics used to compare our algorithms to K-means and K-medoids can relate relevant information with regards to whether the clustering is bad or not.

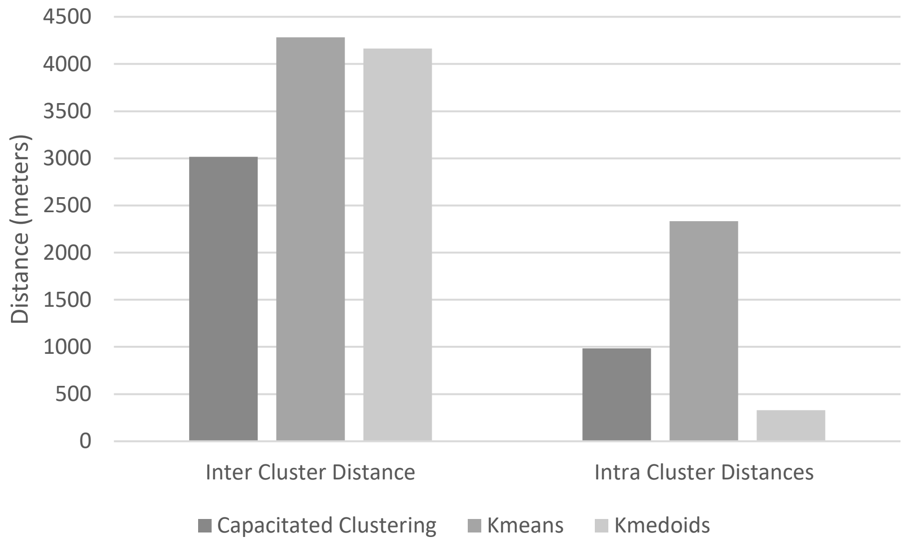

Figure 5.5 Inter/Intra cluster distance

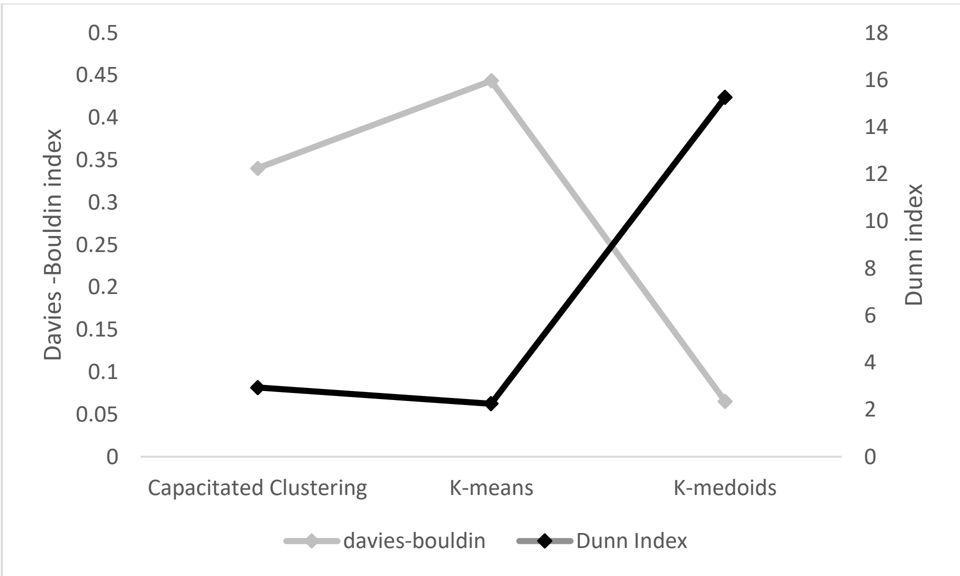

Figure 5.6 Clustering Analysis

In conclusion, we proposed a methodology away for planning the location feeder bus stops from existing conventional bus stops taking into consideration the passenger demand going towards a specified destination (train station). The first part involve determination of potential feeder bus stops by identifying the location of smart card users who are destined to a particular train station through an algorithm developed in this study. This requires a lot of data mining from database management system used in Izmir. The second part, equally follows a set of steps that allows for potential stops to be clustered based on specified capacity of the potential feeder bus stops.

The results of our capacitated clustering procedure was compared with famous clustering algorithms K-means and K-medoids algorithms. The procedure focuses on mainly on high demand stations therefore the final feeder bus stops were compact while on the K-medoids clustering tends to be best in terms average distance of the clusters to the centroids and K-means have the worst average distance from clusters to centroids.

This method shies away from conventional clustering algorithms which does not take into account the demand of the clusters and also from complex optimization techniques presented in literatures. To best of our knowledge little literature exist with regard to using ‘Big Data” in the determination of feeder bus stops.

Clusters	Demand per
1	37
1b	30
2a	37
2b	36
3	42
4a	21
4b	42
5	47
6	48
7	26
8	21
9	44
10	34
11	38
12	32
13a	48
13b	49

Figure 5.7 Capacitated cluster

5.2.2 Solution Module: Genetic Algorithm based on Multiple Travelling Salesman Problem

We present FBRNDP as a representation of MTSP. This problem deals with the optimization of the shortest paths for k cluster. Although, it is possible to generate routes first then and then cluster but this approach performs poorly (Eldrandy,2008). In this study, a modification of Joseph Kirk's Fixed Start (train station) Open Multiple Traveling Salesmen Problem (GAMTSP) was used. The algorithm finds a (near) optimal solution to a variation of the "open" M-TSP by setting up a GA to search for the shortest route (least distance needed for each salesperson to travel from the start location to unique individual cities without returning to the starting location). In Summary, each salesman starts at the first point, but travels to a unique set of cities after that (and none of them close their loops by returning to their starting points) and except for the first, each city is visited by exactly one salesperson.

A potential travelling salesmen problem solution is scored by the length of the route which is to be minimized. The fitness is calculated by scoring various solutions relative to each other and selecting the shortest distance. This algorithm uses a tournament based selection and it is structured in such a way that the population must be divisible 8 because that is the way a good solution in the current solution is propagated into to the next generation. The solutions are grouped 8 randomly at a time, the best one of the 8 is taken and passed on to the next generation. Three different mutation (flip, swap and slide) are then performed because they make modifications that are more likely to make better the best current solution. Note that crossover was not mentioned because it tends to make large changes to a given route and rarely improves decent solutions. These mutations together with the best solution of the 8 randomly selected solutions are then passed on to the next generation. The generational process is repeated until termination criteria is reached. Here the number of iteration was used as the termination criteria. Therefore, tuning parameters such as initial population, number of iteration and minimum number of tour was carried out to check their sensitivities to the solutions so obtained. The Figure 5.8 below shows the effect of population size and number of iteration on time of convergence. Four different population size in the multiples of 8 and number of iteration (100,500,1000,5000,10000) were used as stopping criteria. For all scenarios as the number of population size increases so does the time of convergence. Similarly, as the number of iteration increases so does the time of convergence.

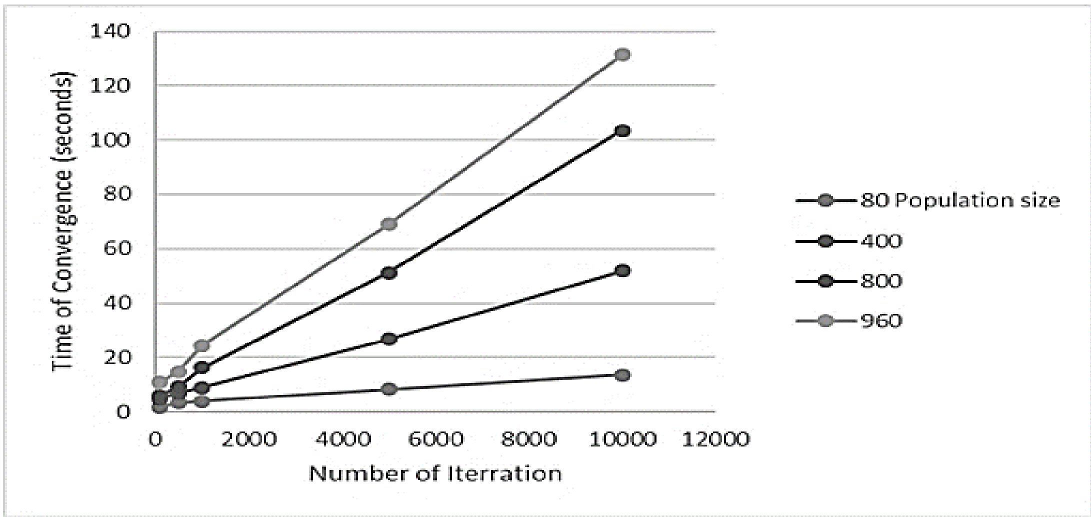

Figure 5.8 The effect of population size and number of iteration on time of convergence

Figure 5.9 shows the effect of constraint minimum tour on the time of convergence. As the number of minimum tour constraint increases so does the time of convergences decreases this is so because the constraint is used to limit the time of solutions that can be found thereby reducing the search space and consequently the time of convergence.

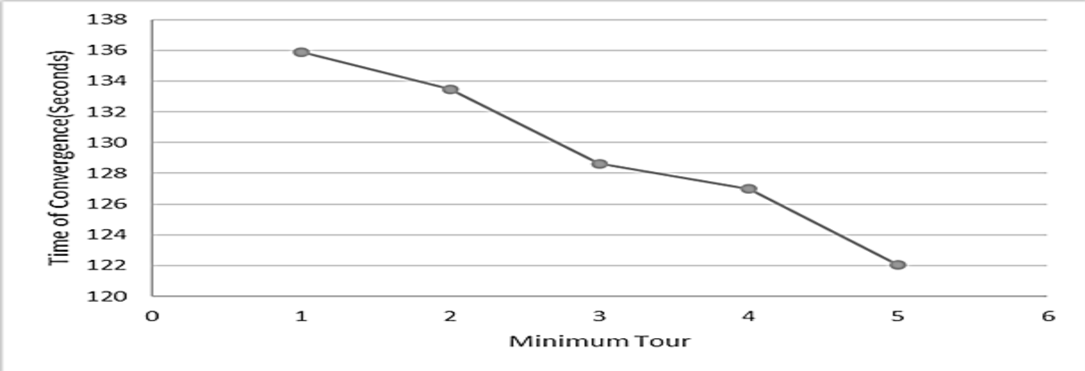

Figure 5.9 The effect of minimum number of tour on time of convergence

The Figure 5.5, shows the effect of initial population size and number of iteration in the attainment minimum distance solution. Clearly, as initial population size increases for all different number of iterations, the algorithm estimates a better minimum distance. For example at Initial population (960 and iteration 10000) yields the minimum distance overall . Of course, the minimum distance so obtained is affected by these parameters and depending on the type of problem and solution needed these parameters may be varied .

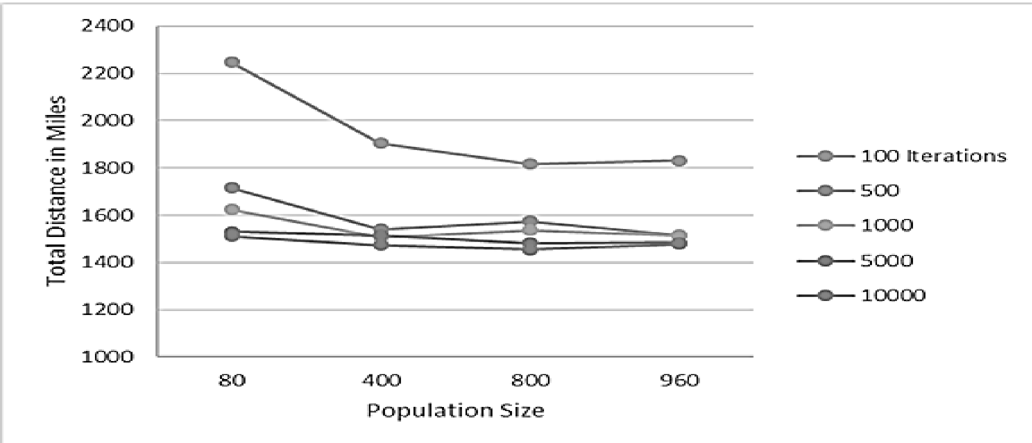

Figure 5.10 Effect of Initial population size and number of iteration on the Distance

Table 5.3 Stability of solutions

	Iteration				
Population Size	**100**	**500**	**1000**	**5000**	**10000**
8	×	×	×	×	×
80	×	×	××	✓	✓
400	××	✓	✓	✓	✓
800	✓	✓	✓	✓	✓
960	✓	✓	✓	✓	✓

Also, the algorithm was run several times to see if the minimum distance so obtained is repeatable because genetic algorithm is stochastic and new solution is likely to emerge every time you run the algorithm. The algorithm was run 10 times each and solutions were ranked as not stable represented by x mark and partially stable by xx mark and more stable by check mark as presented in Table 5.3. Even though the stability does not negate its stochastic nature but a selection of higher parameters will yield a better solution and repeatability of the solutions.

5.2.3 Evaluation Module: Total Cost Evaluation and Its Parameters

The process of producing transit services in the proposed intermodal transit network is defined by a cost function. The objective function is the total cost function, which includes the operating cost, and the user cost. For the proposed transit network optimization, we need a function that can capture the sensitivity of costs to design and operation (Chien & Schonfeld, 1998).

Analysis of transit system borders on the estimation of ridership and direct cost of the system. Also calculation of total cost will require other salient factors like such as various components of time (waiting, riding,), fleet size and other performance measures listed below. Representing system performance requires that one calculate additional parameters. The listed system performance measures are calculated by using the parameters as found in Kuah & Perl (1989). The total exact costs could not be obtained, therefore the concept of value of time (VOT) was used to estimate drivers wage and user costs. Value of time is the ratio gross domestic product (GDP) per person to working time per person.In this study the value of time was estimated using the income model with inputs from TUIK and world bank data. According to TÜİK, 31 March 2016 for the year 2015, the Gross Domestic Product of Turkey (GDP) $1,953,561,000,000. Similarly, the GDP per capita is $ 9,621 Dollars is equivalent to 25,130 Turkish Lira (TL) according to purchasing power parity (PPP) at current prices. In Izmir province, the GDP value and PPP values were estimated by reducing it to the provincial level and dividing the population of the provinces. As of 2001 Izmir's annual contribution percentage to the turkey economy is 6.6%. Therefore, the annual income per person can be estimated as 30,931.57 TL. The average value of time was estimated based on work hours of 2000 was estimated; 30,931.57 divided by 2,000 gives 15.47 TL/hour.

The fuel cost estimated was based on paper by (Topal & Nakir, 2018), the fuel consumption of three the bus types given in Istanbul under real road, trip-time and travel conditions were evaluated at 100% load capacity. They include;

1. 12 m Otokar Kent LF Diesel bus, (0.15 TL/Km)
2. 12 m Karsan Bredamenarini bus CNG bus,(0.97 TL/Km)
3. 10.7 m TCV Bozankaya E-Karat Electric bus,(0.17 TL/Km)

Other unit cost parameters are the unit average waiting cost, bus operating cost and bus riding cost. The unit waiting cost was assumed to be the value of time, bus operating cost is the summation average driver wage(which is assumed to be a function of VOT) and fuel cost, and riding cost was taking as the fare charged for standard user From the above, estimation of unit cost parameters are presented in the Table 5.4.

1. Bus operating cost (B0C): can be defined based on unit of time or distance cost in connection with the transit service provided.

$$BOC = 2 * \lambda_0 l_{ij} f_{ij} \tag{5.1}$$

2. Bus waiting cost (BWC): The waiting cost includes passengers waiting for the buses, which is the product of average wait time, demand, and the value of users time.

$$BWC = \frac{\lambda_W q_i}{2f_{ij}} \tag{5.2}$$

3. Bus riding cost (BRC): the product of demand, in-vehicle time, and value of time can define the user in-vehicle cost. In some literatures it is regarded as running time.

$$BRC = \frac{\lambda_r l_{ij} q_i}{U} \tag{5.3}$$

4. Bus user cost (BUC),

$$BUC = BWC + BRC \tag{5.4}$$

5. Frequency: bus operating frequency

$$f_{ij} = 0.5(\frac{\lambda_w q_i}{\lambda_0 l_{ij}})\wedge 0.5 \tag{5.5}$$

6. Total Vehicle Miles (TVM): is given by multiplying the vehicle hours by the average speed.

$$TVM = Uf_{ij}(l_{ij}/U) \tag{5.6}$$

7. Total Passenger Miles (TPM): summation of (segment length* average volume)

$$TPM = \sum\big((l_{ij} q_i\big) \tag{5.7}$$

8. Total system cost (TSC),

$$TSC = BOC + BUC \tag{5.8}$$

Table 5.4 Estimation of unit cost parameters

SN	Descriptions	Units	Value
1	Unit bus operating cost (λ_0)	TL/veh. Km	20.2
2	Unit bus riding cost ($\boldsymbol{\lambda_r}$)	TL/pass. hr	2.33
3	Unit bus waiting time cost ($\boldsymbol{\lambda_w}$)	TL/pass. hr	15.4
4	Bus capacity	seat	50
5	Average bus operating speed (U)	Km/hr	20
6	Average demand per hour at bus stop i to station j (q_i)	pass./hr	
7	Bus operating frequency (f_{ij})	veh./hr	
8	Distance from stop i to station j (l_{ij})	Km	

5.2.4 Pareto Efficiency

When faced with a challenge of choice in multi-objective optimization, the Pareto efficiency technique may be efficient in this regards. It has wide use in economic and engineering situations, especially, when the resources involved are utilized optimally. This concept can be described as a condition that occurs in multi-objective optimization when one individual objective cannot be made better without making the condition worse for another objective that is, any change would affect the final optimum solution. Therefore yielding a solution that is not dominated by any other feasible solution. The mathematical formulation can be found in (Costa & Lourenço, 2015).

The question of goodness of solution always arise in optimization problems. For example; in a single objective problem (minimization), the selection of optimal solution will be the solution with smaller objective function value. On the other hand, the optimal solution in multi objective can be selected through partial ordering or comparison between the objectives otherwise known as dominance test. In the work of Deb (2001), suggested that, when comparing two solutions in a multi objective problem, a solution dominates another solution if two conditions are met. The first condition is that solution 1 is no worse than solution 2 in all objectives and the second condition is that; solution 1 is better than solution 2 in at least one objective.

Consider an example below which is a minimization problem of two objectives f_1 and f_2; to select the best solution we apply the two conditions.

- Solution 1 dominates solutions 2, 3, and 4 in both objectives
- Solution 2 dominates solutions 3 in both objectives but dominates 4 in only one objective
- Solutions 1 and 2 forms the pareto optimal front/non dominated front and solutions 3 and 4 forms the dominated solutions

The dominance test is a simple way of obtaining the best solution especially for few data points where the pareto frontier can be difficult to obtain. This is the basis for pareto efficiency, and in given a set of solutions, the non-dominated solution set forms the pareto frontier which is a set of all the solutions that are not dominated by any member of the solution set. In this study we opted for the dominance test in comparing our results since there were few data points.

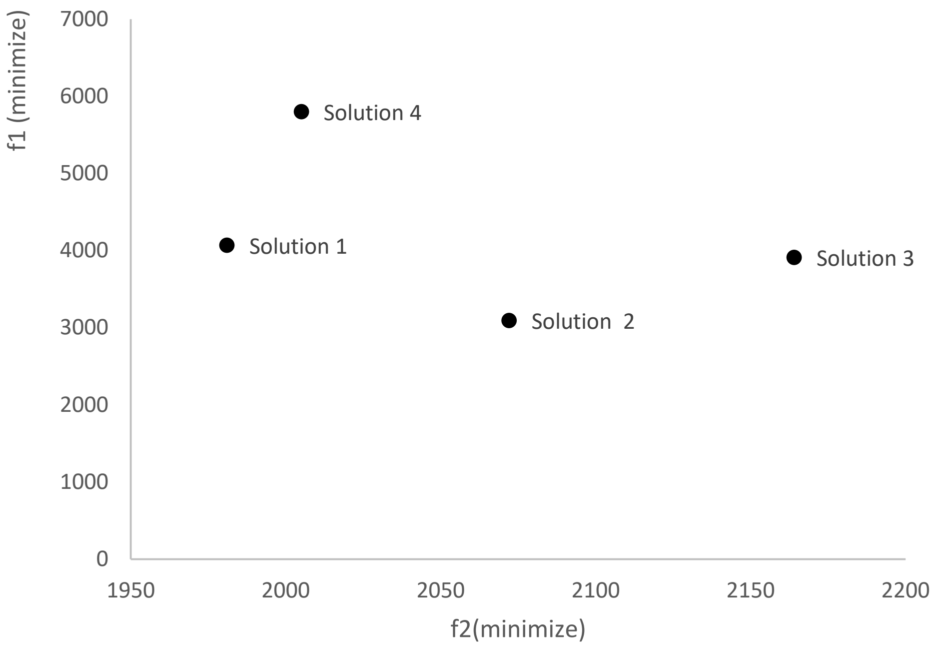

Figure 5.11 Pareto optimal solutions

CHAPTER SIX
APPLICATION TO BENCHMARK STUDY

6.1 Benchmark Study

The methodology was tested in the analysis of transit services, including bus feeder services connecting the rail stations in the case study by Kuah & Perl (1989). They used the network with 59 nodes, which include 55 bus stops (1 to 55), and four rail stations (56 to 59) covering a service area of 2 by 2.5 mile. The bus stop density is 11 stops per square mile with an hourly demand density of approximately 200 passengers per stop. Such a demand density is consistent with that for a typical urban area (Webster & Bly, 1979).

Table 6.1 Base case solution

Route No	**Route Structure**	**Route demand (passengers)**	**Route length (miles)**	**Route frequency (trips/hr.)**
1	1 2 10 24 57	800	1.62	18.14
2	51 46 42 38 56	800	1.08	22.22
3	12 13 18 59	600	0.97	20.31
4	9 16 57	400	1.03	16.09
5	4 6 59	400	0.79	18.37
6	41 32 58	400	0.59	21.26
7	21 26 58	400	0.56	21.82
8	3 5 7 11 17 59	1000	1.29	22.73
9	40 39 35 57	600	0.99	20.1
10	19 23 30 29 34 56	1000	1.37	22.06
11	8 14 15 20 25 57	1000	1.3	22.65
12	52 45 44 49 50 56	1000	1.96	20
13	43 36 31 58	600	0.7	23.9
14	55 54 48 56	600	0.88	21.32
15	33 28 27 22 58	800	0.81	25.66
16	53 47 37 56	600	0.85	21.69
	System-wide performance measures			
	NR=16	TVM=703	BRC=1255	
	AF=21.1	TPM=7926	BUC=3330	
	FS=54	RSH=2706	BOC=211 O	
	TRL=16.8	BWC=2075	TSC=6033	

The number of rail stations is selected such that the ratio of the number of bus stops to the number of rail stations in the range of 12 to 15 stops. The inputs of each node were extracted from Kuah & Perl (1989). All coordinates are in 100 miles. The base case solution is given in Table 6.1.

6.2 Data Module

The same data from the benchmark study was used in this study. The algorithm was tested on the benchmark study by varying the number of stations (1 to 4), and the number of routes to be generated(2,3,4,5 route structure) in the study area (Figure A.1 to A.4). The Figure 6.1 shows that as we vary the number of stations in the study area, the better the algorithm find the shorter distances. Over all, the four station case had marginally lower distance than all other solutions. Therefore, for testing the benchmark study, the algorithm was tested on the four station study area.

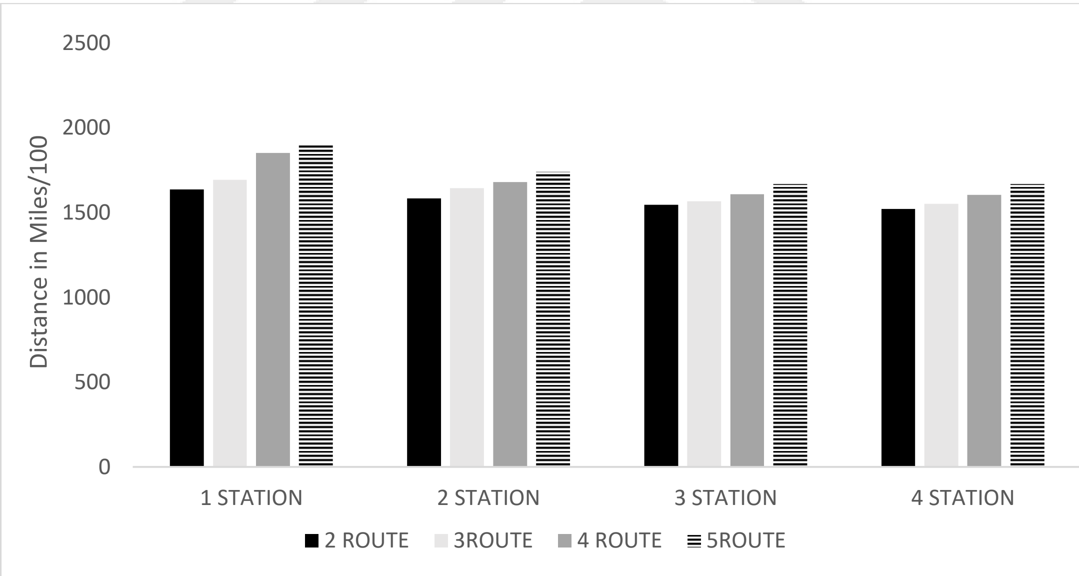

Figure 6.1 Variation of number of station in benchmark study area

Because, it is a four station study area, a form of clustering was carried out to assign bus stops to a particular train station. This was done by calculating the pair wise distance between all bus stops and known train stations and the stops with shortest distances are attached to their nearest train station. The Table 6.2 presents the results of bus stop clustering . This allows feeder bus routes to be estimated based on their proximity to a particular train station.

Table 6.2 Bus stop clustering

SN	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	TR56	TR57	TR58	TR59
2	29	1	14	2
3	34	9	15	3
4	37	10	21	4
5	38	16	22	5
6	42	19	27	6
7	46	20	31	7
8	47	23	32	8
9	48	24	33	11
10	50	25	40	12
11	51	26	41	13
12	53	30	44	17
13	54	35	45	18
14	55	36	52	28
15		39		
16		43		
17		49		

6.3 Solution Module

This module uses genetic algorithm (GAMTSP) was used to generate the feeder bus route structure. The genetic algorithm parameters used were number of iteration, the minimum number of stops that will form a route and most importantly the number of salesman variation (number of routes). As stated earlier it is more beneficial to use no constraints on the minimum number of stops, large enough number of iteration (10000). These parameters ensured that minimum distance was reached within the stopping criteria. As can be seen in the Figure 6.2, the best solution history for 4 route structure fluctuates at smaller number of iteration but above 5000 iteration all solutions becomes stable meaning no new solution is better than the existing result found by the algorithm. The results of the route structure for variable number of routes are presented in the Table 6.3 below. Generally, the 2 route structure solution has more number of stops hence longer routes but overall the length of the route in the system is shorter and number of routes generated were smaller when compared to other solutions. Alternatively, the 5 route structure solution produces shorter routes but again has more numerous routes. Even so, the overall length of the route in the system fluctuates and hence will require economic evaluation. Table A 5 shows the route structure .

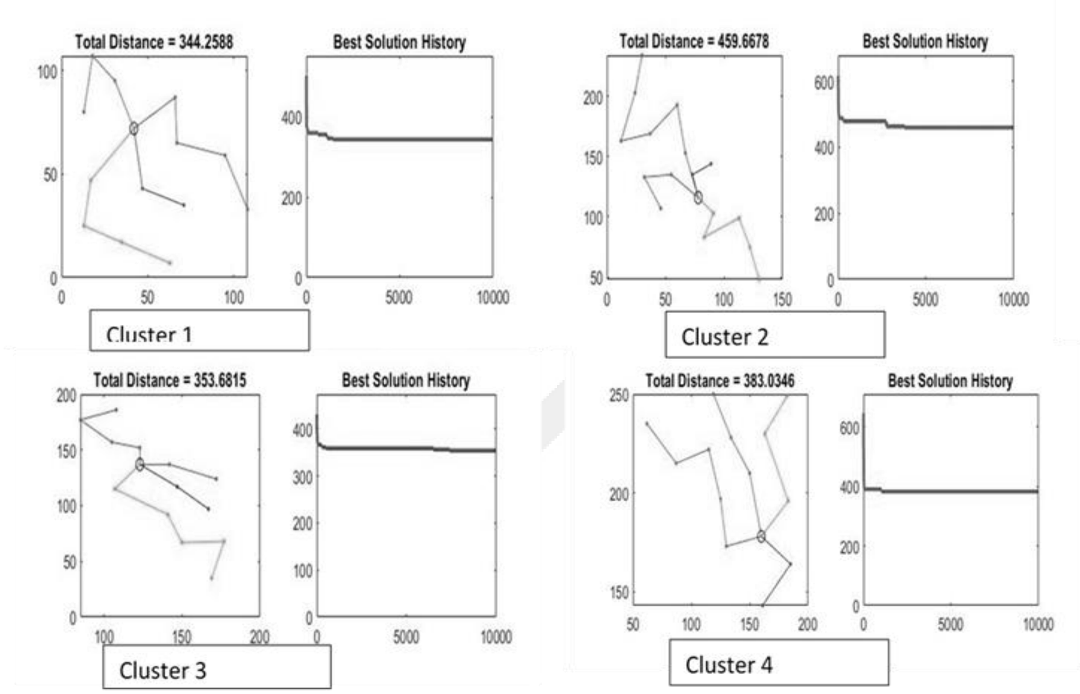

Figure 6.2 Solution convergence

Table 6.3 Number of routes and total route length

Route Structure	**Number of Routes**	**Total Route length (miles)**
2 salesperson	8	15.26
3 salesperson	12	14.94
4 salesperson	16	18.39
5 salesperson	20	16.41
Kuah & Perl, 1989	16	16.8

6.4 Evaluation Module

Measuring the effectiveness of the solutions generated and selecting the best solution requires some sort performance measures for individual routes as well as the overall system . The parameters used here are gotten from Kuah & Perl, 1989 benchmark study presented in the Table 6.4. Similarly, assuming a fixed demand at all stops, for all scenarios the demand were estimated for each route. Similarly, the frequencies and length of all routes were computed. Together with these basic parameters, performance measures were estimated, and consequently evaluation of the system. Table 6.5 presents the fundamental parameters for all five scenarios in Table A.6; D is fixed demand in passenger per hour, L is the length in miles, and F is the frequency of the route veh/hr.

