

ABSTRACT

SENGUPTA, DIBYENDU. Internet Applications for Paratransit Performance Analysis: The Winston-Salem Case Study (Under the direction of Dr. John R. Stone.)

Advanced technologies are utilized to improve transportation systems under the aegis of Intelligent Transportation Systems. In such cases, performance measures, which consist of a set of productivity and service indicators, help decisions about transportation resources, their use and allocation.

This study focuses on the paratransit system at the Winston-Salem Transit Authority (WSTA) for development of an application for retrieving and evaluating paratransit performance data. The study demonstrates the use of Internet connectivity for easy-to-use applications with graphics capabilities, and statistical testing for before-after analysis of the data.

An important result of this work is the prototype development and demonstration of Internet based application for analysis of paratransit productivity data. The application was then used to evaluate the impact of incoming technologies like Interactive Voice Response (IVR) on the performance measures. Using automatic telephone technology, IVR, allows transit customers to use their home telephone to book and cancel transit trips.

While IVR technology has the primary objective of improving customer satisfaction via improved communication, the case study showed that the IVR could have secondary impacts on productivity such as increased cancellations. With the exception of cancellations, all the measures did not show significant changes with the addition of IVR technology and several policy changes

**Internet Applications for Paratransit Performance Analysis:
The Winston-Salem Case Study**

by

Dibyendu Sengupta

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Civil Engineering

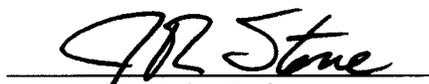
Raleigh

2002

Approved by:


Dr. John W. Baugh, Member


Dr. Nagui M. Roupail, Member


Dr. John R. Stone
Chair of Advisory Committee

BIOGRAPHY

Dibyendu Sengupta was born and raised in the steel city of Jamshedpur, India. He received his elementary and secondary education in Jamshedpur. He then did his undergraduate studies with a major in Civil Engineering from Indian Institute of Technology, Kharagpur, India in 1997 successfully completing the bachelors thesis titled “Time Series Analysis and Forecasting of Hydrological data using Recurrent Neural Networks.” He worked for two years in the field of aviation in New Delhi, India from July 1997 to May 1999, prior to joining North Carolina State University, Raleigh for pursuing MS in Civil Engineering.

ACKNOWLEDGEMENTS

I would first like to thank my thesis committee and professors at North Carolina State University including Dr. John Stone, Dr John W. Baugh Jr and Dr Nagui M. Roupail. It has been a tremendous learning experience working with them. Dr Stone, my thesis advisor, has been a constant source of encouragement and motivation for me.

I am thankful to Suzanne Tellachea and Linda Little of Winston-Salem Transit Authority for all the help in collecting data and other related information. I would also like to thank Christopher Bryant of Trapeze Software for providing invaluable information about the Trapeze database and system at the WSTA facility.

I would also like to thank my parents and my sister, Sushmita for being a constant source of support.

Thanks to all the good friends around me, including Sachin Bammi, Nirmal Govindaraju, Kushal Narayanaswamy, Girish Rao and Meeta Yadav. My due thanks also to all the great people I met during my work at NC State. They include James Colyar, Matt Davis, Ko Sok Chae, Kim Krull, Marco Navarro, Brandon Nevers, Tori Rhoulac, and Leane Wissinger.

TABLE OF CONTENTS

LIST OF FIGURES	v
CHAPTER 1: INTRODUCTION.....	1
1.1 Background	1
1.2 Problem Definition.....	2
1.3 Scope and Objectives.....	7
1.4 Importance of Paratransit System Productivity Research.....	9
1.5 Thesis Overview	10
CHAPTER 2: LITERATURE REVIEW	11
2.1 Paratransit Performance Analysis	11
2.2 Statistical Applications for Paratransit Performance Analysis	13
2.3 Web-based Applications	14
2.4 Decision Support.....	18
2.5 Summary	19
CHAPTER 3: INTERNET DECISION SUPPORT TOOL.....	22
3.1 Introduction.....	22
3.2 Architecture of the Software tool.....	23
3.3 The Winston-Salem Data	26
3.4 Decision Support.....	33
3.5 Development and Implementation Issues	36
3.6 Development of the Tool	38
3.7 The Separate Modules.....	42
3.8 Summary	52
CHAPTER 4: ANALYSIS OF DATA.....	53
4.1 Data Formatting Procedure	54
4.2 Data Verification.....	55
4.3 Statistical Analysis of Productivity Data	55
4.4 Conclusions	69
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS	71
5.1 Summary	71
5.2 Conclusions	72
5.4 Recommendations for Future Research.....	74
CHAPTER 6: REFERENCES	78
APPENDIX A.....	82
Glossary of Terms:.....	83
APPENDIX B	88
1. Sample Reports	89
2. Database Screenshots	94
3. Sample Charts	95
4. Examples of SQL commands.....	96
5. Sample Java Servlet code for extracting data from the database	97

LIST OF FIGURES

Figure 2. 1: Comparison of features of referenced studies	20
Figure 2. 2: Table of the Reporting Techniques	20
Figure 3. 1: 3-tier Architecture Schema	25
Figure 3. 2: Data Flow and Integration Diagram	29
Figure 3. 3: Sample Trips/Miles Productivity Report	30
Figure 3. 4: Sample Provider Productivity Report	31
Figure 3. 5: Sample Section of a raw Ride File	32
Figure 3. 6: Decision Flow Diagram	35
Figure 3. 7: Screenshot of the Opening Page	42
Figure 3. 8: Decision Flow of the Data Display Module	44
Figure 3. 9: Screenshot of Data Display	45
Figure 3. 10: Screenshot of the Performance Analysis Module	45
Figure 3. 11: Decision Flow of the Performance Analysis Module	46
Figure 3. 12: Sample screenshots of the Graphical Displays	47
Figure 3. 13: Decision Flow of the Statistical Analysis Module	48
Figure 3. 14: Sample screenshots of the t-test and ANOVA modules	48
Figure 3. 15: Sample screenshot of the Technology Module	49
Figure 3. 16: Decision Flow of the Technology Module	52
Figure 4. 1: Sample Validated File for the month of March 2000	57
Figure 4. 2: Sequence of Technology and Policy changes	58
Figure 4. 3: Time Series Plots	60
Figure 4. 4: Time Series Plot for Monthly Data	63
Figure 4. 5: Summary of the t-test results	66
Figure 4. 6: Summary of the ANOVA results	68

CHAPTER 1: INTRODUCTION

Paratransit is defined as a demand responsive specialized, shared ride, curb-to-curb service for people with disabilities who are not able to ride fixed-route public transportation, including lift-equipped buses (Lave and Mathias, 1992). The use of paratransit for the elderly and disabled has been increased in conjunction with the Americans with Disabilities Act (ADA) legislation, which strives to give people equal access to public transportation and prohibits discrimination against people with disabilities. This thesis will investigate the state of paratransit system performance analysis with the primary objective of developing stronger statistical testing using the Internet as a medium. A secondary objective will be to determine the possible effects of advanced technologies and information on the system productivity.

The development work and analysis for this thesis was part of the research project titled “Transit Customer Satisfaction with Real-Time Information”. This overall project looked at the types and magnitudes of changes of transit system performance and customer satisfaction derived from implementation of real-time customer information. The research work was funded by the Winston-Salem Transit Authority as a part of the Mobility Management National Demonstration Project, which is funded by the US DOT Federal Transit Administration (Stone, 2000).

1.1 Background

The Winston-Salem Transit Authority introduced the Mobility Manager National Demonstration Project in response to the Federal Transit Administrations’ recognition of WSTA as a national demonstration site for Intelligent Transportation Systems (ITS)

technology. Over the period of 1994 to 2001, the mobility management project has had two major phases. Phase I emphasized system performance improvement and therefore tested different ITS technologies including computerized scheduling, mobile data terminals (MDT), and automatic vehicle locations (AVL).

This project is a part of Phase II and aims to interface ITS components for providing real time information to the paratransit and fixed-route passengers. These include an interactive voice response (IVR), mobile data terminals, Automatic Vehicle Locations and dynamic variable message signs. The overall objective of the Mobility Management is to advance a single point source of information of transportation options through real time information. By the summer of 2001, some of these components like the IVR system had already been installed and some like the MDTs and AVLS were scheduled for 2002.

1.2 Problem Definition

The dictionary meaning of *para* is "beside, near or alongside". Paratransit therefore includes "all public and private mass transportation in the spectrum between private automobile and conventional transit" (Lave and Mathias, 1992). Some of the prime examples of paratransit services in the country are the Maryland MTA, the Valley Transportation Authority (OUTREACH), the Montgomery Area Paratransit Service (MAPS) and the Winston-Salem Transit Authority (TransAID).

Over the years, paratransit systems have adopted many technologies for use in their service. Some of these are digital radio frequency data communication, mobile data terminals, vehicle location devices, Geographic/Mapping Information Systems, card based data storage, automatic telephones and Internet based technologies (Lave and Mathias, 1992). Public transit organizations use performance data coming in from these peripheral technologies, for management, planning and monitoring purposes.

Performance measures are calculated so that federal, local, and state agencies can keep track of the system productivity and efficiency. Some of the commonly used performance service measures include passenger trips per day, vehicle miles per day, on-time performance, service frequency, ride time and cost-based measures like operating cost per passenger trip. Agencies use these measures to monitor and set priorities for sustaining, improving system performance, evaluating system changes, determining staff performance and estimating customer service.

One of the keys to effective transit management is to compare a system against some standards of performance; mostly through historical data or similar such data of peers. This is done, by arriving at a set of performance indicators that look at resource utilization and transit output (Dahlgren, 1998). Some of the important issues that arise in transit performance analysis are data collection, data validation, processing, reporting, decision-making, and the implementation.

Some of the commonly used techniques for assessing transit service efficiency and effectiveness in a transit organization is by using a *Before-and-after Analysis* consisting of evaluating service measures before and after a particular occurrence. This is usually done with the help staff surveys, customer surveys, tabular daily, weekly, monthly and annual data. Statistical methods like averages of data, time series distributions, graphs, charts and t-tests for evaluating measures (Stone, 2000).

The general approach in transit performance analysis is to gather performance data, process it using various simple statistical measures like averages and summations on a daily, weekly or monthly basis. Managers and supervisors report the processed data is then used for reporting to the state and federal agencies in the form of graphs and trends, which are then used to monitor transit performance relative to local and national standards. According to Lave and Mathias (1997), several issues arise in the measuring of performance data –

- Developing an internal database for the agency to compare changes in its own performance.
- Determining the performance of a system with respect to absolute measures of effectiveness.
- Accounting for the variation in geography, local transit conditions and rules and regulations.

Prominently missing in this currently used approach are several important steps like:

- Data Verification – For weeding out bad or corrupted data which may have arisen due to several reasons like incorrect measurement, incorrect data entry etc.
- Real time data monitoring – The data collected by agencies are usually not accessible quickly due to reasons like absence of data in digital format or absence of tools to access the databases.
- Interactive decision support – The data collected are usually not used in any kind of real time decision making due to lack of interactive performance analysis tools.

Thus we see that the current transit performance analysis suffers from lack of real time data monitoring and verification. Use of web-based tools provides the new facet of real time analysis. These tools can combine in them a lot of computational and analytical power like graphical displays, validation and statistical testing which used in conjunction can provide a real time decision monitoring system.

Paratransit like other transportation systems provides benefits to travelers, staff and management. However it also imposes costs on the traveler and the system. Monitoring of system performance can help identify ways of improving performance and therefore increasing the net benefits of the system although the end goals may differ from agency to agency. In some cases, the focus may be on operating costs while in the other it

may be passenger counts. Thus system performance monitoring is closely linked to the core goals of a transit agency.

In the case of Winston-Salem, with the introduction of advanced technologies like the AVLs and MDTs, real time information will be available to the users of the Trans-AID paratransit system including the staff, drivers and the passengers. A *direct impact* of this information availability will be the customer satisfaction. In addition to that, question arises as to whether this availability of real time information will result in any changes in the system productivity. Also pertinent, are the types and magnitudes of operational benefits that are derived from real-time customer information with the addition of new technologies.

1.2.1 Measuring Performance

One of the keys to effective transit management is to compare a system against some standards of performance; mostly through historical data or similar such peer data.

The ATSC¹ has developed an important volume of work on measurement of demand response operational performance. The set of performance indicators that "look at resource utilization and transit output" that were arrived upon is –

- Cost per passenger mile, the cost to transport a passenger one mile.
- Cost per seat mile, the cost to make one seat available for one mile.
- Load factor, the percentage of seats that are filled with paying passengers.
- In-service, the passengers per driver.

¹ See Glossary of Terms, Appendix A

These measures can be used directly to compare different systems, assess changes from one year to another and locate sources of improvement. It is also possible to deconstruct the data with respect to particular routes or regions of a service area, given enough disaggregation of the data. These are, however, difficult for the understanding of relationship among the measures since each of the measures is not on the same scale. (Dahlgren, 1998). ATSC went on to define a model to show these interrelationships. The model utilizes three independent measures to create a single index number. The measures were cost per seat mile, load factor and in-service productivity. The data from each of these measures were transformed into normalized index numbers between zero to one and each of the measures was given a particular weight.

According to Dahlgren (1998), measures for performance monitoring for public transit and passenger transport agencies usually include costs by category, passenger trips, vehicle-revenue-miles etc to determine where resources can best be deployed.

1.2.2 Performance Analysis

System performance and its analysis entail identifying certain measures, which effectively portray the service effectiveness and system productivity. As previously described, some typical productivity measures include passenger-trips/day, vehicle miles/day, vehicle hours/day, passengers/vehicle hour, vehicle miles/passenger trip and vehicle hours/passenger trip.

These productivity measures, which are usually in the form of monthly reports, can be then displayed graphically in the form of tables, histograms, pie charts etc along with the basic statistical measures like mean, median and variance. These basic statistical values give an intuitive idea about the performance of the system and help the transit staff to understand the trends and the changes that occur in these productivity measures.

Intuitive reasoning might, however, be incorrect. Thus stronger statistical tests are required to understand the performance trends and detect significant changes.

Such tests can be applied at regular intervals of time to gauge changes in trends that can occur due to several reasons like changes in the technology, fares etc and assuming the operational methods remain same. (Stone, 2000)

To identify which changes are most effective, “experimental control” is desired i.e. one transit service or operational change at a time is made. Otherwise, confounding effects may occur (Stone, 2000).

1.3 Scope and Objectives

The primary goals of the Mobility Manager Phase II research work are

- (i) to determine the relationship between transit information availability and customer satisfaction; and
- (ii) to determine the effects on transit performance measures with application of advanced technology and availability of real time information for scheduling and communications.

The focus of this thesis is the second topic – the effect of introducing advanced technology on transit productivity as measured by a variety of performance measures.

In the current scenario, system performance monitoring of a paratransit system is done with a certain number of measures and these measures are linked to an agency’s primary goals. As mentioned in the previous section there is, however, a scarcity of real time data retrieval and analysis, to enable quick assessment and decision-making. Also

required are real time analysis tools, for statistical testing of the system measures for stronger detection and verification of changes occurring to the system. In other words, what is needed is a decision support tool that addresses a broad range of concerns of the transit agency from the system performance side. This thesis seeks to address this topic.

Considering the macro level objectives of a paratransit agency and the narrowing down to the goals of the Mobility Manager project of Winston-Salem Transit Authority, assessment of changes in the system is of the ultimate concern. Hence, broadly speaking the main objective of this work is to assess the changes in system productivity with the availability of real time information to the system and the staff. A secondary objective is to apply the productivity assessment methods within the framework of an Internet-based decision support system that can be easily accessed by staff, patrons and external users. Moving away from the traditional methods of assessing system productivity by simple estimation of trends with little or no involvement of data validation and statistics is another objective that we address. The justification for the method that has been adopted to do this and the method itself will be elaborated in the chapters that follow.

The above objectives of this research work were realized by completion of the following tasks:

1. Collect, format, validate and analyze system performance data for the WSTA Trans-AID paratransit system case study.
2. Use the Trans-AID paratransit data to develop and demonstrate the feasibility of Internet based decision support tools for display and statistical analysis of the data for evaluating changes as mentioned in the previous task.
3. Test the effects of the availability of real-time data and performance measures on system productivity.

The above tasks had to be performed within the limitation of certain constraints. These are as follows:

- Absence of real-time data to which the applications would be connected to, since this would require connection to the proprietary Infoserver database. The prototype has been built with the applications connected to a prototype database designed using reports collected from WSTA.
- Since the database is created from reports generated by a pre-processor from the raw ride in the data server, validation of the ride data was not possible.

These limitations prevent the prototype from being a working model that provides the users access to the real time data from the WSTA archives. However, the prototype provides a working demonstration of the feasibility of the technologies used in building a real world implementation of a web-based tool for paratransit decision support.

1.4 Importance of Paratransit System Productivity Research

Paratransit is an important mode of transit as it supplements regular transit service and serves a neglected part of the community – the elderly and the disabled. Efficiency of paratransit performance is of inherent importance, even for a small fleet of buses and taxis, because such service is usually paid for by tax dollars. It is therefore important that the paratransit agencies be able to observe and analyze the overall system performance in an easy and timely fashion.

The Internet provides a convenient way for transit agencies to connect to a database and extract data for processing and analysis from any part of the world. Its use as a tool for building applications is fast increasing due to ease of use and interconnectivity. The use of the Web as a medium for decision support of paratransit data is therefore an obvious choice.

Stronger statistical analysis of paratransit data can be a novel way of finding unseen trends and cycles in data and analyzing changes in the productivity as service and technology improvements occur.

1.5 Thesis Overview

The thesis is composed of five chapters. Chapter 1, the *Introduction* gives a brief insight into the problem, scope and objectives of the proposed work. Chapter 2, the *Literature Review* assesses previous research that has contributed to transit data display and evaluation methods. Also discussed are the novelties of each work and some of their deficiencies and how this work aims to build on these. Chapter 3, *Research Approach* presents the methodology for transit data display and evaluation issues and goes on to discuss the implementation of the prototype decision support tool and the various technologies that have been involved in the making of the tool. Chapter 4, *Data Analysis* presents the data from the Winston-Salem Transit Authority and discusses the data collection, formatting and processing for productivity analysis. Chapter 5, *Conclusions and Recommendations*, presents a final summary of the investigation, its conclusions and recommendations

CHAPTER 2: LITERATURE REVIEW

This chapter presents the current situation of the paratransit industry with respect to different approaches to system performance analysis. The approaches to paratransit system performance analysis are compared according to their advantages and deficiencies. Some previous studies that dealt with different aspects of paratransit system performance and its measures are mentioned to describe the past trends in the field and how this research tackles some of these deficiencies.

2.1 Paratransit Performance Analysis

2.1.1 Paratransit Research

The purpose and therefore the measures, of a transit service can vary from community to community. The American Transit Services Council (ATSC) identified three possible purposes:

- "serving community values (providing transportation alternatives; reinforcing land use patterns)",
- "achieving specific mobility objectives (congestion mitigation, improvement of air quality, compliance with Federal standards)",
- "assuring equity of opportunity (equal access of jobs and services)".

Paratransit research spans many areas – advanced technologies, service and productivity evaluation; customer satisfaction; matching technologies to local problems; and assessing the impacts of system change. Relatively little research has focussed on efficient, real-time decision support methods. Hence, this literature review will examine such methods in paratransit and seek to obtain cross-disciplinary insight from fixed-rate transit, traffic control and a variety of other applications.

A study by Chira-Chavala *et al* (1997) describes the *Advanced Paratransit System: An Application of Digital Map, Automated Vehicle Scheduling and Vehicle Location Systems*, the purpose of which was to automate trip reservations, vehicle scheduling, monitoring service quality, and to perform accounting and reporting tasks. They applied technologies like automated trip scheduling system (ATSS), digital geographical database (DGD), and automated vehicle location (AVL) equipment. Similar to the present work, the primary objective was to evaluate the impacts of the advanced paratransit system on OUTREACH paratransit operation. A secondary objective was to assemble a knowledge base to serve as a bridge between an operational test and the understanding of its impacts and effectiveness for other locales. Some of the findings included estimated savings of \$1.53 per passenger trip, higher percent shared rides and other personnel and cost benefits.

Similarly, Porter *et al* (2000) describe an *Evaluation Plan for the Cape Cod Advanced Public Transportation System*. The Cape Cod Regional Transit Authority (CCRTA) Advanced Public Transportation System (APTS) project was an application of Intelligent Transportation Systems (ITS) to fixed-route and paratransit operations to improve intermodal transportation services. Some of the technologies included GPS based AVL, Mobile Data Terminals (MDT), design and installation of a state of the art Local Area Network (LAN) for the scheduling system and a prototype Internet site for display of vehicles location at real time. A significant objective of the Cape Cod APTS project sponsors is to demonstrate the viability of APTS technologies for rural transit operations. The evaluation goals and measures were safety measures like reduction in incident response time, average travel time; mobility measures like average travel time, efficiency measures like passenger trips per vehicle hour; and productivity measures like cost per vehicle-hour, cost per passenger-trip and number of trips shifted.

Spring *et al* (1996) came up with a framework that followed FTA's APTS Evaluation guidelines, for evaluating a small city's implementation of AVL for its

paratransit system. Using the Winston-Salem paratransit system and its Mobility Management Project as the case study, the key questions like the benefits of implementing AVL, effectiveness of using AVL in small to medium-sized cities and its associated problems were answered.

2.2 Statistical Applications for Paratransit Performance Analysis

2.2.1 Statistical Analysis of Data

With vast amounts of information available in the field of transit, it is very necessary to properly handle, understand and use the information effectively. Computers with their extensive data handling capabilities and speed of computational help in this regard.

Dealing with a problem with the help of statistical data analysis usually involves four basic steps – defining the problem, collecting the data, analyzing the data and reporting the results. All the above four steps have been discussed in the context of this research work in Chapter 4 titled Data Analysis.

2.2.2 Statistical Applications

A modest amount of work involving development of performance analysis tools for transit has been done. While much data is collected at every transit property, the tools necessary for effective use of the data are usually lacking. Lately, a lot of work has been done for developing powerful statistical analysis tools. Many researchers have been involved in building applications to perform statistical analysis using the Web as platform.

A study by Lave and Pozdena (1977) described the application of two multivariate statistical techniques (Farrell efficiency and multiple regression analysis) to financial and operating data from 47 California transit properties collected in 1976-77. This was done to investigate the relationship between variables characterizing transit external environment (geography, demography, organizational structure, economic and labor conditions, etc) on transit operating performance. However, lack of adequate data and small sample size caused the results to be disappointing.

Cryer *et al* (1985) did work on the use of Time Series models to evaluate service and fare policy changes by public transit operators, and used these models to forecast transit system ridership at the route, service, and system level. The work identified the relationships between transit ridership, level of service, travel costs, and market size.

