TCRP Web Document 23 (Project H-28): Contractor's Final Report

Uses of Archived AVL-APC Data to Improve Transit Performance and Management:

Review and Potential

Prepared for:

Transit Cooperative Research Program
TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

Submitted by:

Peter G. Furth
Northeastern University
Boston, Massachusetts

Brendon J. Hemily

Theo H. J. Muller
Delft University of Technology
Delft, the Netherlands

James G. Strathman
Portland State University
Portland, Oregon

June 2003

ACKNOWLEDGMENT

This work was sponsored by the Federal Transit Administration (FTA) in cooperation with the Transit Development Corporation. It was conducted through the Transit Cooperative Research Program (TCRP), which is administered by the Transportation Research Board (TRB) of the National Academies.

DISCLAIMER

The opinions and conclusions expressed or implied in the report are those of the research agency. They are not necessarily those of the TRB, the National Research Council, the FTA, the Transit Development Corporation, or the U.S. Government.

This report has not been edited by TRB.

CONTENTS

PREFACE

Project Objective[1]

In response to growing traffic congestion and consequent passenger demands for more reliable service, many transit operators are seeking to improve bus operations by investing in technology such as automatic vehicle location (AVL) and automatic passenger counters (APCs). The primary application of AVL technology has been in the area of real-time operations monitoring and control; consequently, AVL data have not typically been stored for subsequent analysis. APCs on the other hand can collect passenger activity data compatible with AVL operating data and are beginning to reach the mainstream. Further, APC data can be used for reporting and planning purposes long after opportunities for real-time use have expired. Many operators are planning, implementing, or operating AVL-APC systems.

Beyond the area of real-time operations control, AVL technology holds substantial promise for improving service planning, scheduling, and performance analysis practices. These activities have historically been hampered by the high cost of recovering operating and passenger-activity data; however, AVL systems can capture very large amounts of operating data required for performance analysis and management at a fairly low incremental cost. At the extremes, saving no data equates to missed opportunities, while archives of indiscriminately saved data grow unwieldy. Operators need effective data archiving strategies, techniques, and standard practices in order to capitalize on the potential wealth of extractable information without overburdening their data management infrastructure.

Transit providers have yet to take advantage of low-cost performance and passenger-activity data associated with AVL-APC technology. This is due, in part, to the traditional separation of operations functions (where the AVL-APC technology has been deployed) from the scheduling and service planning functions of transit organizations (where the needs for data and analysis are located). Pockets of excellent AVL-APC data management for specific applications can be found, but integrated approaches that can be applied across the industry do not exist.

Research is needed to assist transit providers in effective collection and use of AVL-APC data for service planning, scheduling, performance evaluation, and system management. This research should provide a coherent framework to coordinate operations, planning, and scheduling functions in the collection and use of AVL-APC data. In addition, this research should include identification of potential applications of the data for analyzing traditional fixed-route services as well as emerging service designs and strategies (e.g., bus rapid transit). Finally, this research should also recognize the varying levels of sophistication and size of transit providers.

Research Working Plan

The research working plan is divided into two phases. Phase I, which culminates with this report, is devoted to a review of needs, practice, and potential in relation to the use of archived AVL-APC data. Phase II focuses on the development of further guidance and tools related to archived AVL-APC data.

[1] As formulated by the project panel.

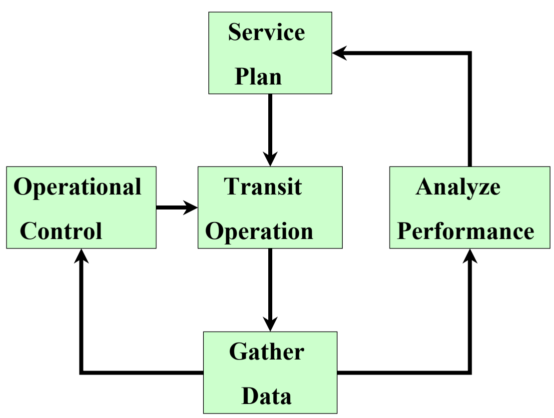
CHAPTER 1

REVIEW OF AUTOMATIC DATA COLLECTION SYSTEMS

Automatic vehicle location (AVL) and automatic passenger counting (APC) systems are capable of gathering an enormous quantity and variety of operational, spatial, and temporal data that, if captured, archived, and analyzed properly, hold substantial promise for improving service planning, scheduling, and performance analysis practices. There is agreement, however, that such data is not being used to its full potential. Most AVL systems in particular have been designed primarily for real-time applications, and often fail to capture and/or archive data items that would be valuable for off-line analysis. For both AVL and APC, technological advances have created new opportunities for improving the quantity, variety, and quality of data captured and archived, and for analyzing it in meaningful ways.