Table 6.4 Parameters

Descriptions	Units	Value
Operating cost unit $\boldsymbol{\lambda_0}$	$/veh.-mile	3
Riding cost unit $\boldsymbol{\lambda_r}$	$/pass.-hr	4
Waiting time Cost $\boldsymbol{\lambda_w}$	$/pass.-hr	8
Max. Allowable route length RL	Mile	2.5
Bus capacity	Seat	50
Average bus operating speed U	mile/hr	20
Average demand per hour at bus stop i to station j q_i	passenger/hour	200
Bus operating frequency f_{ij}	vehicle/hour	
Distance from stop i to station j l_{ij}	miles	

After the fundamental parameters were estimated and the system performance measures such total vehicle mile, total passenger miles and also the feeder bus operating cost, feeder bus user waiting cost, feeder bus user cost, total cost of the system. This evaluation of the various alternatives will aid in selecting the best overall system in terms of minimum total system cost . Considering Table 6.5, the total passenger miles decreases with increase in the number of route this is so because of the individual routes in each route structure. The 2 route structure has minimum number of system routes but generally has longer individual routes implying that some passengers have to travel longer routes while on the other hand, the 5 route structure has shorter individual routes therefore most passengers travel shorter distances. Similarly, the total vehicle miles' changes with change on the number of route structures even though the changes were sometime negative and sometimes positive. This is because the vehicle miles were subject to frequency of the routes. The frequency of the routes is related to other parameters such as unit waiting cost and unit operating cost.

Table 6.5 Cost components

Solutions	Operating Cost ($/hr)	User Cost ($/hr.)	Total Cost ($/hr)
Kuah & Perl (1989)	2110	3330	5440
5 salesman	2072	3092	5164
4 salesman	2164	3909	6073
3 salesman	1981	4066	6047
2 salesman	2005	5797	7802

Table 6.6 System performance measure

Number of Salesman	Total Passenger Miles	Total Vehicle Miles
2 salesperson	13058	668
3 salesperson	9099.6	660
4 salesperson	8282	721
5 salesperson	6878	691
Kuah & Perl, 1989	7926	703

Clearly, from Table 6.6, to select the best solution from the different solutions generated by considering the feeder bus user cost, 3 route structure has the minimum cost while the 5 route structure has the lowest operating cost. A comparison between feeder bus operating cost and feeder bus user cost for all solutions using pareto efficiency was used, and the 5 route structure has the minimum total system cost as shown in the Figures 6.3 and 6.4 below. Therefore, solution with 5 route structure is selected as the best solution considering both user and operators cost. Only the 5 route structure solution generated by the algorithm is slightly better than those generated by Kuah & Perl,1989.

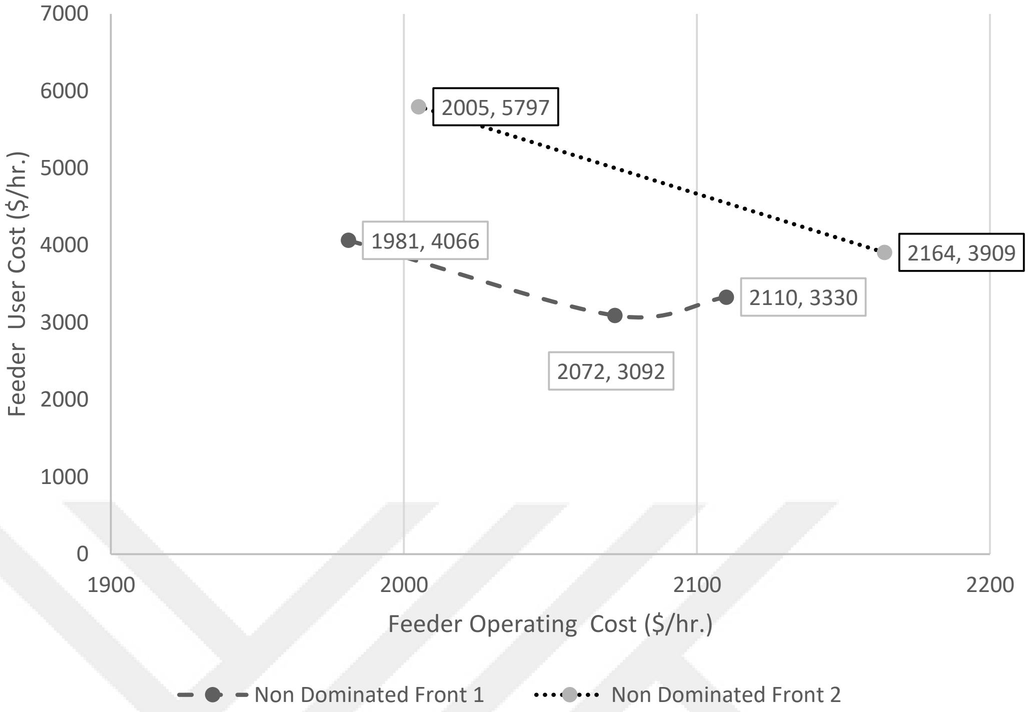

Figure 6.3 Pareto efficiency

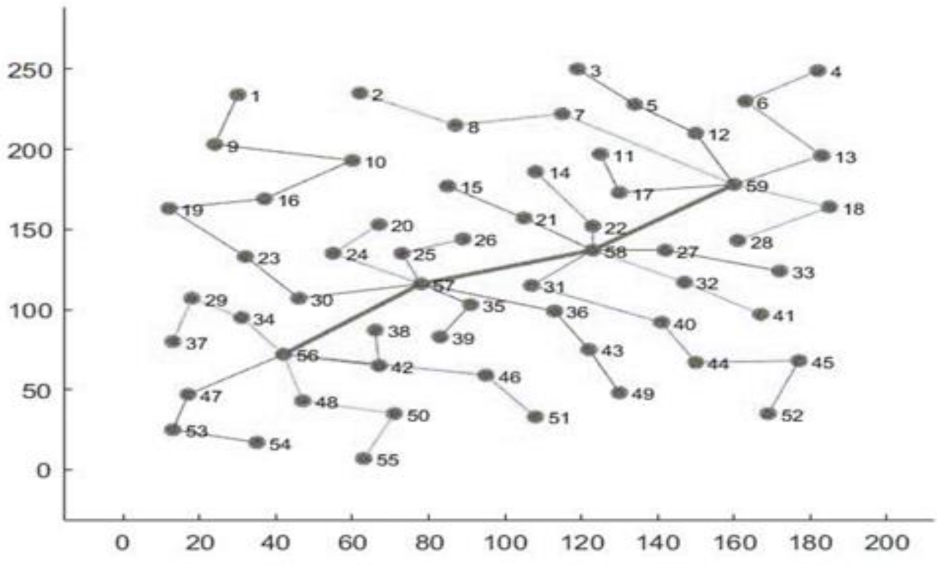

Figure 6.4 5 Route solution

CHAPTER SEVEN
APPLICATION TO BUS ROUTE NETWORK

7.1 Data Module

In this section, existing feeder bus routes to a busy train station is taken as an example to implement the proposed methodology. The data is extracted from Eshot website and database e management system which includes the location of bus stops and coordinates. In this example, the real network connectivity was not used but an assumption that all locations are connected. It is comprised of 5 routes all terminating at the train station. The figure 7.1 shows the feeder bus routes network associated with this particular train station. The network has some bus stops repeated in the different bus routes present. Hence, evaluating the existing route network assuming fixed demand and other local cost parameters is shown in the Table 5.2.

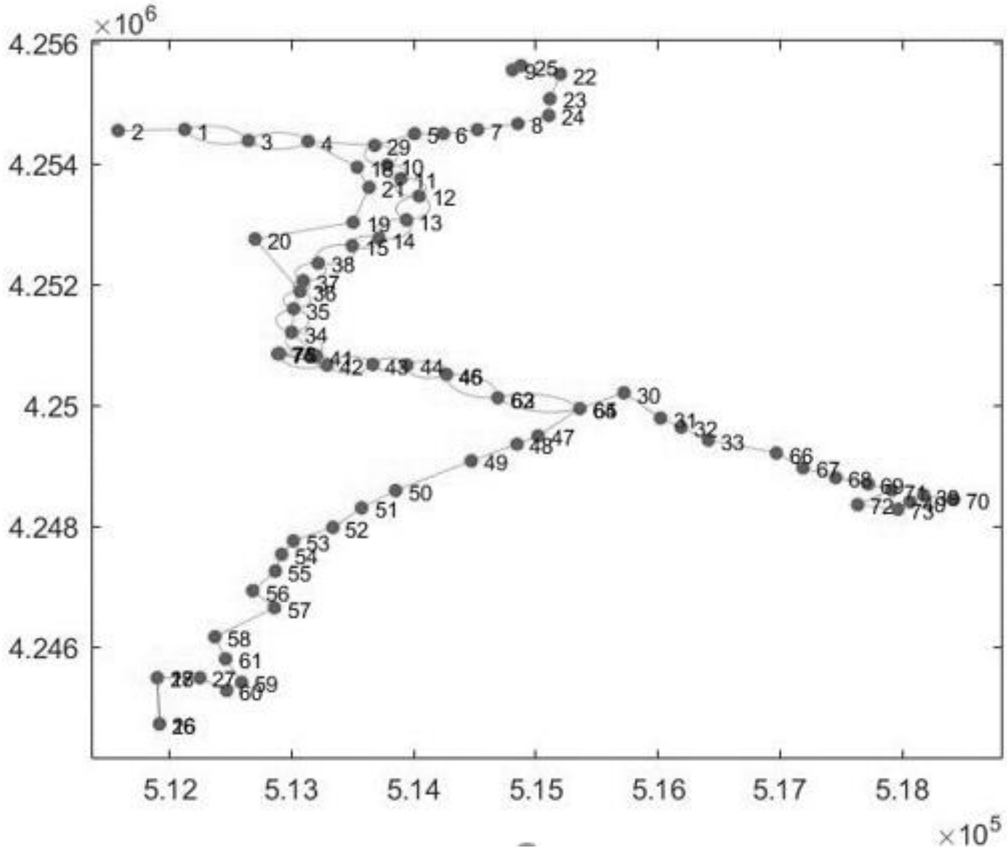

Figure 7.1 Existing sirinyer station feeder bus routes

The existing system is shown in the Table 7.1 below. In this section the existing feeder bus route network will be subjected to the algorithm by testing two cases (before clustering and after clustering was carried out).

Fixed demand was assumed such that each stop capacity is equal 50 passengers per hour and the total demand of the system will be 5250 passengers per hour. Since only a single station is being analysed and the final destination of these feeder routes are to the station there will be no need of clustering the stops to a particular station. Since only a single station is being analysed and the final destination of these feeder routes are to the station there will be no need of clustering the stops to a particular station. For this existing feeder bus route network, we noticed that for all routes some stops are repeated hence making the journey long and expensive for the operator to run. GAMTSP was applied on the existing system without clustering the bus stops, on the other hand these stops are clustered in such a way as to increase the capacity of the of feeder bus stops. If we assume the capacity of each stop is increases to 250 therefore the total number of expected clusters will be 20 feeder bus stops as shown in the Figure 7.2. K-medoid algorithm was used for the clustering

Table 7.1 Existing route structure

SN	Route Structure
1	2 1 3 4 29 10 11 12 13 14 15 39 38 37 36 35 42 75
2	26 28 16 17 27 61 60 62 59 58 57 56 55 54 53 52 51 50 49 48 66 65 6463 47 46 45 44 43 75
3	2 1 3 4 18 21 19 20 37 36 35 42 43 75
4	9 25 22 23 24 8 7 6 5 29 10 11 12 13 14 15 39 38 37 36 35 42 43 75
5	71 40 41 74 73 72 70 69 68 67 34 56 55 54 66 65 64 63 47 46 45 44 43 75

.

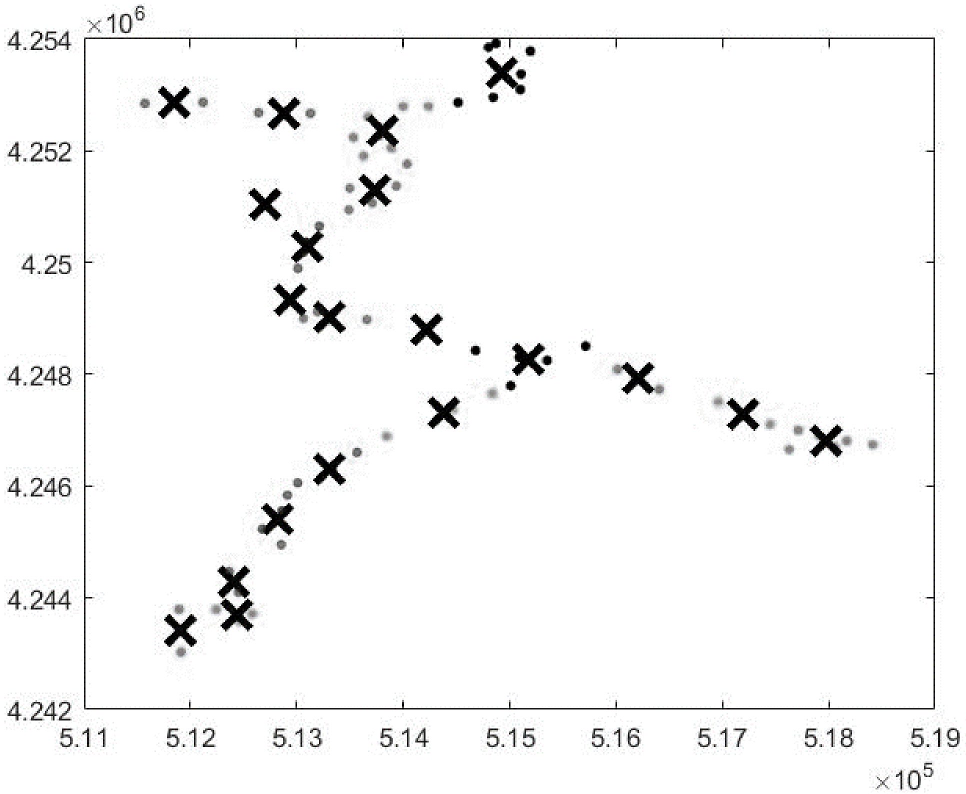

Figure 7.2 Clustering of bus stops

7.2 Solution Module

Before and after clustering, the MTSP algorithm was applied to the problem to generate a 5 route structure solution in order to compare well with existing feeder bus route network. The algorithm was repeated 10 times each, and the solution with minimum total distance representing the best solution was selected. See the table for the statistical analysis. There exists a huge variation in the results for the solutions generated when clustering was not carried out. This is because genetic algorithm is stochastic and also the network connectivity is such that all stops are connected to one another creating a plethora of alternatives. On the other hand, when the algorithm was applied with clustering it can be seen the variations are much milder.

Table 7.2: Statistical analysis

SN	Algorithm with clustering	Algorithm without clustering
Mean distance	26799.9	36847.1
Standard Error	200.7839884	569.5192797
Mediandistancce	26524.5	36871
Mode distance	26347	#N/A
Standard Deviation	634.9347211	1800.978095
Sample Variance	403142.1	3243522.1

In running the algorithm, 10,000 iterations with no constraints and a 5 route structure was selected for running the genetic algorithm. For the selected solutions, the best solution history are depicted in the Figures 7.3 and 7.4 below. Before 5000 number of iteration the solutions largely changes and remain unstable but afterwards the solution remained stable until the stopping criteria is reached at 10000.

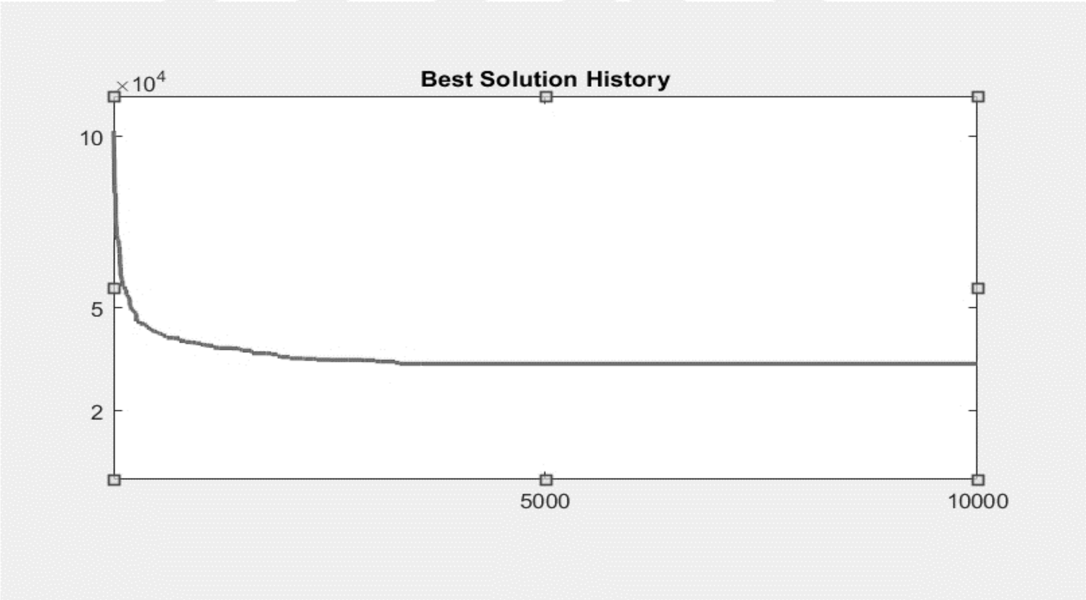

Figure 7.3 Best solution history (without clustering)

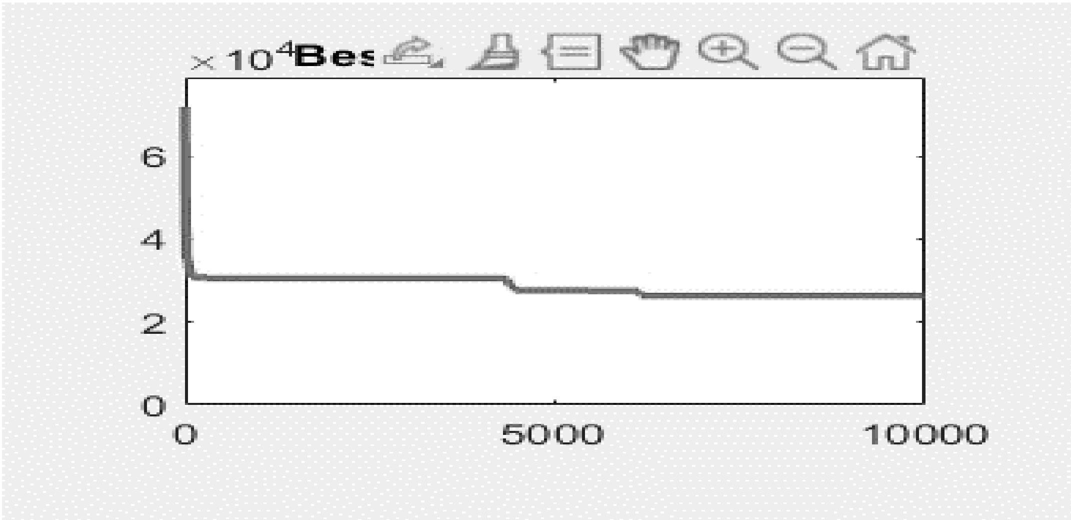

Figure 7.4 Best solution history (With Clustering)

7.3 Evaluation Module

The result of the evaluation the solution is reflected in the Figures 7.5, 7.6, and 7.7 below. When the algorithm was applied to existing condition with and without clustering, the total demand of the system is reduced which is largely due to the assumption that all stops have fixed demand and for the existing feeder bus route systems some feeder bus stops are repeated in the various routes. Most importantly all demand nodes are served which reflects the total service coverage if we assume bus stops provide the total potential demand. The general system performance of total passenger kilometre and vehicle kilometre of the system decreases. A 29% decrease in vehicle kilometre is noticed from existing condition to the solution without clustering while additional decrease of about 12% is seen from algorithm before clustering to algorithm with clustering. A similar pattern can be seen in the Figure 7.7 which shows a sharp decrease of all cost components for both user and operator costs. This is largely so because the route structure was designed based on shortest distance of the system. Bus operating cost and user waiting cost decreases by 29% from existing condition when the algorithm was applied without clustering and it decreases by 36% when applied with clustering algorithm. A similar trend is noticed for bus riding cost which is 44% and 52% respectively.

The sharp decrease in both cost components is largely due to the fact that the total route length is minimized thereby minimizing both cost components. A careful consideration of the results shows that operating cost is much higher than the riding cost which confirms the fact that most public transit services are subsidized.

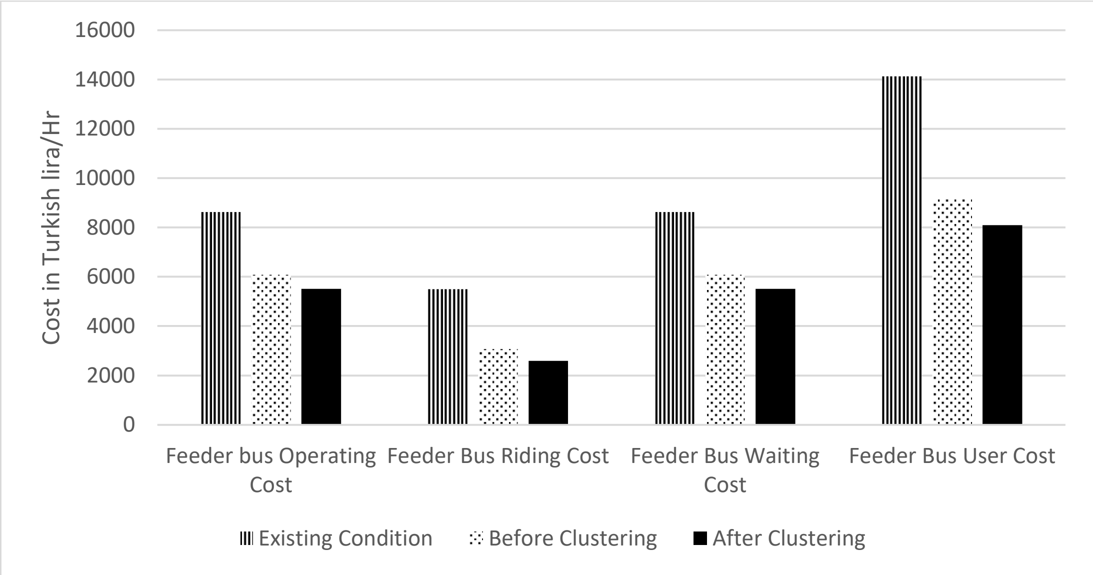

Figure 7.5 Cost components

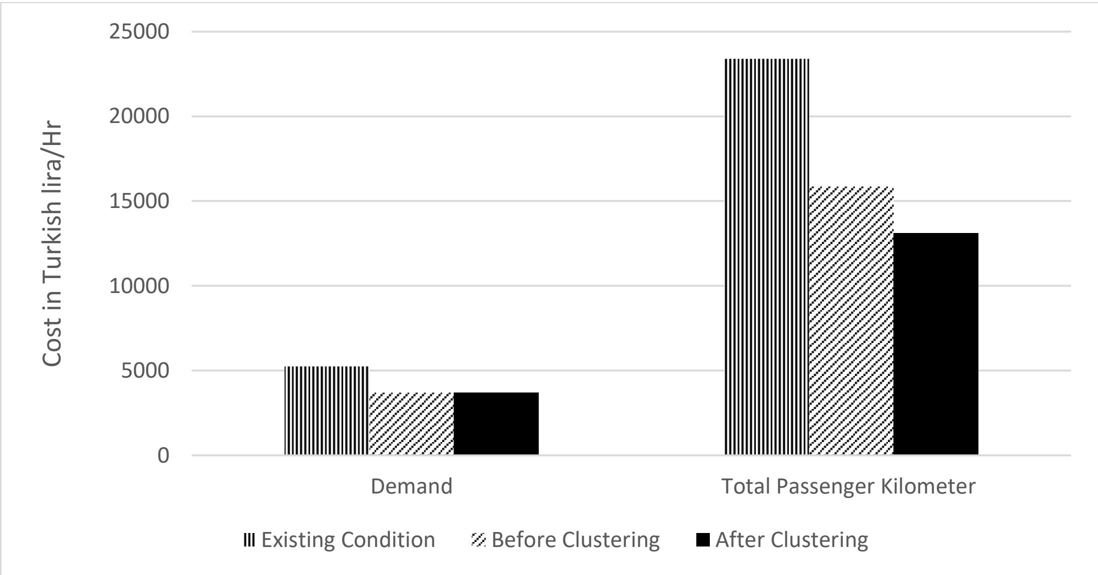

Figure 7.6 Demand and total passenger kilometer

Figure 7.7 Frequency, total vehicle kilometer and total route length

This analysis shows that the algorithm with clustering inclusive provides a better solution when compared against algorithm without clustering and existing base condition. Consequently, the algorithm was implemented with clustering to design feeder bus routes for existing train station by generating alternative solutions.

Five alternative solutions were generated and evaluated to allow for the selection of best route which satisfies both the user and operator of the system. See Figure A.5 for minimum distance and best solution history for all 5 alternatives.

Figure 7.8 shows a bar chart of the cost components of different alternatives. As the number of routes increases the various components there is a comparative reduction in user and operator costs for all alternatives. This reduction witnessed may be as a result overall shorter distance obtained from genetic algorithm based solutions. This also strengthens the fact that operators prefer a bus route that covers the demand points which implies that more ridership and shorter distances since it will reflect in the reduction of operating cost in terms of lower driver wages and lower fuel cost. Similarly, the user will prefer bus routes with minimal cost. The overall system performance measure is measured through passenger vehicle kilometre and vehicle kilometre. In the results also shown in the Table 7.3, only a marginal increase in the total route length as the number routes increases which is also reflected in the other system performance measures.

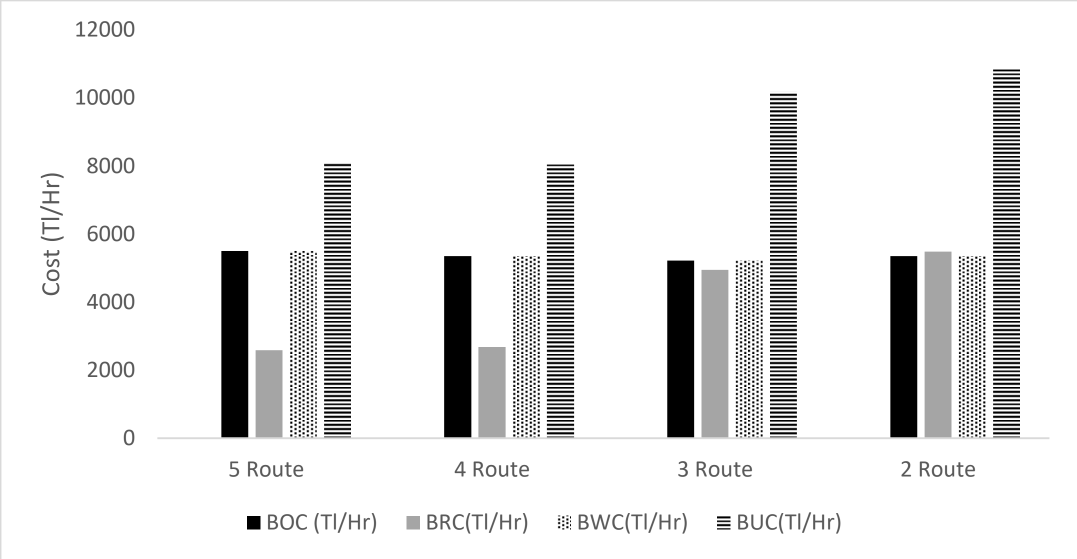

Figure 7.8 Cost components of alternatives

Table 7.3 System Performance Measures

Performance	5 Route	4 Route	3 Route	2 Route
TVK(Veh-Km)	136.2836123	132.52994	129.197	132.46103
TRL(Km)	26.676	25.418	24.131	24.891
TPK(Passenger Km)	13100.95	13013.15	26911.65	30288

Considering the solution that makes profit, the 2 route structure is profitable since riding cost slightly greater than the operating cost but this route structure has more impact on the users since it costs them more in both riding and waiting costs. While in terms of user cost the best solution is one with the minimal user cost which is the 4 route structure solution but the when considering operator cost perspective, the 3 route structure solution has the minimum value. The 4 and 3 route structure solutions are better than the other two solutions but they do not dominate one another. No single solution seems to better in both user and operators cost. See Figure 7.8.

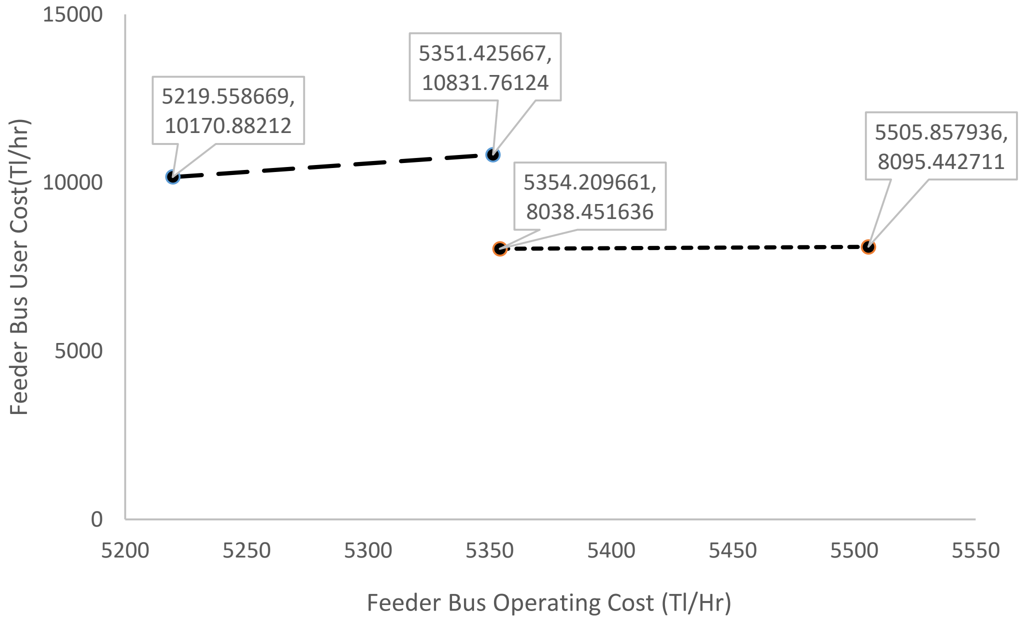

Figure 7.9 Pareto efficiency

Therefore, since no solution dominates, hence ,we look at the single objective of minimizing the total system cost and the 4 route structure has the minimum total system cost as seen in figure 7.9.