On a commercial level, Statware has developed web based software which performs statistical analysis on data by integration of tools into web based applications for use in the biotech industry. With this advent of cutting edge technologies in other fields, there is the promise of improved computational efficiency and real time performance monitoring for transit.

2.3 Web-based Applications

2.3.1 The Web-based approach

The last decade has seen the explosive growth of Internet and it has become an integral part of our lives. Consisting of a global network of computers, the Internet has become especially appealing since the introduction of graphical browsers and increased speed of connection. The World Wide Web (WWW) is one of the most commonly used services of the Internet. Using Hyper Text Transfer Protocol (HTTP) and languages like HTML, web browsers and web servers communicate with each other on the Internet.

Web browsers have become a fairly common tool used by companies for their internal information network. Such use of the Internet is also known as Intranet. Some of the most important reasons for this are -

- It has easy access to multimedia as well as textual information using web browsers.
- Almost all the major operating platforms have standard browsers like Netscape and Internet Explorer and therefore there are no issues of compatibility and cross-platform client development.
- The web is an easy way to access and distribute information since infrastructure like web servers are easy to set up.

With the advent of the Internet, also came the programming language Java and Java-enabled browsers, which promise platform independence and portability. A quick comparison will show some of the advantages such systems have over traditional methods.

- Web-based applications can be used as long as a Web browser and network connection are available. In comparison, most old applications used to be operating system specific with Unix based applications running only on Unix and similarly for other systems like Windows.
- As pointed above, the development of graphical browsers has made development of user interfaces on the Web easy and maintainable. Data manipulation and computation can be performed both on the client and server side with the help of tools like Java, JavaScript and Servlets.
- Databases can be easily hooked to Web-based tools using standard drivers that are widely available and APIs like Java Database Connectivity (JDBC).

From these points, it is therefore evident that a Web-based application compatible on standard browsers is the best approach to developing a decision-making computing tool, for transit as well as other industries.

2.3.2 Web-based Transportation Applications

Byrdia, Turner, Eisele, and Liu (1998) dwelt on the ITS data that are often used for real-time operations and discarded, resulting in analysts and researchers struggling to obtain accurate, reliable data about existing transportation performance and patterns. They presented the development of an ITS data management system that would be used for storing, accessing, analyzing and presenting data from the TMC at Austin, Texas. The researchers came to the conclusion that was a great need for a better understanding of data and processes of “retaining, managing, sharing, and analyzing ITS data for planning and evaluation purposes.”

Bhadha (1998) focussed on the Intranet display of real time traffic information from speed sensors on freeways in Los Angeles and Orange Counties. Java applets were used to allow a continuously refreshed display of updated information. Similarly, Lee (1997) described how the WWW could be used to "provide users with easy access to traffic information." The system he proposed consisted of three subsystems: a client/server interface, a knowledge-based path search system, and a traffic data storage system. Java-based programs provided users with multimedia data accessibility and interactivity. The modules were interconnected to the WWW interface using Common Gateway Interface (CGI) scripting. The system could be extended to an integrated navigation system, which had traffic information in addition to a variety of information displays on the Internet.

On a much larger scale, Mtenga, et al (1999) developed a computerized information system for "management of field data on vehicle impacts with crash

attenuators.” A Java developed Web based interface, using an Oracle database was used to generate for life-cycle costs, attenuator usage and accident outcomes "under a variety of site or crash conditions". The DSS used the historical data to rank attenuators based on the past performances under specific site conditions.

Barrett, et al (1999) developed a website containing machine executable geographic information system called VISTA (Visual Interactive System for Transportation Algorithms). The VISTA framework combines several transportation tools, such that they share a common data specification and user interface, using a Common Object Request Brokerage Architecture (CORBA).

Chang, Chen, and Ran (1999) proposed a 3-tier architecture for implementing GIS database applications on the Internet using a Java frontend for platform independence and simplifying "the deployment issue."

Baugh, Stone, et al (2000) demonstrated the usefulness of Internet based tools for remote data collection, validation and evaluation for transit monitoring and management.

Most of the described works involve work in real time display of information with applications mostly in the field of traffic. Greene et al of *Paratransit Technology Group*, a paratransit solutions provider, have been involved in work geared towards development of customizable reporting systems in an environment using the web as a front end. Web based applications are available to any user having access to a Java-enabled browser and Internet connection. Thus we see that the web has been used as a medium to provide a powerful, flexible, conveniently accessible, multi-platform method of performing computation and a foundation for this research.

2.4 Decision Support

Decision Support can be best described as methods used in management decision making and the tools that go into helping make these decisions are called *Decision Support tools*. A decision support system uses a database, commonly referred to as the *decision support database*, from which data is extracted and analyzed statistically in order to inform business or other decisions.

2.4.1 Requirements of a Decision Support Tool

Development of a decision support tool which can perform the above mentioned tasks namely data display, query and manipulation, production of graphical output and statistical analysis of the queried data, involves the following properties:

- A real time system that is easy to use and maintain.
- A system that can be easily accessible by agencies independent of the computer platform they use.
- A system that is easily integrated with the database management system used at a Transportation Management Center or for that matter any database with similar relations.
- A system that has easy-to-use modules, graphical user interface (GUI) and navigation.

2.4.2 Decision Support Applications

Outside and in the field of transportation, the Internet is being increasingly used for dissemination, querying and manipulation of information. The development of the client-server technologies and user interface like the Hyper Text Markup Language

(HTML) have provided a powerful mechanism for flow of information and therefore several kinds of decision making environments (Mukherjee et al 2000).

Recently, the Internet has been used for supporting real-time interactive decision support applications in many fields including finance where online portfolios of companies and real-time stock prices and market trends have become very common. Often the new type of system called *On-Line Analytical Processing* or OLAP overrides the term decision support. It encompasses a class of technologies that are designed for live ad hoc data access and analysis (*White Paper*, Accrue Software). Courcoubetis & Siris (2000), developed an interface to provide an environment for modifying, saving and executing experiments using a Java enabled web browser. Microsoft Corporations *Digital Dashboard* builds on the concept personalized decision support by displaying a folder *home page* to combine personal, company and external data into a single display. This display might appear in a browser and is especially helpful in consolidating and focussing on critical data with the appropriate software for data display and analysis tools.

The discussion included transportation and non-transportation related works that used Internet and other state of the art technologies. There has, however, been a notable lack of Internet based decision support tools.

2.5 Summary

2.5.1 Summary of Benefits

In summary, although significant work has been done on system productivity analysis, there has been a lack of real time interactive tools for statistical analysis for use with transit and paratransit data. The table below (Figure 2.1) shows a comparison of three works referenced in Section 2.1.1. The comparison illustrates several of the

relatively few earlier works involving impact studies of advanced technologies on paratransit systems lacked real-time display of paratransit data and its analysis.

Figure 2.1: Comparison of features of referenced studies

Work	Use of Advanced Technologies	Evaluation of impact of advanced technologies	Use of Real Time tools
Caltrans' OUTREACH	Yes. Digital geographic database, AVL.	Yes.	No.
Cape Cod	Yes. GPS/AVL, MDT, LAN.	Yes.	A prototype Internet site that displays the location of vehicles.
WSTA - TransAID	Yes. GPS/AVL, MDT, LAN.	Yes.	A prototype Internet site that displays real time system data that is used for analysis.

With the Web being a convenient and increasingly used medium, it serves the purpose of a platform for building such tools.

Figure 2. 2: Table of the Reporting Techniques

Method	Collection	Validation	Tabular displays	Graphical displays	Real Time
Printed reports	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Offline spreadsheets/ databases	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Reporting tools (eg Crystal Reports)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Network based tools (Web based/Applications)	<input checked="" type="checkbox"/>				

As can be seen in the summary table above (Figure 2.2) a comparison of typical reporting techniques shows that most of the traditional techniques like printed reports and spreadsheets/ databases lack the ability to perform real time reporting. This is remedied by connecting the application to the database with the help of a network. This is where

networked or web based applications come into the picture. This is what the thesis aims to accomplish.

The following chapters describe the Winston-Salem paratransit data and the development and implementation of web-based technologies for decision support of this data.

CHAPTER 3: INTERNET DECISION SUPPORT TOOL

This chapter describes the requirements for developing a decision support tool and its architecture. Taking into consideration the WSTA case, the decision support system configuration and various data needs are spelt out and the decision flow determined. Lastly several issues, which are of paramount importance in development of a software decision support tool, are discussed.

3.1 Introduction

3.1.1 Performance Analysis

System performance and its analysis entail identifying certain measures, which effectively portray the service effectiveness and system productivity. Some typical productivity measures include passenger-trips/day, vehicle-miles/day, vehicle-hours/day, passengers/vehicle hour, vehicle-miles/passenger trip and vehicle-hours/passenger trip.

As discussed previously, these productivity measures, which are usually in the form of monthly reports, can be then displayed graphically in the form of tables, histograms, pie charts etc. along with the basic statistical measures like mean, median and variance. These basic statistical values give an intuitive idea about the performance of the system and help the operational staff and transit managers to understand the trends and the changes that occur in these productivity measures. The intuitive reasoning might, however, be incorrect and stronger statistical tests could help interpret the performance trends.

Such tests can be applied at regular intervals of time to gauge changes in trends that can occur due to several reasons like changes in the technology, fares etc. assuming the operational methods remain same (Stone, 2000).

3.1.2 Decision Support

The main goal of a public transportation agency is to serve the public and to promote community mobility. Improvements in public transportation should open opportunities for people, which in turn enable better economic activity. So monitoring the system performance and service effectiveness to achieve these goals is important for increasing the overall net benefits of the transportation system.

A decision support system can be generically described as a combination of tools that can be used to access data and analyze them. Thus, decision support systems give the users the ability to access and manipulate data and avoid lengthy procedures for report generation. The present scenario as illustrated in the literature survey, shows that the transit performance measures and their assessment has suffered from the absence of real time data at the transit managers convenience. Further, statistical analysis is used to a minimal extent in determining the changes occurring in these measures.

3.2 Architecture of the Software tool

The proposed decision support system, which incorporates decision-making ability with the help of computing tools and statistical analysis, has to follow an architecture that allows for incorporation of several modules, each of which performs a set of defined tasks.

3.2.1 System Architecture

Some of the commonly used architectural frameworks like the 2-tier, 3-tier forms have been discussed in detail in the *Glossary of Terms* (see Appendix A).

For the purposes of the tool, a 3-tier architecture is preferred. Some of the main benefits of 3-tier (or an n-tier) architecture, in comparison with a older architecture techniques are:

- Better flexibility which is possible by distributing the load between several servers, and the ability to settle the application layer in different locations;
- Better fault recovery mechanisms possible due to the middle-tier layer distributed on several servers,
- Easy system administration, since the applications can be centrally managed on the server side whereas in 2-tier architecture the application layer could be physically distributed on several client computers,
- Better modularity ensures software reuse and integration of several including older applications.

In recent times for the 3-tier architecture² the middle tier is sometimes replaced with a “Web” layer. In such a case, the scenario becomes as shown below.



3.2.2 The Intended Architecture

The three components of the 3-tier architecture are *presentation*, *business logic* or *functionality*, and *data*.

- The Presentation tier or the **Graphical User Interface** layer - interfaces with the user and consists of hardware such as a PC or workstation. The intended

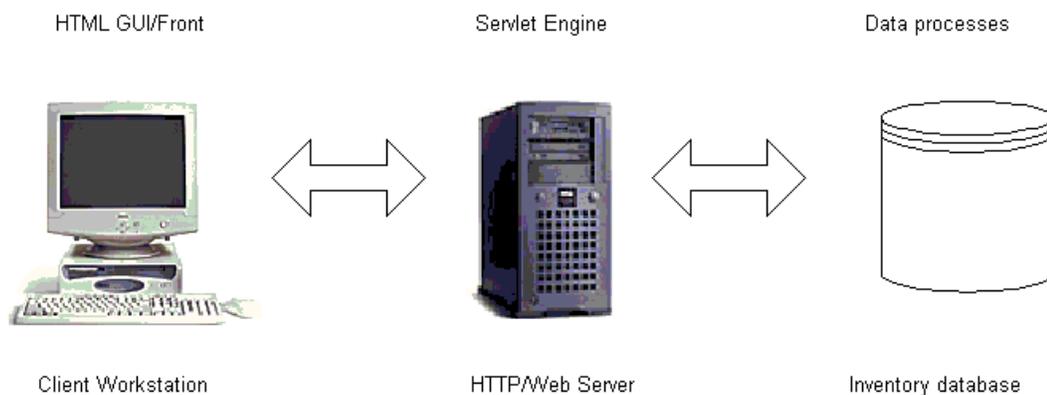
² See *Glossary of Terms* for detailed definition.

technique in this research is HTML or HyperText Markup Language, which provides an easy and robust way to build seamless user interfaces.

- The Functionality tier or the **Object** layer - provides functionality, connectivity, and database servers. In other words it serves as a bridge between the first and third tiers. This work will mostly utilize Java Servlets for this middleware and Java Database Connectivity for establishing connection with the database using a JDBC/ODBC bridge.
- The Data tier or the **Database** layer – for purposes of the prototype being developed an MS Access database will be used, although the original intention was to connect to the commercial scheduling software database which resides on WSTA’s servers. A relational database will allow for access and processing through queries.

The figure below illustrates the inter-connectivity of the three tiers and how they will serve the application.

Figure 3. 1: 3-tier Architecture Schema



3.3 The Winston-Salem Data

3.3.1 Winston-Salem Transit Authority (WSTA)

Winston-Salem Transit Authority is part of the Winston-Salem Department of Transportation and is the provider of coordinated public transportation for the area. Consisting of a fleet of 19 vans³, WSTA services include fixed-route bus service, demand-responsive paratransit (TransAid), downtown circulators, fringe parking with downtown shuttles, park-and-ride lots, carpool matching, and vanpools.

In 1993, WSTA initiated the Mobility Manager National Demonstration Project. The goal of the project is “to provide a single source of information about local travel and mobility options while coordinating operations, trip reservations, and financial transactions.” In the Phase I (1993-1998) of the project, the utility of several ITS technologies were demonstrated including automatic scheduling and dispatch, smart cards, automatic vehicle location, and mobile digital communications for paratransit, for the paratransit system named TransAid. In August 1994, WSTA started using the Paratransit Automated Scheduling System (PASS), which is designed to handle all the functions related to paratransit services. Research efforts have focussed on impacts of APTS technology system productivity and secondarily on customer service.

Phase II of the Mobility Manager project (1999- 2002) has the goal of establishing state of the art traveler information systems. The first installed information system was automated telephone technology so passenger can book and cancel trips, determine the estimated time of arrival of a bus, and check on WSTA protocol without the help of operators. Next, automated signs for bus departure times were installed in the transit terminal and linked to master displays. Underlying the systemwide traveler

³ As told by the WSTA supervisor in late October, 2000.

information systems are fully integrated automatic vehicle locators on all the WSTA vehicles, mobile data communications and computer-aided dispatch and scheduling. All vehicle and passenger information resides on a central server. Besides improving customer satisfaction with better traveler information, WSTA also intends to improve transit system productivity in Phase II.

3.3.2 Advanced Technologies

The aim of the second phase is to provide “real-time” paratransit and fixed-route information to customers by interfacing several ITS components with the existing WSTA database and buses. The technologies that are being used at WSTA are interactive voice response (IVR), automatic vehicle location (AVL), and mobile data terminal (MDT). In addition, WSTA management will install dynamic variable message signs (VMS) for bus bays at the transit center, and a dynamic master arrival and departure board at the transit center.

3.3.3 System Configuration of WSTA Fixed Route & Paratransit Service

The paratransit fleet consists of about twenty vehicles of varying capacity. The dispatch facility resides in the North Trade St office in downtown Winston-Salem and works in tandem with the transit center where the IVR server resides and all the calls are taken. Real-time bus locations are tracked, traveler information provided and trip reservation and cancellations made. By August 2001, the MDT/AVL installation was underway and the commercial scheduling server not completely configured for the additional data.

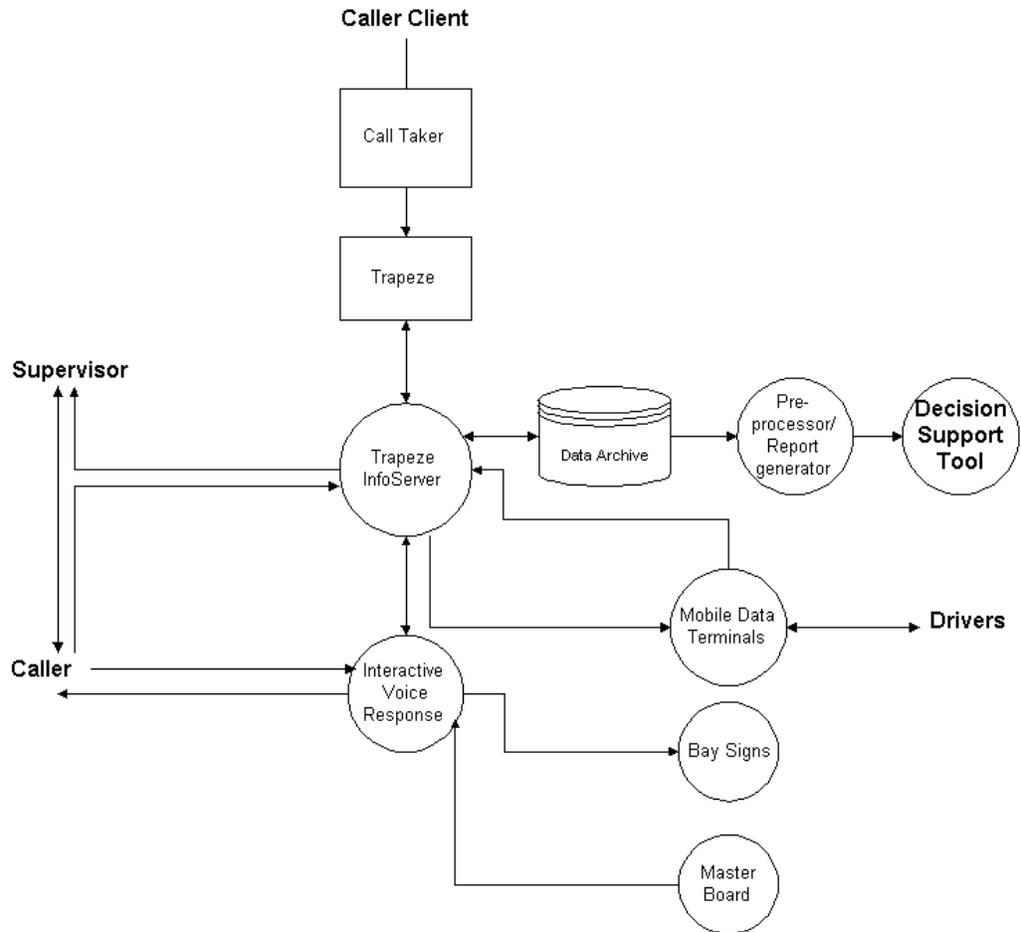
Figure 3.2 shows the overall system configuration, of the WSTA fixed route and paratransit systems and the peripheral devices.

3.3.4 Data Flow

As illustrated in Figure 3.2, the WSTA TransAid system has a variety of data sources, which feeds into the commercial scheduling software Infoserver installed at the paratransit Dispatch Facility at Winston-Salem. The commercial scheduling software Infoserver is a XML/Web-based *application server*⁴ running on Microsoft Transaction Server (MTS). All the AVL/MDT data are stored at the back-end database, which in the WSTA case is Microsoft SQL Server 7. All the data is stored in the MMS Archive and can be accessed through the same interface on the MTS server. This flow of data can be visualized as shown in Figure 3.3 below.

⁴ See *Appendix A: Glossary of Terms* for the description.

Figure 3. 2: Data Flow and Integration Diagram



The data to be used for the purpose of building the decision support tool have been taken from the *Trips Miles Productivity* reports³ (Figure 3.3) and the *Provider Productivity* reports⁵ (Figure 3.4). These reports came from the commercial scheduling software MMS-Archive for evaluation purposes. These reports are generated by a pre-processor, which extracts and compiles the data from the raw ride files (Figure 3.5) generated by the scheduling algorithm. They have the system measures of effectiveness data like miles, trips, cancellations and no shows in a spreadsheet form. The actual measures that have been used are *Trips per Day*, *Passenger per Day*, *Cancellations* and *No Shows* calculated from these basic measures.