The role of automatically collected data in an overall scheme of service quality improvement is illustrated in Figure 1, which shows two quality improvement cycles: one in real-time, and one that is off-line. In the real-time cycle, automatically collected data drives operational control, aiding the transit agency in detecting and responding to deviations from the operational plan. In the off-line cycle, automatically collected data that has been archived drives analyses that aid the transit agency in evaluating and improving its operational plan. Ultimately, good operational performance and high passenger satisfaction follow from having both a good operational plan and good operational control.

Figure 1: Service Quality Improvement Cycles



Source: reference (1)

Historically, AVL design has placed most of its emphasis on the real-time cycle, often including computer-aided dispatching (CAD) tools. APC, in contrast, has always been oriented toward off-line analysis, but has not seen the same widespread adoption as AVL. This report focuses on the off-line cycle, and also examines uses of archived AVL-APC data beyond operations planning.

Chapter 1 provides a review AVL-APC systems and their ability to capture useful data. Chapter 2 examines issues surrounding the analysis of archived AVL-APC data, and offers a list of analysis and decision making tools that use, or could use, archived AVL-APC data to improve management and performance. Based on an analysis of the relationship between the effective use of archived data and the design of automated data collection systems, database archives, analysis tools, and organizational issues, Chapter 3 presents findings and guidance in each of these areas. Chapter 4 identifies topics for further research.

1.1 Survey of Practice: Breadth and Depth

Six information sources were used to review of AVL, APC, and related systems with respect to their ability to capture and archive operational data, and of industry practice in using such archived data. Together, these sources provide both a broad view of historical and current practice and in depth view of practice at transit agencies in the U.S., Canada, and in Europe.

The first source was the literature on intelligent systems in transit and on transit analysis tools. A helpful starting point is a pair of state-of-practice reviews done for the USDOT's Volpe Center *(2, 3).* A second source was a mail survey of U.S. transit agencies with respect to their use of AVL and APC conducted in spring 2001 by Mr. Robbie Bain.

A third source was a wide telephone survey of APC and AVL users. From the first two sources we assembled a preliminary list of some 122 American, 14 Canadian, and 26 European transit agencies using or planning to use AVL or APC. From that list, we telephoned staff members at transit agencies reputed to be advanced in their use of AVL or APC data. We successfully conducted telephone interviews with 20 U.S. and 14 Canadian transit agencies, listed in Table 1. The telephone interviews covered the following topics:

- AVL and APC system description: supplier, age, technology, extent
- Data recording: if data is stored on board, what items are stored, and what events trigger a record
- Communication: if data isn't stored on board, what's the polling interval, and what items are included in the message
- AVL-APC database field descriptions
- Uses of the data, both current and planned
- Requests for samples of reports generated from the data

The fourth source was in depth case studies at nine transit agencies in the U.S., Canada, and the Netherlands, listed in Table 2. The case study reports themselves are attached as appendices. In addition, a partial case study was conducted of Uestra, the transit agency in Hannover, Germany.

The nine case study agencies provide a broad range in many respects. They represent the U.S., Canada, and Europe. Within the U.S., they span the East Coast, Midwest, and West Coast, and include agencies whose operations are statewide, metropolitan, and limited primarily to a central city. Some have focused on AVL, others on APC, others on event recorders, and some on two or more of these functions. They represent a range of vendors, system design, and system age. Some of the selected agencies have well-established practice with archived data, with impacts throughout the organization; others are still in the development stage. These agencies have had failures as well as successes, and we learn from both.

The fifth source of information was a one-day workshop for vendors held on May 28, 2002. An open invitation was extended to vendors of AVL, APC, and related products, with specific invitations sent to known vendors. They were joined by panel members, representatives of several of the case study sites, and members of the project team, providing a good representation of interested transit agency staff and independent researchers. Participants other than project team members are listed in Table 3. We also benefited from direct interaction with vendors referred to us by transit agency staff from the case study sites.

Finally, we used the team's knowledge of the transit industry and related industries, supplemented by information received from members of the project panel.

TABLE 1: Agencies Interviewed in Wide Telephone Survey

STATE / PROVINCE	CENTRAL CITY	AGENCY
CA	San Jose	Santa Clara Valley Transportation Authority (VTA)
CO	Denver	Regional Transportation District (RTD)
FL	Broward County	Broward County Transit
FL	Orlando	Central Florida Regional Transportation Authority (LYNX)
IA	Des Moines	Metro Transit Authority (MTA)
IA	Sioux City	Sioux City Transit
IL	Chicago	Chicago Transit Authority (CTA)
IL	Chicago suburbs	Pace
MA	Cape Cod	Cape Cod Regional Transit Authority (CCRTA)
MD	Baltimore / state	Maryland Transit Administration (MTA)
MI	Ann Arbor	Ann Arbor Transportation Authority (AATA)
MN	Minneapolis	Metro Transit
MO	Kansas City	Kansas City Area Transportation Authority (KCATA)
NJ	New Jersey (statewide)	NJ Transit
NY	Buffalo	Niagara Frontier Transportation Authority (NFTA)
OR	Portland	Tri-Met
TX	Dallas	Dallas Area Rapid Transit (DART)
TX	San Antonio	VIA Metropolitan Transit
WA	Seattle	King County Metro
WI	Milwaukee	Milwaukee County Transit
AB	Calgary	Calgary Transit
AB	Edmonton	Edmonton Transit
BC	Vancouver	TransLink
BC	Victoria	BC Transit
MB	Winnipeg	Winnipeg Transit
NS	Halifax	Halifax Metro Transit
ON	Hamilton	Hamilton Street Railway
ON	London	London Transport Commission
ON	Ottawa	OC Transpo
ON	Toronto	Toronto Transit Commission (TTC)
QU	Hull	Société de Transport de L'Outaouais (STO)
QU	Montreal	Société de Transport de Montréal (STM)
QU	Montreal South Shore	Société de Transport de la Rive Sud de Montréal (STRSM)
QU	Quebec	Société de Transport de la Communaute Urbaine de Quebec (STCUQ)