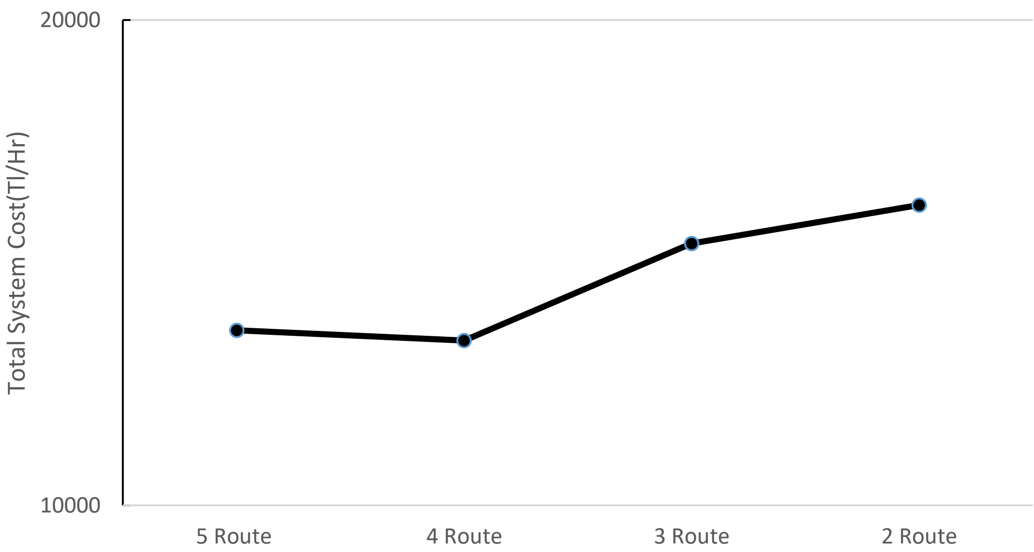

Figure 7.10 Total system cost

In summary, this chapter has discussed the application of the methodology on an existing feeder bus route network comprising 5 bus routes with 75 stops and one train station. The algorithm was applied to this problem in two different cases. The first case is without clustering and the second case includes clustering of the bus stops and capacity of bus stops. The results of the algorithm is discussed along the lines of main cost components: user cost, operation cost, and also other system performance measure; total vehicle kilometre, Passenger kilometre, frequency and total route length

In conclusion, the comparison and discussion of the results of existing feeder bus routes, algorithm application without clustering, and application of algorithm with capacitated clustering were presented accordingly.

CHAPTER EIGHT
DESIGN OF NEW FEEDER BUS ROUTES USING SMART CARD DATA

8.1 Introduction

As one of the major objectives of this study is to design feeder bus route by generating a relatively unbiased origin –destination matrix from smart card data rather than using comparatively tedious conventional method of taking on board surveys which is equally capital intensive. This chapter focuses extracting the origin – destination matrix from smart card data, it also uses coordinate location of existing transit network which includes bus stop location, train station location and existing transit network map in order to model a real life scenario. In the previous sections it was assumed that all stops are totally connected making a complete graph, however, it is nearly impossible to have all nodes completely connected without going through another node indirectly. This was a major concern since when applying genetic algorithm, the solution search space becomes very large and the stochastic nature of the solution method tends to make the optimal solution relatively unstable. Therefore, using superimposed transit network we tried to follow the existing transit network to reduce the effect complete network connectivity. Also, to locate potential feeder bus stop location from existing bus stop location, a clustering methodology was introduced to limit the capacity of the potential feeder bus stops to desired capacity which help in reducing the number of existing bus stops which are not suitable to be used as feeder bus stops.

Furthermore, the application of genetic algorithm based multiple travelling salesman problem used in the previous chapters will be used to establish the route structure and consequently the evaluation of the different route structure alternatives, discussion and selection of the most suitable feeder bus route. Three scenarios will be tested;

1. Design of feeder bus routes(without clustering, with multiple stations, and assuming total network connectivity

2. Design of feeder bus routes(with clustering, multiple stations, and total network connectivity)

3. Design of feeder bus routes(with clustering , single station and arbitrary network)

4. Design of feeder bus routes (six-station problem)

8.2 Design of Feeder Bus Routes (Without Clustering , and Multiple Station)

8.2.1 Data Module

To formulate an origin-destination matrix of the journeys using the public transport system which includes in the city by analysing the smart card information. An algorithm was formulated in the MATLAB programming language within the scope of the study and the flow diagram is presented in Figure 5.2. This exercise rigorous and requires a lot of computer soft wares to prepare the smart data into useable data. It includes cleaning and coding of the smart card data since not all collected data are relevant and there are also missing data. Coordinates of bus stop and train station was extracted and transfer time was calculated for each smart card data. In Izmir, the transfer time is less or equal 90 minutes. Therefore, if transfer time is less than 90 minutes, then we isolate all bus stops that uses that train station and no any other modes. Then the data is recorded. From this algorithm, 321 stops were isolated to imply that these stops usage precede the swiping of the smart card at the train station. In this case we isolate the station at Sirinyer. The Figure 8.1 shows the results of this module. The square dots represent the location of existing bus stops that use this particular station respectively.

Figure 8.1 321 stops associated with Sirinyer station

Not all stations have significant demand therefore the stops were ranked based on the card usage at the various stops. The isolated stops were limited to those stops with at least 3 passengers per hour thereby reducing the stops to 87 stops with a total demand of 633 passengers per hour. See the Figure 8.2 below for the ranking of bus stops. It can be seen that a huge cluster still surrounds the said train station even though there are isolated stops with significant demand that are far away. Feeder bus service differ from conventional bus service in that they are mostly servicing suburban areas with high commuter volumes. This implies most passengers gather in one location to proceed to somewhat common destination i.e. bus stops may be crammed. Therefore, the need locate suitable position for siting feeder bus stops and it is also pertinent to take into consideration the capacity of the stops themselves in terms of passenger.

Figure 8.2: Ranking of stops based on demand

8.2.2 Solution Module

After 87 stops have been identified, four stations were identified that falls within reasonable connectivity to theses stops. Distance between these stops and stations were estimated using euclidean distance and it was found that only 3 stations could be associated to these stops based on nearest neighbours comparism. Therefore, the algorithm was run for single , double and triple station scenario and also varying the number of salesman from 2 to 5. The result of the distance is shown in the Figure 8.3.

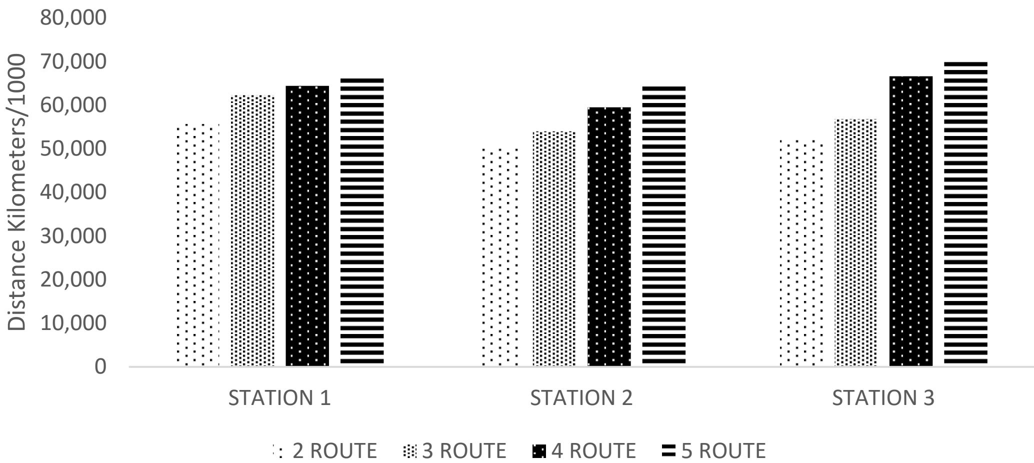

Figure 8.3 Minimum distance for 1, 2, 3 alternative station solutions

In all the scenarios evaluated for all the three station solutions, the two route structure solutions have the minimum distance in all station scenario, hence an evaluation of the two route structure solutions for all stations. The Table 8.1 shows the route structure of all the scenarios. As the number stations increases the more shorter routes are created.

Table 8.1 Route structure

Single station2 route Solution	18 7 83 76 57 1 14 3 62 5 77 27 30 51 6 21 37 33 26 48 78 32 35 10 42 15 12 68 19 28 4 46 8 13 60 81 25 65 88
	17 38 85 43 6134 64 74 47 52 20 79 75 70 55 71 63 49 40 59 87 66 53 29 84 36 54 72 69 11 23 16 80 24 41 31 73 44 9 86 56 67 22 50 82 39 45 88
2 station 2 route solution	36 84 54 21 30 26 48 78 32 89
	23 16 80 11 69 72 37 2 33 24 6 51 41 31 5 62 27 77 35 3 10 42 15 68 89
	17 38 85 43 61 34 64 74 39 47 52 20 79 5067 28 19 12 46 88
	18 7 29 53 83 66 76 59 86 87 9 73 44 14 1 57 56 40 49 63 75 71 55 82 70 22 4 8 13 81 60 25 65 45 88
3 station 2 route solution	18 7 29 53 83 90
	85 38 17 90
	31 41 5 62 3 10 42 15 68 89
	23 16 80 11 69 72 54 84 36 6 51 30 21 24 37 2 33 26 48 78 27 77 35 32 89
	61 43 34 64 88
	74 65 45 39 47 52 20 79 70 82 50 55 71 75 63 49 40 59 76 66 87 86 9 73 44 57 56 1 14 19 12 46 4 28 67 22 8 13 81 60 25 88

8.2.3 Evaluation Module

Evaluating the cost implication of the solutions, it can be seen that as we increase the number stations all cost components decreases. This is as a result of shorter routes been created and also since the demand at these stops are not very large, the bus riding cost for double and triple station solution is significant, while the operating cost (BOC) and waiting cost (BWC) are marginal. See the figure 8.4 below;

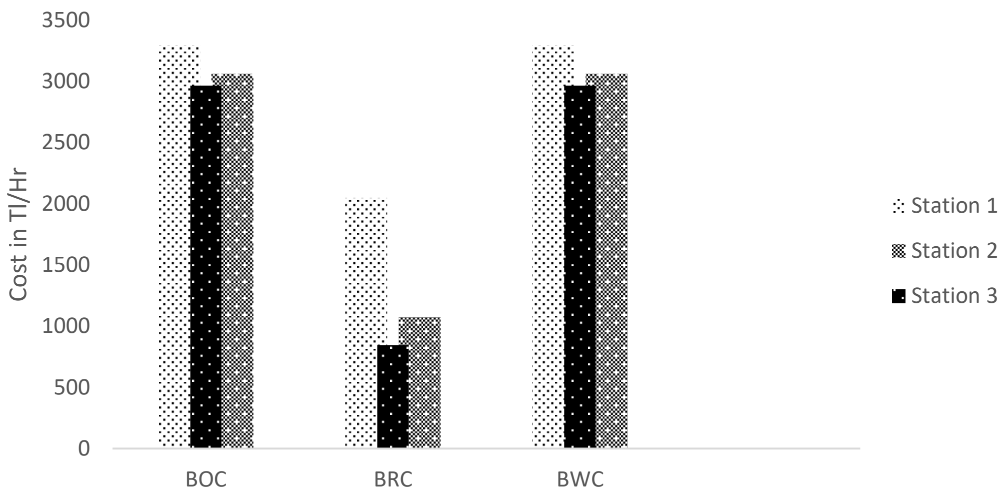

Figure 8.4 Cost comparism between solutions with 1,2,3 stations

A similar trend is also noticeable in Figure 8.5, where the system performances decreases with increase in the number of station included in the solution. The percentage difference in system performance and cost components is shown Figure 8.6 These results are when clustering was not included.

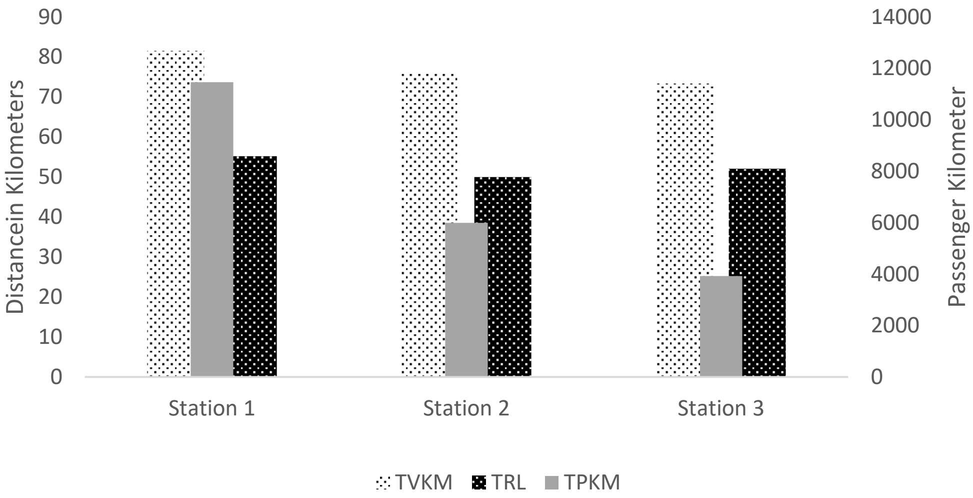

Figure 8.5 Comparism of system performance measures (solutions 1,2,3 stations)

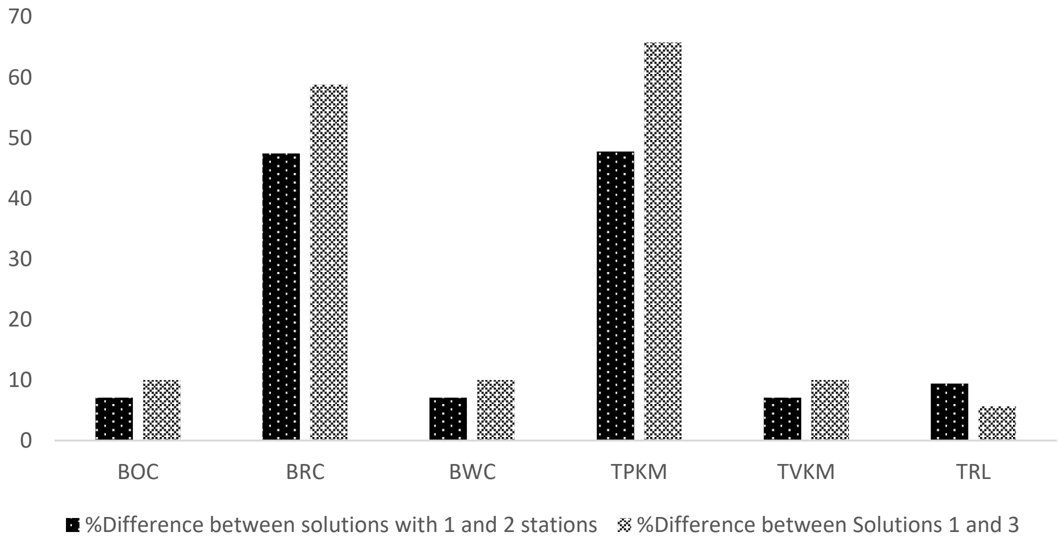

Figure 8.6 Percentage difference between solutions with 1,2,3 stations

8.3 Design of Feeder bus routes with Clustering Algorithm and Multiple station

The clustering is considered has 74 existing bus stops whose demands are known and are distributed in (x ,y) coordinates. These 74 bus stops are grouped to form 13 clusters. Each cluster has n number of bus stops number attached to them with the condition that summation of all bus stops in all clusters is equal the total number of bus stops. This algorithm differs from K Means clustering algorithm since it has a known centroid while K-mean searches for the centroid and find their mean. In this case since the centroid is known, we find the distance for all remainder nodes to the selected initial centroids. Table 13 gives the ranking of all remainder nodes to the initial centroids. The ash colour means that the bus stop has been assigned to the column of that initial centroid. When all stops have been assigned and the capacity of a stop is exceeded, then the stop is split into two to accommodate another cluster.

After the shortest distance to all centroids are assigned, a group of clusters formed around the initial centroids and the table 13 ensued. But we have a constraint that no cluster demand shall exceed 50 passengers per hour. Hence the red coloured cells exceed and must be split into more centroids to accommodate the excessive demand. Centroids 1,2 4 and 13 are split further to accommodate more demand and the final feeder bus nodes their respective demands are shown in Table 8.2.

Table 8.2 Nearest bus stops to centroids

STOPS	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
1	23.85844	2193.637	433.0367	1330.26226	483.8175	1182.942	9952.677	1566.324	1255.057	669.2751	2741.708	1236.925	1825.192
2	1009.733	2796.388	703.0692	687.869441	1338.465	2008.114	10492.13	824.9702	2256.831	400.0469	3512.133	288.6544	1016.942
3	2999.041	1351.932	3131.556	4299.75036	2526.473	1822.615	10838.65	4527.825	2417.87	3426.402	479.7791	4102.566	4776.643
4	5608.899	7806.693	5643.156	4454.37556	6114.852	6807.969	9099.8	4290.826	6210.568	5425.265	8356.626	4843.895	4126.393
5	11525.96	12249.42	11966.19	12040.2209	11612.49	11564.83	4077.282	12198.94	10433.41	12138.61	11736.37	12370.87	12380.59
6	815.277	2768.964	602.9524	621.53432	1207.948	1904.533	10252.09	818.4578	2038.762	351.1	3439.809	411.2419	1053.296
7	2441.581	4655.868	2534.227	1459.50367	2948.369	3631.414	8965.877	1399.251	3075.339	2366.182	5174.873	1927.479	1386.961
8	1550.085	670.5656	1511.045	2722.91091	1049.482	534.2284	10923.99	2927.751	1780.404	1773.414	1383.197	2440.538	3154.596
9	1415.068	3535.47	1361.009	191.381413	1894.886	2605.133	9866.441	269.1093	2435.559	1139.134	4161.617	666.5405	499.3338
10	3268.353	1594.132	3405.896	4571.32491	2798.858	2096.43	10903.66	4800.157	2644.906	3701.126	703.6635	4377.169	5049.676
11	1723.149	517.7865	1728.626	2937.99368	1216.324	581.7263	10890.29	3148.661	1778.006	2001.531	1133.137	2673.62	3380.862
12	2094.696	4052.985	1921.916	800.547067	2528.328	3227.616	10360.3	557.5136	3198.68	1637.773	4757.612	968.1304	291.7008
13	1753.46	709.6087	1603.521	2788.9361	1292.95	945.2439	11331.76	2971.433	2184.273	1817.309	1579.493	2446.086	3176.141
14	1027.255	1375.335	834.9676	2027.33534	653.0406	807.3508	10842.45	2217.86	1784.406	1058.709	2127.082	1709.74	2433.171
15	859.609	2997.071	835.9017	461.792107	1338.929	2049.016	9899.757	704.2183	1949.839	674.3554	3606.452	591.0332	969.3153
16	2617.11	3701.144	3057.683	3378.1965	2702.557	2774.971	7633.025	3589.766	1551.244	3253.531	3574.378	3598.199	3828.359
17	1268.576	1003.214	1174.071	2383.41391	795.6433	564.4252	10874.48	2582.849	1738.923	1424.886	1740.283	2087.228	2805.293
18	587.899	1778.901	872.4054	1900.26495	303.9416	691.4031	9929.794	2137.349	885.2613	1168.193	2220.232	1792.906	2396.613
19	1241.64	2226.064	803.1791	1499.52861	1252.549	1718.137	11122.13	1612.947	2413.584	705.9583	3040.027	1053.893	1761.835
20	1896.099	343.544	1835.349	3045.07051	1397.204	844.5278	11192.8	3243.554	2070.707	2084.674	1170.737	2741.489	3463.457
21	4657.112	6565.747	4483.079	3346.79265	5095.646	5792.324	11180.08	3107.841	5690.198	4190.961	7308.896	3514.195	2848.902
22	828.003	1840.167	472.7614	1589.59762	707.3095	1177.947	10763.29	1768.02	1888.99	619.1159	2588.59	1247.52	1974.968
23	1486.662	910.6101	1594.507	2769.55787	996.1722	286.5088	10465.46	2994.462	1358.303	1890.808	1276.921	2566.197	3241.008
24	2011.014	220.4948	1989.854	3201.70613	1505.179	879.2157	11133.22	3406.604	2051.43	2251.48	952.8464	2916.461	3632.634
25	5958.026	8132.891	5959.307	4754.51271	6460.584	7162.137	9582.323	4573.419	6628.724	5725.224	8718.226	5115.511	4387.183
26	2445.895	4599.516	2425.757	1226.30884	2940.905	3648.685	9592.293	1070.048	3304.667	2197.87	5207.798	1628.616	948.6634
27	888.4203	3006.159	1233.62	1254.28132	1326.689	1921.483	9129.379	1476.545	1294.195	1296.091	3416.143	1483.109	1728.112
28	755.4363	1492.559	889.1662	2037.2383	263.3518	450.0934	10228.14	2264.71	1128.225	1192.916	2009.208	1858.501	2514.651
29	930.3911	2744.248	631.3509	702.256545	1264.888	1939.559	10443.33	858.5229	2177.58	331.4314	3450.435	346.1076	1063.433
30	5155.602	7167.642	5033.712	3848.05993	5621.524	6328.895	10827.85	3620.939	6092.672	4754.623	7870.987	4085.097	3377.932
31	691.6852	1927.096	1052.738	1968.78983	546.2811	828.1585	9690.899	2210.513	644.2701	1332.911	2280.974	1917.949	2474.428
32	2357.212	4454.38	2282.051	1070.71581	2836.25	3546.479	9896.224	872.5898	3313.885	2030.635	5101.284	1414.824	695.3068
33	1440.249	3346.529	1222.28	375.861133	1841.506	2532.132	10394.64	344.2239	2622.549	932.4666	4053.412	261.7056	473.5991
34	2489.117	4675.342	2512.627	1345.66002	2992.051	3693.358	9340.715	1224.32	3262.995	2305.832	5250.309	1780.515	1143.877
35	1700.181	1069.692	1462.894	2591.46595	1306.046	1159.325	11465.61	2754.927	2342.588	1627.837	1949.236	2211.124	2941.639
36	1063.764	3223.249	1361.118	1155.07967	1528.987	2144.951	9061.339	1358.821	1516.302	1375.165	3648.403	1451.293	1597.114
37	1585.261	3775.467	1625.399	573.499933	2087.892	2789.788	9504.473	609.1829	2442.051	1446.567	4347.519	1049.602	748.5224
38	1088.611	1162.433	1168.082	2350.55385	585.271	150.13	10423.83	2572.82	1278.019	1464.041	1684.725	2139.472	2817.306
39	2443.769	4650.641	2509.806	1395.18854	2950.315	3641.693	9113.198	1312.76	3135.461	2326.495	5191.929	1854.119	1277.452
40	2599.615	3844.564	3039.189	3260.63367	2736.222	2876.31	7516.06	3463.004	1632.666	3213.957	3769.376	3511.497	3693.265
41	1711.808	658.8367	1777.154	2972.17973	1210.485	511.2837	10687.2	3191.506	1607.346	2064.487	1061.887	2741.134	3432.276
42	1649.093	3860.737	1743.43	782.677025	2155.936	2845.483	9279.666	833.9693	2396.041	1593.656	4396.082	1254.521	965.4175
43	624.4301	2709.538	1014.184	1317.335	1034.642	1626.192	9329.516	1555.096	1104.518	1132.916	3130.519	1451.223	1817.728
44	642.9082	2544.484	1077.4	1561.70375	935.7442	1448.129	9304.777	1801.917	846.5218	1254.025	2915.614	1656.412	2066.116
45	859.609	2997.071	835.9017	461.792107	1338.929	2049.016	9899.757	704.2183	1949.839	674.3554	3606.452	591.0332	969.3153
46	1121.309	3004.805	1534.711	1710.74608	1441.543	1905.939	8815.191	1931.214	997.5156	1660.512	3298.061	1926.799	2180.112
47	1969.895	3805.844	1732.34	785.408053	2360.653	3041.633	10626.05	580.2367	3145.259	1433.298	4543.829	760.5593	400.994
48	5595.342	7516.994	5431.024	4282.96591	6040.23	6739.229	11493.3	4046.252	6599.558	5140.245	8259.291	4463.638	3790.159
49	437.41	1780.638	538.3157	1682.02825	98.64427	801.1034	10207.47	1907.612	1250.081	842.7315	2355.691	1502.041	2156.447
50	1428.28	3589.204	1702.391	1262.50356	1897.01	2507.118	8808.296	1422.172	1792.377	1682.177	3997.126	1640.965	1624.831
51	4377.521	6217.119	4173.471	3080.1188	4796.753	5483.597	11350.7	2837.597	5459.844	3875.679	6980.643	3200.892	2573.209
52	2336.951	4394.396	2230.578	1029.94029	2804.125	3512.962	10061.1	810.5102	3341.521	1966.918	5062.009	1327.849	596.7641
53	1997.915	3552.699	2428.469	2560.25981	2218.923	2501.376	7971.421	2760.581	1290.076	2573.635	3639	2825.371	2990.245
54	1087.024	3294.302	1185.181	555.817759	1593.015	2288.971	9534.049	758.3435	1963.461	1066.408	3844.563	909.8988	1003.056
55	1020.846	2788.78	705.7601	707.138193	1342.018	2008.164	10515.72	839.8584	2269.957	401.7069	3508.441	297.4101	1027.789
56	2533.934	967.9877	2654.086	3828.50801	2054.329	1347.247	10747.27	4054.815	2054.598	2947.882	355.2129	3624.312	4302.002
57	2019.039	4209.457	2053.526	924.179832	2522.38	3223.186	9396.364	855.4464	2822.12	1860.025	4780.133	1386.424	863.0519
58	1640.812	3856.74	1799.96	986.752414	2143.655	2811.212	9042.225	1073.559	2251.98	1690.849	4345.548	1441.727	1223.224
59	1944.194	746.5251	2053.129	3231.05614	1457.378	748.0361	10602.45	3455.928	1633.664	2347.113	827.3469	3023.438	3702.03
60	701.1751	1902.573	1055.029	1983.60834	536.3924	803.7009	9709.64	2225.074	652.2868	1337.596	2257.352	1927.579	2488.708
61	3134.612	5211.248	3046.197	1839.97541	3611.279	4321.25	10027.16	1626.678	4069.574	2783.764	5873.898	2139.32	1408.699
62	1501.03	3703.264	1713.331	1079.12713	1993.558	2639.649	8946.587	1210.055	2018.417	1644.979	4158.057	1498.556	1392.888
63	1367.74	3157.359	1787.278	1934.27402	1656.092	2064.763	8570.061	2147.105	1026.174	1917.578	3390.776	2172.504	2389.175
64	951.4937	1586.913	678.5183	1838.66066	682.6656	996.9477	10847.99	2020.845	1863.058	866.8711	2348.114	1502.11	2229.416
65	1899.67	1346.022	1597.506	2634.79846	1563.785	1501.902	11751.33	2773.438	2656.394	1705.299	2237.236	2216.722	2935.406
66	2223.407	4436.017	2385.463	1464.07182	2723.344	3379.413	8699.198	1476.063	2718.737	2264.117	4899.516	1939.375	1540.279
67	2760.96	1113.317	2881.165	4056.24153	2282.069	1574.919	10828.84	4282.56	2246.444	3174.385	302.1648	3850.951	4529.642
68	1791.733	3701.36	1585.903	568.387028	2203.635	2895.675	10437.44	361.1994	2945.302	1295.088	4415.852	619.7809	228.1129
69	1897.759	4096.272	1949.886	860.132913	2402.713	3100.554	9354.186	827.1558	2684.203	1768.73	4655.962	1332.508	877.4317
70	2118.888	3579.04	2553.982	2727.94561	2309.957	2547.821	7884.854	2930.647	1314.04	2711.122	3621.95	2983.826	3162.099
71	1586.17	916.2531	1713.024	2878.70928	1104.657	402.7589	10426.75	3105.779	1354.504	2011.293	1186.238	2685.478	3354.385
72	5913.051	8080.957	5906.347	4699.19731	6414.364	7117.611	9676.455	4514.528	6602.753	5668.795	8674.903	5053.61	4324.027
73	873.931	2595.493	1313.682	1789.60356	1087.325	1497.841	9110.668	2026.985	663.6357	1503.071	2890.169	1905.273	2289.059
74	1095.714	2760.123	1534.198	1927.85756	1304.401	1669.69	8888.849	2159.473	669.5134	1713.696	3001.383	2081.471	2416.976

Table 8.3 Assigned stops to clusters

Centroids	C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 10	C 11	C 12	C 13
	1	11	22	9	14	8	5	57	16	6	3	2	4
	27	13	64	15	18	17		69	40	19	10	33	7
	36	20		37	28	23			46	29	56	55	12
	43	24		42	31	38			53		67		21
	44	35		45	49	41			63				25
		59		50	60	71			70				26
		65		54					73				30
				58					74				32
Index				62									34
				66									39
													47
													48
													51
													52
													61
													68
													72
Average Demand (pass/hr.)	67	73	42	63	47	48	26	21	44	34	38	32	97

Table 8.4 Clusters and variable demand at stops

Clusters	x	y	Demand per hour
1	515700.6	4250248	37
1b	515830.5	4250859	30
2a	516831.7	4248341	37
2b	516694.3	4248841	36
3	515372.3	4249954	42
4a	514514.1	4250810	21
4b	514937.8	4251701	42
5	515998.5	4249838	47
6	516563.1	4249407	48
7	519933.4	4259238	26
8	514275.4	4250856	21
9	516905.6	4250606	44
10	515074.1	4250016	34
11	517723.1	4248365	38
12	514477.3	4250335	32
13a	514013.2	4250900	48
13b	511223.6	4252805	49

The average distance for all cluster centres is depicted in table 8.2.3, only cluster 7 and very high average distances of 4 Km and 1.9km.