⁵ These reports are described in greater detail in Chapter 4: *Data Analysis*

Figure 3. 3: Sample Trips/Miles Productivity Report

Trips/Miles Productivity Report

Page No. 2
 Yr: 01/19/00
 Time: 10:47:26

From: 11/01/99 to: 11/30/99

Rte #	Date of	DCW	Trips	Pass.	Pass/Trip	Total Miles	Pass Miles	Service Miles	Calc. DeadHead	Revenue Miles	Actual DeadHead	Pass/ Srvc. Mls	Pass/ Rev. Mils	Pass/ Tot. Mils	Pass Ave Trp Length
11/15/99	MON	195	397	1.0	1735.6	4847.2	1649.2	86.4	1248.4	487.2	0.24	0.32	0.23	12.2	
11/16/99	TUE	389	389	1.0	1612.0	6089.9	1562.6	49.4	1259.1	352.9	0.25	0.31	0.24	15.7	
11/17/99	WED	381	381	1.0	1674.1	5534.1	1583.5	90.5	1272.5	401.6	0.24	0.30	0.23	14.5	
11/18/99	THU	380	382	1.0	1555.7	5159.2	1485.1	70.6	1181.7	374.0	0.26	0.32	0.25	13.5	
11/19/99	FRI	384	384	1.0	1580.7	4992.5	1507.4	73.4	1215.3	365.4	0.25	0.32	0.24	13.0	
11/20/99	SAT	98	98	1.0	277.7	881.1	229.7	48.0	219.8	57.9	0.43	0.45	0.35	9.0	
11/22/99	MON	407	407	1.0	1807.5	4776.5	1731.6	75.9	1314.3	493.3	0.24	0.31	0.23	11.7	
11/23/99	TUE	390	390	1.0	1674.6	5661.9	1610.7	63.9	1205.5	468.2	0.24	0.32	0.23	14.5	
11/24/99	WED	390	390	1.0	1841.1	5204.7	1642.0	199.1	1268.4	572.7	0.24	0.31	0.21	13.3	
11/25/99	THU	99	99	1.0	299.8	1442.4	261.0	38.8	243.8	55.9	0.38	0.41	0.33	14.6	
11/26/99	FRI	141	141	1.0	570.2	1371.6	514.0	56.3	423.4	146.8	0.27	0.33	0.25	9.7	
11/27/99	SAT	101	101	1.0	309.4	1280.7	266.8	40.6	254.3	55.1	0.38	0.40	0.33	12.7	

Figure 3. 4: Sample Provider Productivity Report

Provider Productivity Report																									
For Date : 2/1/00 To 2/29/00																									
Print Date: 10/16/2000																									
Print Time: 12:29:32PM																									
Date	Provider	Count Passengers					Count Of Trips			Hours								Milenge						Trip/RH	Trip/SH
		CLI	PCA	ESC	Others	Total	Trips	NS	Cancel	Total	Service	Rev.	DH	Break	Lunch	Refuel	Others	Total	Service	Rev.	DH	Bls.	Pass.		
2/7/00								0	34																
	CAT	37	0	0	3	40	37	0	7	38.5	31.1	7.4	23.7	0.0	0.0	0.0	0.0	281.1	228.7	138.5	90.2	0.0	148.8	5.03	1.19
	TRA	393	0	0	0	393	393	2	15	118.8	110.5	82.2	28.3	0.0	0.0	0.0	0.0	1746.3	1561.2	1256.0	305.2	0.0	5067.1	4.78	3.56
		430	0	0	3	433	430	2	56	157.3	141.6	89.6	52.0	0.0	0.0	0.0	0.0	2027.4	1789.9	1394.5	395.4	0.0	5215.9	9.81	4.75
2/8/00								0	40																
	CAT	41	0	0	4	45	41	0	7	47.1	45.1	9.7	35.4	0.0	0.0	0.0	0.0	383.3	331.2	199.6	131.7	0.0	227.2	4.21	0.91
	TRA	381	0	0	0	381	381	6	13	118.3	111.7	78.4	33.3	0.0	0.0	0.0	0.0	1825.5	1670.8	1219.9	450.9	0.0	5460.7	4.86	3.41
		422	0	0	4	426	422	6	60	165.4	156.8	88.1	68.7	0.0	0.0	0.0	0.0	2208.8	2002.0	1419.5	582.6	0.0	5687.9	9.07	4.32
2/9/00								0	42																
	CAT	42	0	0	3	45	42	0	6	46.6	39.6	9.5	30.1	0.0	0.0	0.0	0.0	356.4	301.3	183.8	117.5	0.0	213.7	4.43	1.06
	TRA	392	0	0	0	392	392	0	17	108.7	102.1	79.1	23.0	0.0	0.0	0.0	0.0	1732.8	1555.3	1206.8	348.5	0.0	5547.5	4.96	3.84
		434	0	0	3	437	434	0	65	155.3	141.7	88.6	53.1	0.0	0.0	0.0	0.0	2089.2	1856.6	1390.6	466.0	0.0	5761.2	9.39	4.90
2/10/00								0	50																
	CAT	34	0	0	2	36	34	0	11	44.6	42.8	7.9	34.9	0.0	0.0	0.0	20.2	335.7	288.5	159.1	129.4	28.3	179.4	4.30	0.80
	TRA	378	0	2	2	382	378	4	16	117.2	110.7	80.3	30.4	0.0	0.0	0.0	0.0	1747.6	1546.9	1204.3	342.6	0.0	5941.7	4.71	3.41
		412	0	2	4	418	412	4	77	161.8	153.5	88.2	65.2	0.0	0.0	0.0	20.2	2083.3	1835.4	1363.4	472.0	28.3	6121.1	9.01	4.21
2/11/00								0	52																
	CAT	32	0	0	3	35	32	0	16	41.7	40.0	7.5	32.5	0.0	0.0	0.0	0.0	291.9	249.6	137.4	112.2	0.0	150.2	4.28	0.80
	TRA	351	0	2	0	353	351	2	13	108.2	100.0	77.4	22.7	0.0	0.0	0.0	0.0	1617.8	1416.0	1113.4	302.7	0.0	5119.7	4.54	3.51
		383	0	2	3	388	383	2	81	149.9	140.0	84.9	55.2	0.0	0.0	0.0	0.0	1909.7	1665.6	1250.8	414.9	0.0	5269.9	8.82	4.31
2/12/00								0	11																
	CAT	16	0	0	0	16	16	0	17	15.9	15.2	2.6	12.6	0.0	0.0	0.0	0.0	102.4	85.6	50.4	35.2	0.0	50.4	6.15	1.05
	TRA	95	0	0	0	95	95	0	2	21.9	19.8	18.2	1.6	0.0	0.0	0.0	0.0	306.6	250.4	250.4	0.0	0.0	1137.6	5.22	4.80
		111	0	0	0	111	111	0	30	37.8	35.0	20.8	14.2	0.0	0.0	0.0	0.0	409.0	336.0	300.8	35.2	0.0	1188.0	11.37	5.85

Figure 3. 5: Sample Section of a raw Ride File

BUS	LAB	RIDERS	EQUIP	ID	CALL	ETA	ETACAL	APPTIME	PC	ADDRESS	APARTMENT	CITY	BUSCOLOR	DATESTAMP	ADDOACTIM	COLLECT	XCOORD	YCOORD	ROUTER	VEHICLE	DRIVER	COMMENTS	DI	CAPCALC
100	10001900	1	AM	1000		11:40	11:40	12:00	*	1628 N CHESTNUT ST		WINSTON-SALEM	10	100			80243781	36115436						5
100	10002100	1	AM	1000		11:58	12:00		-	0 UNIVERSITY PKWY		WINSTON-SALEM	25	100			80277343	36171259						5
100	10003900	1	AM	1001		20:30	20:30		+	403 S HAWTHORNE RD		WINSTON-SALEM	10	100			80269921	36089330				MAIN TOWER		5
100	10003800	1	AM	1001		20:39			-	1000 E 4TH ST		WINSTON-SALEM	25	100			80233825	36099538				WONDERLAND ACADEMY		5
100	10003700	1	AM	1002		20:45	20:39		+	1000 E 4TH ST		WINSTON-SALEM	10	100			80233825	36099538				WONDERLAND ACADEMY		5
100	10003900	1	AM	1002		20:50			-	13 SALEM GARDEN RD	E	WINSTON-SALEM	25	100			80195743	36081376						0
100	10002800	1	AM	1003		16:55	16:42	17:30	+	184 STANLEY AVE		WINSTON-SALEM	10	100			80199975	36127374				SCHEDULE VARIES 9PM-4AM		0
100	10002900	1	AM	1003		17:05	17:30		-	113 CASSELL ST		WINSTON-SALEM	25	100			80258540	36055586						0
100	10003000	1	AM	1004		17:00	17:30		+	1900 REYNOLDA RD		WINSTON-SALEM	10	100			80279899	36119289				GRAYLYN CONF CEN		0
100	10003100	1	AM	1004		17:29			-	1121 E 24TH ST		WINSTON-SALEM	25	100			80232777	36121049						0
100	10001500	1	AM	1005		09:30	09:29	09:45	+	727 VOSS ST		WINSTON-SALEM	10	100			80232231	36161730				ONLY ON SUNDAY		2
100	10001700	1	AM	1005		09:45	09:45		-	200 Brookdown Ave		Winston Salem	25	100			80259195	36097887						0
100	10003400	1	AM	1006		18:15	18:31		+	200 Brookdown Ave		Winston Salem	10	100			80259195	36097887						0
100	10003500	1	AM	1006		18:48			-	727 VOSS ST		WINSTON-SALEM	25	100			80232231	36161730				ONLY ON SUNDAY		0
100	10004900	1	AM	1007		23:45	23:45		+	425 N Cherry St		Winston Salem	10	100			80247126	36098471						0
100	10004700	1	AM	1007		23:52			-	519 LIBERIA ST		WINSTON-SALEM	25	100			80255503	36083876						0
100	10004800	1	AM	1008		24:15	24:15		+	519 LIBERIA ST		WINSTON-SALEM	10	100			80255503	36083876						0
100	10004900	1	AM	1008		24:25			-	1109 CHARLES ST		WINSTON-SALEM	25	100			80213373	36055739						0
100	10002900	1	AM	1009		15:00	15:00		+	7910 N Point Blvd		Winston Salem	10	100			80279897	36193827						0
100	10002700	1	AM	1009		15:14			-	2615 PENCLETON DR	APTS	WINSTON-SALEM	25	100			80232375	36124536						0
100	10004000	1	AM	1010		22:30	22:30	23:00	+	708 POWELL ST		WINSTON-SALEM	10	100			80234754	36080242						0
100	10004500	1	AM	1010		22:42	22:42	23:00	-	3333 SILAS CREEK PKWY		WINSTON-SALEM	25	100			80289738	36079541						0
100	10003200	1	AM	1011		18:00	18:00		+	3964 UNIVERSITY PKWY		WINSTON-SALEM	10	100			80274731	36172068						0
100	10003300	1	AM	1011		18:21			-	595 BARRY CT	B	WINSTON-SALEM	25	100			80191990	36193446						0
100	10004000	1	AM	1012		21:05	21:04	21:20	+	0 UNIVERSITY PKWY		WINSTON-SALEM	10	100			80277343	36171259						0
100	10004100	1	AM	1012		21:20	21:20	21:20	-	1628 N CHESTNUT ST		WINSTON-SALEM	25	100			80243781	36115436						0
100	10004200	1	AM	1013		22:20	22:00	11:00	+	582 FREEDOM CT	A	WINSTON-SALEM	10	100			80224962	36052625				10:50PM (3RD SHFT) (AM PICKUPS ONLY ON WEEKENDS)		0
100	10004300	1	AM	1013		22:26	22:26	11:00	-	901 BETHESDA RD		WINSTON-SALEM	25	100			80284308	36079589						0
100	10001800	1	AM	1014		11:30	11:30		+	13 SALEM GARDEN RD	E	WINSTON-SALEM	10	100			80195743	36081376						0
100	10002000	1	AM	1014		11:42			-	1000 E 4TH ST		WINSTON-SALEM	25	100			80233825	36099538				WONDERLAND ACADEMY		0
100	10002100	1	AM	1015		11:45	11:40	12:00	+	1000 E 4TH ST		WINSTON-SALEM	10	100			80233825	36099538				WONDERLAND ACADEMY		0
100	10002200	1	AM	1015		11:54	12:00		-	403 S HAWTHORNE RD		WINSTON-SALEM	25	100			80269921	36089330				MAIN TOWER		0
100	10001900	1	AM	1016		08:05	08:05		+	1121 E 24TH ST		WINSTON-SALEM	10	100			80232777	36121049						0
100	10001100	1	AM	1016		08:14			-	1900 REYNOLDA RD		WINSTON-SALEM	25	100			80279899	36119289				GRAYLYN CONF CEN		0
100	10001200	1	AM	1017		08:40	08:40	15:00	+	2615 PENCLETON DR	APTS	WINSTON-SALEM	10	100			80222375	36124536						0
100	10001300	1	AM	1017		08:54	15:00		-	7910 N Point Blvd		Winston Salem	25	100			80279897	36193827						0
100	10001400	1	AM	1018		09:00	09:00		+	595 BARRY CT	B	WINSTON-SALEM	10	100			80191990	36193446						0
100	10001500	1	AM	1018		09:30			-	3964 UNIVERSITY PKWY		WINSTON-SALEM	25	100			80274731	36172068						0
100	10002800	1	AM	1019		06:26	06:26		+	2815 FREDMONT CIR		WINSTON-SALEM	10	100			80229031	36123967						0
100	10002900	1	AM	1019		06:33			-	809 WELLINGTON RD		WINSTON-SALEM	25	100			80287376	36112741						0
100	10002400	1	AM	1020		13:30	13:10	13:20	+	809 WELLINGTON RD		WINSTON-SALEM	10	100			80287376	36112741						0
100	10002500	1	AM	1020		13:23	13:20		-	2815 FREDMONT CIR		WINSTON-SALEM	25	100			80229031	36123967						0
0	0	1	AM	1021		04:00	04:00	05:00	+	113 CASSELL ST		WINSTON-SALEM	25	100			80258540	36055586						C
0	0	1	AM	1021		04:00	05:00		-	184 STANLEY AVE		WINSTON-SALEM	10	100			80199975	36127374				SCHEDULE VARIES 9PM-4AM		C
0	0	1	AM	1022		07:00	07:00	07:30	+	901 BETHESDA RD		WINSTON-SALEM	10	100			80284308	36079589						C

In the following chapter titled *Data Analysis*, the data collected from the above mentioned reports (Figure 3.4, 3.5) will be examined in greater detail in accordance with the first task mentioned in the *Scope and Objectives* sub-section of the Chapter 1. This will include collection of the data, formatting it for the final usage in the database and offline analysis for determining the initial trends.

3.4 Decision Support

Designing a decision support system for a particular transportation system involves several steps including identification of goals, determination of the state of research in advanced transportation systems and assessment of the strengths and weaknesses of the tools being developed.

In the case of this work, the main focus is monitoring paratransit system performance and determining the deficiencies and therefore zeroing in on the decisions that need to be taken for the overall benefit of the paratransit system and its users.

3.4.1 Decision Flow

The transit manager has a variety of issues to take care of. The overall architecture of the system is integrated to include the IVR and AVL/MDT data coming in from the buses and dispatch facility into the database, which in turn performs scheduling on the unscheduled data.

The way the decision support tool has been designed in such a way that the transit facility can view the Measures of Effectiveness (MOE) that have been generated from the data stored in the database on which scheduling has been performed. In this prototype

version only data from September 1999 to May 2001⁶ can be processed. The prototype is also capable displaying and analyzing weekly and monthly data. The program can, however, be modified for including daily and yearly data.

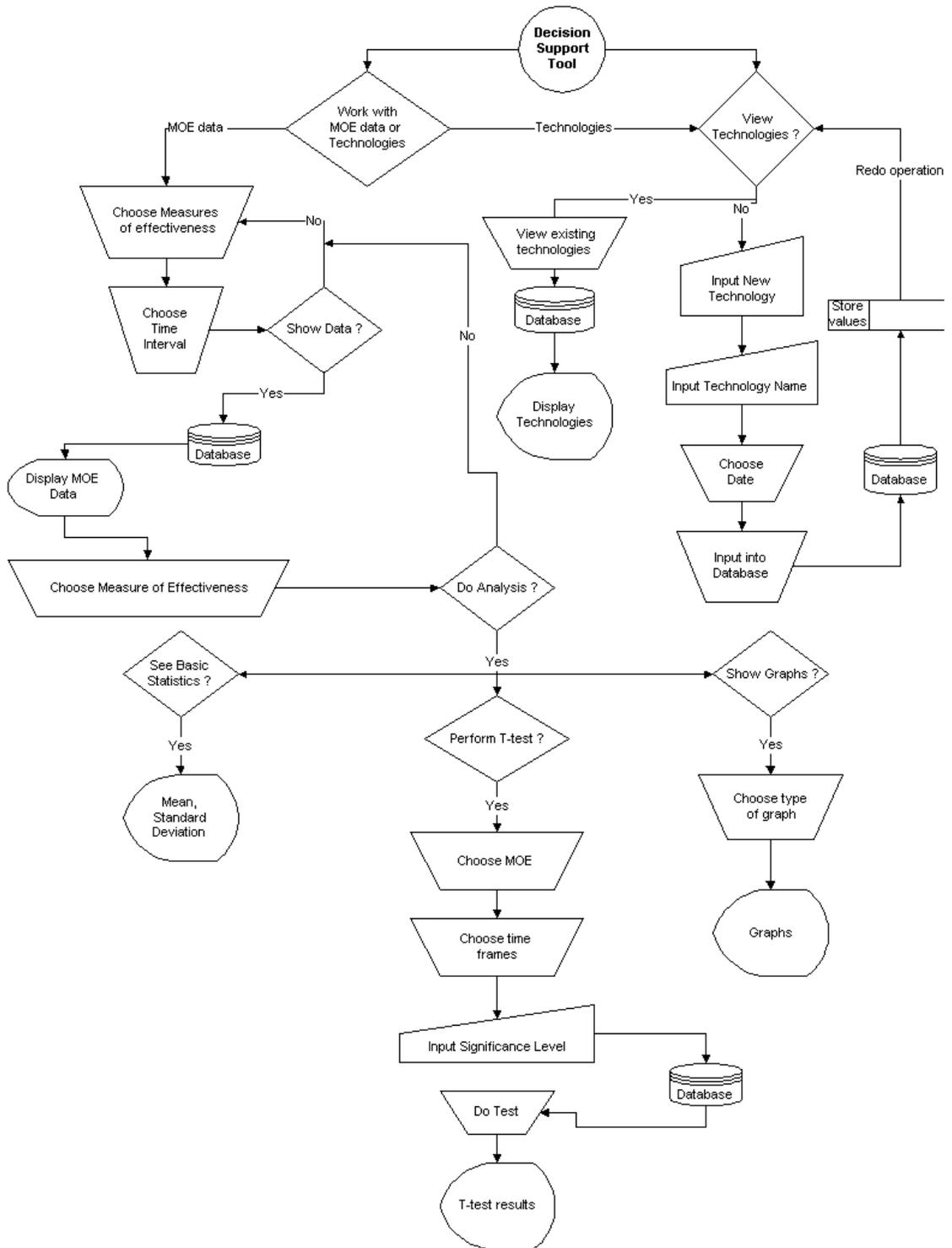
Using the tool the transit manager can view any measure of effectiveness, for any length of time, to assess how the paratransit system is performing. This can be done either by a visual assessment of the MOE data trends or by an in-depth statistical assessment of the data. The user can go to the various modules, which will allow him/her to choose a measure of effectiveness and display them in tabular or graphical formats, calculate averages, and perform various statistical tests.

3.4.2 Decision Flow Diagram

Understanding the flow of decision information is important for conceptualizing a decision support tool. With this in mind a flow of data and manual operations needed for performing the tasks needed for proper functioning of the tool was envisaged. The diagram (Figure 3.7) is a logical follow up of Figure 3.3, which shows the data flow and integration. It aims to capture all the different modules that will be needed for functioning of the tool and integrates them to ensure flow of data necessary for the processing and analysis.

⁶ Date earlier than Sep 1999 could not be incorporated into the database because of incompatible format of the earlier DOS based PASS reports. Research data collection stopped in May 2001.

Figure 3. 6: Decision Flow Diagram



3.5 Development and Implementation Issues

Software is a constantly evolving and development of any software tool involves choosing the best current technologies and overcoming various issues relating to performance and reliability of these technologies for development purposes. Some of these important issues are discussed below.

3.5.1 Cost Issues in Development and Implementation

Cost is a very important factor that has to be considered at every stage of a software development cycle and the prime cost input is mainly the programming time.

Java makes use of software reuse with the help of *classes* and *objects*. This facilitates modular programs, which are efficient to build, compile and run. A whole new generation of Java-based development and database connection environments has made the development of web and database related development relatively easy and efficient.

Another cost issue that needs to be considered is the actual real world implementation of the prototype. The client or the involved agency (in this case, the Winston-Salem Transit Authority) will have little or no additional investment in implementing a Java-based full-scale version of the developed prototype. Most of the programs that have been built will be a plug and play fit into a standard database. Little modifications will be required depending on the data definition language used in the database. However, some capital investment will be required in obtaining a Web server that will be running at the Transit Management Center. With respect to WSTA and similar agencies, the commercial scheduling software server running at the transit facility will be able to handle this additional load.

The use and maintenance of a Web-based decision support system does not entail a major expenditure for training of staff. Besides, it essentially means working on the same computer systems and therefore saves money on acquiring new software and hardware.

3.5.2 Speed of Operation and Computation

Java is essentially interpretive in nature and therefore cannot match the speed and performance of the native C compiler code. However, the recent introduction of Just-In-Time (JIT) compilers by Sun has increased the speed of Java executables by a huge margin.

Some of the graphical displays developed in the prototype take some time to load before they can be visible to the user. Use of a faster web server and more computation power will resolve this issue.

3.5.3 Security

Implementation of a critical application which involves working with proprietary technologies and databases which demand secure transfer of information to and from the user location. Intrusion of remote systems by hackers and unauthorized users is a major concern for many commercial organizations.

There have been quite a few developments in the area of Internet security, such as the development of secure, stronger encryption techniques, secure servers and better data communication protocols. In addition, this prototype has been developed with the module to provide the facility that it will be accessible only to concerned members of the

particular agency with the help of valid login and passwords. The availability of a network connection and a web browser will enable the user access to the site.

3.5.4 Obsolescence of Technology

Technological obsolescence is of high concern in the technology industry especially since there have been rapid changes in the web-based technologies. This increases the risk of a given technology becoming obsolete.

The Internet as has been described in the previous chapter has shown a remarkable increase in popularity over the last decade. Along with it has come a host of technologies for web-based development. Java is one such technology and over the past few years has shown tremendous promise as a robust platform for Internet related development.

Most of the Java related technologies have been embraced and continue to be, by institutions and organizations of repute. As of now therefore, Java seems to have established a firm grip on the area of Internet computing and will continue to be accepted as a standard in the area of Internet related development in the coming years.

3.6 Development of the Tool

The main attempt of this research work is performance evaluation for a paratransit system with respect to its effectiveness and productivity.

The approach followed in this work is relatively new to the area of transit. As mentioned in Chapter 2, there have been several non-transit applications, which dealt with statistical analysis of data using the Internet as a medium. Significant work has also

been done in the area of transit with respect to analysis of performance data. Relatively less research has been done in the area of real time performance analysis and validation of transit performance data. There have also been a few instances of works related to real time display of transit and paratransit data (Bhadha, Barrett et al) as mentioned in section 2.3.2. However the combined applications of validating and statistical processing of performance data for further analysis with Web-based tools, has not occurred before this research.

This work follows the approach of extracting measures of effectiveness data from the database with the help of SQL querying and displaying it to the user using Java Servlets. The user can then manipulate the data for further graphical displays and statistical analysis.

The decision support tool has been envisioned as a general-purpose tool to be used by the management of a transit agency for gauging the overall performance of the paratransit system. This calls for development of a tool, which can seamlessly blend into the existing architecture of the paratransit system and enable the operational staff to perform the needed functions.

In such a situation, the ease of use and availability of a web browser makes it the natural choice for designing the front-end user interface. Java is an object-oriented language that was specifically developed by Sun Microsystems as an ideal language for development of secure, distributed, network-based end-user applications in environments as diverse as networked embedded devices to the World-Wide Web and the desktop. It was therefore chosen as the implementation language.

3.6.1 Database Module

It is also important that the data that is used can be easily accessed and processed. The database needs as described earlier, especially that of modularity, calls for use of a relational database for storage of data. This can facilitate easy querying and updating of the data. The tool employs MS Access as the relational database program. Access⁷ is a relational database running under Windows NT.