Table 8.5 Mean distance from stops to Cluster

Cluster	Average distance (Meters)
C1	648.6762733
C2	707.667521
C3	575.6398601
C4	781.941526
C5	400.2752901
C6	408.2225132
C7	4077.281935
C8	841.3010783
C9	1143.10811
C10	462.8299083
C11	460.2050715
C12	282.5900331
C13	1988.608075

8.3.1 Data Module

From the 17 clusters we obtained we introduced four new train station and checked their proximity to the new bus stop(centroids) as we tested in the precious chapter. Considering pairwise distance it can be seen from the table 16. above, that, 10 stops were assigned to station 1 and the remainder were assigned to station 2 based on their proximity the stops. station 3 and station 4 were quite far away from the stops. This gives credence to the initial algorithm which identified the stops with respect to their usage of the train station 1. This was quite different from the results from the previous chapter and it is because every transportation problem has its own peculiarities. Hence the route structure was designed based 2 station and 1 station and the results were compared. Because the number of stations is few, a four route structure solution was explored for both cases. Keeping genetic algorithm parameters as before, because the number of nodes are reduced the algorithm was able attain the optimum distance even when repeated.

Table 8.6 Assignment of stops on introduction of Stations

Centroid	Station 1	Station 2	Station 3	Station 4
C1	2791.3	2830.2	4556.4	3766.1
C2	2837.7	3058.6	4088.6	4193.5
C3	1525.1	1798.5	3611.1	3158.5
C4	2091.9	2630.8	2915.9	4111.0
C5	10825.9	11572.2	7985.4	13142.0
C6	1283.8	1606.5	3515.7	3037.5
C7	3925.3	4065.2	4938.0	5004.7
C8	1598.2	1623.3	4066.2	2800.4
C9	1020.0	1409.7	3434.7	2916.5
C10	2579.3	3198.5	2854.4	4464.0
C11	4628.9	4315.3	6770.2	4464.3
C12	4249.3	4010.1	6266.1	4338.6
C13	2570.6	2498.5	4690.1	3340.2
C14	3196.7	3130.6	5057.4	3870.8
C15	3880.3	3745.3	5709.4	4292.3
C16	2272.1	2198.4	4530.4	3104.6
C17	5379.5	5136.6	7212.6	5354.7

The route structure obtained for both solutions are presented in the table below. The route structure is quite similar and only two routes go through station 2 to station 1. The average frequency required for both solutions marginally changes from 5 vehicles/hr for solution 1 to 4 vehicles per hour for solution 2. See Tables 8.7 and 8.8 for route structure and fundamental parameters for solution 1 and solution 2 respectively.

Table 8.7 Fundamental parameters(single station)

Solution1/Route structure	Demand	Length(Km)	Frequency(Vehicle/Hr)
7 12 1	68	13.4	2
5 8 11 1	106	1.5	7
16 3 15 18 10 1	204	3.52	7
19 13 14 17 9 4 6 1	254	7.67	5

Table 8.8 Fundamental parameters (two station)

Solution2 /Route Structure	Demand	Length (Km)	Frequency (Vehicle/Hr)
7 12 1	68	13.4	2
15 18 2 1	82	2.5	5
19 13 14 17 16 2 1	184	5.8	5
9 3 4 6 5 10 8 11 1	260	6.41	6

8.4 Comparism Between Solutions Before And After Clustering

The results of solutions with single station and double station were compared to two scenarios; before clustering and after clustering algorithm is implemented. Figures 8.7 and 8.8 shows cost component and system performance for single station solution respectively. In all cost components and system performance there is significant reduction. This is as result reduction in the number of stops and consequently reduction in the length of routes in all solutions. A similar result is noticeable in the 2 station solution in terms of cost and system performance, see Figures 8.9 and 8.10.

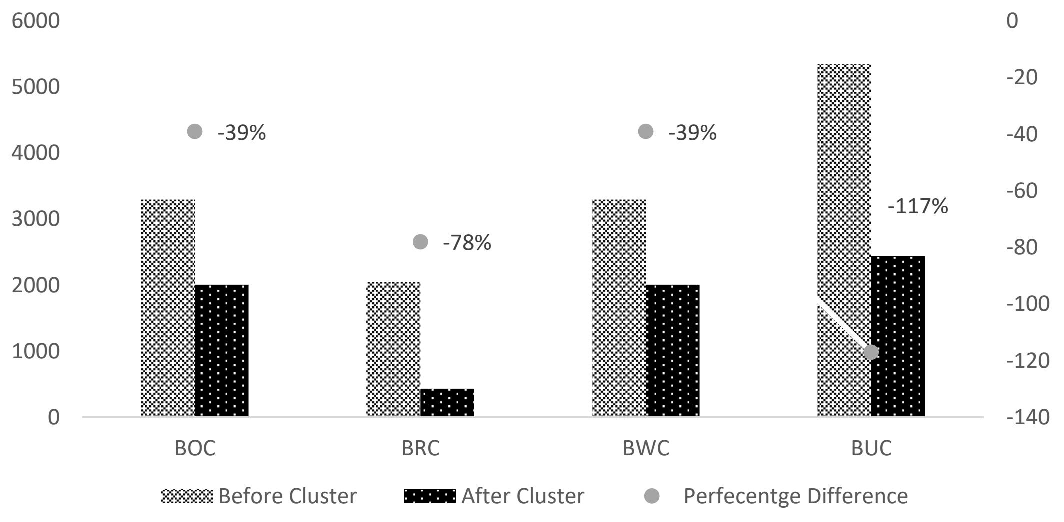

Figure 8.7 Cost Component Comparism before and after clustering (single station)

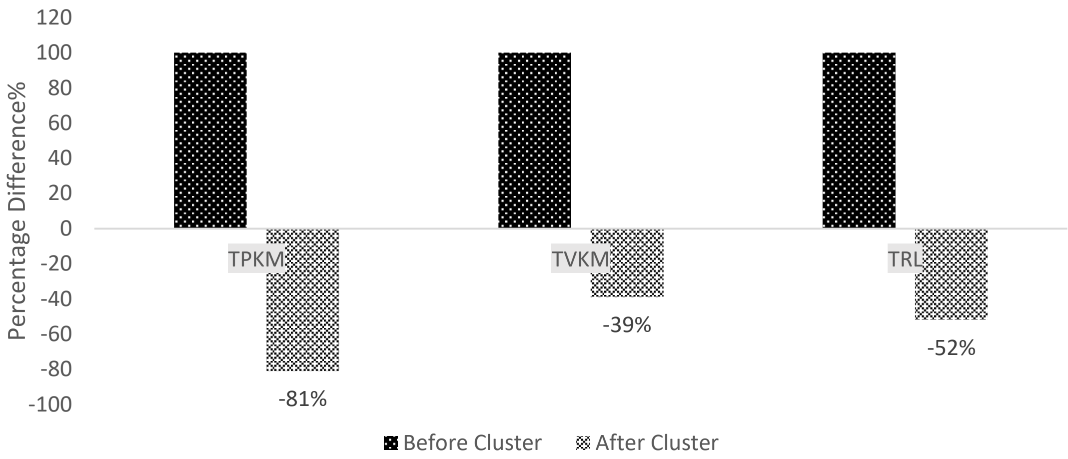

Figure 8.8 Performance measure before and after clustering (single Station)

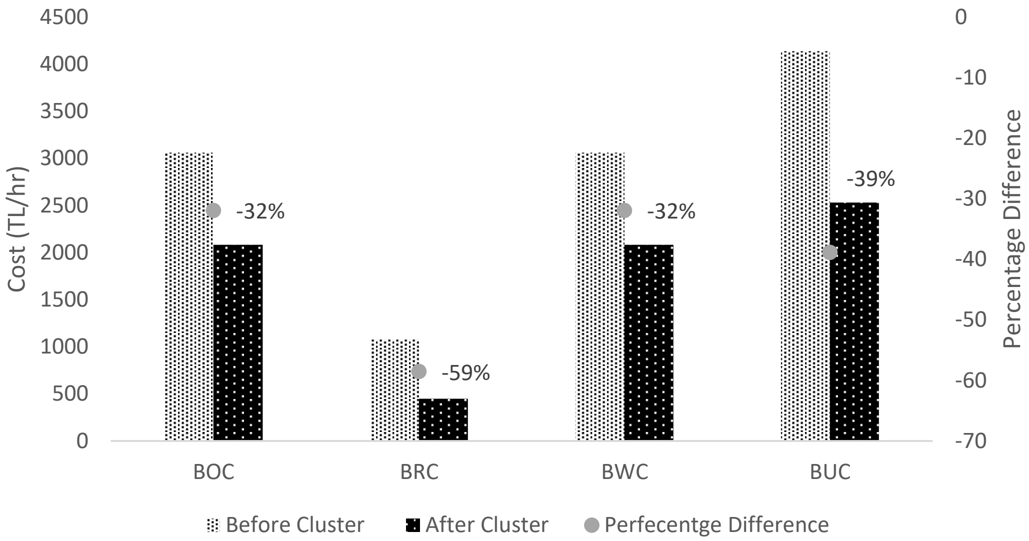

Figure 8.9 Cost component comparism before and after clustering (two station)

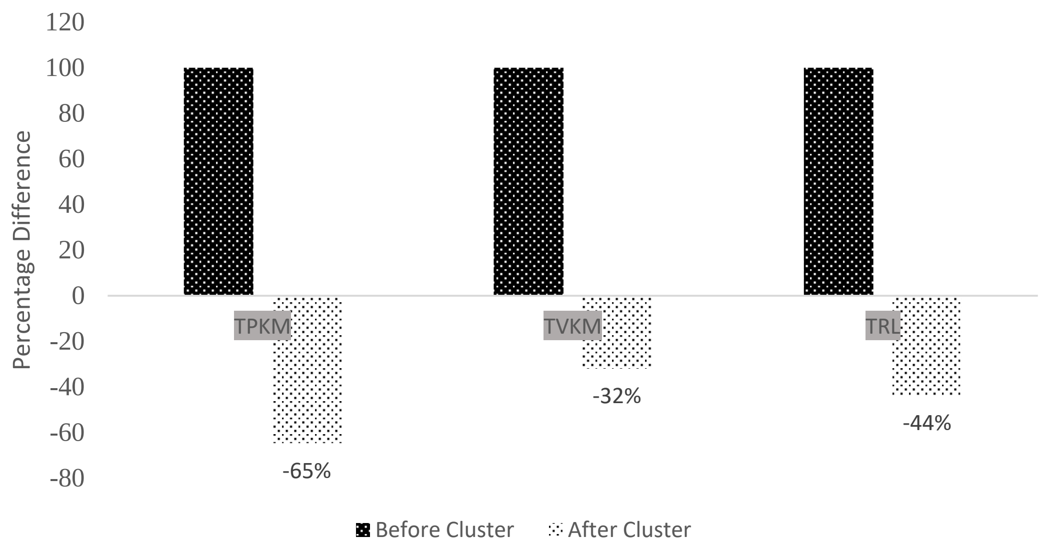

Figure 8.10 Performance measure before and after clustering (two Station)

Evaluating both solutions in terms of cost components and transit system performances measures are depicted in the Figures 8.7 and 8.8 below. The total passenger kilometre, total route length and total vehicle kilometre travelled changes by 2%, 8% and 5% when additional train station was introduced. Also, the bus operating cost, riding cost as well as waiting cost increases by 4%, 3%, 4% respectively (Figure 8.9). This increase in cost is expected even though the cost of travelling distance from train station 2 to station 1 was not included. This result is quite different from the result gotten from the benchmark study, which indicated that, as the number of train station increase so does the cost component of the user and operator cost. This is so because the benchmark problem has the number of stops evenly distributed amongst the station while for this case, they were unevenly distributed and since the number of stops were fewer.

Therefore, from the results, the solution with one station slightly edges out solution with two station because it will cost more for all stakeholders. Again in this section it is assumed that all nodes are connected. Therefore, in the next section. the result with one station will be explored further and trying to mimic real transit network and its connectivity since it is nearly impossible for all nodes to be connected directly.

8.5 Design for single Station with Clustering and Arbitrary Network

In this study, to copy the real network, the generated feeder bus stops are not directly linked necessitating the use of intermediary intersections. Major intersections were selected by superimposing the real bus network and highlighting the intersections to make a connection. The Figure 8.11 highlights the additional intersection nodes and feeder bus stops. Therefore, to find the shortest path to all nodes, an algorithm resembling Floyd Warshal was used (See Appendix A.3). It is an algorithm for finding the shortest path between all the pairs of vertices in a weighted graph. This algorithm works for both the directed and undirected weighted graphs. But, it does not work for the graphs with negative cycles (where the sum of the edges in a cycle is negative).

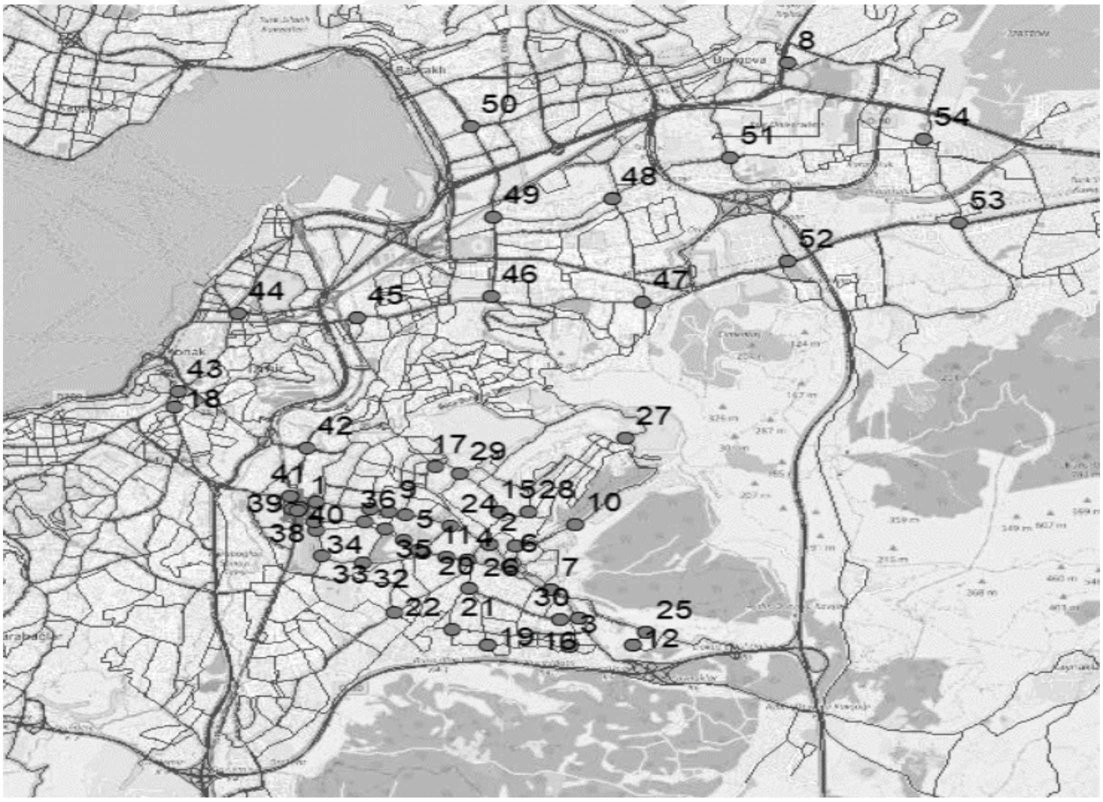

Figure 8.11: Feeder bus stops and intersections

From the above network an arbitrary network was estimated since real distances were not measured but euclidean distances between the nodes. See the Figure8.12 and Table 8.9 for the representative network and distance matrix respectively. The network connectivity as well distance matrix now serves as input for the MTSP algorithm. The parameters used for genetic algorithms are a number of salesmen, minimum tour, and number of iterations. Figure 8.13 given below shows the route structure for each case of the salesman The same parameters for the algorithm were retained that is a large stopping criteria and no constraints on the number of stops to be included in each route.

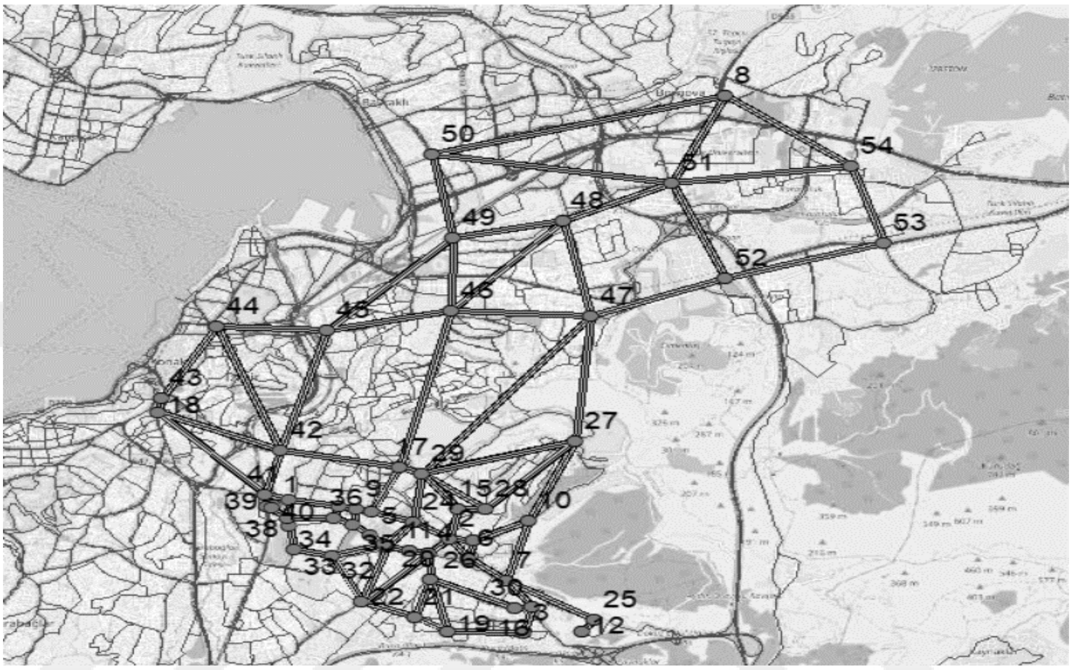

Figure 8.12 Generated network

Table 8.9 Feeder bus origin-destination distance matrix

	SIRINYER	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17
SIRINYER	0	3137.104	5479.221	2721.559	1816.382	3643.994	4354.261	12047.42	1573.254	4432.026	2417.071	6252.999	1721.237	1307.426	3761.534	4672.891	2803.135	2709.142
C1	3137.104	0	2427.445	440.7166	1320.722	506.8903	1217.158	11312.54	1563.849	1294.922	745.2046	3115.896	1441.039	1829.678	624.4301	2152.159	1888.804	5345.534
C2	5190.677	2427.445	0	2469.119	3748.166	1920.555	1210.287	12560.62	3991.294	2456.92	2773.607	3109.025	3469.441	4257.123	3051.875	517.7865	4316.249	7772.979
C3	2721.559	440.7166	2757.662	0	1761.438	947.6069	1657.874	11753.26	1661.113	1735.638	304.488	3556.612	1000.322	1905.728	1065.147	1951.332	2329.521	5327.215
C4	1816.382	1320.722	3748.166	1761.438	0	1827.612	2537.879	10543.12	243.1279	2615.643	1599.753	4436.617	903.9187	508.9563	1945.152	3472.881	986.7524	4024.813
C5	3643.994	506.8903	1920.555	947.6069	1827.612	0	710.2673	11405.63	2070.74	1301.923	1252.095	2609.005	1947.929	2336.568	1131.32	1645.269	2395.695	5852.425
C6	4354.261	1217.158	1210.287	1657.874	2537.879	710.2673	0	11350.34	2781.007	1246.633	1962.362	1898.738	2658.196	3046.835	1841.588	935.0019	3105.962	6562.692
C7	12047.42	11312.54	12560.62	11753.26	10543.12	11405.63	11350.34	0	10786.25	10103.7	11861.64	13249.08	11447.04	11052.08	10688.11	12285.34	9556.37	11883.12
C8	1573.254	1563.849	3991.294	1661.113	243.1279	2070.74	2781.007	10786.25	0	2858.771	1356.625	4679.745	660.7908	265.8284	2188.28	3612.445	1229.88	3781.685
C9	4432.026	1294.922	2456.92	1735.638	2615.643	1301.923	1246.633	10103.7	2858.771	0	2040.126	3145.371	2735.96	3124.6	1919.352	2181.634	3183.726	6640.456
C10	2417.071	745.2046	3062.15	304.488	1599.753	1252.095	1962.362	11861.64	1356.625	2040.126	0	3861.1	695.834	1601.241	1369.635	2255.82	2323.098	5022.727
C11	6252.999	3115.896	3109.025	3556.612	4436.617	2609.005	1898.738	13249.08	4679.745	3145.371	3861.1	0	4556.934	4945.574	3740.326	2833.74	5004.7	8461.43
C12	1721.237	1441.039	3757.984	1000.322	903.9187	1947.929	2658.196	11447.04	660.7908	2735.96	695.834	4556.934	0	905.4065	2065.469	2951.654	1890.671	4326.893
C13	1307.426	1829.678	4257.123	1905.728	508.9563	2336.568	3046.835	11052.08	265.8284	3124.6	1601.241	4945.574	905.4065	0	2454.108	3857.061	1495.709	3515.856
C14	3761.534	624.4301	3051.875	1065.147	1945.152	1131.32	1841.588	10688.11	2188.28	1919.352	1369.635	3740.326	2065.469	2454.108	0	2776.589	1264.374	5160.12
C15	4672.891	2152.159	806.3303	1951.332	3472.881	1645.269	935.0019	12285.34	3612.445	2181.634	2255.82	2833.74	2951.654	3857.061	2776.589	0	4040.964	7278.547
C16	2803.135	1888.804	4239.739	2329.521	986.7524	2319.185	3029.452	9556.37	1229.88	2856.704	2323.098	4928.19	1890.671	1495.709	1264.374	3964.454	0	3895.746
C17	2709.142	5345.534	7772.979	5327.215	4024.813	5852.425	6562.692	11883.12	3781.685	6640.456	5022.727	8461.43	4326.893	3515.856	5160.12	7278.547	3895.746	0

Similarly, for having alternative solutions the number of routes were varied by varying the number of salesmen from 1 to 5. Figure 8.13 given below shows the route structure for each case of the salesman and from the results of the algorithm implementation all solutions converge thereby giving the near optimal solution. This shows the stability of the results, even though the algorithm implementation was repeated several times to obtain a more favourable result out of the many repetitions because genetic algorithm is a stochastic method (Table 8.10).

Table 8.10 Statistical analysis

Description	Single-Route Solution	Two-Route Solution	Three-Route Solution	Four-Route Solution	Five-Route Solution
Mean	30739	29256	31242	33863	34920
Standard Error	2.4393	46.7789	36.8421	28.2943	23.92747
Median	30743	29188	31203	33796	34874
Mode	30743	29188	31203	33796	34874
Standard Deviation	8.0903	155.1483161	122.1917271	93.84164611	79.35845031
Sample Variance	65.4545	24071	14930.8181	8806.2545	6297.7636

Referring to the figure, the best solution history for all solutions converges early because large stopping criteria is chosen to ensure the best solution is obtained and from the graph, the line appears straight for all solutions which affirms that the best solution is obtained. From Table 8.10 of the statistical analysis, solutions with minimum distance were taken and evaluated. The route structure of all the solutions are presented in the Table 8.11. As the number routes increases from a single route solution to 5 route solutions, the shorter routes are generated even though the entire stops are covered and all demand are catered for.

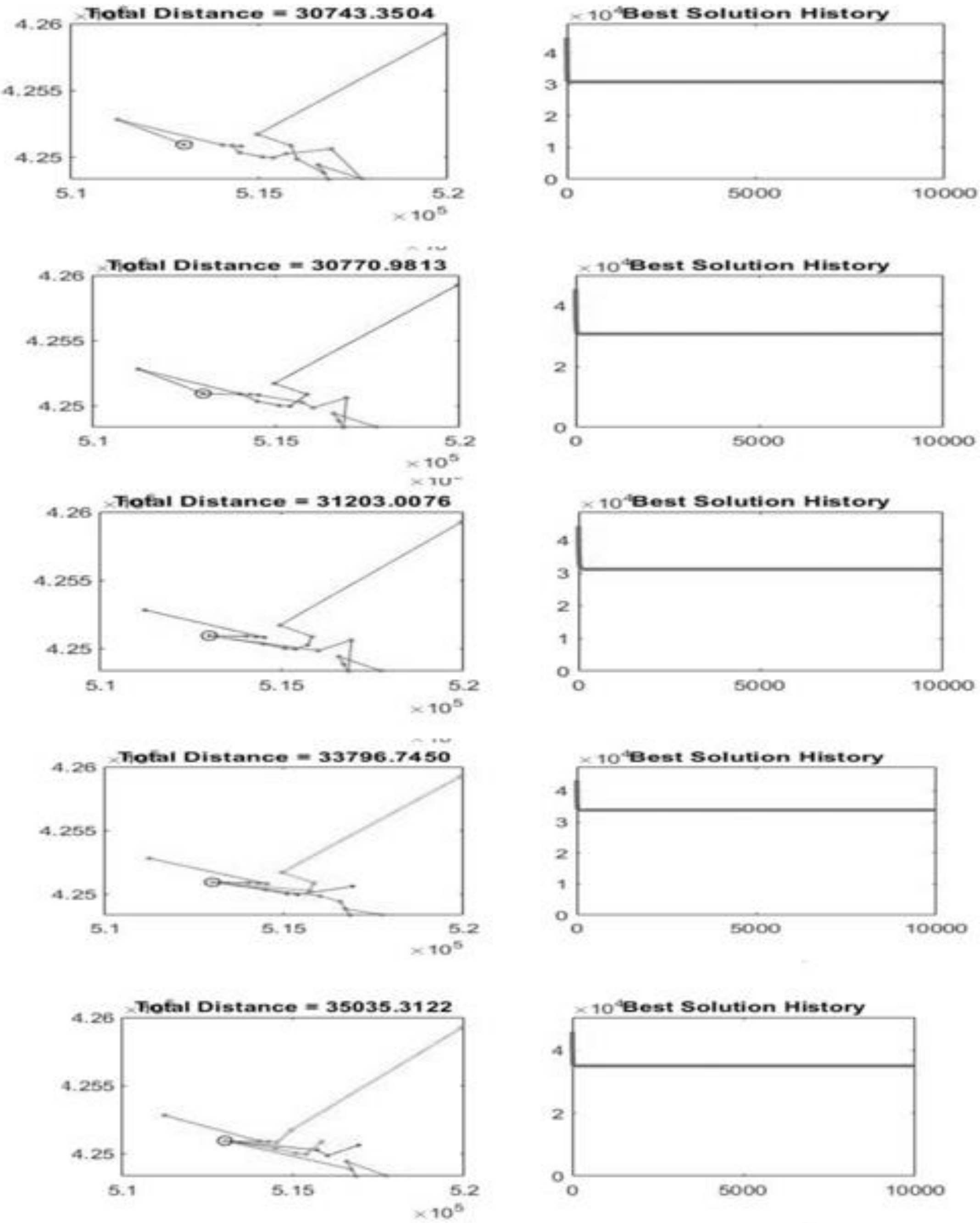

Figure 8.13: GA solutions for 5 alternatives

From the route structure, the fundamental parameters were estimated and the system performance measures such total vehicle mile, total passenger miles and also the feeder bus operating cost, feeder bus user waiting cost, feeder bus user cost, total cost of the system. This evaluation of the various alternatives will aid in selecting the best overall system in terms of minimum total system cost.

Table 8.11 Route structure

Five-Route Solution	Four-Route Solution	Three-Route Solution	Two-Route Solution	Single-Route Solution
8 17 15 1	8 17 15 2 1	8 17 15 2 4 11 13 1	8 17 15 4 11 13 14 18 1	8 17 15 6 16 3 7 12 10 2 4 11 13 13 9 5 14 18 1
10 6 2 1	10 4 11 13 1	12 7 16 3 10 6 1	12 7 16 3 10 6 2 5 9 1	
12 7 3 16 1	12 16 3 7 6 1	18 5 9 14 1		
15 4 11 13 1	18 5 9 14 1			
18 14 9 1				

The fundamental parameters of frequency, length of routes and demand are presented in the Table 8.12. In table 8.12, D is demand, L is the length in miles, and F is the frequency of the route veh/hr. In all solutions the demand was totally satisfied, but frequency varies depending on the demand the route is carrying as well as the length of the route, therefore leading to the change in average frequency of the solution. The total route length of the solutions also varied and it increases with increase in the number of routes in the solution.

Table 8.12 Fundamental parameters

5-Route Solution			4-Route Solution			3-Route Solution			2-Route Solution			1-Route Solution		
D	L	F	D	L	F	D	L	F	D	L	F	D	L	F
126	12.6	2.8	165	14.6	2.9	194	11	3.6	318	20	3.5	632	31	3.9
127	4.9	4.4	155	4.4	5.1	272	14	3.7	314	11	4.7			
104	8.3	3.1	146	9.2	3.4	166	5.8	4.6						
145	3.7	5.4	166	5.8	4.6									
130	5.3	4.3												

Considering the Figure 8.17 the total passenger kilometres decreases with increase in the number of route this is so because of the length of the individual routes in each route structure. The 2 route structure has minimum number of system routes but generally has longer individual routes implying that some passengers have to travel longer routes while on the other hand, the 5 route structure has shorter individual routes therefore most passengers travel shorter distances. Similarly, the total vehicle kilometres changes with change on the number of route structures even though the changes were sometime negative and positive. This is because the vehicle kilometre is subject to frequency of the routes. The frequency of the routes is related to other parameters such as unit waiting cost and unit operating cost.

Furthermore, the algorithm is based on MTSP problem whereby all demand points are only served once thereby eliminating the aforementioned problem. Using variable demand obtained from smart card data the number of routes was varied between 1 to 5 routes with other constraints kept constant.