In relational database systems such as Access, Oracle and Sybase data is stored in tables (or *relations*) made up of one or several columns (also referred to as *fields* or *attributes*). The data stored in each column is of a particular data type like character, text or number. A particular collection of values of all the fields is usually referred to as a *record*, *row* or *tuple*. See Appendix B for a screenshot of the database table containing MOE data that was implemented.

Database systems are accessed from different interfaces using a standard called Open Database Connectivity (ODBC), which can be used to submit statements in Structured Query Language (SQL) to the database.

SQL is a standard language for creating, updating, querying and retrieving data from a database. Most of the SQL commands are standard according to the American National Standard Institute (ANSI) but common database systems have their own proprietary extensions to this language. The SQL commands that have been used to accomplish data retrieval and querying include the “Select”, “Insert” and “Update” statements. See Appendix B for example SQL query commands that have been used.

⁷ **MS Access** is usually used for prototyping and importing/exporting data with, but its performance falls when attempted in a multi-user scenario. It cannot match the performance, recovery and online backup capabilities of a client/server database like Oracle or Sybase, but it can provide a robust solution with good performance for limited applications and small number of users. Hence its use in this prototype.

Java Database Connectivity (JDBC), which is a part of the Java Development Kit as an Application Programming Interface (API), is used in this tool for creating connections and access to the database. The driver used is the JDBC/ODBC Bridge for MS Access and is available with the Windows NT operating system. Concurrency issues do not arise because the users do not update the database at any point of time.

3.6.2 Front-End Module

The user interfaces (as shown in Figure 3.8, 3.10, 3.11) are designed using HTML and are connected to the database using JDBC connection and different transactions are performed using Java Servlets, which make possible the various requests and their responses from the users and back.

3.6.3 Middleware Application

The middleware application has been designed using Java applications containing servlets and JDBC for processing the transactions. The queries written in SQL used for these transactions are embedded in the Servlets classes.

The Java developed Servlet technology is used to develop modules, which extend servers capable of receiving requests and sending responses, like web servers. For example, Servlets can take data from an HTML form and use it according to the business logic to get data or update the database.

In order to process these requests and responses coming in from the user forms, the server of choice would be one which does this processing quickly and reliably. Java Server Web Development Kit (JSWDK) is a servlet engine that runs an HTTP server for

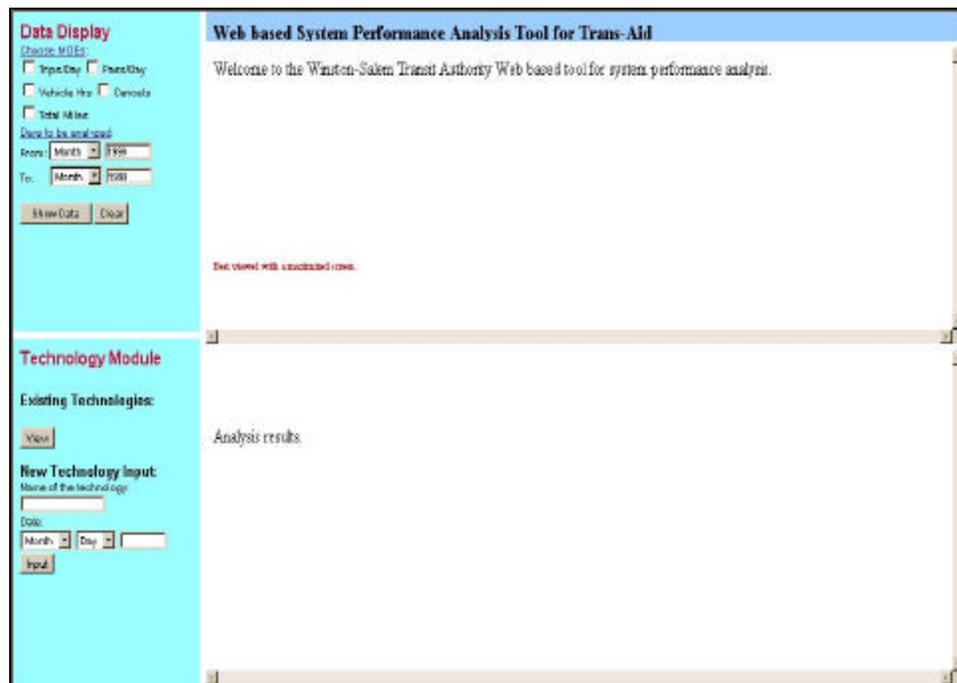
running and testing the servlets and contains the servlet APIs and classes. The tool was developed using JSWDK as the HTTP server for the testing and running. JSWDK however, proved to be slow and unwieldy at times, especially with large volumes of data. Therefore, it was decided to switch to another web server, which would give faster and better performance with larger volumes of data. The Resin web server, distributed by Caucho, was chosen for this purpose. Using the above mentioned technologies a tool has been developed which is capable of retrieving transit data and information for decision making and assessment purposes.

3.7 The Separate Modules

There are several modules in this tool; the Data Display module, the Performance Analysis module that includes the Graphical Display and Statistical Testing sections.

Figure 3.8 below is a screenshot of the opening page and shows some of the modules that appear on the front page. Each separate module is discussed below.

Figure 3. 7: Screenshot of the Opening Page

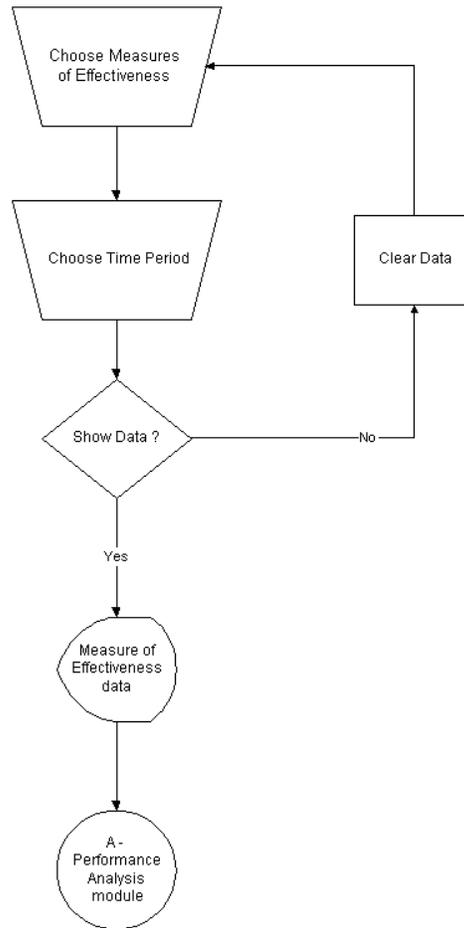


These modules are discussed below and the method of designing and implementing each of them described.

3.7.1 Data Displays

The measures of effectiveness data stored in the database can be viewed with the help of the user interface. The user gets to choose the measures he wants to see by checking the particular checkboxes. The evaluation time frame desired can be chosen with the help of the drop down menus for each of the beginning and end dates. The data can then be seen, by clicking the button for showing data. This can be encapsulated in the data flow diagram as shown below in figure 3.5.

Figure 3. 8: Decision Flow of the Data Display Module



The data is displayed in the form of tables with the headers showing the names of the measures and also linked to the help pages of each measure for the uninitiated user. The displayed data also has a series of radio buttons for each of the measures, of which the user can choose one for analysis, as shown in Figure 3.4.

Figure 3. 9: Screenshot of Data Display

Month	Trips_Per_Day	Vehicle_Hours	Total_Miles
Choose to Analyze	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1	309.88	88.97	1378.98
2	321.0	98.08	1421.94
3	313.34	96.21	1372.94
4	298.55	89.46	1268.07
5	250.12	75.5	1072.19
6	325.72	95.03	1450.46
7	330.88	98.14	1490.87

Clicking on the button named “Do Analysis”, brings up the user interface for performing several kinds of statistical analysis on the lower left frame. Analysis can then be done, by going to the various sections for seeing trends, viewing basic statistics or performing statistical tests.

Figure 3. 10: Screenshot of the Performance Analysis Module

Performance Analysis

Trends

Line Bar Both

Statistical Tests

3.7.2 Graphical Displays

Graphical displays of line and bar charts have been used in the tool with the intention of giving the user a visual idea of the performance trends. Java2 combines in it a powerful API called Java 2D for advanced line, art, text, and image rendering.

The application makes use of the Swing and Java 2D APIs for developing the statistical displays. The charts are developed on a Swing based frame, according to the chosen measure of effectiveness and its values for the time period and then this frame is finally converted to a gif image, using a gif encoder, to display it on the browser as an embedded image. The user is given a choice of seeing either line or bar chart or both.

Figure 3. 11: Decision Flow of the Performance Analysis Module

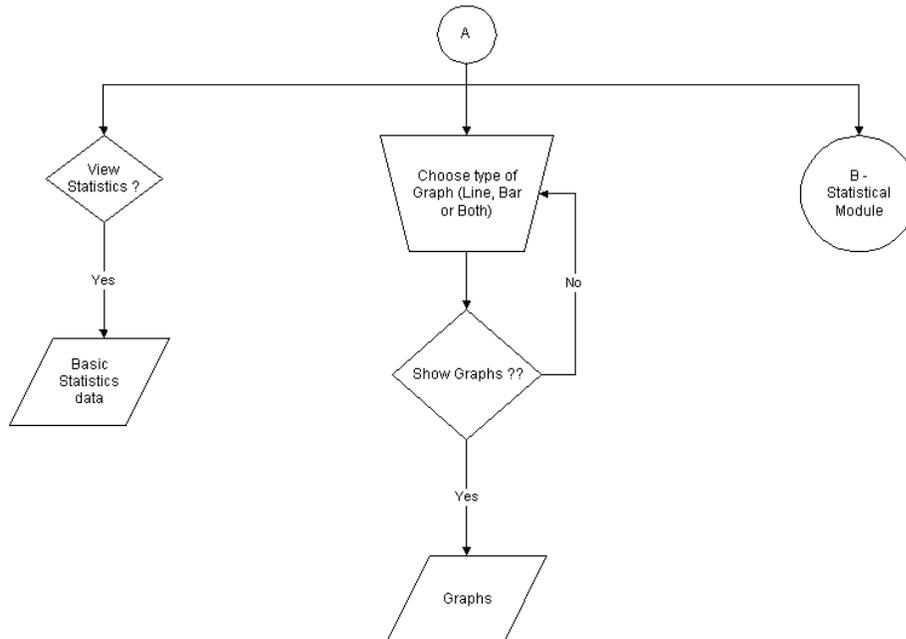
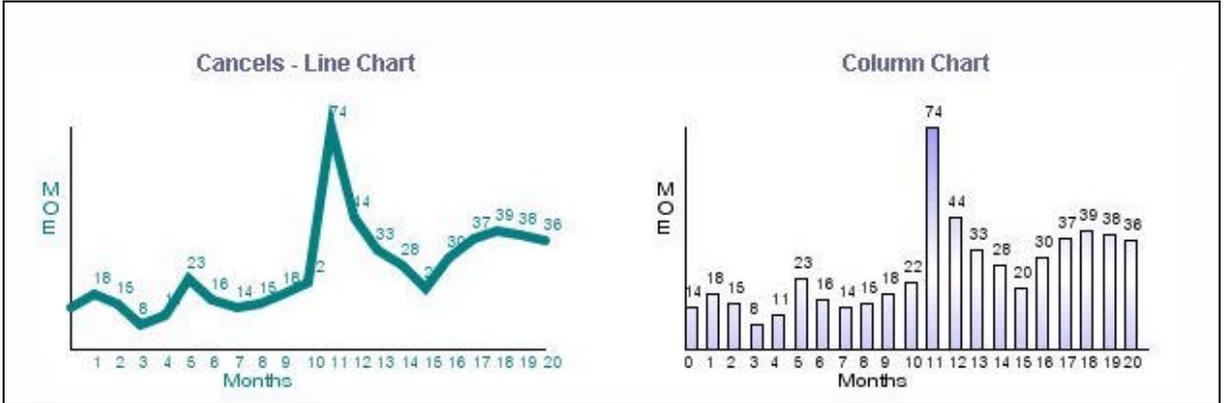
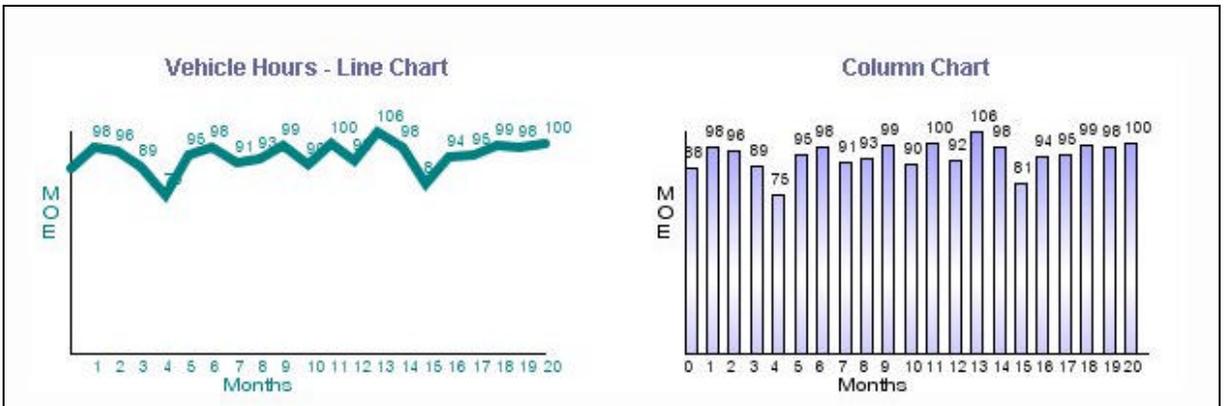


Figure 3. 12: Sample Screenshots of the Graphical Displays

a) Cancellations



b) Vehicle Hours



3.7.3 Statistical Analysis

The user can perform statistical analysis on the MOE data to gain additional insight into the performance trends and their significance. Basic statistics can be

performed on a chosen MOE at the click of a button to see the values of the mean, median and variance.

Figure 3.14 illustrates how the tool can be used for significance testing performed on two different time ranges for a particular measure of effectiveness in the form of t-tests. As shown in Figure 3.15, the user can enter the significance level as input and also choose the two time ranges. Using these and the MOE as the parameters, the user performs the test which give a set of results that includes the basic statistical results of the two groups of the particular MOE and the probability value. The probability value is then compared to the given value of the significance level or alpha value, and information given on whether the null hypothesis is true or false. An explanation of t-tests and the results are available for users who need an introduction to these tests as shown in Figure 3.15c and 3.15d.

Figure 3. 13: Decision Flow of the Statistical Analysis Module

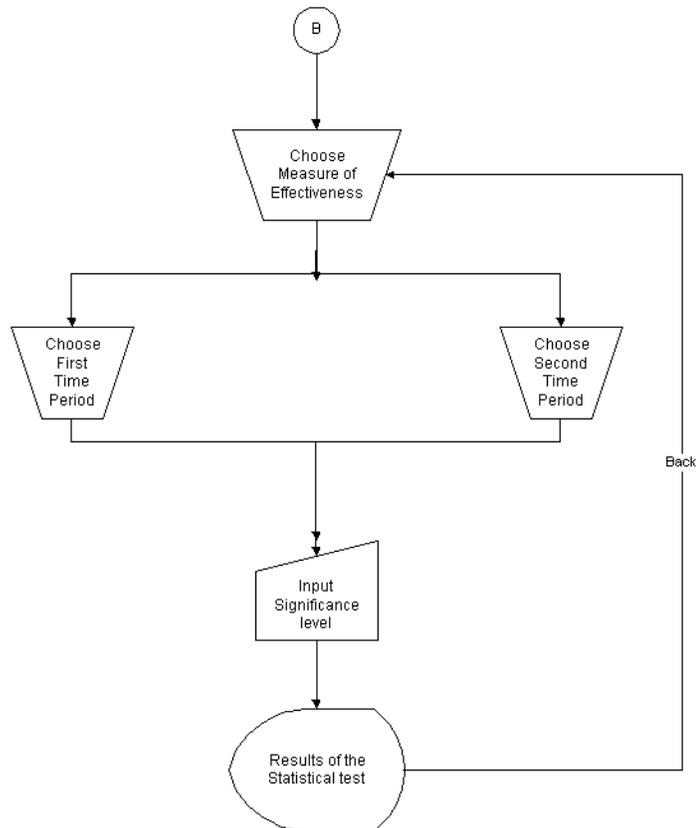


Figure 3. 14: Sample Screenshots of the t-test and ANOVA modules

a) The t-test Module Interface

Choose the t-test parameters:

Choose an MOE	Choose the time frames	
<input type="radio"/> Trips_Per_Day	1st Set: From: <input type="text" value="Month"/> <input type="text" value="1999"/>	To: <input type="text" value="Month"/> <input type="text" value="1999"/>
<input type="radio"/> Pass_Per_Day	2nd set: From: <input type="text" value="Month"/> <input type="text" value="1999"/>	To: <input type="text" value="Month"/> <input type="text" value="1999"/>
<input type="radio"/> Vehicle_Hours		
<input type="radio"/> Cancellations		
<input type="radio"/> Total_Miles		

Alpha or Significance level [Click here for details](#)

b) The ANOVA Module Interface

Choose the Anova parameters:

Choose an MOE	Choose the time frames	
<input type="radio"/> Trips_Per_Day	Start date: From: <input type="text" value="Month"/> <input type="text" value="1999"/>	
<input type="radio"/> Pass_Per_Day	End date: From: <input type="text" value="Month"/> <input type="text" value="1999"/>	
<input type="radio"/> Vehicle_Hours		
<input type="radio"/> Cancellations		
<input type="radio"/> Total_Miles		

Alpha or Significance level [Click here for details](#)

c) The Results Interface

t-test results:	Explanation
<p>Unrelated t-test performed at - Sat Aug 11 18:07:50 EDT 2001:</p> <p>Value of $t_{stat} = 0.2399302$ Difference in means = 3.6844654</p> <hr/> <p>First group: Number of Months = 4 Mean = 310.6925 Variance = 87.11818 Median = 309.88</p> <hr/> <p>Second group: Number of Months = 5 Mean = 307.008 Variance = 1070.2175 Median = 314.16</p> <hr/> <p>For a confidence level of 0.05 and Degrees of Freedom 7 The t_{crit} is 1.8946</p>	<p>$t_{stat} < t_{crit}$ Null hypothesis accepted! This means that the Typo_Per_Day scores has not shown any significant changes over time for a Confidence level of 95.0%. For an explanation on t-tests, please click here</p>
<input type="button" value="← Back"/>	
<input type="button" value="Print"/>	

d) The t-test Explanation Interface

Explanation of t-test

The t-test is used to assess the statistical difference in the means of two groups. This analysis is appropriate whenever you want to compare the means of two groups. Mostly in cases when some important change occurs between the period.

The t-test compares one variable (in our case, one **measure of effectiveness**) between two groups. And in the process, provides answers to two important questions:

1. Is the observed difference statistically significant?
2. Secondly, if the difference is statistically significant, then is the difference meaningful or substantial?

The t-test's formula is a ratio of the difference of the averages of the two groups and dispersion of the scores of the groups. The first of the ratio is the difference of the means or averages of the groups. The bottom part measures the variation or dispersion between the scores, also referred to as the **standard error of the difference**.

From the formula, we see that the t-value will be positive if the first mean is larger than the second and negative if it is smaller. After the t-value calculation, a risk/significance level (α) is used, according to the input, to look up a significance table using another variable called **degrees of freedom** (ν). So the t-value depends on the alpha level, and the degrees of freedom.

The null hypothesis, $H_0: \mu_1 - \mu_2 = 0$, is proposed to determine if the two groups display any changes in the measures. In the two-tailed test, H_0 will be rejected if either $t \geq t_{\alpha/2}$ or $t \leq -t_{\alpha/2}$.

3.7.4 Technology Module

The technology module has been incorporated into the tool to facilitate addition of newer technologies into the database, as and when they are added to the paratransit systems. These technologies may be already existing like the Interactive Voice Response (IVR) or the AVL/MDT technology.

The existing technologies can be viewed with a click of a button. The module is in form of a User Interface, which gives the user the facility to write down the name of the technology and choose the data when the technology has been incorporated into the system. When the button for inputting the technology is clicked, the addition is made into the database. Any errors in the process of addition, for example adding the same technology or a wrong date will be reported in the form of error messages.

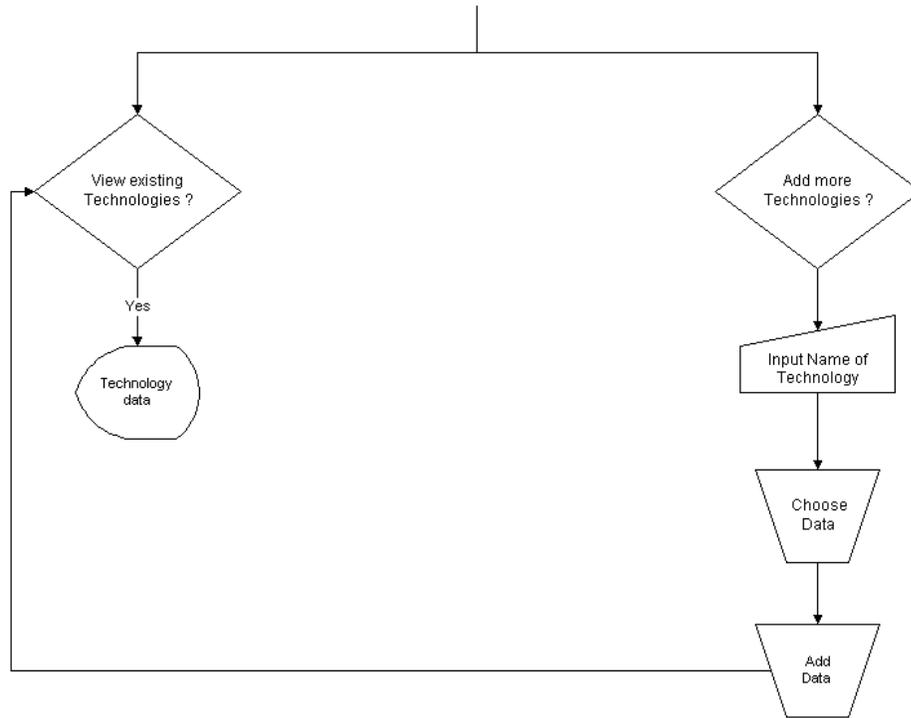
The current implemented version of the tool does not use the Technology Module for demarcating the incoming of a particular technology in the graphical displays. Hence, its use for updating the database for further technologies has been disabled.