Figure 8.14 and 8.15 shows a bar chart of the cost components of different alternatives. As the number of routes increases there is a comparative reduction riding and a reverse is noticed for operator costs for all alternatives. As it were, no single route could make profit if we consider the riding cost as the revenue generated, however, it buttresses the point that public transportation is usually is subsidized by government. Based on this the single route may be more beneficial to the operator since there is a little difference between the operators cost and revenue generated. But for the user, considering the riding the cost only the 5 route structure solution has minimal cost. However, when considering the waiting cost as against the riding cost for the user, 5 route structure has the higher waiting cost even though the overall user cost (which is summation of riding cost and waiting cost) is still minimum across all alternatives (Figure 8.14 and 8.15). For system performance measures the total vehicle kilometre and total route length increases with increase in the number of routes, this explains the increase in operating cost (Figure 8.16). The opposite can be said of total passenger kilometre; it decreases with increase in the number of routes which leads to the reduction in the user cost (Figure 8.17).

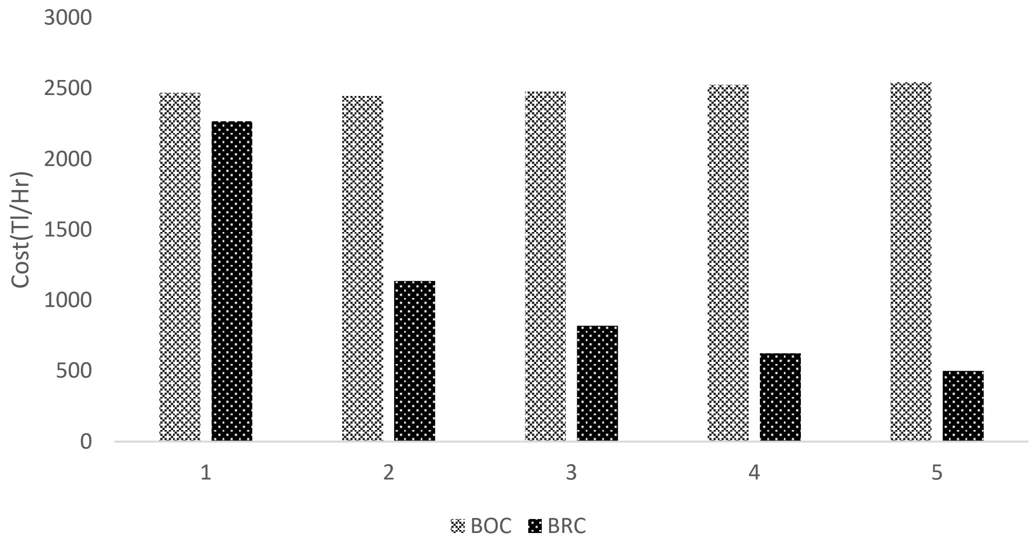

Figure 8.14 Cost components

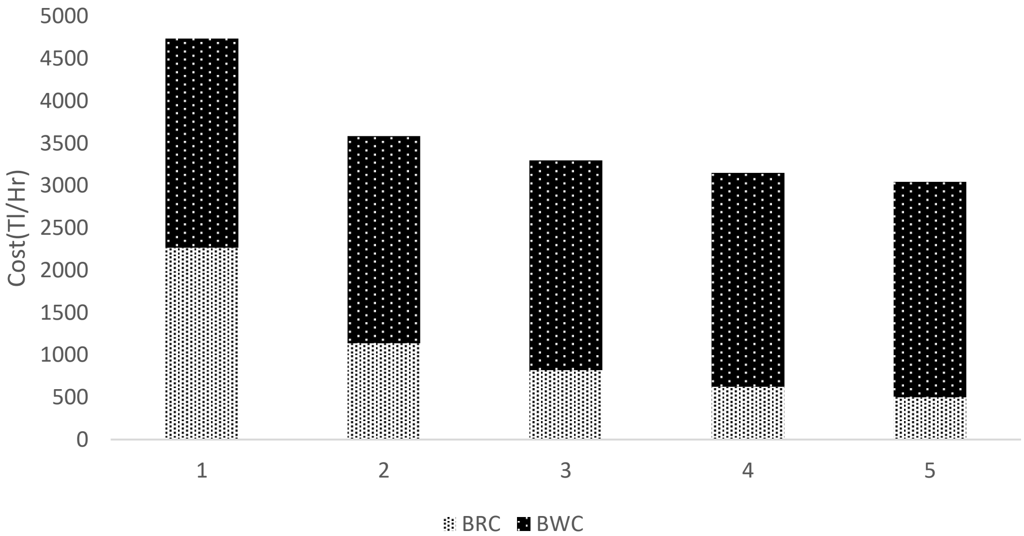

Figure 8.15 Cost components

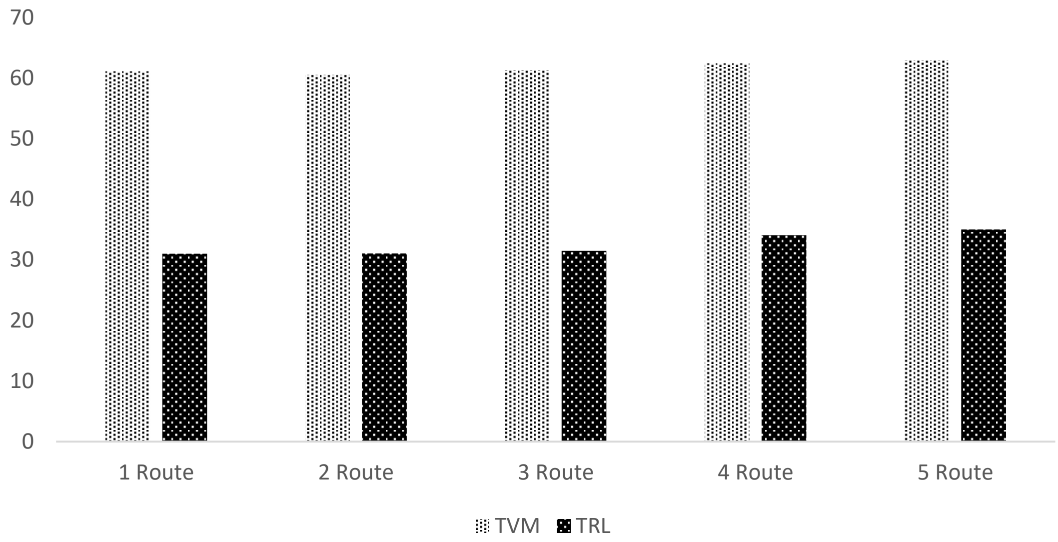

Figure 8.16 Total vehicle kilometer and total route length

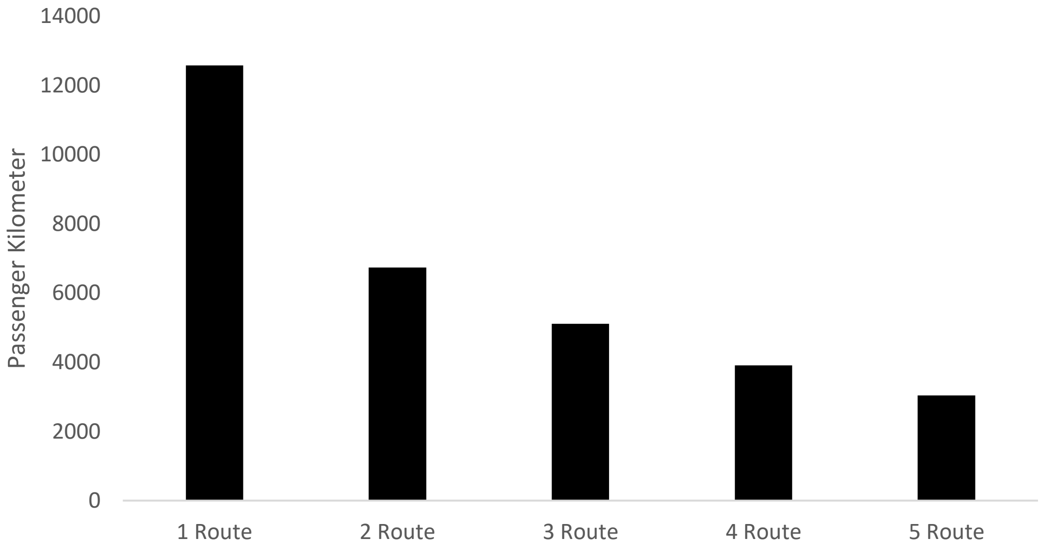

Figure 8.17 Total passenger kilometer

The results of this algorithm are largely so because the feeder routes were designed based on the shortest distance, which tends to favour both the user and operator. Additionally, since all stops are visited, all the demand at the stops are catered for. The users will prefer a minimized travel time and cost while the transit operators will focus more on the maximization of profit. Like most optimization problems, FBRNDP strives to create optimal routes taking into consideration the cost incurred by the operator, user costs. When faced with a challenge of choice in multi-objective optimization, the Pareto efficiency technique may be efficient in this regards. It has wide use in economic and engineering situations, especially, when the resources involved are utilized optimally. This concept can be described as a condition that occurs in multi-objective optimization when one individual objective cannot be made better without making the condition worse for another objective that is, any change would affect the final optimum solution. Therefore, yielding a solution that is not dominated by any other feasible solution.

Figure 8.18 shows the use of Pareto efficiency theory application on our solutions. As the number routes changes, the better the solution. The solutions that lie on the Pareto line all have at least a minimum user and/or operator cost when compared to other solutions but between Pareto solutions themselves, no solution is out rightly better depicted in Figure 8.19. Therefore, the best solution is selected based on total minimum cost which is a summation between user and operators cost. As can be seen in Figure 8.20, Solution 5 has total minimum cost.

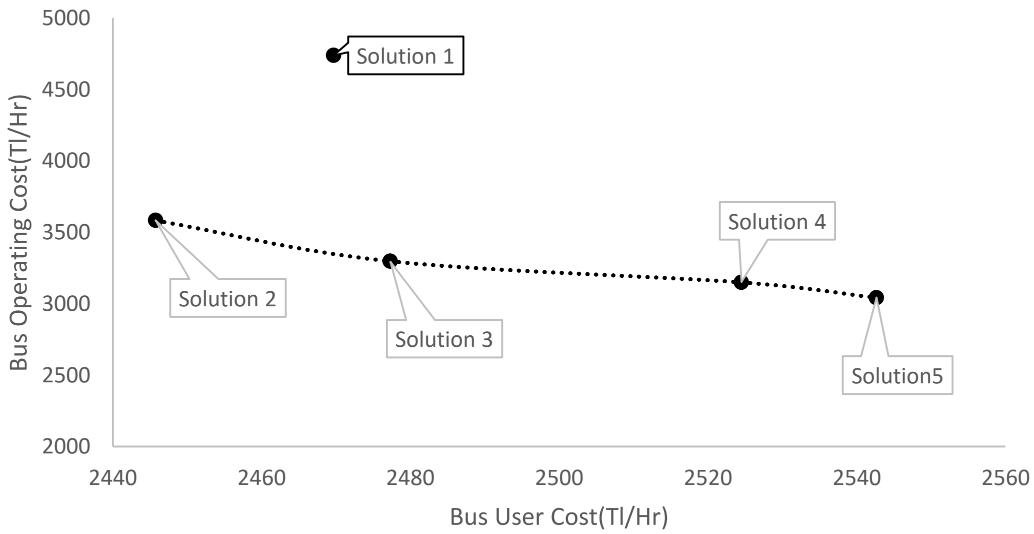

Figure 8.18 Pareto efficiency

In summary, this chapter has discussed the application of the methodology in the design of feeder bus route using smart card data for variable demand estimation, feeder bus stops location determination taking consideration the capacity of the bus stops, utilization of real transit network and finally the evaluation of alternative solutions in terms of user and operator cost and final selection of best solution.

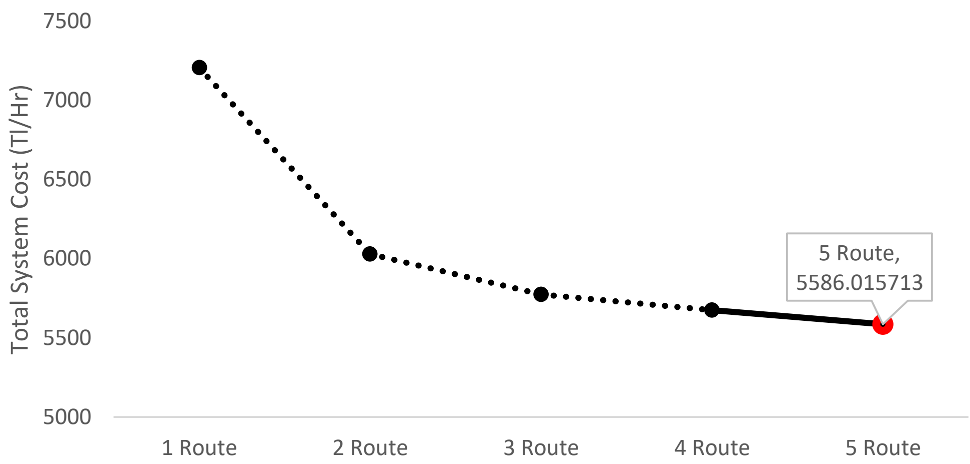

Figure 8.19 Total system cost

8.6 Design of Feeder Bus Routes (Six Station Solution)

8.6.1 Data Module

For this problem, see the figure below. The study area comprises of 195 stops which were associated to 10 train stations. When the pairwise distances were calculated amongst these 10 station and the195 stops, four stations were found to have less impact that has less than 5 stops associated with them, therefore, they were removed. The remainder of 6 station with 195 stops can be seen in Figure 8.20. The orange coloured dots represents the train stations while the blue coloured dots represents the potential bus stops.

To find the potential feeder bus stops, two clustering algorithms were implemented. The first was KMedoids algorithm which is a clustering algorithm that uses real data points as centroids for the clusters unlike the KMeans algorithms which uses new centroids for the clusters. This is relevant to our problem when we are considering fixed demand at the centroids. On the other hand, our proposed capacitated method for clustering was used since we are considering a variable demand at the stops and also with the constraint on the capacity of the stops. The total demand at 195 stops is 2172 passengers/hour, therefore if we expect the capacity of each stop to be 50 passenger/hour, then the number of expected clusters will be 2172/50 which yields approximately 45 clusters. Using this number of clusters we run the KMedoids and the result can be seen in Figure 8.21 Similarly, for the capacitated clustering, each stop is assigned to a particular station based on its proximity to the station(Figure 8.22). Furthermore these bus stops are now clustered based on the condition that the demand at these stops shall not exceed 50 passengers/hour. Figure 8.22 and 8.23 shows the result of the both clustering algorithms.

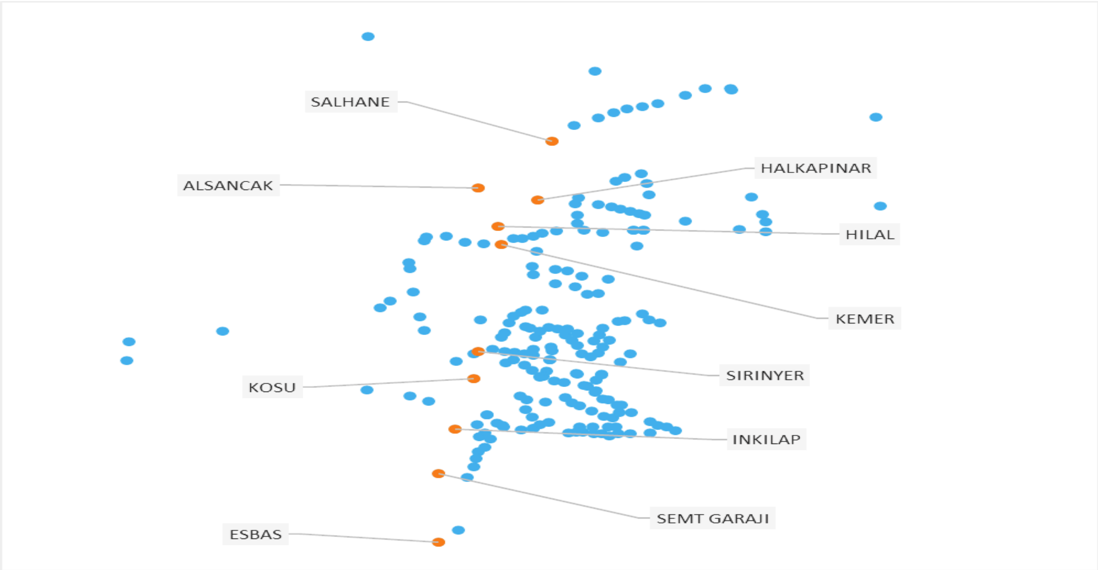

Figure 8.20 195 stops associated 10 train stations

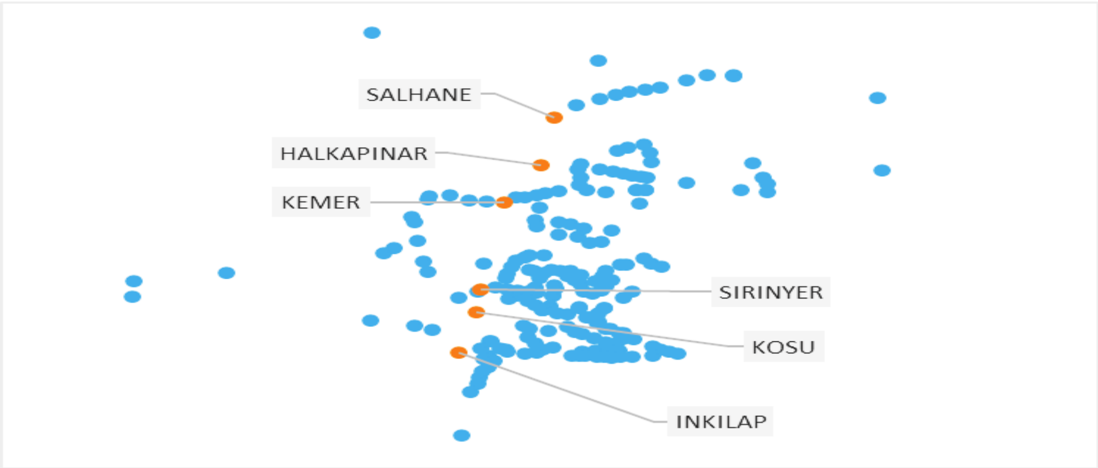

Figure 8.21 195 stops associated 6 train stations

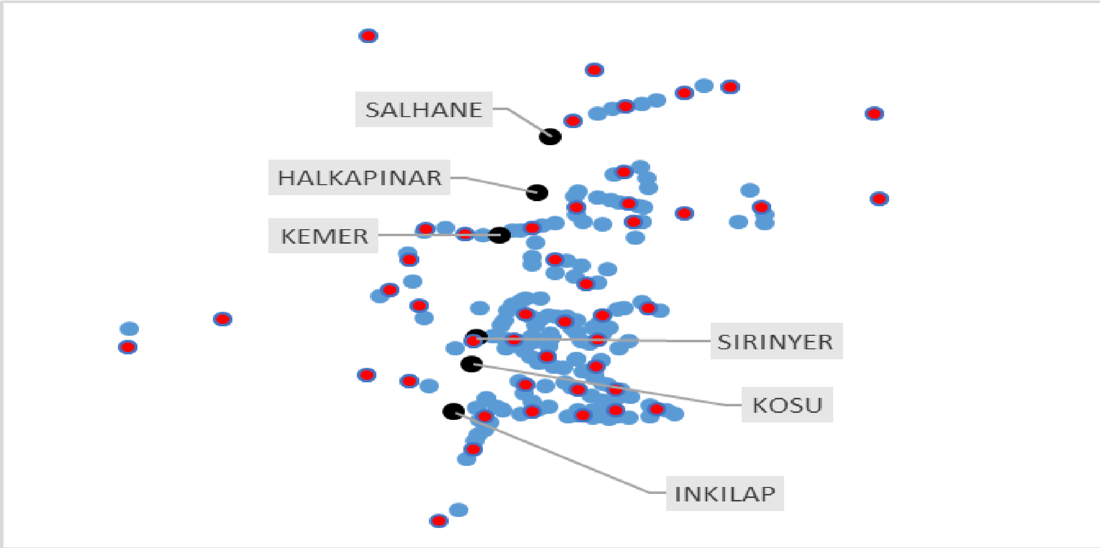

Figure 8.22 K Medoids clustering

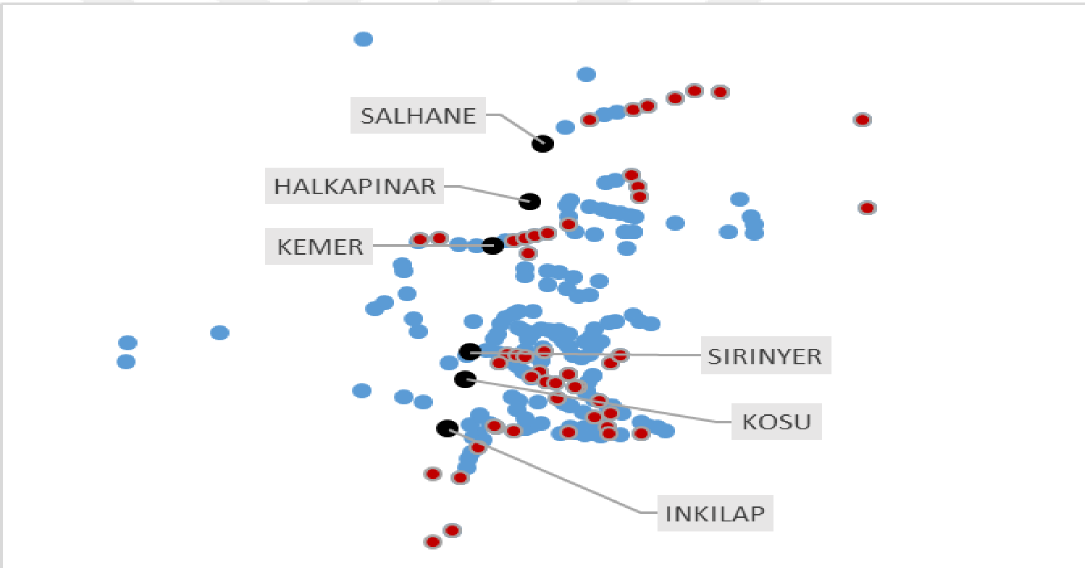

Figure 8.23 Capacitated clustering

8.6.2 Solution Module

Both problems were implemented by the modified fixed start multiple salesman algorithm that is without clustering and with clustering (K-medoids and Capacitated clustering) See Figures A.6 to A.8 for best solution history and the route structure depicted in the Figures A.9 to A.11.

8.6.3 Evaluation Module

Evaluating the cost implication of the solutions and comparing them to when the algorithm was implemented without clustering and with clustering (KMedoids clustering and capacitated clustering). See the Figures 8.24 to 8.28 below;

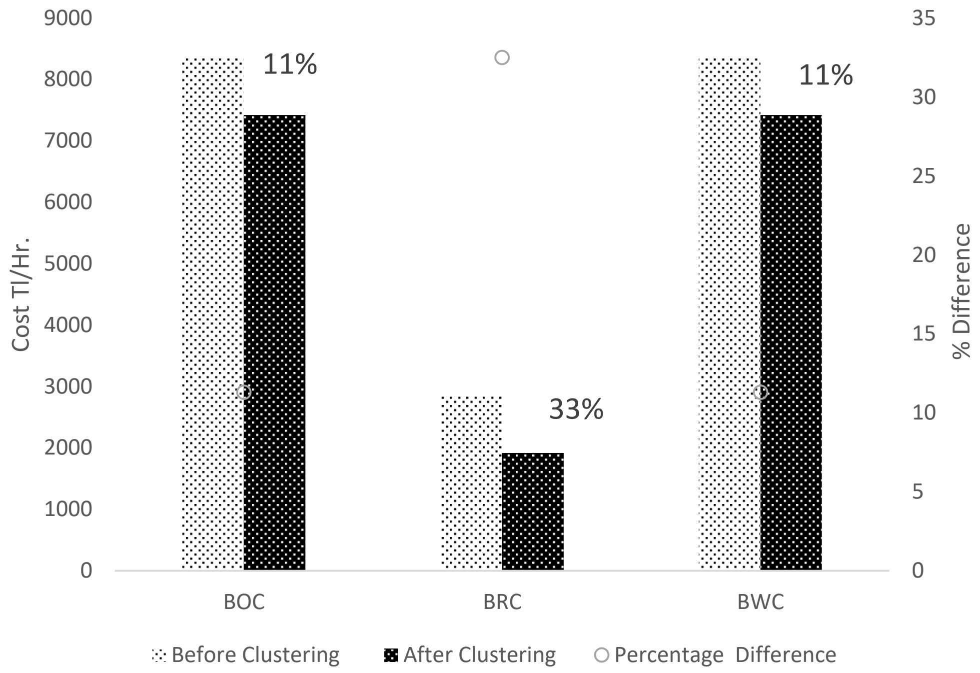

Figure 8.24 Cost components (K-medoids)

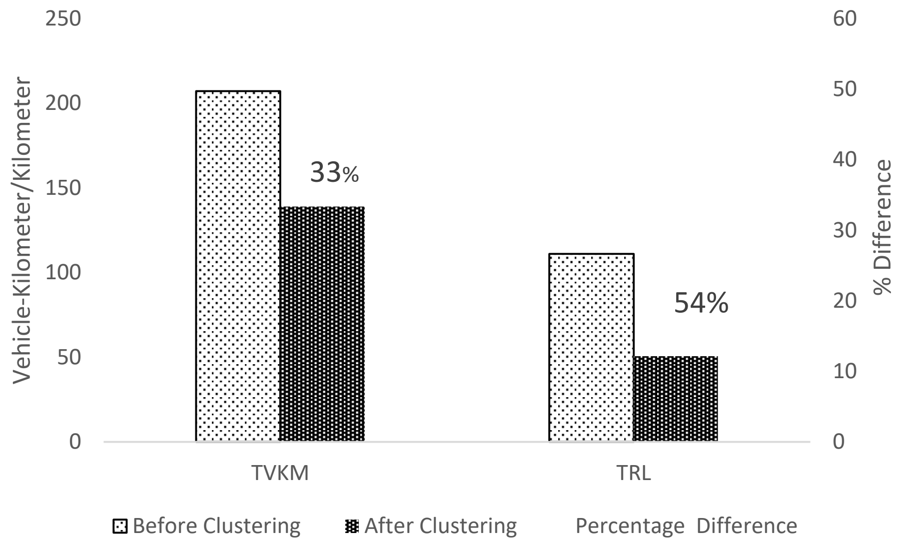

Figure 8.25 Total vehicle–kilometer and total route length (K-medoids)

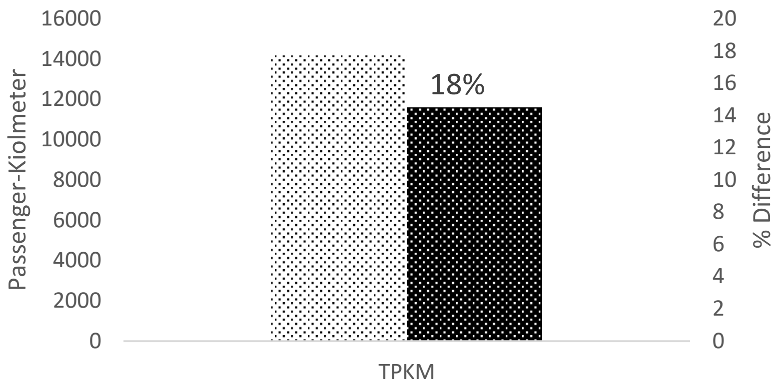

Figure 8.26 Total passenger-kilometer (K-medoids)

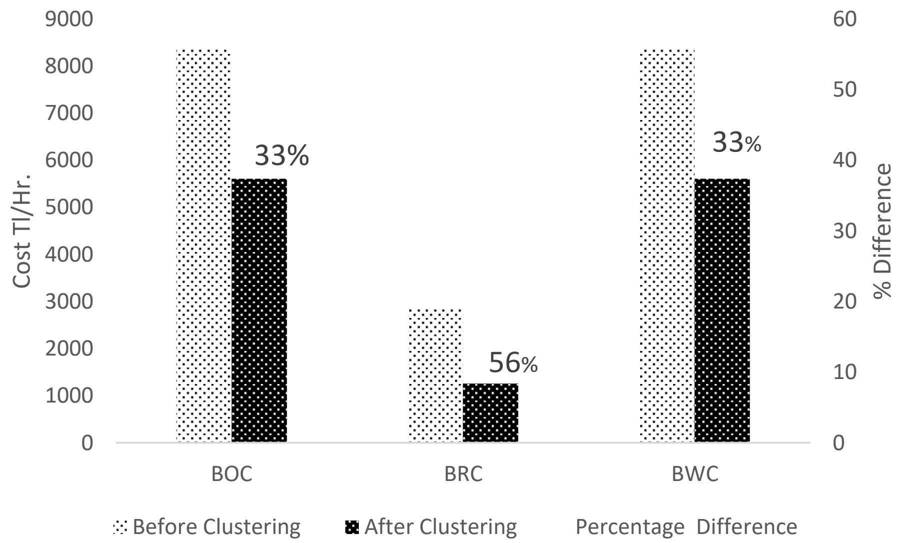

Figure 8.27 Cost components (capacitated)

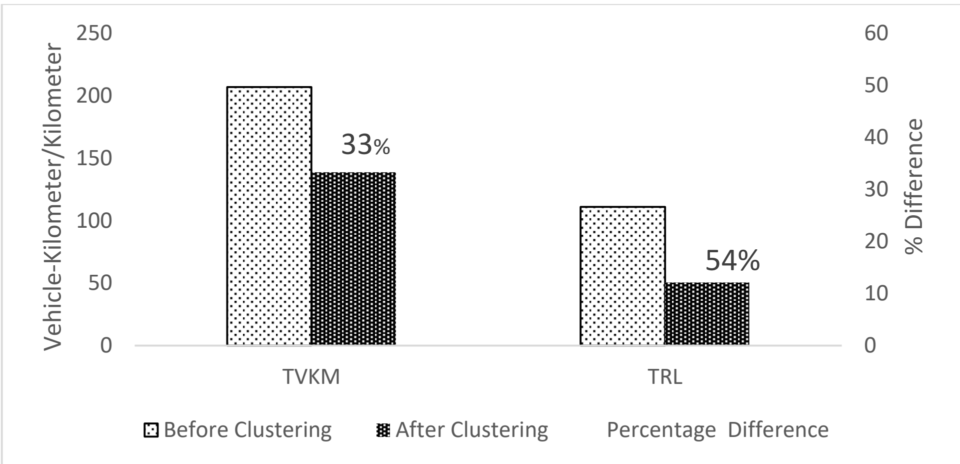

Figure 8.28 Total vehicle –kilometer and total route length (capacitated)

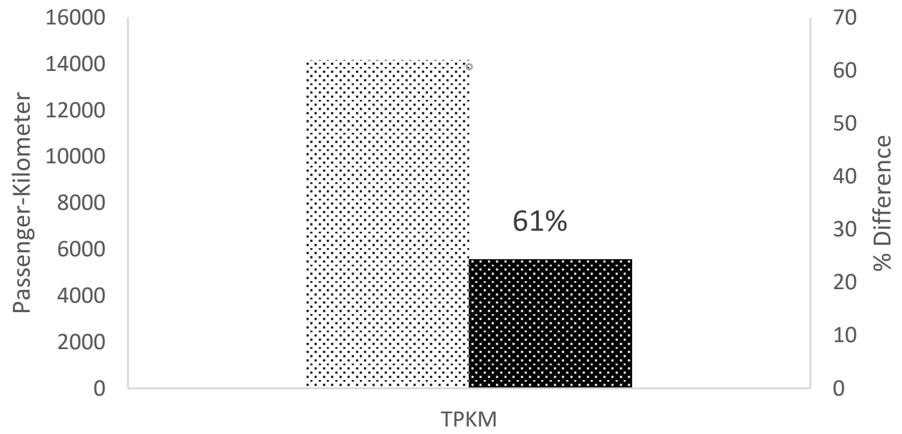

Figure 8.29 Total passenger-kilometer (capacitated)

Overall, while the K-medoid clustering was more evenly spread across the entire service area dictated by the location of the existing demands, because the centroids were chosen are among the existing stops, however, the capacity at these centroids are fixed. While for capacitated clustering, the spread does not appear evenly, since the centroids were chosen based on their proximity to the station and the demand which exists at those centroids. Figure 8.29 shows the comparism between the cost components and system performance measures for both cases of clustering. The capacitated clustering yields lower values for all categories, because the routes are of shorter distances even though they carry the same demand.

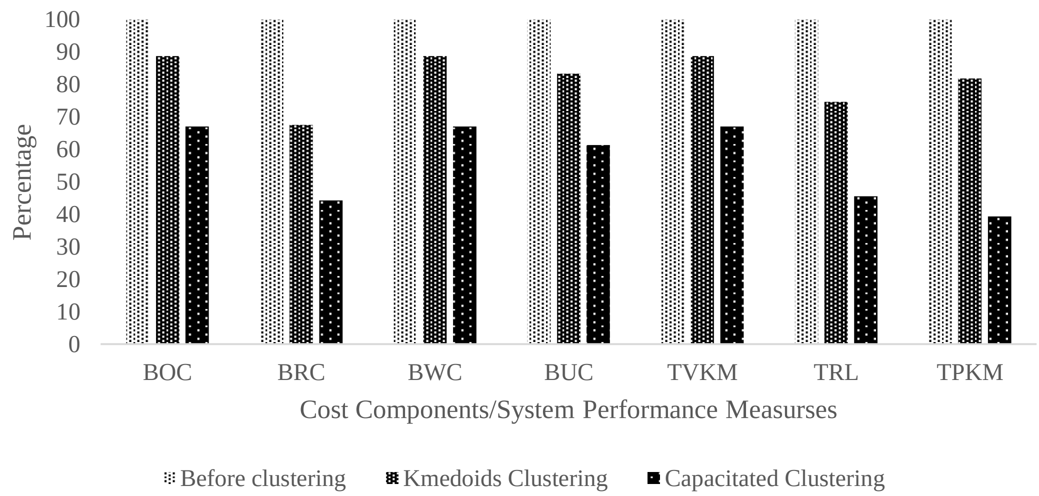

Figure 8.30 Cost and system performance comparism (K Medoids, capacitated)

8.7 Sensitivity Analysis

Sensitivity analysis was carried out on the total system cost equation. This was done by varying the weighted percentage shares both bus operating costs and bus user costs. Figure 8.31, 8.32, and 8.33 shows the effect of the weighted percentage shares of operating cost and user cost on the total system cost for the 3 major examples (benchmark study, existing bus routes system, and an extraction from smart card data). At the right section of this graph, bus operator cost has higher weights than the bus user cost and vice versa.

As the weight of bus operating cost increases (the lesser the user cost) and consequently the lesser the total system cost. This is a typical impact since it is expected that the more investment the operator is willing to make the more positive impact can be felt by the user thereby minimizing the total system cost. Also as the weight of operating cost decreases (weight of user cost increases), the number of route or alternative solutions becomes apparent and vice versa on the other side of the graph i.e. the difference between alternative solutions is minimized. This discussions cuts across all the examples mentioned above, even though, these examples were quite different. For example, the benchmark study which is a four station problem is more sensitive to different alternative routes going from right to left of the graph. This is because the study area was uniform and has a near equal number of stops attached to each station. In the remain two cases however, this phenomenon is not quite apt because the selected stops in the study area are not uniform, the problems are related to the same location and hence the similarity of the results. A look from another perspective shows that on the left side of the figures, the total system cost decreases with the decrease in the number of route structure solutions except in the case of existing route system which is the reverse. This is so because, the existing bus system used in this example comprised of five initial bus routes which influences the ensuing solutions. In all cases, the impact of cost components on weighted total system cost is clear but the extent of this impact is site dependent.

It is imperative to note that, optimization was not carried out on the total system cost since it was not used as objective function . It will be interesting to generate these alternative solutions based weighted objective function of total system cost based on user and operator cost and even more parameters. The focus of this study is not optimization of this objective function but rather this cost relationship was used as a simple medium for selecting the best route solutions from alternative solutions.

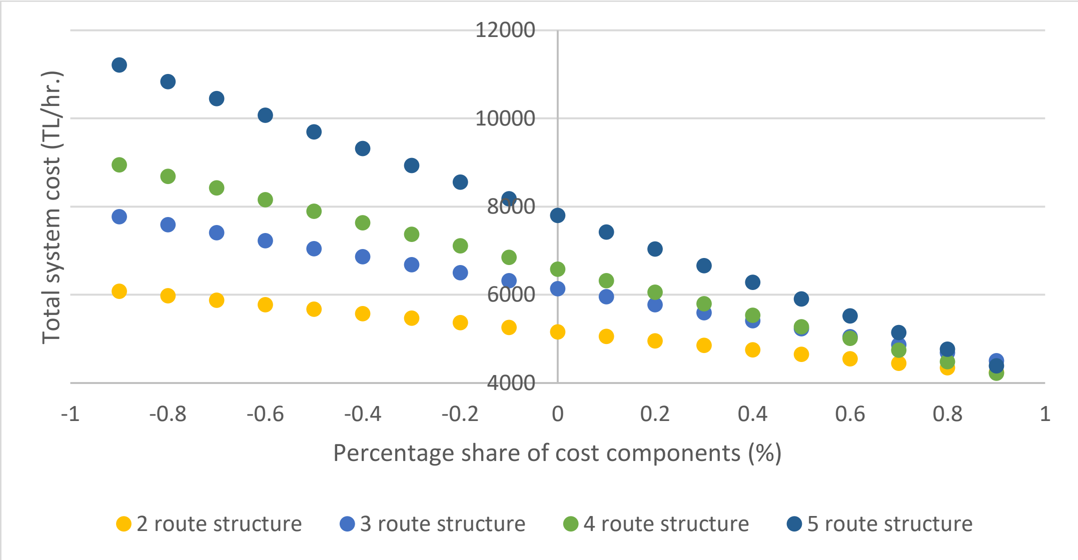

Figure 8.31 Sensitivity analysis (benchmark study)

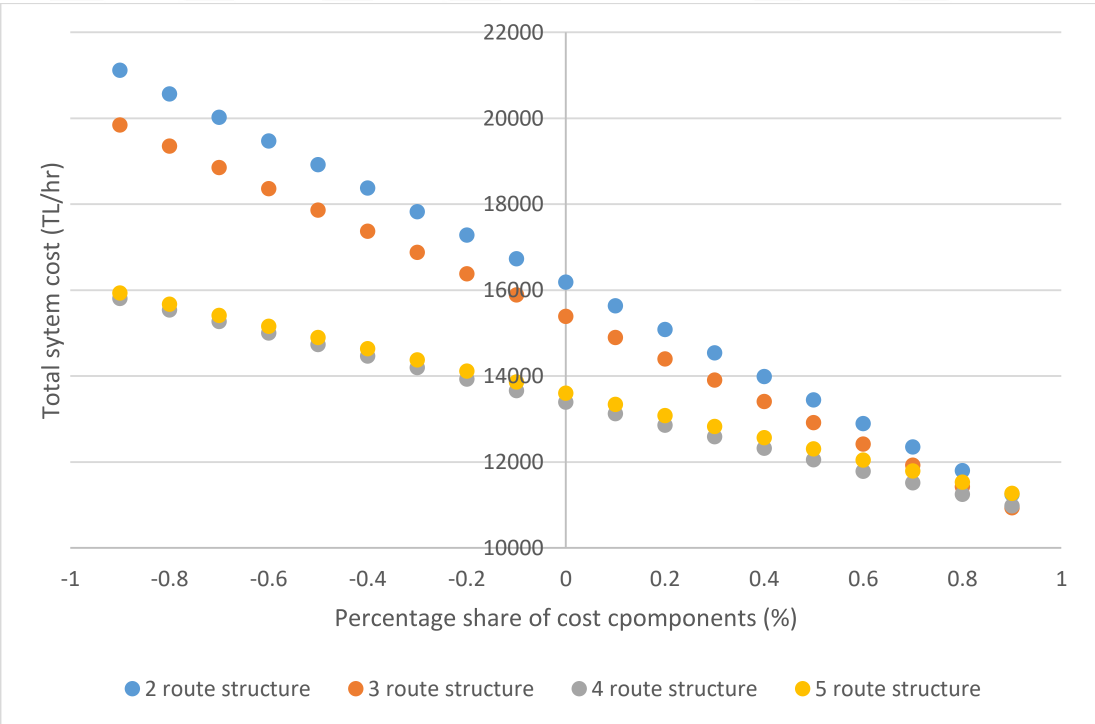

Figure 8.32 Sensitivity analysis (existing bus route)

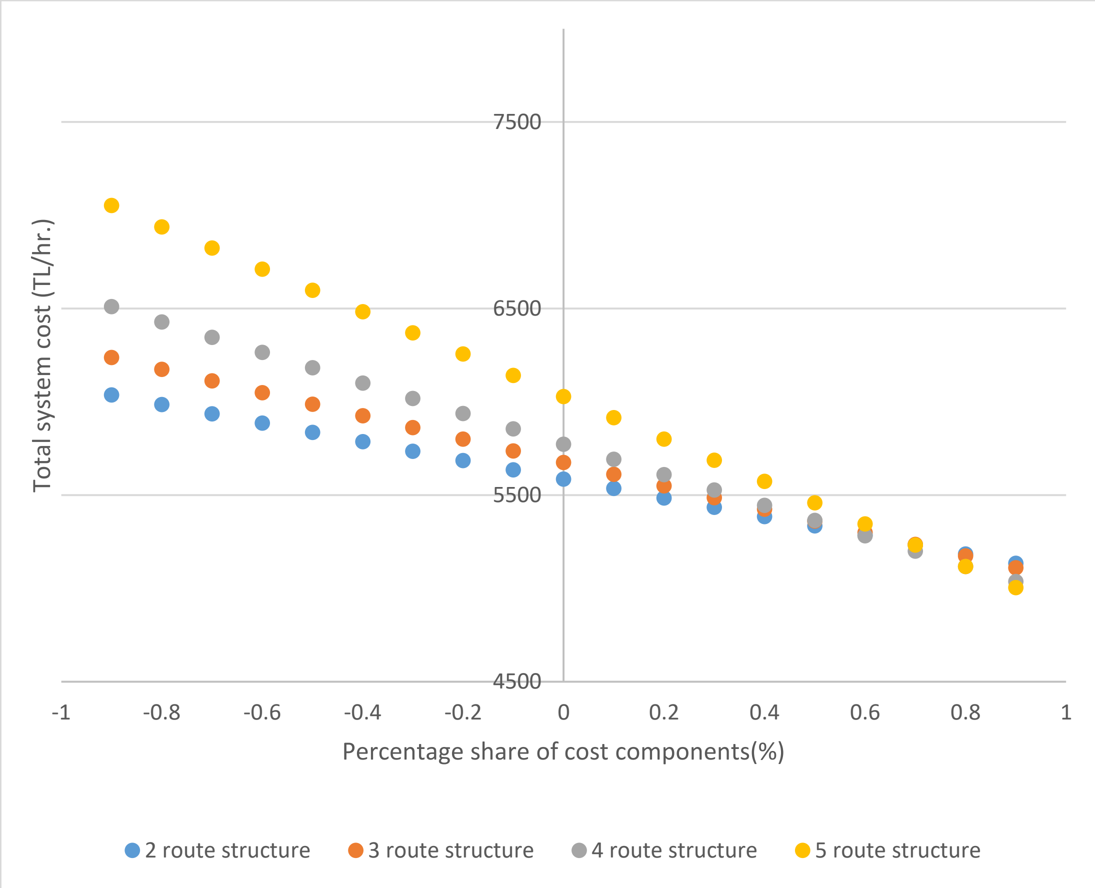

Figure 8.33 Sensitivity analysis (smart card data extraction)

CHAPTER NINE
CONCLUSION AND RECOMMENDATION

9.1 Conclusions

Attempts have been made of over decades to design a more efficient and yet sustainable transportation system by integrating the independent operations of multi modal transportation through the coordination between main trunk lines and feeder services. The FBRNDP has a lot potentials and researches are still ongoing because of many reasons such as; new policies by operators, level of details and the development of more sophisticated computing power. These reasons create new requirements, challenges and potentials for planners and researchers. The main purpose of this research is to develop a methodology for the planning of public feeder bus routes by using smart card data, transit network and existing bus stop coordinates. To meet up with the aforementioned aim, the objectives were achieved;

1. **Components of FBRNDP:** the first objective is to provide an understanding on current state of practice of feeder bus route design taking into cognisance the available data sources, solution methodologies. There are many factors that affects feeder bus route network design problem but depending on the nature of the problem, study area, expected target amidst of limitations will shape the way the problem is solved (Almasi et al., 2014) . In this regards, we highlighted important aspects of feeder bus route network design as discussed by relevant literature. This will aid researchers and operators alike when trying to solve a particular feeder bus routes network problem.

2. **Data Preparation And Feeder Bus Stop Location:** the second objective is to highlight the potential of using smart card data for feeder bus route configuration. The output of this objective is to carry exploratory data analysis on smart card data for use in feeder bus route configuration and mining of data from smartcard to obtain OD matrix. Planning of bus services often requires conventional household travel surveys which are expensive to conduct and

tedious to analyse. With smart card data having details embedded in them, the potential to access them for continuous unbiased origin destination surveys for planning purpose can be achieved. Feeder bus service can be simplified as a known single destination (such as train station) and multiple origins (bus stops), therefore, the size of the origin destination is limited to many origins to a single destination. This work presents the application of smart card data in the planning and design feeder bus system serving a train station as a destination. It demonstrates a method for the determination of potential feeder bus stop location from combined data sources of smart card data and existing coordinates of bus stops and train station as well as real transit network map. It includes cleaning and coding of the smart card data since not all collected data are relevant and there are also missing data. Coordinates of bus stop and train station was extracted and transfer time was calculated for each smart card user. In Izmir the transfer time is less or equal 90 minutes. Therefore, if transfer time is less than 90 minutes, we isolate all bus stops that uses a station and no any other modes. With this the algorithm was able to identify potential feeder bus stops however, these stops were numerous and cannot also be used as feeder bus stops so we employed a capacitated clustering of the bus stops based on the demand. This was particularly important to ensure the capacity of the generated feeder bus stop locations were not exceeded.

3. **Solution Methodology:** the third objectivity is to develop a strategy for optimal feeder bus route construction. Tour construction problem which has mainly travelling salesman problem and vehicle routing problem as sub problems can be represented by multiple travelling salesman problem as a generalization or relaxation respectively. MTSP is quite similar to real-life problem FBRNDP where we have many origins and one destination. An algorithm for solving MTSP was adopted and modified to accommodate multiple stations and clustering of the nodes. According to the need of the designer, the number of the salesperson (number of routes) may be varied to show its sensitivity to other factors such as frequency, route length, and route structure. The algorithm was tested on three scenarios; Kuah & Perl ,1989

benchmark study, existing feeder bus route network, and was used to design a new feeder bus routes. In all scenarios the algorithm was able minimized the total distance in reasonably fast computational time and good tuning of genetic algorithm parameters ensured the solutions were repeatable with reasonable statistics. The evaluation module for all scenarios were all similar in that it was able to obtain minimized operator and user cost for all 3 scenarios and best solution could be selected via pareto efficiency and minimized total system cost.

Planning of bus services often requires conventional household travel surveys which are expensive to conduct and tedious to analyse. With smart card data having details embedded in them, the potential to access them for continuous unbiased origin destination surveys for planning purpose can be achieved. Feeder bus service can be simplified as a known single destination (such as train station) and multiple origins (bus stops), therefore, the size of the origin destination is limited to many origins to a single destination. This research contributes to knowledge in the following aspects;

1. The methodology developed in this study can contribute to very scanty literature with regards to planning of feeder bus routes.
2. It also demonstrates the use of smart card data and existing bus stop station location in the determination of potential feeder bus stop locations by using capacitated clustering based on passenger demand.
3. It also uses real demand at the stops as an extraction from smart card data rather than employing conventional survey.
4. It can also be adapted to real public transit network.
5. The evaluation aspect allows for planning and design of new bus routes based on minimum total system cost.

9.2 Recommendations

Because of the ensuing challenges and continuous developments in the fields of mathematics and computers analysis, there are still areas of untapped opportunities for future research. Other areas that can be explored are listed below;

1. As discussed in earlier section, most studies rely on a many to 1 pattern for simplification purposes as was the case for our research even though most transit passengers in developed countries have varying origin and destination. It will be interesting to explore many to many travel pattern
2. Travel patterns, demand and capacity at the stations of main trunk line may change if a new feeder service is implemented. Aspects such as walking activities within transfer stations and capacity of the stations themselves can create serious concerns. Thus feasibility and post evaluation of the system might be necessary.
3. Modelling Aspects: conflicting perspectives of stake holders, the fact that a sustainable transit system should be multimodal, makes route generation problem a multi objective problem as suggested by researchers.
4. Nature of problem: in most researches transit services are planned and designed for normal daily operations but aspects like disasters management, seasonal/mega events or even emergencies for hospitals, owl services etc. will change the nature of problem and will open a new frontier in urban transit planning. Thus making objectives a situation based and cannot be the same with the planning and design daily operating transit services. Also developing solution methods from previous models of daily operating transit systems and building new ones to fit the particular problem can be a good area of research.
5. In most of the researches railways and feeder buses are the most researched even though different access modes or mode combinations exists. Modes like cycling, walking, can also be used to access main trunk lines station. Therefore, inclusion of different modes or mode combinations can be looked into. Diverse conditions of a transit network with regard to demand, distance of area from rail stations, etc., invoke different modes with various operating performance characteristics for every region of the network. Multimodal networks have a

greater effect on reducing user costs. They are more likely to attract private vehicle users to use transit. This in turn means more profit for operators, and thus using multimode in the feeder network design is appropriate for operators, too. Similarly, if there are multi mainline movements in a given location, providing feeder bus routes will be more complex since the capacity of the transfer station will impacted, will increase the mode choice of the users depending on their destination. These considerations can help model the problem to be much closer to the real world.

6. Operators normally uses fare setting as a strategy for profit making while keeping in mind the total welfare in mind. This creates some sort of opposing objectives thereby leading to different fare strategies and products. Multi modal transportation also offers different variety of fare products depending on the characteristics of the mode and the intent of the planners. Thus integrating the different fare structures into one single one which is a necessary tool in modern transit system will be quite difficult to implement and model.
7. The cost function used for evaluation could developed further to include to other cost components such as social and environmental cost.

REFERENCES

Almasi, M. H., Mirzapour Mounes, S., Koting, S., & Karim, M. R. (2014). Analysis of feeder bus network design and scheduling problems. *The Scientific World Journal*, *2014*, 1-11.

Alsger, A., Assemi, B., Mesbah, M., & Ferreira, L. (2016). Validating and improving public transport origin–destination estimation algorithm using smart card fare data. *Transportation Research Part C: Emerging Technologies*, *68*, 490–506.

Arostegui Jr, M. A., Kadipasaoglu, S. N., & Khumawala, B. M. (2006). An empirical comparison of tabu search, simulated annealing, and genetic algorithms for facilities location problems. *International Journal of Production Economics*, *103*(2), 742–754.

Arya, V., Goyal, A., & Jaiswal, V. (2014). An optimal solution to multiple travelling salesperson problem using modified genetic algorithm. *International Journal of Application or Innovation in Engineering & Management*, *3*(1), 425–430.

Audouin, M., Razaghi, M., & Finger, M. (2015). *How Seoul used the 'T-Money'smart transportation card to re-plan the public transportation system of the city; implications for governance of innovation in urban public transportation systems*. *8th TransIST Symposium in Istanbul*. Retrieved July 27, 2018, from https://www.researchgate.net/profile/MaximeAudouin/publication/290574722.

Baaj, M. H., & Mahmassani, H. S. (1991). An AI-based approach for transit route system planning and design. *Journal of Advanced Transportation*, *25*(2), 187–209.

Baaj, M. H., & Mahmassani, H. S. (1995). Hybrid route generation heuristic algorithm for the design of transit networks. *Transportation Research Part C: Emerging Technologies*, *3*(1), 31–50.

Bagchi, M., & White, P. R. (2005). The potential of public transport smart card data. *Transport Policy*, *12*(5), 464–474.

Bagchi, M., & White, P. R. (2004). What role for smart-card data from bus systems? *Proceedings of the Institution of Civil Engineers-Municipal Engineer*, *157*(1), 39–46.

Barry, J. J., Newhouser, R., Rahbee, A., & Sayeda, S. (2002). Origin and destination estimation in New York City with automated fare system data. *Transportation Research Record*, *1817*(1), 183–187.

Bektas, T. (2006). The multiple traveling salesman problem: an overview of formulations and solution procedures. *Omega*, *34*(3), 209–219.

Blythe, P. T. (2004). Improving public transport ticketing through smart cards. *Proceedings of the Institution of Civil Engineers-Municipal Engineer*, *157*(1), 47–54.

Briem, L., Buck, S., Ebhart, H., Mallig, N., Strasser, B., Vortisch, P., Dorothea, W., & Zündorf, T. (2017). Efficient traffic assignment for public transit networks. In *16th International Symposium on Experimental Algorithms 20,* 1–14.

Byrne, B. F., & Vuchic, V. (1972). Public transportation line positions and headways for minimum cost *Traffic flow and Transportation*, *1972,* 347-60.

Carter, A. E., & Ragsdale, C. T. (2006). A new approach to solving the multiple traveling salesperson problem using genetic algorithms. *European Journal of Operational Research*, *175*(1), 246–257.

Ceder, A. (2013). Integrated smart feeder/shuttle transit service: simulation of new routing strategies. *Journal of Advanced Transportation*, *47*(6), 595–618.

Ceder, A., & Wilson, N. H. M. (1986). Bus network design. *Transportation Research Part B: Methodological*, *20*(4), 331–344.

Chatterjee, S., Carrera, C., & Lynch, L. A. (1996). Genetic algorithms and traveling salesman problems. *European Journal of Operational Research*, *93*(3), 490–510.

Chien, S., & Schonfeld, P. (1998). Joint optimization of a rail transit line and its feeder bus system. *Journal of Advanced Transportation*, *32*(3), 253–284.

Chien, S., Yang, Z., & Hou, E. (2001). Genetic algorithm approach for transit route planning and design. *Journal of Transportation Engineering*, *127*(3), 200–207.

Chowdhury, S. M., & I-Jy Chien, S. (2002). Intermodal transit system coordination. *Transportation Planning and Technology*, *25*(4), 257–287.

Ciaffi, F., Cipriani, E., & Petrelli, M. (2012). Feeder bus network design problem: A new metaheuristic procedure and real size applications. *Procedia-Social and Behavioral Sciences*, *54*, 798–807.

Costa, N. R., & Lourenço, J. A. (2015). Exploring pareto frontiers in the response surface methodology. *In Transactions on Engineering Technologies*, *2014*, 399–412. Retrieved February 3, 2019 from https://link.springer.com/book/10.1007/978-94-017-9804-4

Deb, K. (2011). Multi-objective optimisation using evolutionary algorithms: an introduction. In *Multi-Objective Evolutionary Optimisation for Product Design and Manufacturing*, 3-34. Retrieved August 29, 2017 from https://link.springer.com/chapter/10.1007/978-0-85729-652-8_1

Deri, A., & Kalpakci, A. (2014). Efficient usage of transfer based system in intracity bus transit operation: sample of Izmir. *Procedia-Social and Behavioral Sciences*, *111*, 311–319.

DiJoseph, P., & Chien, S. I. (2013). Optimizing sustainable feeder bus operation considering realistic networks and heterogeneous demand. *Journal of Advanced Transportation*, *47*(5), 483–497.

Eldrandy, K. A., Ahmed, A. H. N., & AbdAllah, A. F. (2008). Routing Problems: A Survey. *The 43rd Annual Conference on Statistics, Computer Sciences and Operations Research* , 51-70. Retrieved November 4, 2018 from https://www.researchgate.net/publication/236009566

Eom, J. K., Song, J. Y., & Moon, D.-S. (2015). Analysis of public transit service performance using transit smart card data in Seoul. *KSCE Journal of Civil Engineering*, *19*(5), 1530–1537.

Fan, W. & Machemehl, R. B.(2004). *Optimal transit route network design problem: algorithms, implementation and numerical results* . Retrieved July 19, 2019, from, http://static.tti.tamu.edu/swutc.tamu.edu/publications/technical reports.

Farahani, R. Z., Miandoabchi, E., Szeto, W. Y., & Rashidi, H. (2013). A review of urban transportation network design problems. *European Journal of Operational Research*, *229*(2), 281–302.

Ghaziri, H. (1996). Supervision in the self-organizing feature map: Application to the vehicle routing problem. *Meta-heuristics*, 651–660. Retrieved March 17, 2016 from https://link.springer.com/content/pdf.

Goldberg, D. E. (1989). *Genetic algorithms in search optimization, and machine learning.* Boston: Addison-Wesley Longman Publishing Co.

Gschwender, A., Munizaga, M., & Simonetti, C. (2016). Using smart card and GPS data for policy and planning: The case of Transantiago. *Research in Transportation Economics*, *59*, 242–249.

Guihaire, V., & Hao, J.-K. (2008). Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, *42*(10), 1251–1273.

Huili, D., Haode, L., & Xiaoguang, Y. (2007). OD matrix estimation method of public transportation flow based on passenger boarding and alighting. *Computer and Communications*, *25*(2), 79.

Improvements, T. M. (2011). Improving methods for evaluating the effects and value of transportation system changes. *Victoria Transport Policy Institute.* Retrieved May 25, 2017, from http://www. Vtpi. Org/Tdm/Tdm125.

Jakimavičius, M., & Burinskiene, M. (2009). Assessment of Vilnius city development scenarios based on transport system modelling and multicriteria analysis. *Journal of Civil Engineering and Management*, *15*(4), 361–368.

Jayasutha, R. and Zoraida, B. S. (2013). The optimizing multiple travelling salesman problem using genetic algorithm. *International Journal For Scientific Research and Development*, *1*(5), 1200-1203.

Jie, Y. U., & Yang, X. G. (2006). Estimation a transit route od matrix using on/off data: an application of modified bp artificial neural network. *System. Engineering*, *24*, 89–92.

Joseph, K. (2020). *Fixed start open multiple traveling salesmen problem genetic algorithm.* Retrieved January 15, 2020, from https://ch.mathworks.com/matlabcentral/fileexchange/21198-fixed-start-open traveling-salesman-problem-genetic-algorithm.

Jung, J., & Sohn, K. (2017). Deep-learning architecture to forecast destinations of bus passengers from entry-only smart-card data. *IET Intelligent Transport Systems*, *11*(6), 334–339.

Karakatič, S., & Podgorelec, V. (2015). A survey of genetic algorithms for solving multi depot vehicle routing problem. *Applied Soft Computing*, *27*, 519–532.

Kepaptsoglou, K., & Karlaftis, M. (2009). Transit route network design problem. *Journal of Transportation Engineering*, *135*(8), 491–505.

Kieu, L.-M., Bhaskar, A., & Chung, E. (2015). A modified density-based scanning algorithm with noise for spatial travel pattern analysis from smart card AFC data. *Transportation Research Part C: Emerging Technologies*, *58*, 193–207.

Kim, K. S., Cheon, S., & Lim, S. (2011). Performance assessment of bus transport reform in Seoul. *Transportation*, *38*(5), 719–735.

Király, A., & Abonyi, J. (2011). Optimization of multiple traveling salesmen problem by a novel representation based genetic algorithm. *Intelligent computational optimization in engineering* 241–269. Retrieved January 25, 2016 from https://link.springer.com/chapter/10.1007/978-3-642-21705-0_9.

Kuah, Geok K, & Perl, J. (1988). Optimization of feeder bus routes and bus-stop spacing. *Journal of Transportation Engineering*, *114*(3), 341–354.

Kuah, Geok K, & Perl, J. (1989). The feeder-bus network-design problem. *Journal of the Operational Research Society*, *40*(8), 751–767.

Kuan, S. N., Ong, H. L., & Ng, K. M. (2004). Applying metaheuristics to feeder bus network design problem. *Asia-Pacific Journal of Operational Research*, *21*(04), 543–560.

Kuan, S. N., Ong, H. L., & Ng, K. M. (2006). Solving the feeder bus network design problem by genetic algorithms and ant colony optimization. *Advances in Engineering Software*, *37*(6), 351–359.