Figure 3. 15: Sample screenshot of the Technology Module



The screenshot shows a web-based user interface for the Technology Module. The title is "Technology Module" in red. Below the title, there are two main sections. The first section is "Existing Technologies:" and contains a "View" button. The second section is "New Technology Input:" and contains a text input field for "Name of the technology:", a "Date:" label, two dropdown menus for "Month" and "Day", and a text input field for the day. Below these fields is an "Input" button.

Figure 3. 16: Decision Flow of the Technology Module



3.8 Summary

This chapter evaluated the decision support technologies and methods of approaching the problem of constructing the decision support tool that performs real-time statistical analysis of paratransit system performance data. Computing technologies like the Internet and Java combine to give a robust environment for a dynamic database oriented Internet related development work. In this context, the several advantages of these new techniques in comparison to traditional methods were also discussed. Finally the details of the development of the prototype tool for a paratransit performance evaluation and its different input and output modules were discussed.

CHAPTER 4: ANALYSIS OF DATA

This chapter documents the first task as described in the *Objectives* sub-section of Chapter 1 i.e. collection, formatting, validating and analyzing the data before they are used as the base data in the modules of the decision support tool. The chapter also addresses the primary concern of the research work, which is to determine if introduction of real time information to the paratransit users (management and customers) – like easy booking and cancellation of rides with the automatic telephone system, electronic data transfer between the transit center and vehicles – causes changes to the system productivity. The proposed tools would enhance, support and display productivity measures at any location and time, and help management evaluate fluctuations in the productivity measures.

The idea of a decision support tool for a paratransit system was conceived to give operational staff real time analysis and visualization of the data. Therefore, the actual real world implementation of the prototype would have to connect to the database at the transit facility for the data extraction. For the purposes of the prototype however, digital data acquired from WSTA in the form of concise reports. These reports are what most paratransit agencies prepare on monthly, weekly and daily basis containing basic measures of system productivity like ridership, mileage and cost data. In a future study the prototype can be connected to the server at the transit center for real-time data evaluation.

The data come from three kinds of reports, as described below:

- *Trips/Miles Productivity Reports* – These are DOS based text reports, generated in the older PASS system. They have productivity measures like Trips, Passengers, Total Miles and Deadhead in the form of daily values collected in a monthly file. (See Figure 3.4 for a sample Trips/Miles Productivity Report).

- *Provider Productivity Reports* – These are Windows based reports generated by Trapeze (the latest Windows version of PASS) in a proprietary file format having .chr extension. They have most of the measures that are there in the Trips/Miles Productivity reports and in addition have other count measures like No Shows and Cancellations. They also have Hourly data in the form of Total Hours, Service Hours, Revenue Hours etc. (See Figure 3.5 for a sample Provider Productivity Report).
- *Monthly Service Reports* – These are Windows based spreadsheet reports that are manually generated at the end of each month. They have cost measures like Operating Revenues, Expenses, Net Cost and Farebox Revenues. (See Appendix B for a sample Monthly Service Report).

The collected data has to undergo several processes including proper formatting, processing, and validation to remove unnecessary data to ensure correct final analysis. These steps are described below.

4.1 Data Formatting Procedure

The WSTA Provider Productivity reports⁸ (Figure 3.2) are in the proprietary .chr format. Some of the data, that are present in these files and are important for system productivity, include counts of Passengers, Trips, Cancellations, Hours, and Mileage data.

The .chr files contained daily data in comma-delimited form. These files were converted to Excel compatible form and then imported to an Access database file in the form of tables. The Access tables were then converted to the final concise form by eliminating the fields such as breaks hours, refuel hours, deadhead, and fixed route data

⁸ Provided by the WSTA supervisor.

that are not used. These edited Access tables provided the basis for analysis and graphical displays.

4.2 Data Verification

The data that is collected and formatted for analysis purposes has to be subjected to validation to weed out errors and their possible resulting effect on the analysis.

In the course of formatting, the data was queried in the Access database to check whether the values for each of the measures falls between the usual values and that there are no absurdly high or low values. This was done by querying, into each of the measures of the data files (or *attributes* in database terminology), to check if any of the values are above or below a certain value. In almost all the cases, the queries returned no results. This suggests that the measures had no unusually high or low values, which might have been due to a data input error. The files, which returned invalid results, were checked again and almost always the root was found to be bad data entries. These were replaced with data from more recently acquired and corrected WSTA reports. These validated data sets were then used for identifying trends and other analysis using the measures like ridership, mileage etc. Figure 4.1 shows such a set of validated data for one month (March 2000). For purposes of the analysis, it was assumed that the files are free of errors like bad data entries, inadvertent deletions etc that are present in the ride data.

4.3 Statistical Analysis of Productivity Data

4.3.1 Background on Statistical Analysis

A preliminary offline analysis, using spreadsheets provided graphical trends and basic statistical tests. This served several purposes. Firstly, it provides an idea of the trends that the data follows. This in turn suggests whether the availability of advanced

technology has any effect on the system productivity before we proceed to the development of tools that would enable display and analysis of such data at real time. Secondly, this serves the overall research objective of finding productivity changes in the system with coming of new technologies⁹. These changes could be tested as to whether they occurred due to chance variations or are statistically significant. Lastly, the formatted and already analyzed data is a more compact and concise *database module*¹⁰ for use in the prototype.¹¹

As mentioned in the *Manual of Transportation Engineering Studies* (Hummer et al) statistical analysis always involves several terms. Some of the terms that are discussed include *units*, *measures of effectiveness*, *factors*, and *treatment*. The *units* for the data to be used are the months as they are the subjects for the measurement of change. The *Measures of Effectiveness*' (MOEs) that are to be measured during the course of the experiment are taken from the reports. As can be seen in Figure 4.1, they include counts of passengers, trips, no shows, cancellations, hours and mileage. The *factors* or the variables manipulated in the experiment are the advanced technology or the incoming of real time information in the form of IVR, AVL/MDT and variable message signs. *Treatment* in this case would be the application of different technologies over a period of time. The different technologies are being introduced at separate times and every combination of the technologies would serve as a different treatment. In this case, the major treatment would be the application of the IVR technology and the introduction of the dollar fare in the month of August 2000. MDTs and AVLs were installed eventually in February 2002.

⁹ Till the time period the data was analyzed, only the IVR system was introduced.

¹⁰ The structure, design and development of the database module of the implemented tool has been described in detail in the sub-section titled *Implementation of the Decision Support Model* in the following chapter.

¹¹ Such an offline analysis would not be necessary in the final operational decision support tool.

Figure 4. 1: Sample Validated File for the month of March 2000

Date	Passenger	Trips	NS	Cancel	TotalHours	ServiceHours	RevHours	TotalMileage	ServiceMileage	RevMileage	PassengerMileage	TripsPerRH	TripsPerSH
3/1/00	384	384	3	0	159	108.1	80.8	1812.7	1594	1179.4	4012.7	4.75	3.55
3/2/00	355	353	2	0	103.2	94.3	75.1	1678.3	1511.9	1174.3	4995.3	4.7	3.74
3/3/00	343	341	0	0	107.4	98.2	69	1565.7	1373.9	1073.9	4105	4.94	3.47
3/4/00	89	89	0	0	21.7	19.2	17.8	324.9	258.2	258.2	1131.7	5	4.64
3/6/00	389	389	1	0	119.7	113.1	83.6	1896	1732.2	1317.8	4663.3	4.65	3.44
3/7/00	367	367	0	0	107.8	103.3	78.9	1785.2	1655.6	1241.2	4845.5	4.65	3.55
3/8/00	400	400	2	0	112.6	104.4	81.5	1762.9	1587.5	1224.1	4550.4	4.91	3.83
3/9/00	383	380	0	0	124.3	117.2	82.1	1863.4	1663.2	1286.3	5762.7	4.63	3.24
3/10/00	348	346	0	0	99.3	92.1	70.3	1501.6	1306.2	1041.6	4698.9	4.92	3.76
3/11/00	93	93	0	0	23.1	20.7	19.1	322.8	254.6	254.6	1143.4	4.88	4.5
3/13/00	347	347	7	0	105.4	99.1	72.5	1542	1374.3	1097.1	4346.8	4.79	3.5
3/14/00	369	369	0	0	110.6	104.8	80.2	1726.6	1569.2	1214.7	4562.7	4.6	3.52
3/15/00	374	374	3	0	97	91.6	72.7	1471.9	1330.8	1050.3	4716	5.14	4.08
3/16/00	404	402	0	0	113.1	106.1	81	1805.8	1611.2	1256.1	5357.2	4.96	3.79
3/17/00	353	353	1	0	107.7	99.2	73	1658.4	1455	1137.4	5031.3	4.83	3.56
3/18/00	96	96	0	0	23.2	20.7	19.3	329.8	261.1	261.1	1259.5	4.97	4.64
3/20/00	377	376	1	0	115.5	107.1	78	1777.6	1582.5	1218.1	4495.6	4.82	3.51
3/21/00	385	385	4	0	109	104.1	77.8	1636.1	1501.9	1170.6	5252.4	4.95	3.7
3/22/00	400	400	1	0	118	110.9	82.5	1772.6	1581.9	1208.5	4503.6	4.85	3.61
3/23/00	384	382	2	0	113	105.6	76.2	1838.7	1637.7	1230.6	4673.6	5.01	3.62
3/24/00	327	327	7	0	96.4	89.3	65.1	1446.2	1257.8	975.3	4383.9	5.03	3.66
3/25/00	95	95	0	0	23.8	21.3	19.9	379.7	312.8	312.8	1569.2	4.78	4.46
3/27/00	413	413	5	0	115.5	108	86.4	1851.5	1693.1	1349.7	6004.7	4.78	3.82
3/28/00	376	376	1	0	111.9	106.2	80.1	1704.2	1535.3	1217.3	5467.6	4.69	3.54
3/29/00	398	397	3	0	115.6	109.4	84	1739.9	1572	1269.4	5156.2	4.73	3.63
3/30/00	359	357	0	0	97.7	91	70.2	1617.6	1440.1	1142	5461.9	5.09	3.92
3/31/00	343	343	2	0	98.2	90.4	68.8	1431.3	1224.2	987.5	4523.5	4.99	3.79
Avg	331.5	330.9	1.7	0.0	98.1	90.2	68.4	1490.9	1328.8	1042.6	4321.3	4.9	3.8

During the study the Trans-AID paratransit system remained relatively constant except for changes in the latter quarter of 2000. Some of the major characteristics of the years of which the data is present and has been analyzed are:

- **1999:**
 1. In September the data server was changed from an old DOS version called *PASS* to the latest Windows version called *Trapeze*. This is the start of the time period for the data collection.
- **2000:**
 1. The *Interactive Voice Response (IVR)* system was introduced on *10th of July 2000*.
 2. The *Dollar fare* for the passengers was introduced beginning the month of *August 2000*.
 3. The *Penalty for No-shows* was introduced in the following month, that is *September 2000*.
- **2001:**
 1. No changes.
- **2002:**
 1. *MDT* and *AVL* were installed and functioning over the month of *January 2002*.

This sequence of events is graphically represented in the Figure 4.2.

Figure 4. 2: Sequence of Technology and Policy changes

	1999	2000	2001	2002
Technology implementation	Sep -Trapeze installed/ <i>PASS</i> phased out	Jul - IVR installed		Jan - MDT/AVL installed
Policy changes		Aug - Dollar fare Sep - Penalty for no-shows		

During the 1999 to 2000 period, the system characteristics which remained constant are – fleet size, service area, service hours and staff size.

The major changes, for which the data is available, happened in the short space of 3 months. Effectively, the changes did not take place in a longer, controlled time period that would help distinguish impacts given the monthly aggregate form of the data. Data, on a daily or weekly basis¹² could be used in future applications of the prototype.

4.3.2 Graphical Analysis

Graphical analysis gives a visual picture of the trends attained by the data in the chosen time period (daily, weekly and monthly periods during 1999-2001).¹³

4.3.2.1 Daily Data

Graphical analysis of the daily data arranged as months will give a picture of the changes that occur in the measures in each month on a day-to-day basis. Figure 4.2 shows the Time Series plot of the measures for the month of September 1999 and 2000. Each point represents a daily count. This has been done by constructing time series distributions consisting of bar graphs of a combination of similar measures into groups. These groups included the following combinations of measures of effectiveness:

- Passenger and trip counts.
- Cancellations.
- Total hours, service hours, and revenue hours.
- Total, service, revenue and passenger mileage.

¹² Daily/weekly data could not be obtained from the reports for the cancellations

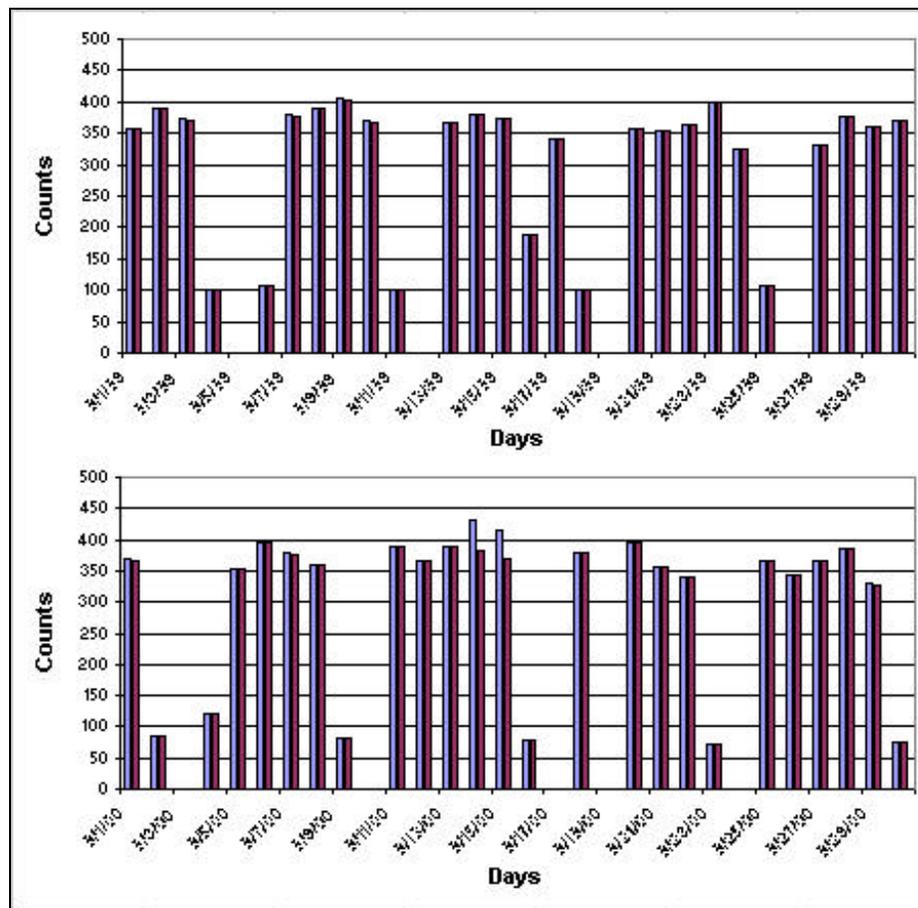
¹³ In the working version of the prototype, the features for displaying the daily and weekly data were disabled, as this data was unavailable.

Some generic observations about the daily data are as follows:

Passengers and Trips

- The daily data arranged by months appears to be consistently similar with regular weekly ridership cycles (Figure 4.2(a)). The counts reach peak at the middle of the week before tapering off towards the end of the week. This trend is however not followed on all the months. Notable exception is November 1999.
- The ridership data averages around 300 passengers per day and the Saturday ridership averages around 100.
- Some of the months have a few days of very low ridership. These low counts can be attributed to holidays and inclement weather.

Figure 4. 3(a): Time Series Plot for Passenger and Trip Counts – Sep 1999 and 2000.



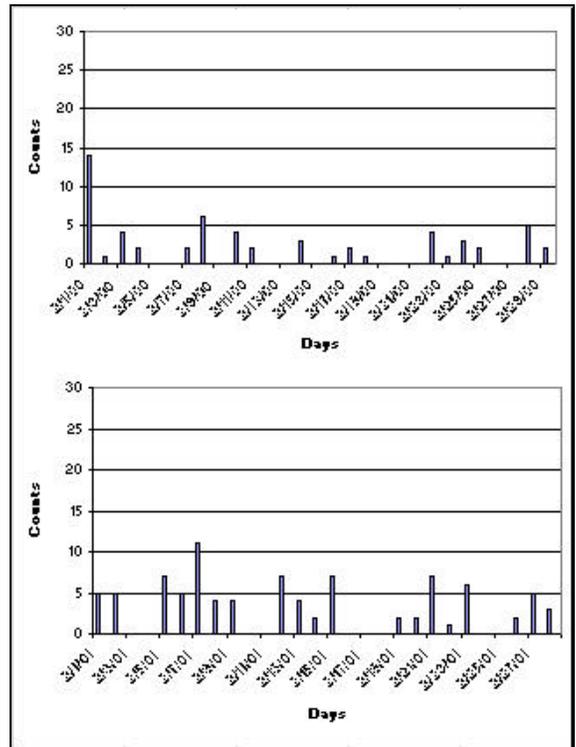
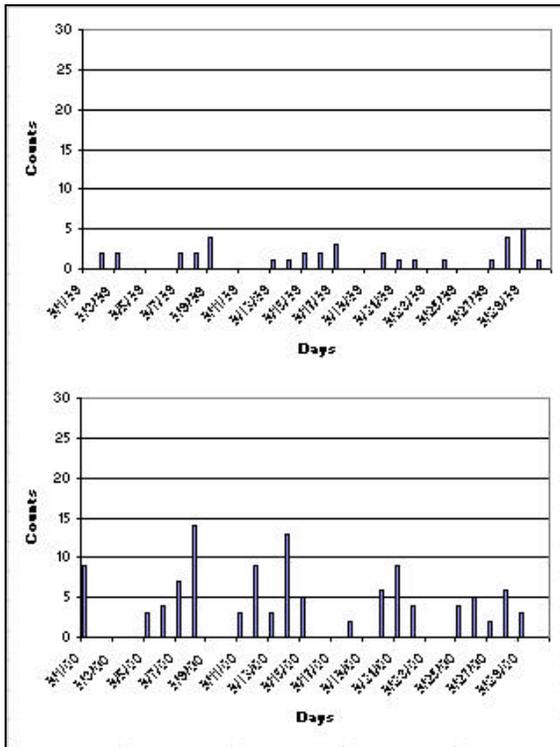
Cancellations

- Most of the months have cancellations scattered throughout the month, with daily highs of more than seven cancellations. The highs occur mostly in the mid-week.
- The month of August 2000 shows a jump in the number of cancellations per day.
- A comparison of the months of September 1999,2000 and Feb 2000, 2001 (Figure 4.2(b)) shows that the average number of cancellations increased in the one-year period.

Figure 4. 3(b): Time Series Plot for Cancellations

1. Sep 1999 and 2000

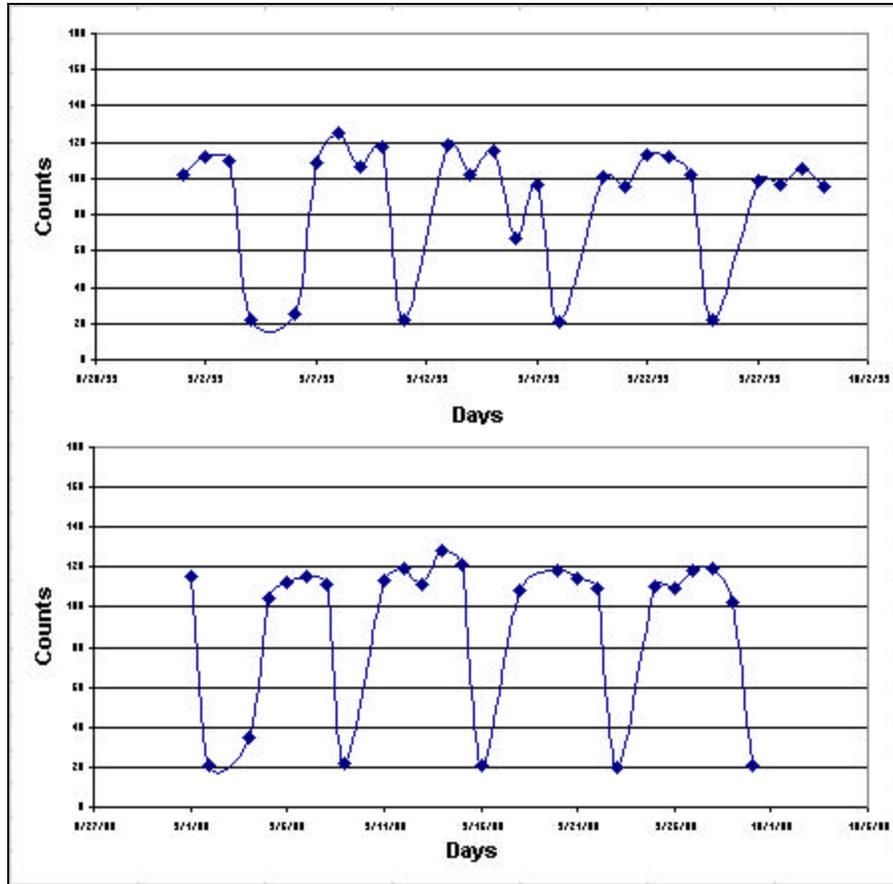
2. Feb 2000 and 2001



Vehicle Hours

- Show consistent trends throughout the whole period. Depending on the number of weekends, the peaks vary from three to four each month depending on the number of weekends, as shown in Figure 4.2(c). Each week consists roughly of a double peaked cycle with a slight crest in the middle of the week.

Figure 4. 3(c): Time Series Plot for Vehicle Hours – Sep 1999 and 2000.