Laporte, G. (1992). The traveling salesman problem: An overview of exact and approximate algorithms. *European Journal of Operational Research*, *59*(2), 231–247.

Lee, Y.-J., & Vuchic, V. R. (2005). Transit network design with variable demand. *Journal of Transportation Engineering*, *131*(1), 1–10.

Li, D., Lin, Y., Zhao, X., Song, H., & Zou, N. (2011). Estimating a transit passenger trip origin-destination matrix using automatic fare collection system. *International Conference on Database Systems for Advanced Applications*, *6637*, 502–513.

Li, T., Sun, D., Jing, P., & Yang, K. (2018). Smart card data mining of public transport destination: a literature review. *Information*, *9*(1), 18.

Lin, J.-J., & Wong, H.-I. (2014). Optimization of a feeder-bus route design by using a multiobjective programming approach. *Transportation Planning and Technology, 37*(5), 430–449.

Lund, H. M., Cervero, R., & Willson, R. (2004). Travel characteristics of transit-oriented development in California. *Sacramento, California Department of Transportation*. Retrieved March 8, 2016 from http://staging.community-wealth.org/sites/clone.community-wealth.org/files/downloads/report-lund-cerv-wil.

Ma, X. (2013). Smart card data mining and inference for transit system optimization and performance improvement. *Sustainability, 54,* 1587–1601.

Mandl, C. E. (1980). Evaluation and optimization of urban public transportation networks. *European Journal of Operational Research*, *5*(6), 396–404.

Martins, C. L., & Pato, M. V. (1998). Search strategies for the feeder bus network design problem. *European Journal of Operational Research*, *106*(2–3), 425–440.

Matai, R., Singh, S. P., & Mittal, M. L. (2010). Traveling salesman problem: an overview of applications, formulations, and solution approaches. *Traveling Salesman Problem, Theory and Applications*, *1*, 1-25.

Mohaymany, A. S., & Gholami, A. (2010). Multimodal feeder network design problem: ant colony optimization approach. *Journal of Transportation Engineering*, *136*(4), 323–331.

Munizaga, M., Devillaine, F., Navarrete, C., & Silva, D. (2014). Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies*, *44*, 70-79.

Nohl, K. (2007). Mifare, little security, despite obscurity. *The 24th Congress of the Chaos Computer Club in Berlin*. *December 2007*. Retrieved January 26, 2018 from https://ci.nii.ac.jp/naid/10027609030.

Owais, M. (2015). Issues related to transit network design problem. *International Journal of Computer Applications*, *975*, 8887.

Pan, S., Yu, J., Yang, X., Liu, Y., & Zou, N. (2015). Designing a flexible feeder transit system serving irregularly shaped and gated communities: determining service area and feeder route planning. *Journal of Urban Planning and Development*, *141*(3), 4014028.

Park, Y.-B. (2001). A hybrid genetic algorithm for the vehicle scheduling problem with due times and time deadlines. *International Journal of Production Economics*, *73*(2), 175–188.

Pattnaik, S. B., Mohan, S., & Tom, V. M. (1998). Urban bus transit route network design using genetic algorithm. *Journal of Transportation Engineering*, *124*(4), 368–375.

Pau, S. A. (2014). *Using smart card technologies to measure public transport performance: data capture and analysis*. Master Thesis, Polytechnic University of Catalonia, Spain

Saadia, T. (2016). *Feeder network design to access an existing bus rapid transit system in Lahore*. PhD Thesis, Yokohama National University, Japan

Sedighpour, M., Yousefikhoshbakht, M., & Mahmoodi Darani, N. (2012). An effective genetic algorithm for solving the multiple traveling salesman problem. *Journal of Optimization in Industrial Engineering*, *8*, 73–79.

Shrivastav, P., & Dhingra, S. L. (2001). Development of feeder routes for suburban railway stations using heuristic approach. *Journal of Transportation Engineering*, *127*(4), 334–341.

Shrivastava, P., & O'Mahony, M. (2005). Modeling an integrated public transportation system-a case study in Dublin, Ireland. *European Transport 41*, 28-46.

Shrivastava, P., & O'Mahony, M. (2007). Design of feeder route network using combined genetic algorithm and specialized repair heuristic. *Journal of Public Transportation*, *10*(2), 7.

Shrivastava, P., & O'Mahony, M. (2009). Use of a hybrid algorithm for modeling coordinated feeder bus route network at suburban railway station. *Journal of Transportation Engineering*, *135*(1), 1–8.

Singh, A. (2016). A review on algorithms used to solve multiple travelling salesman problem. *International Research Journal of Engineering and Technology*, *3*(4), 598–603.

Smart, M., Miller, M. A., & Taylor, B. D. (2009). Transit stops and stations: transit managers' perspectives on evaluating performance. *Journal of Public Transportation*, *12*(1), 4.

Turkish statistical institute (2014). *Information Society Statistics*. Retrived February, 5, 2016 from http://Www. Tuik. Gov. Tr/PreIstatistikTablo. Do.

Tang, L., Liu, J., Rong, A., & Yang, Z. (2000). A multiple traveling salesman problem model for hot rolling scheduling in Shanghai Baoshan Iron & Steel Complex. *European Journal of Operational Research*, *124*(2), 267–282.

Topal, O., & Nakir, İ. (2018). Total cost of ownership based economic analysis of diesel, CNG and electric bus concepts for the public transport in Istanbul city. *Energies*, *11*(9), 2369.

Trepanier, M., & Chapleau, R. (2006). Destination estimation from public transport smartcard data. *IFAC Proceedings Volumes*, *39*(3), 393–398.

Utsunomiya, M., Attanucci, J., & Wilson, N. (2006). Potential uses of transit smart card registration and transaction data to improve transit planning. *Transportation Research Record*, *1971*(1), 118–126.

Van Breedam, A. (2001). Comparing descent heuristics and metaheuristics for the vehicle routing problem. *Computers & Operations Research*, *28*(4), 289–315.

Van Nes, R., Hamerslag, R., & Immers, L. H. (1988). The design of public transport networks . *Transportation Research Record,1202*, 74-83.

Yap, M., Nijënstein, S., & Vanoort, N. (2017). Improving predictions of the impact of disturbances on public transport usage based on smart card data. *Proceedings of the 96th TRB Annual Meeting, Washington, DC*, 8–12.

Yu, J., & Yang, X. (2006). Estimating a transit route od matrix from on-off data through an artificial neural network method. In *Applications of Advanced Technology in Transportation 2006*, 467–472.

Yuan, S., Skinner, B., Huang, S., & Liu, D. (2013). A new crossover approach for solving the multiple travelling salesmen problem using genetic algorithms. *European Journal of Operational Research*, *228*(1), 72–82.

Zhang, F., Yuan, N. J., Wang, Y., & Xie, X. (2015). Reconstructing individual mobility from smart card transactions: a collaborative space alignment approach. *Knowledge and Information Systems*, *44*(2), 299–323.

Zhang, T., Gruver, W. A., & Smith, M. H. (1999). Team scheduling by genetic search. *Proceedings of the Second International Conference on Intelligent Processing and Manufacturing of Materials*, *2*, 839–844.

Zhao, J., Qu, Q., Zhang, F., Xu, C., & Liu, S. (2017). Spatio-temporal analysis of passenger travel patterns in massive smart card data. *IEEE Transactions on Intelligent Transportation Systems*, *18*(11), 3135–3146.

APPENDICES

APPENDIX A: Clustering Algorithms, Floyd Warshal, and Analysis of Results

A.1 K- Means Clustering Algorithm

This is an algorithm that groups coordinate points into a specified "k" number sub groups. It is done in a such a way that each data point belongs to only one group and the data points so assigned to any particular cluster must be such that the arithmetic mean of all data points to that cluster is kept at a minimum. The following steps are used in this algorithm;

- Specify number of clusters *K*.
- Initialize centroids by first shuffling the dataset and then randomly selecting *K* data points for the centroids without replacement.
- Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.
- Compute the sum of the squared distance between data points and all centroids.
- Assign each data point to the closest cluster (centroid).
- Compute the centroids for the clusters by taking the average of the all data points that belong to each cluster.

A.2 K –Medoids Clustering Algorithm;

This algorithm is also similar to the Kmeans algorithms because both can be used to group datasets. While K means tries to minimizes the total squared error, K Medoids on the other hand tries it minimizes a sum of general pairwise dissimilarities instead of a sum of squared Euclidean distances. In essence K Medoids chooses data points as centres, i.e. a medoid of a finite dataset is a data point from this set, whose average dissimilarity to all the data points is minimal i.e. it is the most centrally located point in the set.

The *k*-medoid clustering is as follows:

- Specify the number of Medoids “k”
- Initialize randomly select *k* of the *n* data points as the medoids
- Assign each data point to the closest medoid.
- For each medoid and each data point associated to the medoid, swap the medoid and the data point and compute the average dissimilarity of the data point to all the data points associated to medoid).
- Select the medoid with the lowest cost of the configuration.
- Repeat alternating steps 3 and 4 until there is no change in the assignments.

A.3 Floyd Warshal Algorithm

This is a method to find all-pairs shortest paths in a graph. It does this by finding the shortest path between every pair of vertices in a graph. The following

For a graph with N vertices, the following steps;

- Initialize the shortest paths between any 2 vertices with Infinity.
- Find all pair shortest paths that use 0 intermediate vertices, then find the shortest paths that use 1 intermediate vertex and so on.. until using all N vertices as intermediate nodes.
- Minimize the shortest paths between any 2 pairs in the previous operation.

- For any 2 vertices (i,j) , one should actually minimize the distances between this pair using the first K nodes, so the shortest path becomes.

    min(dist[i][k]+dist[k][j],dist[i][j]).

    dist[i][k] represents the shortest path that only uses the first K vertices,

    dist[k][j] represents the shortest path between the pair k,j.

As the shortest path will be a concatenation of the shortest path from i to k, then from k to j.

Table A.1 Route structure

S/N	5 salesman	4 salesman	3 salesman	2salesman
1	37 29 34 56	37 29 43 56	37 29 34 56	55 50 48 54 53 47 37 39 34 56
2	54 53 47 56	55 54 53 47 56	51 46 42 38 56	51 46 42 38 56
3	55 50 48 56	50 48 26	47 53 54 55 50 48 56	1 9 10 16 19 23 30 24 20 26 25 57
4	51 46 56	51 46 42 38 56	1 9 10 16 19 23 30 57	49 43 36 39 35 57
5	38 42 56	1 9 19 16 10 20 57	26 20 24 25 57	52 45 44 40 41 33 32 27 31 58
6	16 19 23 30 57	20 25 57	49 43 36 39 35 57	14 15 21 22 58
7	1 9 10 20 24 57	30 23 24 57	52 45 44 40 31 58	2 8 7 11 17 59
8	26 25 57	49 43 36 39 35 57	41 33 32 27 58	3 5 12 6 4 13 18 28 59
9	39 35 57	14 15 21 22 58	14 15 21 22 58	
10	49 43 36 57	33 27 58	2 8 7 3 5 11 17 59	
11	15 21 58	41 32 58	4 6 12 13 59	
12	14 22 58	52 45 44 40 31 58	28 18 59	
13	33 27 58	2 8 7 11 17 59		
14	41 32 58	3 5 12 59		
15	52 45 44 40 31 58	4 6 3 59		
16	2 8 7 59	28 18 59		
17	3 5 12 59			
18	4 6 13 59			
19	11 17 59			
20	28 18 59			

Table A.2 Fundamental parameters

S/N	5 salesman			4 salesman			3 salesman			2 salesman		
	D	L	F	D	L	F	D	L	F	D	L	F
1	600	0.7	23.7	600	1.8	21.1	600	0.7	23.7	1800	2.4	21.3
2	600	0.8	22.4	800	1.2	22.2	800	1.1	22.2	800	1.2	21.9
3	600	0.8	22.0	400	0.5	21.4	1200	1.6	22.5	2200	3.1	22.9
4	400	0.8	17.8	800	1.2	19.4	1400	2.3	20.3	1000	1.3	21.7
5	400	0.5	23.6	1200	2.1	26.2	800	0.8	25.3	1800	2.6	24.8
6	800	1.3	20.7	400	0.4	22.0	1000	1.3	22.7	800	0.9	21.6
7	1000	1.6	20.3	600	0.8	22.8	1000	1.5	20.8	1000	1.4	20.5
8	400	0.4	26.5	1000	1.3	24.8	800	0.9	23.9	1600	2.5	21.3
9	400	0.4	25.8	800	0.9	22.7	800	0.9	24.8			
10	600	1.2	18.3	400	0.5	21.3	1400	2.1	21.2			
11	400	0.6	22.0	400	0.6	20.7	800	1.2	21.4			
12	400	0.5	22.7	1000	1.6	21.5	400	0.6	21.1			
13	400	0.5	22.9	1000	1.4	21.7						
14	400	0.6	21.3	600	0.9	12.5						
15	1000	1.6	20.7	600	2.6	20.7						
16	600	1.2	18.0	400	0.6	21.1						
17	600	0.9	21.7									
18	600	0.9	20.6									
19	400	0.5	22.2									
20	400	0.6	20.9									

Table A.3 Network connectivity

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58
1	0																																							1																		
2		0																																					1																		1	
3			0		1		1									1									1		1																															
4				0													1			1											1																											
5			1		0							1									1		1																																			
6						0				1								1							1																																	
7			1				0	1																			1																															
8							1	0			1															1					1																											
9									0																																										1	1			1			
10						1				0					1																					1																						
11								1			0																1	1																														
12					1							0												1																																		
13													0													1																																
14														0																		1				1																						
15										1					0																						1	1																				
16			1													0													1	1																												
17																	0				1										1																											
18						1												0												1													1				1											
19																			0																							1	1	1														
20				1																0	1	1																																				
21					1												1			1	0	1																																				
22																				1	1	0	1																																			
23					1																		0		1								1																									
24												1											1	0	1							1	1																									
25			1			1																		1	0					1																												
26								1					1													0																																
27			1				1				1																0																															
28											1																	0	1	1																		1										
29																1												1	0	1																												
30																1		1							1		1	1		0																		1										
31				1				1									1														0																											
32														1										1								0																							1	1		
33																							1	1									0	1																								
34																																	1	0	1																							
35																																		1	0				1																			
36										1				1																						0	1																					
37															1																					1	0		1																			
38															1																							0	1			1																
39		1																																	1		1	1	0	1	1																	1
40																																							1	0	1	1																
41	1																																						1	1	0																	
42																			1																			1		1		0	1															
43																		1	1																							1	0		1	1												
44																			1																									0	1													
45																																											1	1	0	1												
46																																											1		1	0	1			1								
47																		1																												1	0	1	1	1								
48																												1		1																	1	0	1				1					
49																																															1	1	0	1		1						
50																																														1	1		1	0	1							
51									1																																									1	0	1						
52									1																																								1		1	0	1		1			
53																																																1				1	0	1				
54									1																																												1	0	1			
55																						1										1																				1		1	0			
56																																1																								0	1	
57		1																																					1																	1	0	1
58																		1																																							1	0

Table A.4 Distance matrix

Bus Stop No.	X-Coordina	Y-Coordinate	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BS8	BS9	BS10	BS11	BS12	BS13	BS14	BS15	BS16	BS17	TR1	TR2
BS1	515700.6	4250248.36	0	3137.104	5479.221	2721.559	1816.382	3643.994	4354.261	12047.42	1573.254	4432.026	2417.071	6252.999	1721.237	1307.426	3761.534	4672.891	2803.135	2709.142	4339.525
BS2	515830.5	4250859.13	3137.104	0	2427.445	440.7166	1320.722	506.8903	1217.158	11312.54	1563.849	1294.922	745.2046	3115.896	1441.039	1829.678	624.4301	2152.159	1888.804	5345.534	2058.683
BS3	514514.1	4250809.92	5190.677	2427.445	0	2469.119	3748.166	1920.555	1210.287	12560.62	3991.294	2456.92	2773.607	3109.025	3469.441	4257.123	3051.875	517.7865	4316.249	7772.979	1165.875
BS4	514937.8	4251701.06	2721.559	440.7166	2757.662	0	1761.438	947.6069	1657.874	11753.26	1661.113	1735.638	304.488	3556.612	1000.322	1905.728	1065.147	1951.332	2329.521	5327.215	1617.966
BS5	519933.4	4259238.05	1816.382	1320.722	3748.166	1761.438	0	1827.612	2537.879	10543.12	243.1279	2615.643	1599.753	4436.617	903.9187	508.9563	1945.152	3472.881	986.7524	4024.813	3379.404
BS6	514275.4	4250856.23	3643.994	506.8903	1920.555	947.6069	1827.612	0	710.2673	11405.63	2070.74	1301.923	1252.095	2609.005	1947.929	2336.568	1131.32	1645.269	2395.695	5852.425	2565.573
BS7	516905.6	4250606.01	4354.261	1217.158	1210.287	1657.874	2537.879	710.2673	0	11350.34	2781.007	1246.633	1962.362	1898.738	2658.196	3046.835	1841.588	935.0019	3105.962	6562.692	2376.162
BS8	514477.3	4250334.87	12047.42	11312.54	12560.62	11753.26	10543.12	11405.63	11350.34	0	10786.25	10103.7	11861.64	13249.08	11447.04	11052.08	10688.11	12285.34	9556.37	11883.12	13371.23
BS9	514013.2	4250900.17	1573.254	1563.849	3991.294	1661.113	243.1279	2070.74	2781.007	10786.25	0	2858.771	1356.625	4679.745	660.7908	265.8284	2188.28	3612.445	1229.88	3781.685	3279.079
BS10	511223.6	4252805.08	4432.026	1294.922	2456.92	1735.638	2615.643	1301.923	1246.633	10103.7	2858.771	0	2040.126	3145.371	2735.96	3124.6	1919.352	2181.634	3183.726	6640.456	3353.605
BS11	516831.7	4248341.36	2417.071	745.2046	3062.15	304.488	1599.753	1252.095	1962.362	11861.64	1356.625	2040.126	0	3861.1	695.834	1601.241	1369.635	2255.82	2323.098	5022.727	1922.454
BS12	516694.3	4248840.6	6252.999	3115.896	3109.025	3556.612	4436.617	2609.005	1898.738	13249.08	4679.745	3145.371	3861.1	0	4556.934	4945.574	3740.326	2833.74	5004.7	8461.43	4274.9
BS13	515372.3	4249954.34	1721.237	1441.039	3757.984	1000.322	903.9187	1947.929	2658.196	11447.04	660.7908	2735.96	695.834	4556.934	0	905.4065	2065.469	2951.654	1890.671	4326.893	2618.288
BS14	515998.5	4249838.23	1307.426	1829.678	4257.123	1905.728	508.9563	2336.568	3046.835	11052.08	265.8284	3124.6	1601.241	4945.574	905.4065	0	2454.108	3857.061	1495.709	3515.856	3523.695
BS15	516563.1	4249407.34	3761.534	624.4301	3051.875	1065.147	1945.152	1131.32	1841.588	10688.11	2188.28	1919.352	1369.635	3740.326	2065.469	2454.108	0	2776.589	1264.374	5160.12	2683.113
BS16	515074.1	4250015.72	4672.891	2152.159	806.3303	1951.332	3472.881	1645.269	935.0019	12285.34	3612.445	2181.634	2255.82	2833.74	2951.654	3857.061	2776.589	0	4040.964	7278.547	1972.205
BS17	517723.1	4248365.47	2803.135	1888.804	4239.739	2329.521	986.7524	2319.185	3029.452	9556.37	1229.88	2856.704	2323.098	4928.19	1890.671	1495.709	1264.374	3964.454	0	3895.746	3947.487
TR18	512993.6	4250928.99	2709.142	5345.534	7772.979	5327.215	4024.813	5852.425	6562.692	11883.12	3781.685	6640.456	5022.727	8461.43	4326.893	3515.856	5160.12	7278.547	3895.746	0	6945.181
TR19	512876.3	4250066.67	4089.109	2058.683	1165.875	1617.966	3379.404	2565.573	2376.162	13371.23	3279.079	3353.605	1922.454	4274.9	2618.288	3523.695	2683.113	1683.661	3947.487	6694.765	0

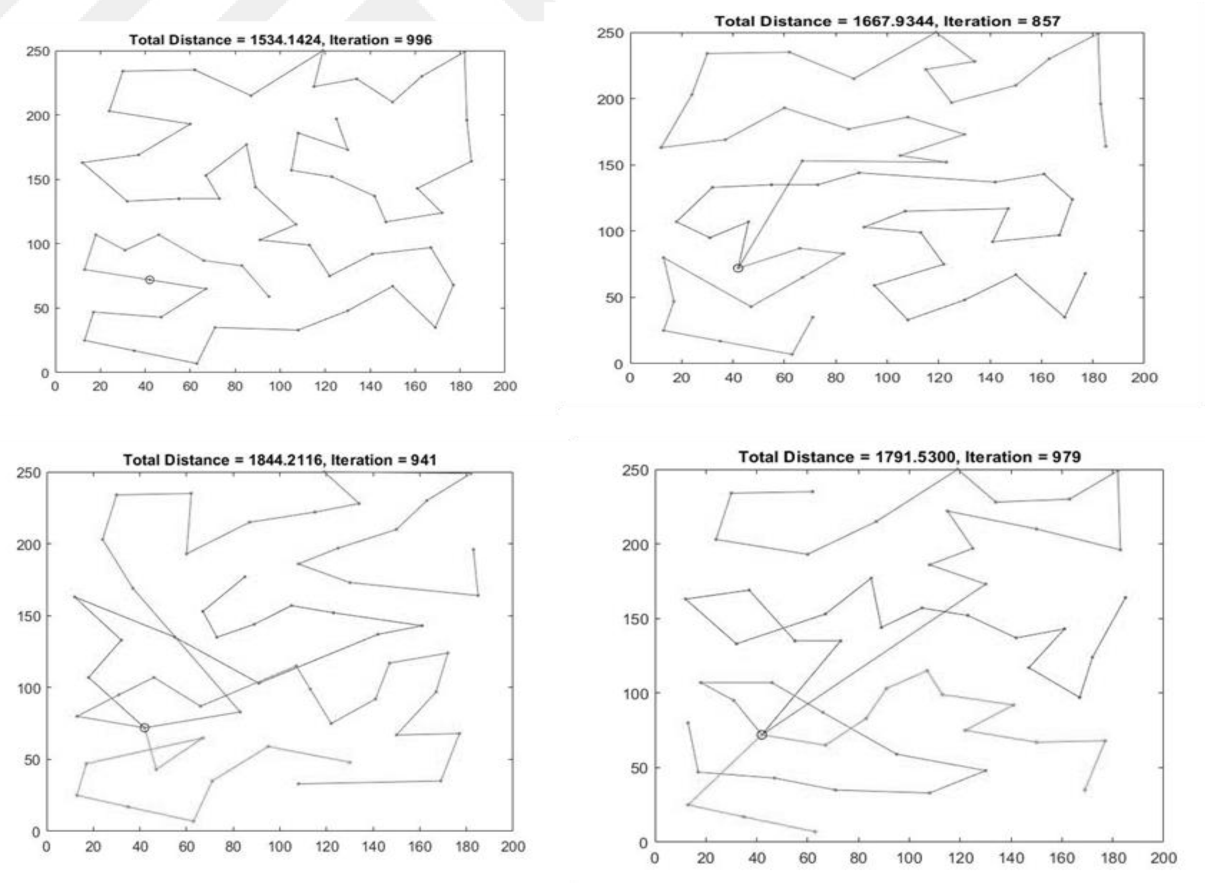

Figure A.1 Two, three, four, and five route structure solutions for single station problem

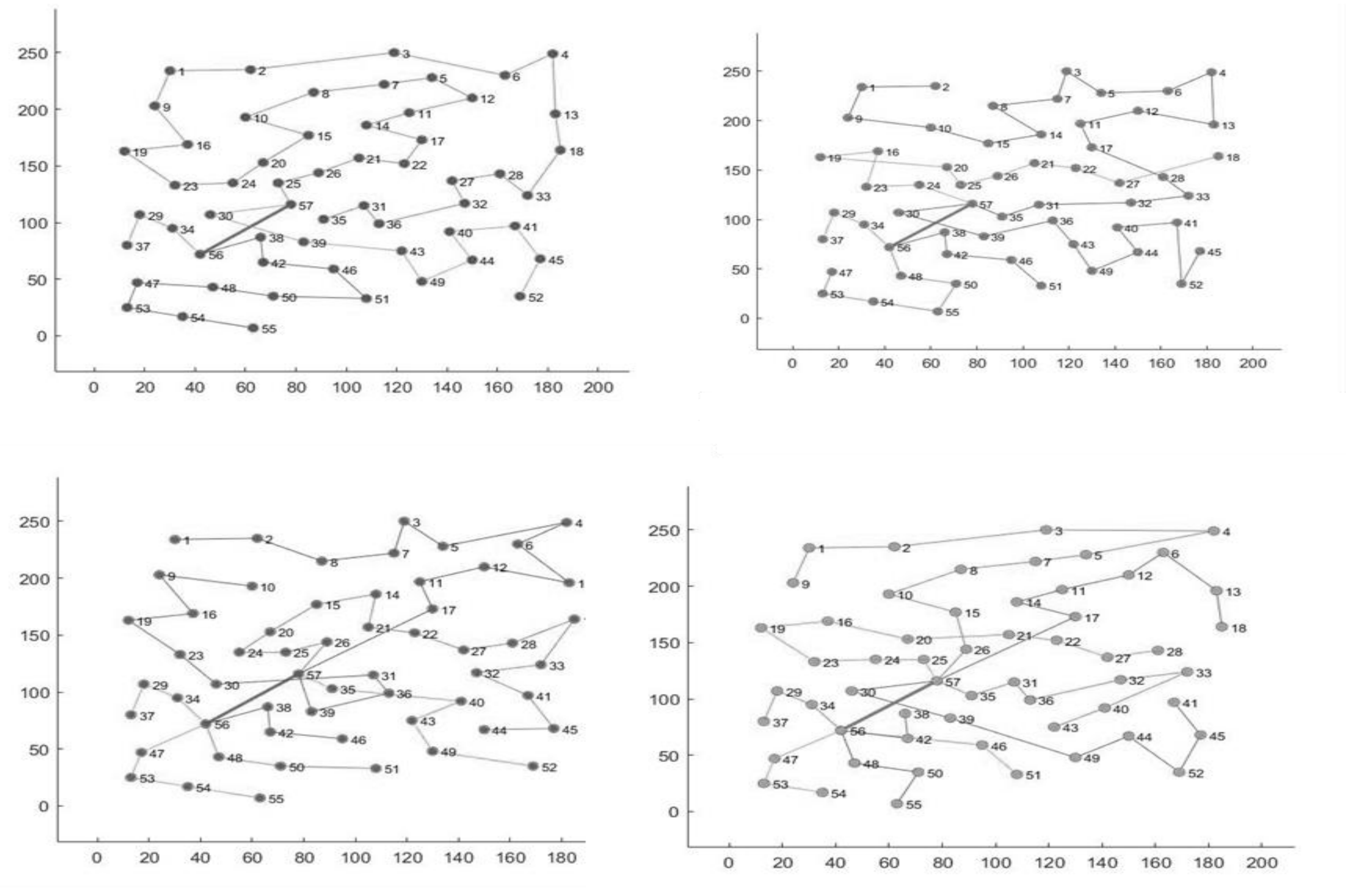

Figure A.2 Two, three, four, and five route structure solutions for two station problem

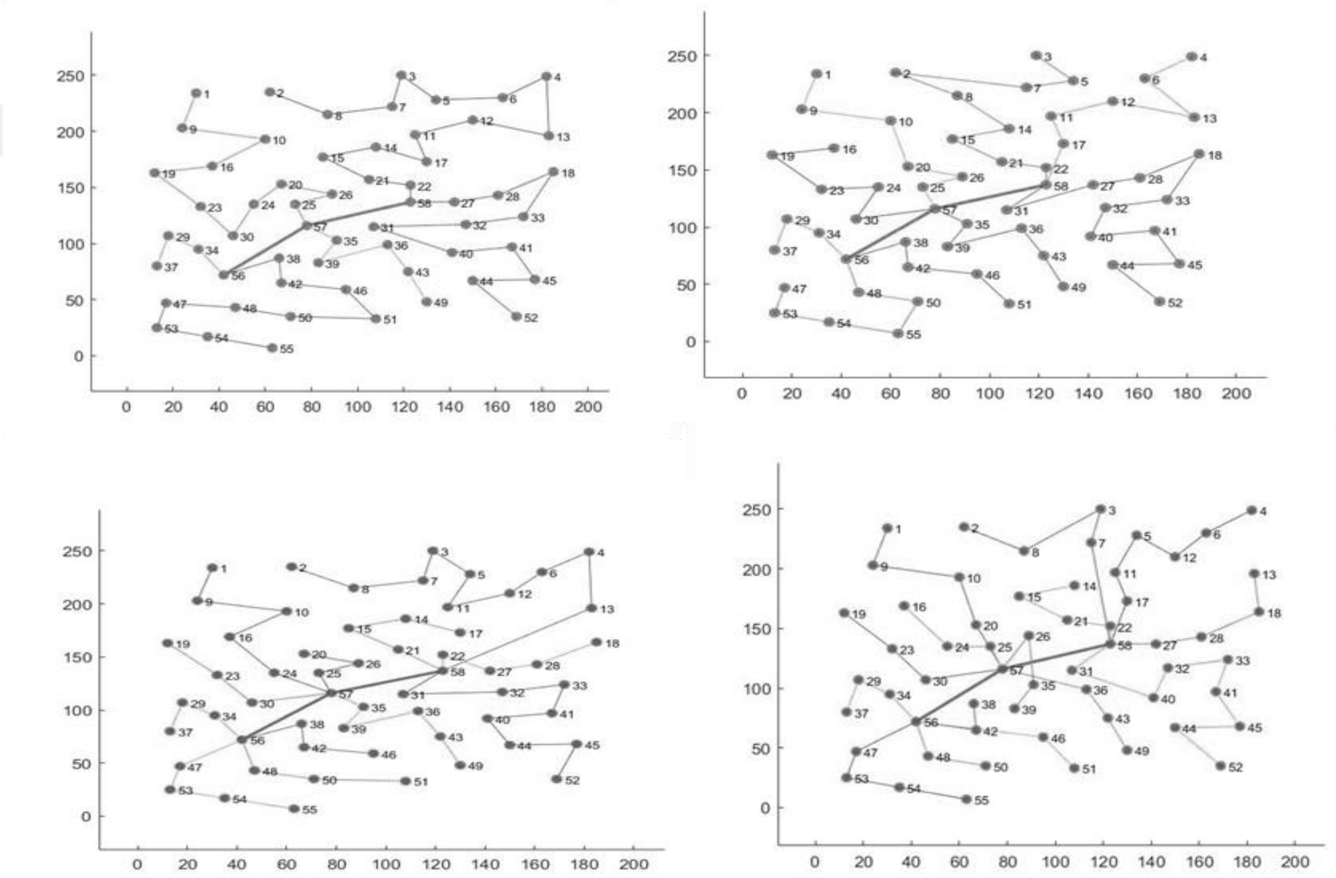

Figure A.3 Two, three, four, and five route structure solutions for three station problem

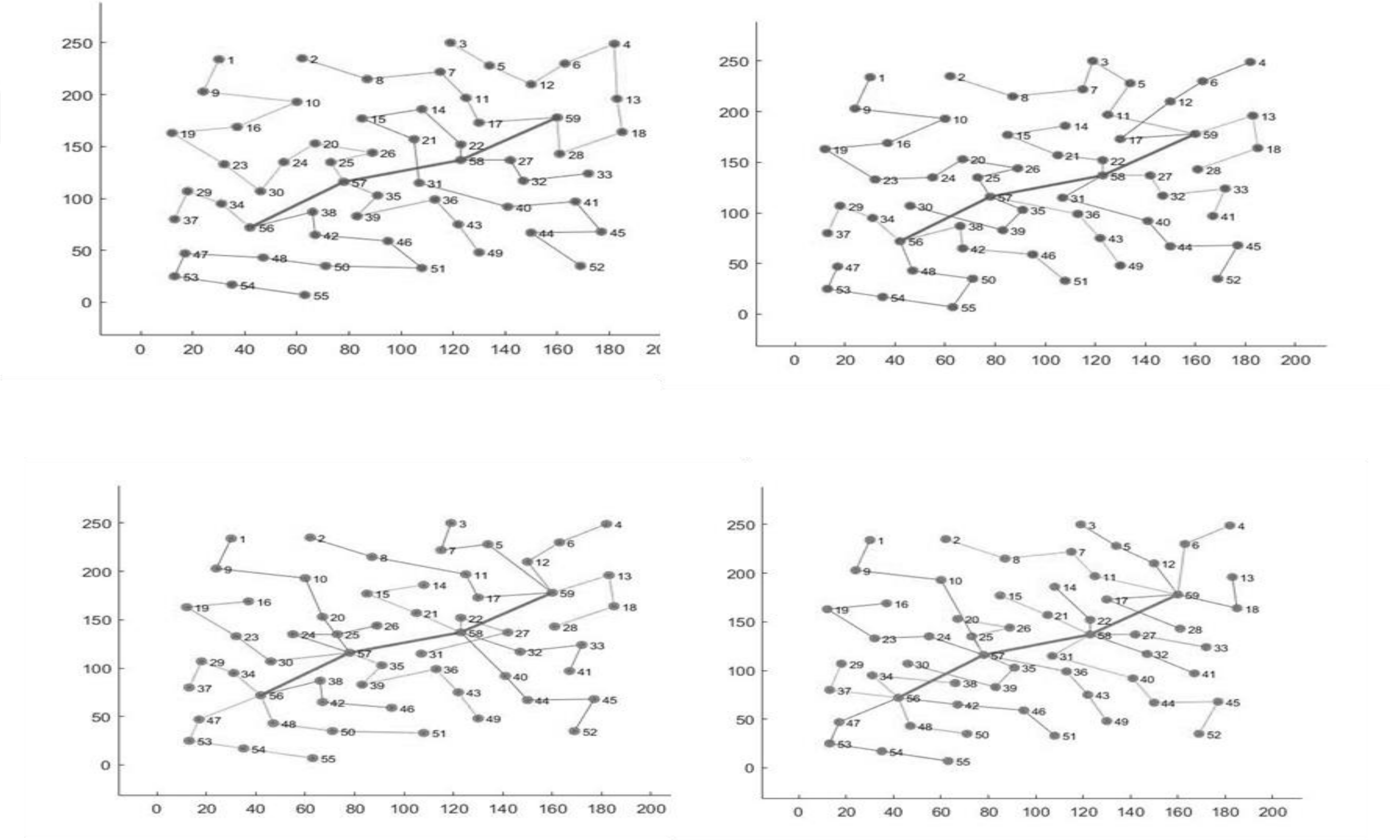

Figure A.4 Two, three, four, and five route structure solutions for four station problem

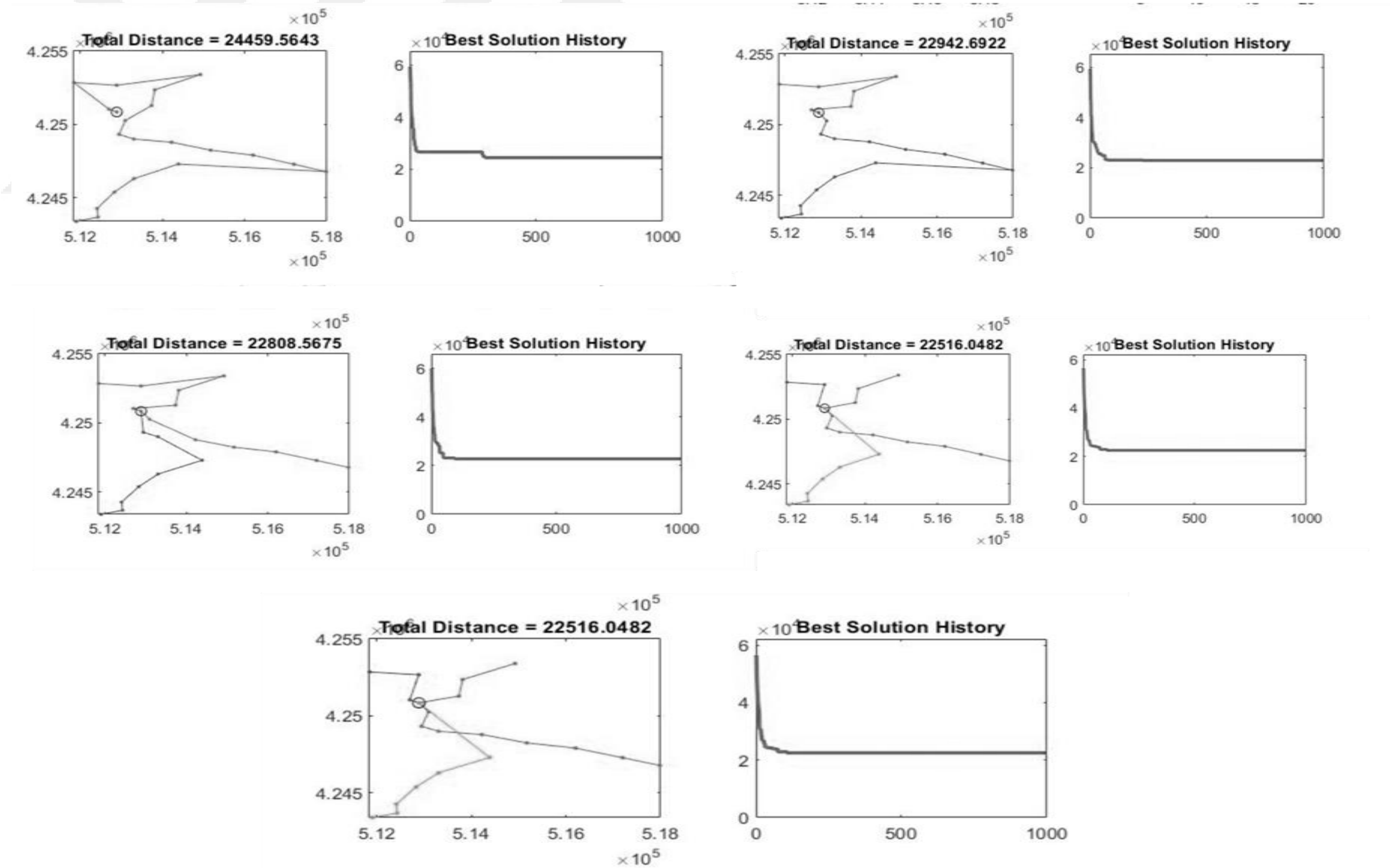

Figure A.5 Total distance and solution history for existing feeder bus route

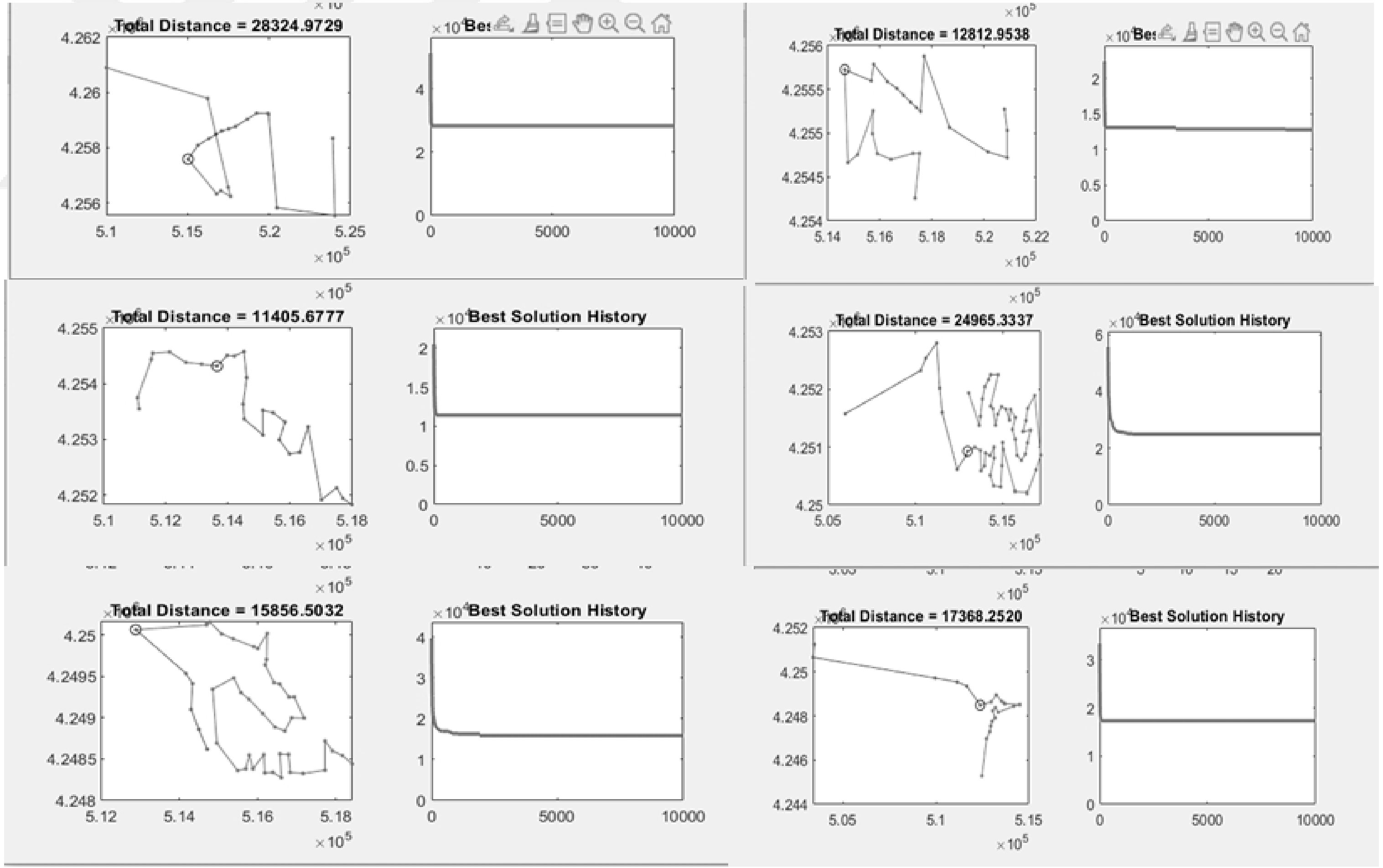

Figure A.6 Total distance and solution history for 6–station problem (without clustering)

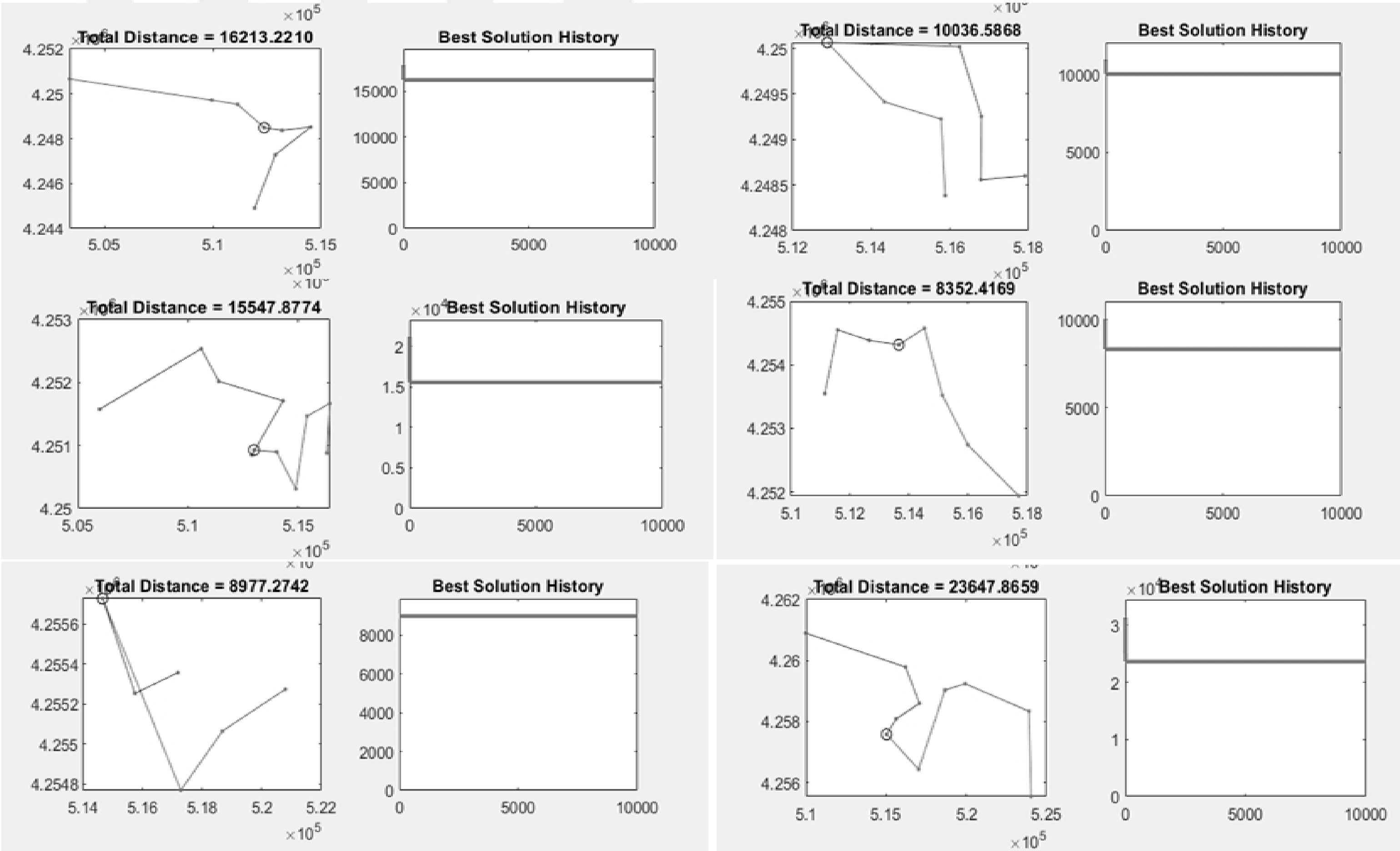

Figure A.7 Total distance and solution history for 6 –station problem (k-medoids clustering)

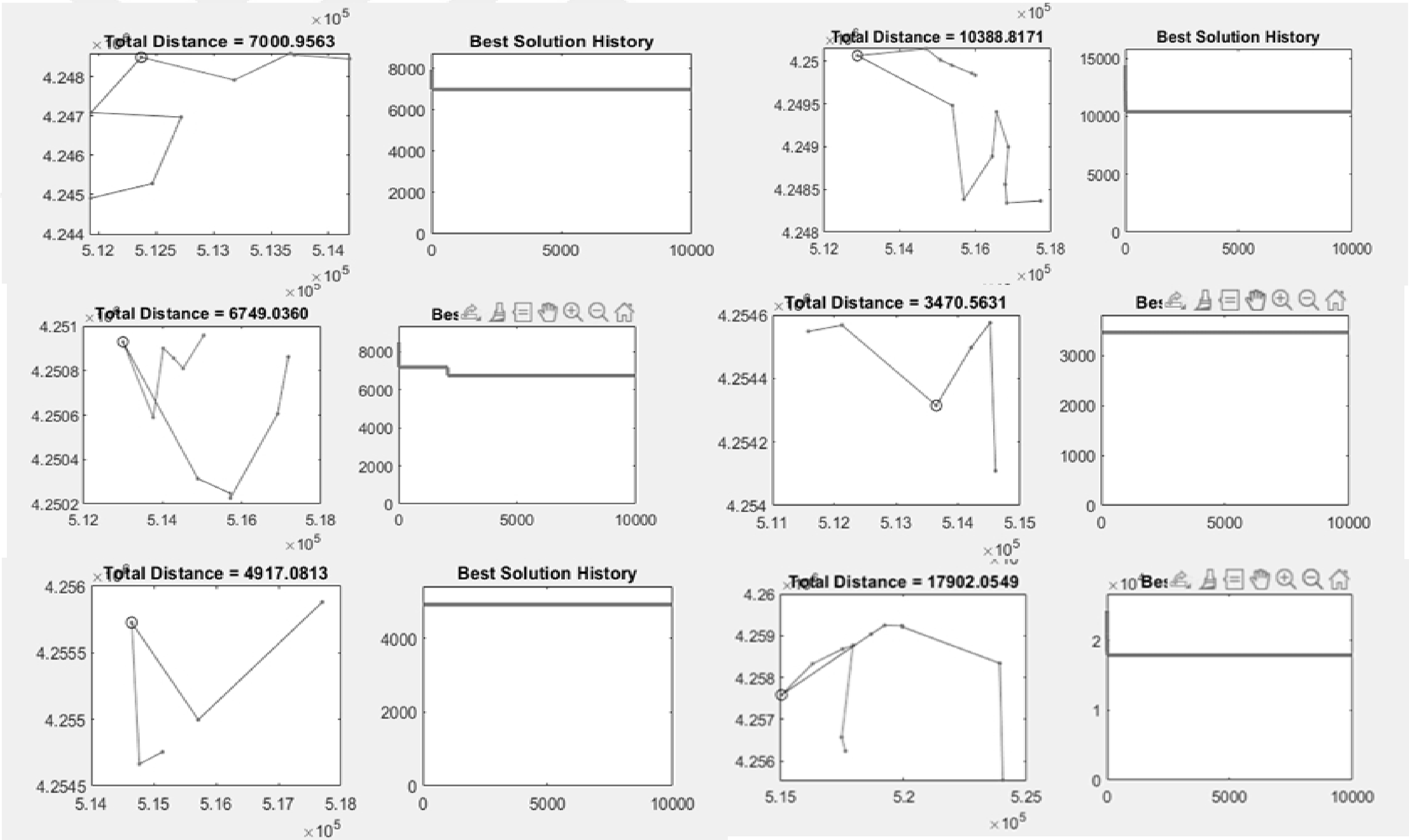

Figure A.8 Total distance and solution history for 6 station problem (capacitated clustering)

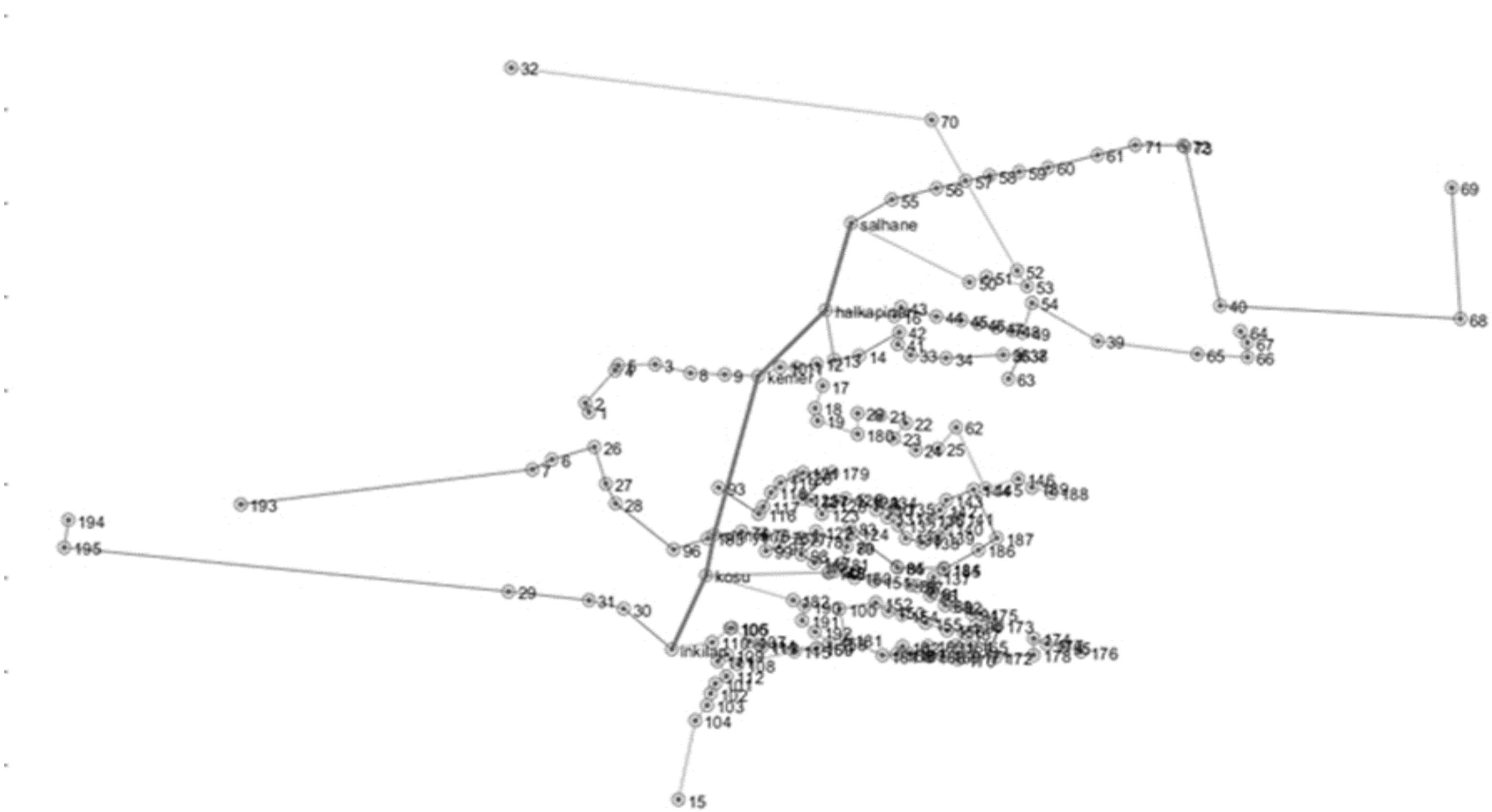

Figure A.9 Two route structure solution for 6–station problem (without clustering)

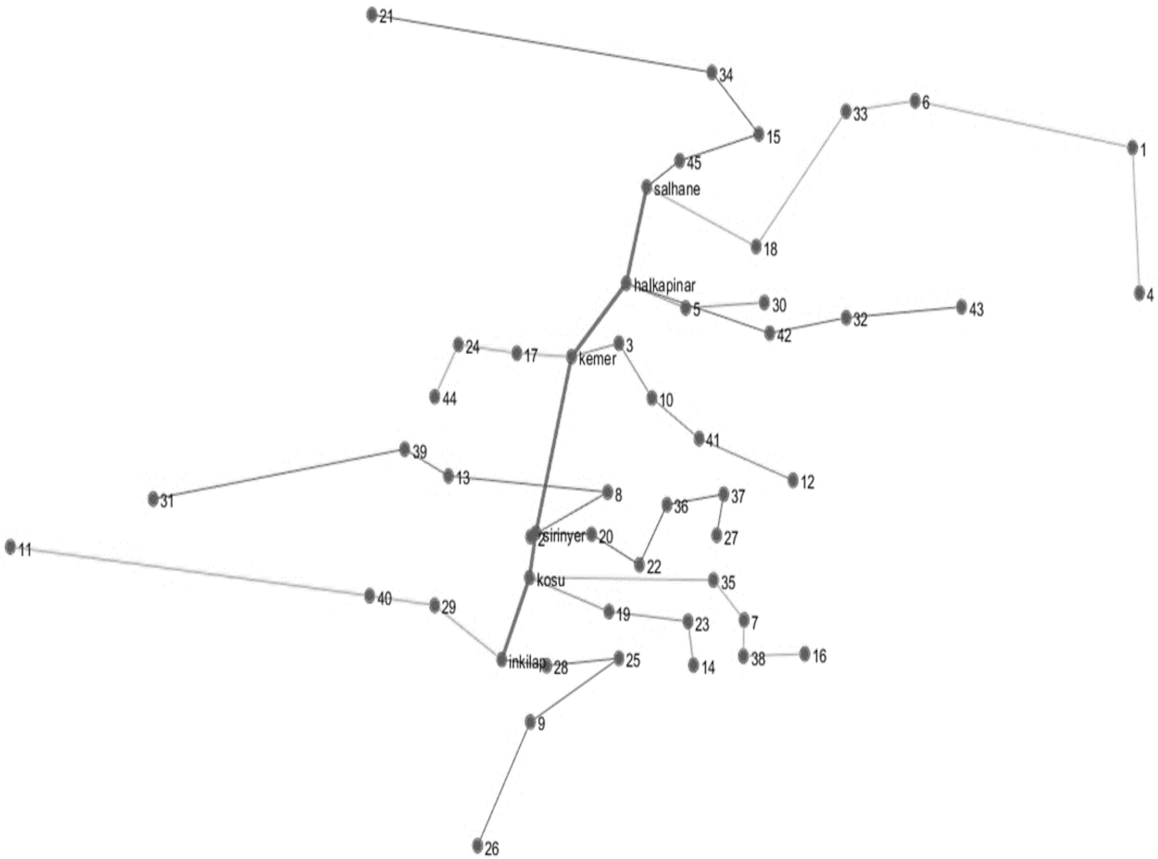

Figure A.10 Two route structure solution for 6–station problem (K-medoids clustering)

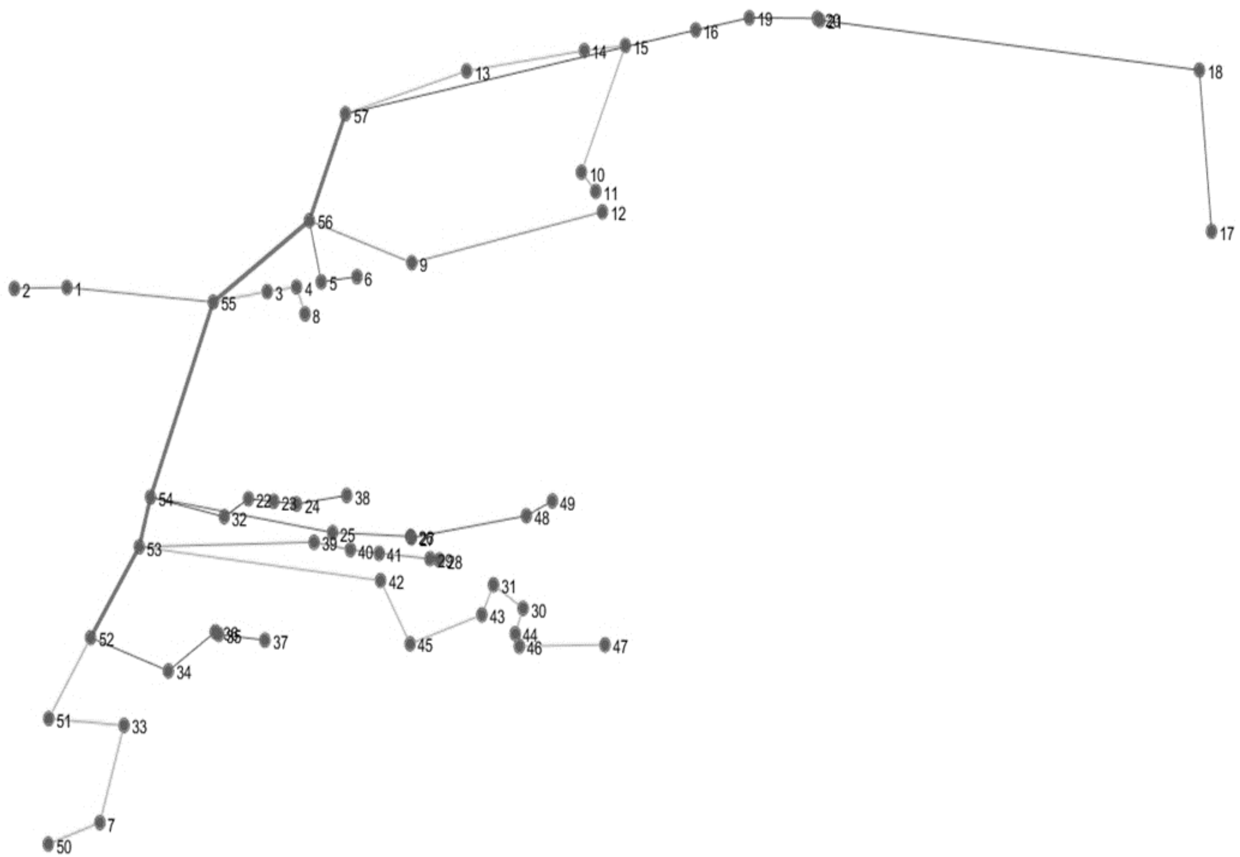

Figure A.11 Two route structure solution for 6–station problem (Capacitated clustering)