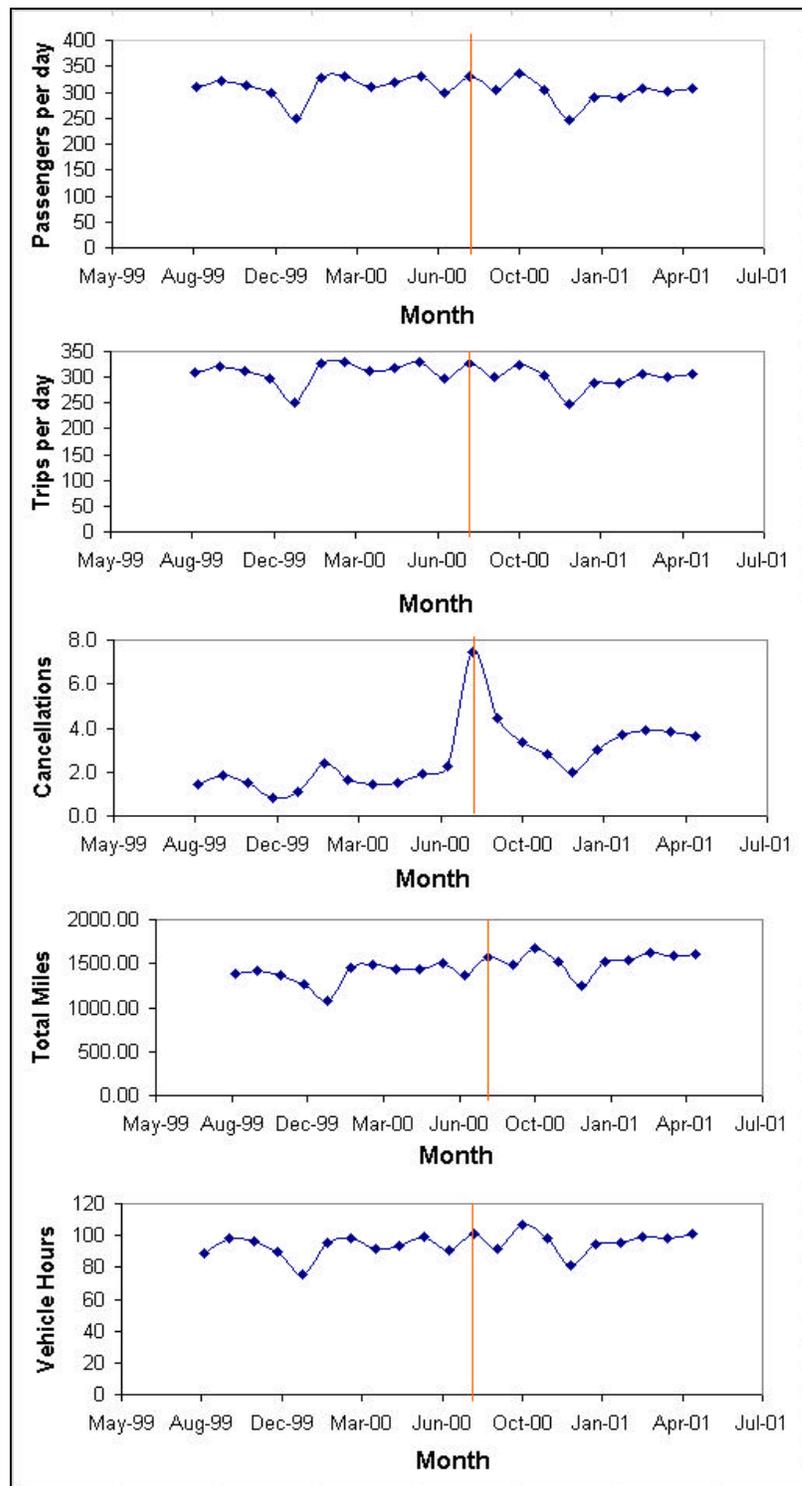


On a *daily* basis, Vehicle Hours seems to be the most consistent due to consistent routes and staffing while Cancellations are erratic because of the changes in passenger perceptions.

4.3.2.2 Monthly Data

The above daily data were then averaged over the monthly period. Figure 4.3 shows the Time Series plot for all the measures of effectiveness.

Figure 4. 4: Time Series Plot for Monthly Data



Each point represents a daily average over a month.

Some generic observations about the monthly data (Figure 4.3) are as follows:

- It is noticed that the trends of all the measures are almost similar, with the exception of the trend for Cancellations.
- The Cancellations show a peak on the month of August 2000 and the supervisor of WSTA attributes this to the dollar fare that was introduced that month, resulting in a lot of last-minute no shows and cancellations. Overall, it shows an increase.
- The lowest count of Cancellations occurs in December 1999.
- The average value of Cancellations appears to be increasing after the month of August 2000, perhaps because of the ease of IVR dialing and automatically canceling without speaking to the operator.
- The average value of the all the other measures, (passengers, trips, vehicle-hours, and vehicle-miles) do not appear to be changing over the entire period.

These observations above which were noted from daily and monthly graphical data, however, need to be verified by implementing stronger statistical tests. This is explained in further detail, in the next section.

4.3.3 Statistical Analysis

4.3.3.1 Results of the t-test

The t-test is a method to compare the means of samples from two different populations to determine whether or not they are significantly different. This is done by a comparison of the means or medians.

An unpaired t-test assuming equal variances for two samples was performed on the measures. This was done for pairs of time periods chosen over a whole range of time period to observe significant changes, if any, in the values of the means. This is in accordance to the recommendations in *The Manual of Transportation Engineering Studies*.

The two sets of time periods used for performing t-test, consider August 2000 i.e. when IVR became operational, to be the critical month. The first time period was from September 1999 to July 2000. The second time period was from August 2000 to May 2001. Values taken from all the months for each of the measures were subjected to the tests for a significance value of 0.05 or 95% *confidence level*¹⁴ and a significance value of 0.10 or 90% *confidence level*.

This is done to detect significant changes occurring in the measures especially in the period after July-August 2000, as mentioned in Section 4.3.1, to assess productivity changes. August 2000 was the month when several changes took place in the overall paratransit system, in particular the introduction of the new *dollar fare* and the *IVR system*. These changes make the period between August 2000 to October 2000 the time of interest for detecting changes in the system performance measures.

¹⁴ The 5% level, as mentioned in the *Glossary of Terms*, is a widely used level. This is, the usual test for experimental data in controlled environments where there is less possibility of variations. For a real world case the 10% level can be used as benchmark.

Figure 4. 5: Summary of the t-test results

MOE	Test	Period		Confidence Level	Null Hypothesis Result
		To	From		
Trips per day	t-test	Sep-99	Jul-00	95%	TRUE
		Aug-00	May-01	95%	
	t-test	Sep-99	Jul-00	90%	TRUE
		Aug-00	May-01	90%	
Passengers per day	t-test	Sep-99	Jul-00	95%	TRUE
		Aug-00	May-01	95%	
	t-test	Sep-99	Jul-00	90%	TRUE
		Aug-00	May-01	90%	
Vehicle Hours	t-test	Sep-99	Jul-00	95%	TRUE
		Aug-00	May-01	95%	
	t-test	Sep-99	Jul-00	90%	FALSE
		Aug-00	May-01	90%	
Cancellations	t-test	Sep-99	Jul-00	95%	FALSE
		Aug-00	May-01	95%	
	t-test	Sep-99	Jul-00	90%	FALSE
		Aug-00	May-01	90%	
Total Miles	t-test	Sep-99	Jul-00	95%	FALSE
		Aug-00	May-01	95%	
	t-test	Sep-99	Jul-00	90%	FALSE
		Aug-00	May-01	90%	

From the above summary table (Figure 4.5), we can draw the following conclusions:

- The measures *Trips/Day* and *Passengers/Day* do not show any change between the two time intervals (Sep 1999 - July 2000 and Aug 2000 - May 2001).
- *Vehicle Hours* shows a change (increase) only at the lower confidence level of 90%, not at 95% confidence level. A change at a lower confidence level indicates less reliability of the result.¹⁵
- *Cancellations* and *Total Miles* show a change (increase) at the both 90% and 95% levels.

¹⁵ Since 90% confidence indicates a significance of 10% i.e. a 10% chance of error.

4.3.3.2 Results of the ANOVA

The purpose of Analysis of Variance (ANOVA) is to test for significant differences between means. ANOVA extends the idea of t-test to testing whether the means of all the samples of the population are equal. If we are only comparing two means, then ANOVA will give the same results as the t-test for independent samples.

As an initial start, the test of ANOVA was performed on one of the measures from each of the above given groups to see if there was any significant difference in the means possibly as a result of any service or technology change including the new fare and IVR automatic telephone access for passengers. Values taken from all the months for each of these measures were subjected to ANOVA tests for a significance value of 0.05 or 95% *confidence level*² and 0.10 or 90% *confidence level*. As in the t-test, two sets of time periods were used for performing ANOVA. The first time period was from September 1999 to July 2000. The second time period was from September 1999 to May 2001. So the ANOVA measured the changes in the means of the monthly values.

The above summary table (Figure 4.5) shows the following:

- *All the measures* unanimously show *no change* in the first time period i.e. from September 1999 to July 2000. This proves that the average value of all the measures show no significant changes in this period for the confidence levels of both 95% and 90%. This means that the WSTA Trans-AID system was stable and that no operational or service changes were made. WSTA staff confirmed this.
- The measures *Trips/Day*, *Passengers/Day*, and *Vehicle Hours* do not show any change over the entire period (September 1999 – May 2001) for both the confidence levels of 90% and 95%. This indicates there has been no statistically significant change during this period. This means that the Trans-AID service measured by vehicle hours is constant and passenger demand is constant in spite of the introduction of the \$1 fare and the automatic IVR communication. That no

Figure 4. 6: Summary of the ANOVA results

MOE	Test	Period		Confidence Level	Null Hypothesis Result
		To	From		
Trips per day	Anova	Sep-99	Jul-00	95%	TRUE
		Sep-99	May-01	95%	TRUE
	Anova	Sep-99	Jul-00	90%	TRUE
		Sep-99	May-01	90%	TRUE
Passengers per day	Anova	Sep-99	Jul-00	95%	TRUE
		Sep-99	May-01	95%	TRUE
	Anova	Sep-99	Jul-00	90%	TRUE
		Sep-99	May-01	90%	TRUE
Vehicle Hours	Anova	Sep-99	Jul-00	95%	TRUE
		Sep-99	May-01	95%	TRUE
	Anova	Sep-99	Jul-00	90%	TRUE
		Sep-99	May-01	90%	TRUE
Cancellations	Anova	Sep-99	Jul-00	95%	TRUE
		Sep-99	May-01	95%	FALSE
	Anova	Sep-99	Jul-00	90%	TRUE
		Sep-99	May-01	90%	FALSE
Total Miles	Anova	Sep-99	Jul-00	95%	TRUE
		Sep-99	May-01	95%	TRUE
	Anova	Sep-99	Jul-00	90%	TRUE
		Sep-99	May-01	90%	FALSE

significant demand change occurs is reasonable because the elderly and handicapped clientele represent captive riders with no other transit options.

- *Cancellations* show a statistically significant change (increase) in the time period September 1999 to May 2001 for both the confidence levels 90% and 95%. This indicates to a fair degree that the number of cancellations has increased over the time period, especially after the months of August-October 2000, when IVR automatic telephone cancellations became possible. It is likely that the availability of easy dial-in cancellation helped increase the cancellation rate.
- *Total Vehicle Miles* shows a statistically significant change (increase) in the time period September 1999 to May 2001 only for the confidence level 90% and not at a higher confidence level of 95%.

4.4 Conclusions

Some overall conclusions that can be arrived at are:

- From the results of the ANOVA and t-tests it can be positively stated that there occurred *no statistically significant change* in the measures in the time period of *September 1999 to July 2000*. (Since the null hypothesis is *true* for both the confidence levels 90% and 95%).
- The measure *Vehicle Hours* shows a *statistically significant change* over the time periods before and after August 2000 on a t-test at a confidence level of 90%. However, on an ANOVA testing over the whole period it does not show any statistically significant change in monthly means. The significant difference in the t-test can be attributed to a wider confidence level causing a *False* null hypothesis¹⁶. This can be also possibly be attributed to the presence of unequal numbers of Saturdays for a month that causes differences in narrower testing like the t-test with a smaller confidence level but evens out in the case of a wider range test like ANOVA. The t-test results for a lower confidence level also go against the eye assessment done from the graphical plots that suggested a no-change.
- Similar is the case with the measure *Total Miles* for a confidence level of 95%. However the *False* null hypothesis is probably misleading because the t-statistic is only approximately equal to the t-critical value.
- *Cancellations* is the only measure that shows difference with time, when compared in the two time intervals. The sudden peak occurring in the month of August 2000 causes this. The cause of this peak in number of cancellations has been discussed in 4.3.2.2. It is also the measure that shows an overall trend that is different from the rest of the measures that have a similar trend over the time period.
- The *Interactive Voice Response (IVR)* or the automated telephone system was started on August 2000. The analysis of the data shows that the performance data having statistically significant changes to a large extent for the measure of Cancellations and

¹⁶ In statistical terms, this is commonly referred to as a *Type I error*, where a null hypothesis is rejected as false.

to some extent for Vehicle Hours and Total Miles. However, the month of August 2000 was the time of another change coming into the paratransit system i.e. the introduction of the *dollar fare*. Hence, it cannot be concluded firmly if the changes showing in the performance data are due to the introduction of the IVR system. This is especially so, because of the fact that introduction of the dollar fare induced a huge amount of last-minute cancellations and therefore the peak in cancellations for the month of August 2000. This in itself is probably responsible for causing a *false null hypothesis* in the statistical tests.

- Analysis of the *Cancellation* data shows that the average mean increases after the occurrence of peak in the month of August 2000 and continues till May 2001 (end of the period). This indicates to the fact that the introduction of the IVR system has *increased* the amount of *cancellations* occurring in the system.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary

There has been ongoing work for analysis of transit data as illustrated in Chapter 2. This work describes an interactive method of visualizing and statistically analyzing paratransit performance data using the Internet as a medium. It is based upon two concepts:

- Use of the Web as the medium for building a real time interactive decision support tool for paratransit data.
- Use of web-based statistical tests for gaining insight into data, discovering trends, and determining policy-related impacts.

The study focused on Winston-Salem Transit Authority (WSTA) that runs a twenty-vehicle paratransit fleet geared for Winston-Salem, NC and Forsyth County. Over a period, new technologies were added to enhance the productivity and service of the WSTA system. These included the Interactive Voice Response (IVR) system, Automatic Vehicle Locators (AVL), Mobile Data Terminals (MDTs) and Bay Signs. The IVR system began operation in the month of August 2000 while the other systems are still in the process of being tested and integrated into the system.

The study required collecting, formatting and analyzing paratransit data from Winston-Salem Transit Authority for incorporation into a database for the years 1999 to 2001. The initial analysis of the data allowed an investigation into trends that the WSTA data follows, for the time period in which the data is available. The study involved development of applications for display of data and graphs, development of programs that perform the tasks of display and statistical testing of the data.

The findings and conclusions are described in the next section.

5.2 Conclusions

The conclusions of the study that were arrived at during this thesis work are described below. They are organized as follows – the Internet decision support tool, changes in transit performance measures, and impacts of advanced technology on transit productivity. The data spanned a period from September 1999 to May 2001. From September 1999 to July 2000, no service or operational policy changes occurred. In August 2000 WSTA installed IVR and instituted a \$1 fare. From September 2000 to May 2001 no other service or technology changes occurred. After that, the MDT/AVL has been installed in January 2002 and are running, but lies outside the period in which the data was collected.

5.2.1 Internet Decision Support Tool

The following conclusions can be made about the Internet based tool:

- The World Wide Web (the Internet) can be an easy and effective medium for building interactive multi-user Java-based applications with a lot of interactivity., for robust decision support tools. The ease of interaction can be vital for detecting changes in measures related to efficiency of a system.
- The prototype Internet decision support tool that was built for the purpose of this study is ready for field-testing.

The prototype decision support tool analyzed the transit data for the period of September 1999 to May 2001. The first advanced technology (IVR) was installed in August 2000. Hence the period of September 1999 to July 2000 represents the “before” data.

5.2.2 Transit Performance

The primary conclusions regarding the impact of IVR on WSTA productivity are as follows:

- Data including Trips, Vehicle Hours and Cancellations were consistent and accurate, except for Vehicle Miles. There were no anomalous values that may have resulted from WSTA staff data input or WSTA calculations. Based on t-tests and ANOVA, the performance data is *stable* i.e. there are no statistically significant WSTA productivity changes from *September 1999* to *July 2000*, before IVR became operational. Furthermore, there were no service or operational changes in this period.
- After the technology was installed in August 2000, *Cancellations* show a significant change over the period of September 1999 to May 2001 (i.e. the time that the data is available). The abrupt change in cancellations occurred in the month of August 2000, under the simultaneous effect of the introduction of the dollar fare, the no-show penalty and the IVR system. Changes in *Cancellations*, however, cannot be attributed to any incoming technology like the IVR because cancellations most likely increased because of public dissatisfaction with the introduction of the *dollar fare* and introduction of *penalty* for no-shows. The increase in the average number of cancellations also indicates that the addition of the IVR technology into the paratransit system facilitated easy dialing up and canceling of rides.
- The measure *Vehicle Hours* also shows changes (increase) but no conclusive statements can be made about the accuracy and significance of these results. More after-data will probably be helpful as far as the tests and results are concerned.

5.3 Software limitations in the Prototype

Some of the bugs that are present are as follows:

- The Performance Analysis module cannot analyze an unlimited number of Months and Measures of Effectiveness. The web server gives an *Error 414*, which means the length of the URL is too long. Using a better web server that is capable of handling larger data and therefore larger URLs will solve this problem.
- A bug in Java 1.2.2 causes the display to disappear sometimes when the statistical tests are being performed. It throws an error called “*Function Sequence Error*” in the server console. It does not appear at all times. Pressing the ‘Show Data’ button again will cause the data to be displayed again. The interface also disappears occasionally in the “Statistical Analysis” module when the pressing of the buttons for “ANOVA” or “t-test” causes the module interface to disappear. Pressing the buttons a second time makes the interface reappear.

5.4 Recommendations for Future Research

5.4.1 WSTA Data

The following recommendations can be made about the WSTA data:

- Obtain access to the proprietary information server so that real-time data can be aggregated daily or weekly.
- Mileage data was not incorporated in the prototype analysis because the WSTA supervisor confirmed that the mileage data were manually taken from odometer readings and therefore do not reflect the correct (“revenue”) mileage, free from breaks, rest periods etc. As MDT, AVL and odometer sensors are implemented, This data deficiency will be overcome with inclusion of accurate mileage data including total vehicle miles, revenue miles (with passengers) and deadhead miles.
- Inclusion of PASS data (i.e. data before Sep 1999) for analysis provided they can be extracted in proper format, so that a larger time period of *before* data is available.

- Statistical analysis of the data used a time period scale of a month. Analysis on a daily or weekly basis to reveal important fluctuations in the data.
- Institute an archival system so that long-term before/after data sets can be analyzed.
- Select a set of performance and service measures that can be evaluated over extended time periods. Such data can be used to show impacts of change in policy, service, technology, routes etc.

5.4.2 WSTA Performance Evaluation

The following recommendations can be made about the performance evaluation techniques:

- Stagger changes in policy, technology, etc. in order to separate evaluate them. This improves experimental control. Incoming of technologies, as indicated in the preceding chapters have had dates that are not very far from each other. This makes controlled study of the effect of one technology difficult to study, since the period of after data is very less.
- Analyze Saturday service separately. On a larger time frame these low count days cancel out, but their presence shows in a statistical test like a *t-test* where the mean of a particular time period is compared to another. For example, the measure Vehicle Hours shows a statistically significant change for the periods before and after the month of August 2000 but the overall time period shows a statistically insignificant change on an ANOVA test.

5.4.3 Prototype Tool

The following recommendations can improve the methodology for paratransit system performance analysis:

- Use a larger database management system (DBMS) like Oracle or SQL Server instead of the Access database, which has been used for the purposes of building the prototype. A robust multi-user enterprise DBMS, would help resolve many issues like tool performance, recovery and online backup capabilities. This would ensure faster performance and permit more users.
- Replace the prototype JSWDK and Resin servers with a bigger and better web server that would cure bugs, improve performance, and reduce response time of web pages and concurrency. JSWDK and Resin are capable of handling only low loads and have several known bugs.
- Access to the vendors' repository of data containing archives of AVL, IVR and other ride data. This would enable a real world implementation of the prototype that would give the transit facility access to the database and enable real-time evaluation of all performance measures. Since a data dictionary and schematic of the database relations (or tables) are available, it would be possible to redesign the Java applications to connect to the actual database at WSTA and query data from it. See Appendix B for the schematics of the paratransit tables contained in the Trapeze database.
- Modify the tool such to analyze daily data. This would require enhancing the interface to include inputs for *Days* and changing of the Java applications to accept this extra data.
- Develop a *Technology Module* for the tool. Such a module would serve the purpose of updating the database of the new technologies that would be added to the paratransit system. New applications would have to be developed that display these technologies and their dates of implementation in the graphical trends and the tabular displays.
- Something about the other measures like cost etc that have been used from reports but are probably not available in the database.
- Something about inclusion of goal figures in the trends for purposes of comparison.

- Modify the prototype version to make it an Intranet version, which would allow only users who have a valid login and password access into the interface. This would mean making a separate database relation to include the user names, identifications and administration of the database to include new members and update existing member information.

CHAPTER 6: REFERENCES

1. Stone, John R. (2000) "Transit Customer Satisfaction with Real-Time Information", Department of Civil Engineering, North Carolina State University, Raleigh, NC.
2. Chira-Chavala, T., Christoffel Venter, and Geoff Gosling (1997) "Advanced Paratransit System: An Application of Digital Map, Automated Vehicle Scheduling and Vehicle Location Systems". Located:
http://www.itsdocs.fhwa.dot.gov/jpodocs/repts_te/2bf01!.pdf
3. Turner, S. M., Robert E. Byrdia, Jyh C. Liu, and William L. Eisele (1997) "ITS Data Management System: *Year One Activities*". Located:
http://www.itsdocs.fhwa.dot.gov/jpodocs/repts_te/6f901!.pdf
4. Porter, C., Tom Buffkin and Laurie Hussey (2000) "Evaluation plan for the Cape Cod Advanced public transportation system", *Federal Transit Administration*.
5. Byrdia, R., S. Turner, W. Eisele, J. Liu (1998) "Development of an ITS Data Management System", *Transportation Research Board*.
6. Cryer, J., M. Kyte, J. Stoner (1985) "Development of Time-Series based Transit patronage models. Volume II-Handbook for applying Time Series Models to Transit Ridership Forecasts", *National Technical Information Services*, Springfield, VA.
7. Lave, R., C. A. Pozdena (1977) "Statistical Analysis of Transit Performance." *Institute of Transportation Studies*, University of California, Irvine.
8. "Statware Speeds the Adoption of Web-Based Statistical Analysis to Support Enterprise-wide Quality Improvement Initiatives." Located at:
<http://www.statware.com/statware/news/August%203%202000.htm>

9. Shuler, R. "How does the Internet Work?" Located at:
http://rus1.home.mindspring.com/whitepapers/internet_whitepaper.html
10. Bhadha, Raymond J. (1998) "Java Technology in Transportation - The LA County-Caltrans freeway Traffic Intranet display." *Institute of Transportation Engineers*, University of California, Berkeley.
11. Barrett, C., S. T. Waller, and A. K. Ziliaskopoulos (1999) "VISTA - Visual Interactive System for Transportation Algorithms." Department of Civil Engineering, Northwestern University, Evanston, IL.
12. Lee, Seokhoon (1997) "Toward the next generation - Traffic Information System using Internet." *IEEE Computer Society Press*, Los Alamitos, CA.
13. Mtenga, P. V., J. Sobanjo, L., and K. Spainhour (1999) "Multicriteria DSS with historical database for attenuator selection." *Journal of Computing in Civil Engineering*, ASCE, Reston, VA.
14. Chang, B. P, J. Chen, and B. Ran (1999) "Architecture development for Web-based Geographic Information System applications in Transportation." *Transportation Research Board*.
15. Evans, A., R. Kingston, S. Carver, and I. Turton (1999) "Web-based GIS used to enhance public democratic involvement" Centre for Computational Geography, School of Geography, University of Leeds. Located at:
<http://www.geog.leeds.ac.uk/research/papers/99-1/>
16. "Measuring Demand-Response Operational Performance: A Model Measuring Transit's Output and Resource Utilization", *16th National Conference on Accessible Transportation & Mobility*, March 1999.

17. Lave, R., and R. Mathias, "State of the Art of Paratransit" *Transportation Research Board.*"
18. ITS Decision Support & Planning. Located at:
http://www.path.berkeley.edu/~leap/itsdecision_resources/decision.html
19. 1998 California Transportation Plan: Transportation System Performance Measures - Final Report. Located at: <http://svhqsg4.dot.ca.gov/hq/tpp/tspm/tspm.htm>
20. Dahlgren, Joy. (1998) "Definition and Measurement of Transportation System Performance" *California PATH Research Report*, Institute of Transportation Studies, University of California, Berkeley.
21. Dahlgren, J., M. AlKadri, and R. Picallo (1998) "Decision Support Systems for Managing and Applying ITS research" *California PATH Working Paper*, Institute of Transportation Studies, University of California, Berkeley.
22. Moore, D. S., and G. P. McCabe (1999) "Introduction to the Practice of Statistics" *W.H. Freeman and Company*, New York.
23. Hummer, J. E., H. Douglas Robertson, and D. C. Nelson (1994) "Manual of Transportation Engineering Studies", *Institute of Transportation Engineers*.
24. "An Introduction to OLAP Multidimensional Terminology and Technology," *White Paper*, Accrue Software. Located at: <http://www.pilotsw.com/olap/olap.htm>
25. Baugh, J.W., J.R. Stone, S. Chakravarty, M. Surasky (2000) "The Winston-Salem Mobility Manager: Remote Data Collection, Validation and Performance Evaluation", Department of Civil Engineering, North Carolina State University, Raleigh, NC.

26. Mukerjee, S., K. Ozbay (2000) “Web-Based Expert Geographical Information System for Advanced Transportation Management Systems” *Transportation Research Board, Paper No. 00-1333*.
27. Courcoubetis, C., V. S. Siris (2000) “A Web-based tool for advanced statistical analysis of network traffic measurements”, *Foundation for Research and Technology, Greece*.
28. Spring, G. S., John Collura (1996) “Evaluation of Automated Vehicle Location Technologies used in Automatic Scheduling & Dispatching Systems for small urban areas in North Carolina”, *Urban Transit Institute, Project No. 4-47031*

Appendix A

- **Glossary of terms**

Glossary of Terms¹⁷:

2-tier Architecture - In this form of architecture, the data logic resides on the server and the presentation function on the client. The business logic however runs both in the client and the server, with each processor doing a bit of both; processing and exchanging the inter-process communication messages across the network.

3-tier Architecture - In this form of architecture, functions are separated into different presentation, business application and data components. Due to this it is more flexible and can support both changes in the business environment and changes in technology or in other words changes in the three layers.

The description 3-tier refers to different scenarios but mostly it refers to the fact that users communicate with three computers (or servers) to get and manipulate data. In normal circumstances, the model can be envisaged as:



Anova – A statistical test used to evaluate the significant difference in means over a period of more than two intervals.

Application Programming Interface (API) – It is an interface, by which an application program accesses the operating system.

Application Server – These are servers which handle operations between a user interface and a backend database mostly for complex transaction based operations. They comprise of a suite of software, which perform a multitude of functions. In other words, a program that handles all the applications and databases needed for transactions.

¹⁷ Most of the computation related definitions used have been taken from online dictionaries like the FOLDOC and Webopedia. Some of the database related definitions have been taken from *Fundamentals of Database Systems – Elmasri & Navathe*.

Automatic Vehicle Location (AVL) – A technology that uses Global Positioning System (GPS) or radio beacons to track the position of transit vehicles and transmit the positions to a central terminal. This position information can be used to show real time positions of vehicles to customers, to give customers the Estimated Time of Arrival (ETA) of transit vehicles, and to give transit managers more system performance data for analytic purposes. Other than real-time operations control, the AVL/GPS technology can be used to a great extent for improving service planning, scheduling and performance analysis practices.

Back-end – Refers to any software programming tools whose tasks are not apparent to a front user. In the context of the tools developed the term refers to the database which stores the data that is used for manipulation and analysis.

Common Gateway Interface (CGI) – A standard for running applications on a web server, for dynamic interaction with a user with forms. Common scripting languages like Perl and C can be used for the scripting.

Client-Server Architecture - This widely used model of distributed computing distributes the functions of the application to combinations of workstations, networks and shared computers. It is based on the idea of storing and processing data on a remote machine (server) and displaying it on another machine (client). This idea of a client/server model is commonly implemented in the Web, where the web browser of a computer displays documents, which are stored and processed on a remote server.

Concurrency – It is the ability of the database to be used by more than one user at the same time. Concurrency issues arise when these multiple users attempt to update the database at the same time. These concurrent transactions give rise to problems like the *lost update problem* and the *temporary update problem*.

Database – A filing system consisting of sets of persistent data for querying and manipulating.

Data Dictionary – Stores catalog information about the database schema and constraints. Other than this it also stores design decisions, usage standards, application program descriptions and user information.

Driver – A program that controls a device. In the context of the tool, an application that helps connect the JDBC programs to the database for manipulation of the data.

Decision Support Systems (DSS) – Software used to aid management decision-making.

Front-end – An interface that filters information to the rest of the applications. Usually done to make a software user friendly.

Hyper-Text Markup Language (HTML) – A language for writing text and graphical content on web pages.

Intelligent Transportation Systems (ITS) – Refers to the integrated application of modern technologies and management strategies to surface transportation systems.

Interactive Voice Response (IVR) – Any telephone-based application which interactively takes input from callers and returns output in the form of voice or auditory information. In paratransit applications, the customer calls in and is able to access general and personal transit information and book, change or cancel service requests.

Intranet – A network, that utilizes the web browsers and Internet interconnectivity to share proprietary material among its users.

Java Database Connectivity (JDBC) – An API for accessing databases from Java application.

Java Server Web Development Kit (JSWDK) – Sun's web server toolkit for developing and testing server-side Java applications.

Mobile Data Terminal (MDT) – An MDT allows information contained in base computer systems to be accessed and used by transit drivers, ensuring accuracy and saving company time in processing passengers as they board and leave the bus. MDTs send the information to the base computer to support updating customer records, invoicing, supplying progress/ status reports and many other functions. The information received and sent are text and numerical data, in the form of radio signals.

On-Line Analytical Processing (OLAP) – A category of software tools that can perform query and analysis of data. This enables analysts to gain quick insight into the data.

Relational Database Management System (RDBMS) – A type of database management systems where data is stored in the form of tables or *relations*, which in turn contain records or *tuples*.

Resin – Web server distributed by Caucho Technology, Inc.

Significance level – The significance level of a statistical hypothesis test is a minimum fixed value of probability for rejecting the null hypothesis H_0 . In other words, significance levels show how likely a result is possible due to chance. The significance level is usually denoted by α (alpha). The higher the value of α , less reliable it is as an indicator of the relation. In scientific research areas, the 5% level is a widely accepted criterion for meaningful evidence.

Servlet – A Java program which resides and executes on a server to provide functionality or processing of data on the server. According to the FOLDOC dictionary, a Servlet is “a Java program that runs as part of a network service, typically an HTTP server and responds to requests from clients.” i.e an application fragment used to code business

logic. They therefore are most commonly used for dynamic web content generation and do not directly create user interfaces.

T-test – A statistical test used to evaluate the significant difference in means between two groups.

Traffic Management Center – The center for an Advanced Traffic Management System (ATMS) to facilitate the movement of persons and goods, with minimum delay by maximizing throughput and minimizing impact of incidents.

Variance – Variance is a measure of how spread the distribution is. It is given by the formula

$$s = \frac{\sum(x - m)^2}{N}$$

Web Server (or HTTP Server) – A server that helps in receiving and sending requests from remote web browsers. Some common web servers are IIS, Apache and Resin.

World Wide Web (WWW) – A system of servers to support specially formatted text, graphics and multimedia files that are written in HTML. They are viewed through applications called web browsers.

Appendix B

- *Sample reports*
- *Database screenshots*
- *Sample SQL queries*
- *Sample Java servlet code*

1. Sample Reports

Figure 1(a): Sample Trip/Miles Productivity report

Trips/Miles Productivity Report

Page No. 2
 Date: 01/19/00
 Time: 10:47:26
 From: 11/01/99 To: 11/30/99

Kte #	Date of	Route	DOW	Trips	Pass.	Pass/Trip	Total Miles	Pass Miles	Service Miles	Calc. DeadHead	Revenue Miles	Actual DeadHead	Pass/ Srvc. Mile	Pass/ Rev. Mile	Pass/ Tot. Mile	Pass Ave Trp Length
	11/15/99	MON		395	397	1.0	1735.6	4047.2	1649.2	86.4	1248.4	487.2	0.24	0.32	0.23	12.2
	11/16/99	TUE		389	389	1.0	1612.0	6089.9	1562.4	49.4	1259.1	352.9	0.25	0.31	0.24	15.7
	11/17/99	WED		381	381	1.0	1674.1	5538.1	1583.5	90.5	1272.5	401.6	0.24	0.30	0.23	14.5
	11/18/99	THU		380	382	1.0	1555.7	5159.2	1485.1	70.6	1181.7	374.0	0.26	0.32	0.25	13.5
	11/19/99	FRI		384	384	1.0	1580.7	4992.5	1507.4	73.4	1215.3	365.4	0.25	0.32	0.24	13.0
	11/20/99	SAT		98	98	1.0	277.7	881.1	229.7	48.0	219.8	57.9	0.43	0.45	0.35	9.0
	11/22/99	MON		407	407	1.0	1807.5	4776.5	1731.6	75.9	1314.3	493.3	0.24	0.31	0.23	11.7
	11/23/99	TUE		390	390	1.0	1674.6	5661.9	1610.7	63.9	1206.5	468.2	0.24	0.32	0.23	14.5
	11/24/99	WED		390	390	1.0	1841.1	5204.7	1642.0	199.1	1268.4	572.7	0.24	0.31	0.21	13.3
	11/25/99	THU		99	99	1.0	299.8	1442.4	261.0	38.8	243.8	55.9	0.38	0.41	0.33	14.6
	11/26/99	FRI		141	141	1.0	570.2	1371.6	514.0	56.3	423.4	146.8	0.27	0.33	0.25	9.7
	11/27/99	SAT		101	101	1.0	309.4	1280.7	268.8	40.6	254.3	55.1	0.38	0.40	0.33	12.7

Figure 1(b): Sample Provider Productivity report

Provider Productivity Report For Date : 2/1/00 To 2/29/00																									
		Count Passengers					Count Of Trips			Hours								Mileage							
Date	Provider	CLI	PCA	ESC	Others	Total	Trips	NS	Cancel	Total	Service	Rev.	DH	Break	Lunch	Refuel	Others	Total	Service	Rev.	DH	Bl.	Pass.	Trip/RH	Trip/SH
2/7/00	CAT	37	0	0	3	40	37	0	7	38.5	31.1	7.4	23.7	0.0	0.0	0.0	0.0	281.1	228.7	138.5	90.2	0.0	148.8	5.03	1.19
	TRA	393	0	0	0	393	393	2	15	118.8	110.5	82.2	28.3	0.0	0.0	0.0	0.0	1746.3	1561.2	1256.0	305.2	0.0	5067.1	4.78	3.56
		430	0	0	3	433	430	2	56	157.3	141.6	89.6	52.0	0.0	0.0	0.0	0.0	2027.4	1789.9	1394.5	395.4	0.0	5215.9	9.81	4.75
2/8/00	CAT	41	0	0	4	45	41	0	7	47.1	45.1	9.7	35.4	0.0	0.0	0.0	0.0	383.3	331.2	199.6	131.7	0.0	227.2	4.21	0.91
	TRA	381	0	0	0	381	381	6	13	118.3	111.7	78.4	33.3	0.0	0.0	0.0	0.0	1825.5	1670.8	1219.9	450.9	0.0	5460.7	4.86	3.41
		422	0	0	4	426	422	6	60	165.4	156.8	88.1	68.7	0.0	0.0	0.0	0.0	2208.8	2002.0	1419.5	582.6	0.0	5687.9	9.07	4.32
2/9/00	CAT	42	0	0	3	45	42	0	6	46.6	39.6	9.5	30.1	0.0	0.0	0.0	0.0	356.4	301.3	183.8	117.5	0.0	213.7	4.43	1.06
	TRA	392	0	0	0	392	392	0	17	108.7	102.1	79.1	23.0	0.0	0.0	0.0	0.0	1732.8	1555.3	1206.8	348.5	0.0	5547.5	4.96	3.84
		434	0	0	3	437	434	0	65	155.3	141.7	88.6	53.1	0.0	0.0	0.0	0.0	2089.2	1856.6	1390.6	466.0	0.0	5761.2	9.39	4.90
2/10/00	CAT	34	0	0	2	36	34	0	11	44.6	42.8	7.9	34.9	0.0	0.0	0.0	20.2	335.7	288.5	159.1	129.4	28.3	179.4	4.30	0.80
	TRA	378	0	2	2	382	378	4	16	117.2	110.7	80.3	30.4	0.0	0.0	0.0	0.0	1747.6	1546.9	1204.3	342.6	0.0	5941.7	4.71	3.41
		412	0	2	4	418	412	4	77	161.8	153.5	88.2	65.2	0.0	0.0	0.0	20.2	2083.3	1835.4	1363.4	472.0	28.3	6121.1	9.01	4.21
2/11/00	CAT	32	0	0	3	35	32	0	16	41.7	40.0	7.5	32.5	0.0	0.0	0.0	0.0	291.9	249.6	137.4	112.2	0.0	150.2	4.28	0.80
	TRA	351	0	2	0	353	351	2	13	108.2	100.0	77.4	22.7	0.0	0.0	0.0	0.0	1617.8	1416.0	1113.4	302.7	0.0	5119.7	4.54	3.51
		383	0	2	3	388	383	2	81	149.9	140.0	84.9	55.2	0.0	0.0	0.0	0.0	1909.7	1665.6	1250.8	414.9	0.0	5269.9	8.82	4.31
2/12/00	CAT	16	0	0	0	16	16	0	17	15.9	15.2	2.6	12.6	0.0	0.0	0.0	0.0	102.4	85.6	50.4	35.2	0.0	50.4	6.15	1.05
	TRA	95	0	0	0	95	95	0	2	21.9	19.8	18.2	1.6	0.0	0.0	0.0	0.0	306.6	250.4	250.4	0.0	0.0	1137.6	5.22	4.80
		111	0	0	0	111	111	0	30	37.8	35.0	20.8	14.2	0.0	0.0	0.0	0.0	409.0	336.0	300.8	35.2	0.0	1188.0	11.37	5.85

Figure 1(c): Sample Ride File

BUS	LAB	RIDERS	EQUIP	ID	CALL	ETA	ETACAL	APPTIME	PC	ADDRESS	APARTMENT	CITY	BUSCOLOR	DATESTAMP	ADDOACTIV	COLLECT	XCOORD	YCOORD	ROUTER	VEHICLE	DRIVER	COMMENTS	DI	CAPCALC
100	10001900	1	AM	1000		11:40	11:40	12:00		+ 1629 N CHESTNUT ST		WINSTON-SALEM	10	102			80243791	36115436						
100	10002300	1	AM	1000			11:58	12:00		- 0 UNIVERSITY PKWY		WINSTON-SALEM	25	102			80277343	36171259						
100	10003800	1	AM	1001		20:30	20:30			+ 403 S HAWTHORNE RD		WINSTON-SALEM	10	102			80298921	36088330				MAIN TOWER		
100	10003900	1	AM	1001			20:39			- 1000 E 4TH ST		WINSTON-SALEM	25	102			80233823	36089536				WONDERLAND ACADEMY		
100	10003700	1	AM	1002		20:45	20:39			+ 1000 E 4TH ST		WINSTON-SALEM	10	102			80233823	36089536				WONDERLAND ACADEMY		
100	10003900	1	AM	1002			20:50			- 13 SALEM GARDEN RD	E	WINSTON-SALEM	25	102			80185743	36081376						
100	10002800	1	AM	1003		16:55	16:42	17:30		+ 184 STANLEY AVE		WINSTON-SALEM	10	102			80189975	36127374				SCHEDULE VARIES 5PM-4AM		
100	10002900	1	AM	1003			17:05	17:30		- 113 CASSBELL ST		WINSTON-SALEM	25	102			80236540	36055566						
100	10003000	1	AM	1004		17:00	17:30			+ 1900 REYNOLDA RD		WINSTON-SALEM	10	102			80279890	36119289				GRAYLYN CONF CEN		
100	10003100	1	AM	1004			17:29			- 1121 E 24TH ST		WINSTON-SALEM	25	102			80232777	36121049						
100	10001500	1	AM	1005		09:30	09:29	09:45		+ 727 VOSS ST		WINSTON-SALEM	10	102			80232231	36181730				ONLY ON SUNDAY		
100	10001700	1	AM	1005			09:45	09:45		- 200 Brookstone Ave		Winston Salem	25	102			80259195	36097957						
100	10003400	1	AM	1006			18:15	18:31		+ 200 Brookstone Ave		Winston Salem	10	102			80259195	36097957						
100	10003500	1	AM	1006			18:48			- 727 VOSS ST		WINSTON-SALEM	25	102			80232231	36181730				ONLY ON SUNDAY		
100	10004600	1	AM	1007		23:45	23:45			+ 425 N Cherry St		Winston Salem	10	102			80241138	36093471						
100	10004700	1	AM	1007			23:52			- 519 LIBERIA ST		WINSTON-SALEM	25	102			80236503	36083678						
100	10004800	1	AM	1008		24:15	24:16			+ 519 LIBERIA ST		WINSTON-SALEM	10	102			80235503	36083678						
100	10004900	1	AM	1008			24:25			- 1109 CHARLES ST		WINSTON-SALEM	25	102			80213375	36055729						
100	10002900	1	AM	1008		15:00	15:00			+ 7910 N Point Blvd		Winston Salem	10	102			80278967	36193827						
100	10002700	1	AM	1009			15:14			- 2615 PENCLETON DR	APTS	WINSTON-SALEM	25	102			80222375	36124536						
100	10004400	1	AM	1010		22:30	22:30	23:00		+ 708 POWELL ST		WINSTON-SALEM	10	102			80234754	36080242						
100	10004500	1	AM	1010			22:42	23:00		- 3333 SILAS CRESH PKWY		WINSTON-SALEM	25	102			80286738	36070541						
100	10003200	1	AM	1011		18:00	18:00			+ 3964 UNIVERSITY PKWY		WINSTON-SALEM	10	102			80274731	36172068						
100	10003300	1	AM	1011			18:21			- 505 BARRY CT	B	WINSTON-SALEM	25	102			80191990	36193448						
100	10004000	1	AM	1012		21:05	21:04	21:20		+ 0 UNIVERSITY PKWY		WINSTON-SALEM	10	102			80277343	36171259						
100	10004100	1	AM	1012			21:20	21:20		- 1629 N CHESTNUT ST		WINSTON-SALEM	25	102			80243791	36115436						
100	10004200	1	AM	1013		22:20	22:00	11:00		+ 582 FREEDOM CT	A	WINSTON-SALEM	10	102			80224962	36052635				10:50PM (3RD SHIFT) (AM PICKUPS ONLY ON WEEKENDS)		
100	10004300	1	AM	1013			22:26	11:00		- 901 BETHESDA RD		WINSTON-SALEM	25	102			80294308	36079509						
100	10001800	1	AM	1014		11:30	11:30			+ 13 SALEM GARDEN RD	E	WINSTON-SALEM	10	102			80185743	36081376						
100	10002000	1	AM	1014			11:42			- 1000 E 4TH ST		WINSTON-SALEM	25	102			80233823	36089536				WONDERLAND ACADEMY		
100	10002100	1	AM	1015		11:45	11:48	12:00		+ 1000 E 4TH ST		WINSTON-SALEM	10	102			80233823	36089536				WONDERLAND ACADEMY		
100	10002200	1	AM	1015			11:54	12:00		- 403 S HAWTHORNE RD		WINSTON-SALEM	25	102			80298921	36088330				MAIN TOWER		
100	10001900	1	AM	1016		08:05	08:05			+ 1121 E 24TH ST		WINSTON-SALEM	10	102			80232777	36121049						
100	10001100	1	AM	1016			08:14			- 1900 REYNOLDA RD		WINSTON-SALEM	25	102			80279890	36119289				GRAYLYN CONF CEN		
100	10001200	1	AM	1017		08:40	08:40	15:00		+ 2615 PENCLETON DR	APTS	WINSTON-SALEM	10	102			80222375	36124536						
100	10001300	1	AM	1017			08:54	15:00		- 7910 N Point Blvd		Winston Salem	25	102			80278967	36193827						
100	10001400	1	AM	1018		09:00	09:09			+ 505 BARRY CT	B	WINSTON-SALEM	10	102			80191990	36193448						
100	10001600	1	AM	1018			09:30			- 3964 UNIVERSITY PKWY		WINSTON-SALEM	25	102			80274731	36172068						
100	10002800	1	AM	1019		08:20	08:20			+ 2815 PIEDMONT CIR		WINSTON-SALEM	10	102			80229031	36123967						
100	10003900	1	AM	1019			08:33			- 809 WELLINGTON RD		WINSTON-SALEM	25	102			80297376	36112741						
100	10002400	1	AM	1020		13:30	13:10	13:20		+ 809 WELLINGTON RD		WINSTON-SALEM	10	102			80297376	36112741						
100	10002500	1	AM	1020			13:23	13:20		- 2815 PIEDMONT CIR		WINSTON-SALEM	25	102			80229031	36123967						
0	0	1	AM	1021		04:00	04:00	05:00		+ 113 CASSBELL ST		WINSTON-SALEM	25	102			80236540	36055566						C
0	0	1	AM	1021			04:00	05:00		- 184 STANLEY AVE		WINSTON-SALEM	10	102			80189975	36127374				SCHEDULE VARIES 5PM-4AM		C
0	0	1	AM	1022		07:00	07:00	07:30		+ 901 BETHESDA RD		WINSTON-SALEM	10	102			80294308	36079509						C

Figure 1(d): Sample Validated File of a Month – Mar 2000

Date	Passenger	Trips	NS	Cancel	TotalHours	ServiceHours	RevHours	TotalMileage	ServiceMileage	RevMileage	PassengerMileage	TripsPerRH	TripsPerSH
3/1/00	384	384	3	0	159	108.1	80.8	1812.7	1594	1179.4	4012.7	4.75	3.55
3/2/00	355	353	2	0	103.2	94.3	75.1	1678.3	1511.9	1174.3	4995.3	4.7	3.74
3/3/00	343	341	0	0	107.4	98.2	69	1565.7	1373.9	1073.9	4105	4.94	3.47
3/4/00	89	89	0	0	21.7	19.2	17.8	324.9	258.2	258.2	1131.7	5	4.64
3/6/00	389	389	1	0	119.7	113.1	83.6	1896	1732.2	1317.8	4663.3	4.65	3.44
3/7/00	367	367	0	0	107.8	103.3	78.9	1785.2	1655.8	1241.2	4845.5	4.65	3.55
3/8/00	400	400	2	0	112.6	104.4	81.5	1762.9	1587.5	1224.1	4550.4	4.91	3.83
3/9/00	383	380	0	0	124.3	117.2	82.1	1863.4	1663.2	1286.3	5762.7	4.63	3.24
3/10/00	348	346	0	0	99.3	92.1	70.3	1501.6	1306.2	1041.6	4698.9	4.92	3.76
3/11/00	93	93	0	0	23.1	20.7	19.1	322.8	254.6	254.6	1143.4	4.88	4.5
3/13/00	347	347	7	0	105.4	99.1	72.5	1542	1374.3	1097.1	4346.8	4.79	3.5
3/14/00	369	369	0	0	110.6	104.8	80.2	1726.6	1569.2	1214.7	4562.7	4.6	3.52
3/15/00	374	374	3	0	97	91.6	72.7	1471.9	1330.8	1050.3	4716	5.14	4.08
3/16/00	404	402	0	0	113.1	106.1	81	1805.8	1611.2	1256.1	5357.2	4.96	3.79
3/17/00	353	353	1	0	107.7	99.2	73	1658.4	1455	1137.4	5031.3	4.83	3.56
3/18/00	96	96	0	0	23.2	20.7	19.3	329.8	261.1	261.1	1259.5	4.97	4.64
3/20/00	377	376	1	0	115.5	107.1	78	1777.6	1582.5	1218.1	4495.6	4.82	3.51
3/21/00	385	385	4	0	109	104.1	77.8	1636.1	1501.9	1170.6	5252.4	4.95	3.7
3/22/00	400	400	1	0	118	110.9	82.5	1772.6	1581.9	1208.5	4503.6	4.85	3.61
3/23/00	384	382	2	0	113	105.6	76.2	1838.7	1637.7	1230.6	4673.6	5.01	3.62
3/24/00	327	327	7	0	96.4	89.3	65.1	1446.2	1257.8	975.3	4383.9	5.03	3.66
3/25/00	95	95	0	0	23.8	21.3	19.9	379.7	312.8	312.8	1569.2	4.78	4.46
3/27/00	413	413	5	0	115.5	108	86.4	1861.5	1693.1	1349.7	6004.7	4.78	3.82
3/28/00	376	376	1	0	111.9	106.2	80.1	1704.2	1535.3	1217.3	5467.6	4.69	3.54
3/29/00	398	397	3	0	115.6	109.4	84	1739.9	1572	1269.4	5156.2	4.73	3.63
3/30/00	359	357	0	0	97.7	91	70.2	1617.6	1440.1	1142	5461.9	5.09	3.92
3/31/00	343	343	2	0	98.2	90.4	68.8	1431.3	1224.2	987.5	4523.5	4.99	3.79
Avg	331.5	330.9	1.7	0.0	98.1	90.2	68.4	1490.9	1328.8	1042.6	4321.3	4.9	3.8

Figure 1(e): Sample Monthly Service Report – June 2000

10/13/00
17:08:11
Page No: 1

MONTHLY SERVICE REPORTS

06/01/00 thru 06/30/00

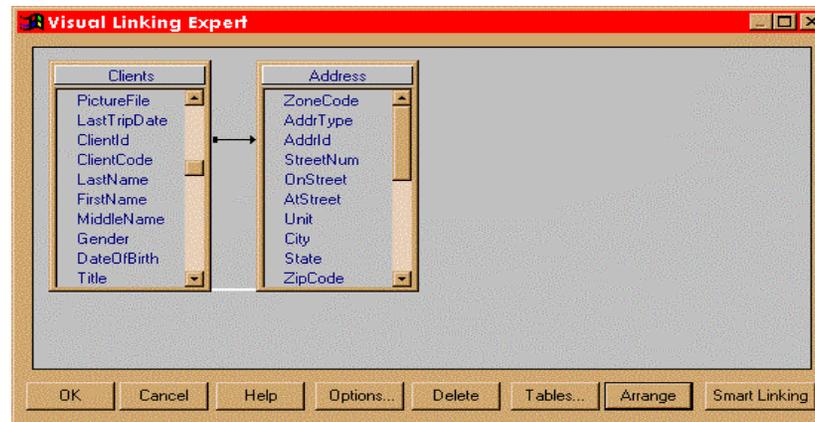
PASSENGER CLASS	WEEKDAY	SATURDAY	SUNDAY	TOTAL
	47	0	0	47
01	1241	0	0	1241
02	1400	94	0	1494
03	4820	130	0	4950
31	43	0	0	43
32	674	111	0	785
34	22	0	0	22
42	16	0	0	16
CAT	685	56	51	792
EDBLI	18	0	0	18
EDREC	72	0	0	72
EDSS	30	0	0	30
LFL	6	0	0	6
T	9	0	0	9
Total:	9083	391	51	9525
PROVIDER				
AT	739	56	51	846
RA	8344	335	0	8679
Total:	9083	391	51	9525
PURPOSE				
	633	76	0	709
10	786	0	0	786
540	1253	92	0	1345
560	1083	0	0	1083
561	84	0	0	84
562	48	0	0	48
563	668	0	0	668
564	1207	0	0	1207
CAT	685	56	51	792
DIA	1048	129	0	1177
MED	1394	38	0	1432
REC	155	0	0	155
SSS	39	0	0	39
Total:	9083	391	51	9525
VEHICLE HOURS				
CAT	213.23	77.93	77.00	368.17
RA	2491.97	81.90	0.00	2573.87
Total:	2705.20	159.83	77.00	2942.03

2. Database Screenshots

Figure 2(a): Screenshot of the Prototype Database table containing MOE data

Month	Year	Trips Per Day	Pass Per Day	Total Miles	Pass Per TotalMiles	Pass Per RevMiles	Pass Per SerMiles	Vehicle Hours	Cancel
1	1998	249.07	293.55	170120.7	0.041	0.098	0.0668432234010725	2457	
2	1998	277.48	283.68	296534	0.024	0.065	0.0408244379387185	2488	
3	1998	279.92	279.96	114988.1	0.065	0.099	0.0703681879819515	2629	
4	1998	282.07	284.39	465291.8	0.015	0.019	0.0163539097544635	2526	
5	1998	264.26	267.96	174218.2	0.039	0.087	0.0541820164233648	2000	
6	1998	289.07	303.07	311685.6	0.025	0.327	0.251197011138101	2587	
7	1998	321.28	324.35	180169.1	0.05	0.382	0.119523400982065	2554	
8	1998	317.58	318.25	111111.1	0.034	0.008	0.00620451656307774	2258	
9	1998	339.8	340	737614.9	0.009	0.299	0.24338115738989	2277	
10	1998	319.92	320.19	490967.4	0.017	0.098	0.0671213990276488	2498	
11	1998	319.69	320.07	34945.1	0.238	0.317	0.260458210316278	2574	
12	1998	302.46	302.66	41492.1	0.218	0.29	0.22913727960174	2796	
0	0							0	

Figure 2(b): Sample paratransit Table from the Trapeze database



3. Sample Charts

Figure 3(a): Passengers per day

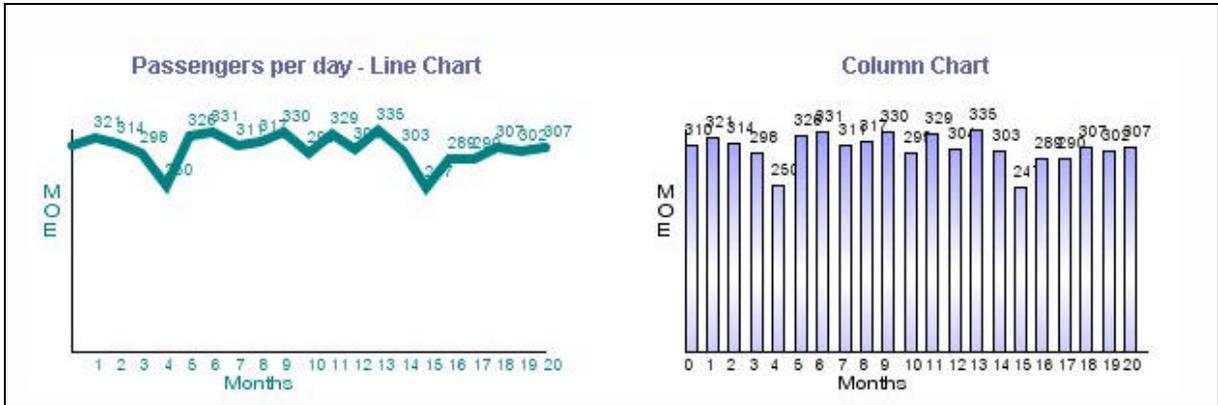


Figure 3(b): Trips per day

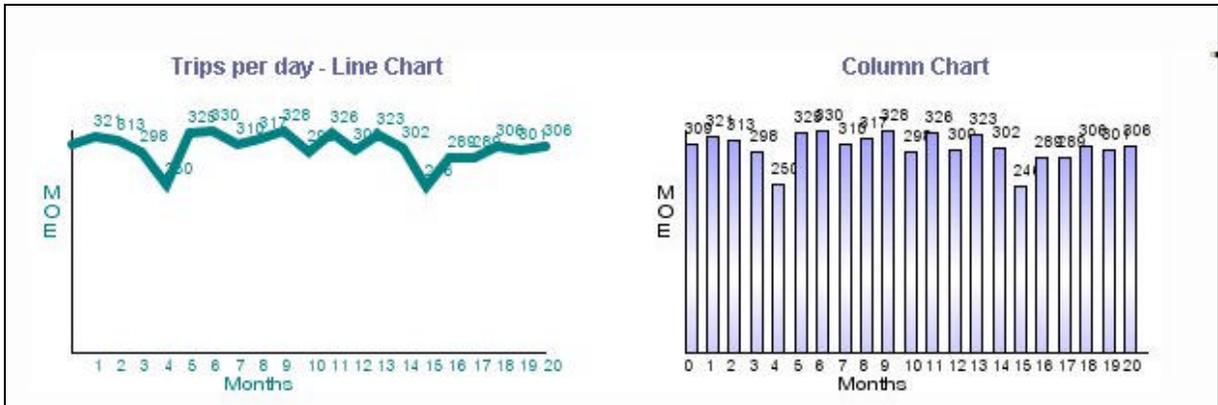
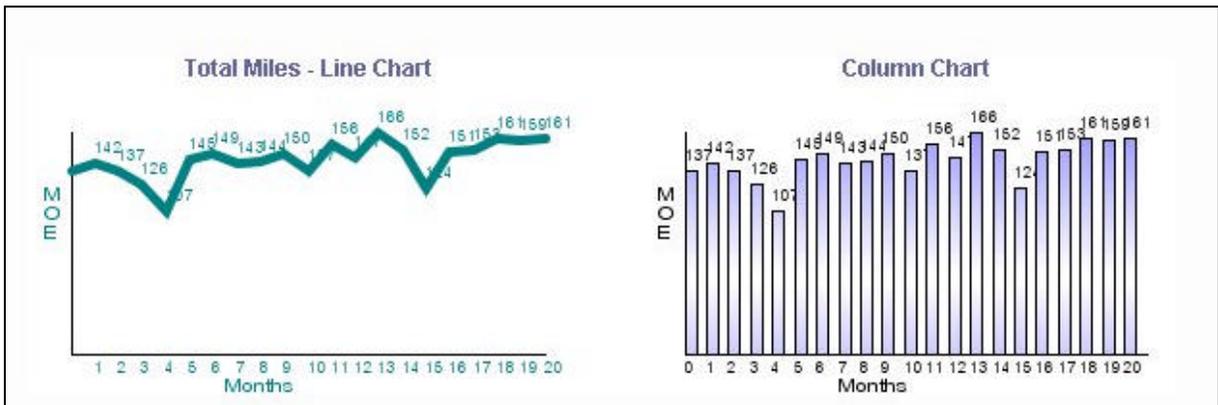


Figure 3(c): Total Miles



4. Examples of SQL commands

1. SQL query to extract MOE data from the Access database for a particular time period and ordered by dates:

```
SELECT query FROM MOES WHERE MOES.Date >= # " + FromMonth + "/" + FromYear + " AND MOES.Date <= # " + ToMonth + "/" + ToYear + " ORDER BY MOES.Date;
```

where, *query* = List of MOEs checked off on the main menu.

MOES= name of the table in the database which contains the relevant MOE data.

MOES.Date=the attribute Date in the table MOES.

FromMonth, *FromYear*, *ToMonth*, *ToYear*=dates extracted from the main menu

choices.

2. SQL query to obtain the number of tuples returned in the query to extract MOE data for a particular time period and ordered by dates:

```
SELECT count (*) FROM MOES WHERE MOES.Date >=# " + FromMonth + "/" + FromYear + " AND MOES.Date <#" + ToMonth + "/" + ToYear + "#;
```

5. Sample Java Servlet code for extracting data from the database

```
import java.io.*;
import java.text.*;
import java.util.*;
import javax.servlet.*;
import javax.servlet.http.*;
import java.net.*;
import java.sql.*;
import java.io.*;
import java.lang.*;

public class TestServlet extends HttpServlet {

    Connection con = null;

    public void doGet(HttpServletRequest req, HttpServletResponse res) throws ServletException, IOException
    {
        dbConnectAccess dbc = new dbConnectAccess();
        Connection con = dbc.getConnect();

        System.out.println("Inside Inbox just after getting connection");

        res.setContentType("text/html");
        PrintWriter out = res.getWriter();

        out.println("<html><title>Test Servlet</title><head>");

        Statement stmt = null;
        ResultSet result = null;
```

```

String MOE[]= new String[5];

for (int i=0;i<5 ;i++ )
{
    MOE[i] = (String) req.getParameter("MOE"+i);
}

String FromMonth = (String) (req.getParameter("FromAnalyzeMonth").trim());
String ToMonth = (String) (req.getParameter("ToAnalyzeMonth").trim());
String FromYear = (String) (req.getParameter("FromAnalyzeYear").trim());
String ToYear = (String) (req.getParameter("ToAnalyzeYear").trim());

try
{
    String query="";
    int p=0, counter=0, m, n;
    do
    {
        if ((MOE[p]!=null) && (counter==0))
        {
            ++counter;
            query=query+MOE[p];
        }
        else if (MOE[p]!=null)
        {
            ++counter;
            query=query+" "+MOE[p];
        }
        p++;
    }
}

```

```

        m=counter;
    }
    while (p<5);
    System.out.println("m="+m);
    System.out.println(("query =" +query));

    String index[] = new String[m];
    int q=0, counter2=0;
    do
    {
        if (MOE[q]!=null)
        {
            index[counter2]=MOE[q];
            ++counter2;
        }
        q++;
    }
    while (q<5);

    // ***** QUERY *****
    String SQLQuery, SQLQuery2;

    int FM=0, TM=0;
    int FY=0, TY=0;
    int j=0;

    if ( (FromMonth.equals("Month")) || (ToMonth.equals("Month")) )
    {
        out.println("<font face='arial, sans-serif' size=3 color=#CC0033 style='font-weight:BOLD'>");
        out.println("Please choose correct values for the Months");
    }

```

```

else if ((FromMonth!="Month") && (ToMonth!="Month"))
{

    FM = Integer.parseInt(FromMonth);
    TM = Integer.parseInt(ToMonth);
    FY = Integer.parseInt(FromYear);
    TY = Integer.parseInt(ToYear

    System.out.println("FM="+FM+", FY="+FY);
    System.out.println("TM="+TM+", TY="+TY);

    String tempo = Integer.toString(FM);
    System.out.println("String FM="+tempo);

    SQLQuery="Select "+query+" from MOES where MOES.Date_Month>=#" + FromMonth+"/1/" +
FromYear.substring(2,4) + "# AND MOES.Date_Month<=#" + ToMonth+"/1/" +ToYear.substring(2,4) + "# order by
MOES.Date_Month";
    System.out.println(SQLQuery);
    result = dbc.executeQuery(SQLQuery);
}

out.println("<font face='veranda, courier, arial, sans-serif' size='-6' style='font-weight:BOLD'>");

out.println("<table width='98%' border='1' bordercolor='black' CELLSPACING=0 CELLPADDING=1>");
out.println("<caption><b> MOE data:</b> From "+FromMonth+"-01-"+FromYear+" to "+ToMonth+"-01-
"+ToYear+"</caption>");

out.println("<FORM name = ChooseStatForm method = POST ACTION = 'http://152.1.55.114:8080/servlet/
menu2ServletRefresh' TARGET='Menu2' onSubmit='>");

```

```

SQLQuery2 = "Select count(*) from MOES where MOES.Date_Month>=#" + FromMonth + "/1/" +
FromYear.substring(2,4) + "# AND MOES.Date_Month<#" + TM + "/1/" + ToYear.substring(2,4) + "#";
ResultSet result2 = dbc.executeQuery(SQLQuery2);
while (result2.next())
{
j = Integer.parseInt(result2.getString(1));
}

String MOEValue[][]=new String[j+1][m]; //-----stats

if (MOE[0]==null && MOE[1]==null && MOE[2]==null && MOE[3]==null && MOE[4]==null)
{
out.println("Please choose an MOE before pressing Show Data");
}

else if ( (FM<9) && (FY<2000) )
{
out.println("<font face='arial, sans-serif' size=3 color=#CC0033 style='font-weight:BOLD'>");
out.println("Please choose a start date on or after the month of <b>September, 1999</b>");
}

else if ( (TM>5) && (TY>2000) )
{
out.println("<font face='arial, sans-serif' size=3 color=#CC0033 style='font-weight:BOLD'>");
out.println("Please choose an end date on or before the month of <b>May, 2001</b>");
}

else if ( (FM>TM) && ((FY>TY) || (FY==TY)) )
{
out.println("<font face='arial, sans-serif' size=3 color=#CC0033 style='font-weight:BOLD'>");
}

```

```

        out.println("Please choose an end date on or before the start date");
    }

else if ( (FM<TM) && (FY>TY) )
    {
    out.println("<font face='arial, sans-serif' size=3 color=#CC0033 style='font-weight:BOLD'>");
    out.println("Please choose an end date on or before the start date");
    }
else
    {
    int MonthNo=Integer.parseInt(FromMonth);

    out.println("<tr bgcolor=#99cccc>");
    out.println("<td align='center' width='10'><b> Month </b></td>");

    for (int z=0;z<m ;z++ )
    {
        System.out.println("MOE is =" +index[z]);
        out.println("<td align='center' width='10'><b><a href='http://152.1.55.114:8080/' + index[z] + ".html"
target='TrendWindow'>" +index[z]+"</a></b></td>");
        out.println("<INPUT TYPE=hidden NAME=moe"+z+" VALUE="+index[z]+">");
    }

    out.println("<tr><td align='center' bgcolor=#99cccc' width='10'>Choose to Analyze</td>");
    for (int z=1;z<m+1 ;z++ )
    {
        out.println("<td align='center' width='10'><b><INPUT TYPE='radio' NAME='ChooseStat'
VALUE="+z+"></td>");
    }
    out.println("</tr>");
    out.println("</tr>");

```

```
//-----beginning of initializing 'MOEValue' array
int counterMonth=1;
String MonthName="";
while(result.next())
{
    if (MonthNo==1)
        MonthName="Jan";
    else if (MonthNo==2)
        MonthName="Feb";
    else if (MonthNo==3)
        MonthName="Mar";
    else if (MonthNo==4)
        MonthName="Apr";
    else if (MonthNo==5)
        MonthName="May";
    else if (MonthNo==6)
        MonthName="Jun";
    else if (MonthNo==7)
        MonthName="Jul";
    else if (MonthNo==8)
        MonthName="Aug";
    else if (MonthNo==9)
        MonthName="Sep";
    else if (MonthNo==10)
        MonthName="Oct";
    else if (MonthNo==11)
        MonthName="Nov";
    else if (MonthNo==12)
        MonthName="Dec";
}
```

```

out.println("<tr>");
out.println("<td align='center' bgcolor='#99cccc' width='10'><b> "+counterMonth+" </b></td>");

for (int z=1;z<(m+1) ;z++)
{
MOEValue[counterMonth-1][z-1]=result.getString(z);
int temp1=(int) (Float.parseFloat(MOEValue[counterMonth-1][z-1])*1000);
if (counterMonth%2==0)
{
out.println("<td align='center' bgcolor='#ccccff' width='10'> "+temp1/1000.0+"</td>");
}
else {
out.println("<td align='center' width='10'> "+temp1/1000.0 + "</td>");
}
String hidden= "hidden"+counterMonth+""+z;
out.println("<input type=hidden name="+hidden+" value="+MOEValue[counterMonth-1][z-1]+ ">");
System.out.println(hidden);
}
out.println("</tr>");
MonthNo++;
counterMonth++;
}

out.println("<input type=hidden name=m value="+m+">");
out.println("<input type=hidden name=counterMonth value="+counterMonth+">");

out.println("</table><br></font>");
out.println("<center><INPUT TYPE='submit' NAME='StatSubmit' VALUE='Do Analysis'> </center>
</FORM>");
}
con.close();

```

```
    }  
    catch(Exception e)  
    {  
        e.printStackTrace();  
    }  
    finally  
    {  
        if (stmt != null)  
        {  
            try  
            {  
                stmt.close();  
            }  
            catch(SQLException e)  
            {  
                e.printStackTrace();  
            }  
        }  
    }  
    out.flush();  
    out.println("</body></html>");  
    out.close();  
    }  
}
```