TABLE 2: Case Study Sites

CENTRAL CITY	AGENCY
New Jersey (statewide)	NJ Transit
Chicago	CTA
Minneapolis	Metro Transit
Portland, OR	Tri-Met
Seattle	King County Metro
Montreal	STM
Ottawa	OC Transpo
Eindhoven, the Netherlands	Hermes
The Hague, the Netherlands	HTM

Our first four sources pertain to transit agencies that are using, or have begun planning for using, archived AVL and APC data. Such agencies naturally have the strongest motivation to think of ways to use that data. In addition, both in the vendor workshop and as a project team we thought beyond current practice to explore the further potential of technology to capture and archive data and of further potential uses for the data.

TABLE 3: Vendor Workshop Participants

Vendor Representatives

Dirk van Dijl, ACIS
Alain de Chene, Infodev
Carol Yates, Orbital
George Mount, NextBus
Andreas Rackebrandt, INIT
Neil Odle, IRIS
Anil Panaghat, US Holdings
Vijay Raganath, consultant with Delaware Transit
Hershang Pandya, US Holdings
Rohit Patel, Intellect Corporation
Mike Kushner, Logic Tree

Panel Members and Friends

Jim Kemp, NJ Transit (panel chair)
Wei-Bin Zhang, Univ. of California PATH program
Fabian Cevallos, Broward County Transit
Gerald Pachucki, Utah Transit Agency
Tom Friedman, King County Metro
Kimberly Slaughter, S.R. Beard
Erin Mitchell, Metro Transit (Minneapolis)
Yuko Nakinishi, Polytechnic University
Sarah Clements, FTA
Bob Casey, USDOT Volpe Center
Stephan Parker, TRB
Eric Bruun, consultant

Case Study Site Representatives (in addition to panel members already listed)

Steve Callas, Tri-Met
Kevin O'Malley, Chicago Transit Authority
Michel Thérer, Societé de Transport de Montréal
Glenn Newman, NJ Transit

1.2 Automated Data Collection Systems: Historical Perspective

1.2.1 AVL

Okunieff's TCRP synthesis *(4)* provides an insightful description of AVL systems and review of AVL's history. Historically, AVL was developed for real-time applications such as emergency response and computer aided dispatch (CAD), and is often acquired as part of a radio system upgrade. Less expensive systems simply inform display maps and dispatchers as to where buses are. More advanced systems track buses against their schedule, and can therefore determine schedule deviation and whether a bus is off route.

In traditional AVL systems, data is transmitted through a wide area network (WAN) by radio to a central computer, where it can be used for real-time applications, and can be archived. Most applications use round-robin polling in which the central computer queries each bus in turn and then waits for its response. On the bus, real-time data (e.g., GPS coordinates, odometer count, alarm status) is held in registers, and at each poll the data is sent as a message. The polling interval depends on the number of buses being monitored, the number of radio channels available, and the message length. Radio channels, a scarce resource in metropolitan areas, are allocated in the U.S. by the FCC. To keep the polling interval short, message length is usually limited to identifiers, location parameters, event codes, and status of various alarms. At some agencies, the effort to reduce message length led to a design in which an onboard computer tracks the location against the schedule, and a bus that is on route and on schedule (usually defined as less than 5 minutes late) sends no location data, only an OK-status code. The polling interval is typically 40 to 120 s, though cases of polling intervals as short as 12 s and as long as 4 minutes have been reported.

Polling provides "location-at-time" data – i.e., the location of the bus at the arbitrary time at which it is polled. This contrasts with the much more useful "time-at-location" data, i.e., time at which a bus passes a point of interest such as a stop or timepoint, based on which schedule adherence and running time is analyzed. "Time-at-location" can further be generalized to the category "event data" – a record indicating the time and location of events of interest that might include, besides passing a location of interest, mechanical events such as door opening and closing, lift use, and crossing a speed threshold. While people have long recognized the potential for using poll data for off-line analysis, interpolating between polls to estimate time at which points of interest were passed, the inherent limitation of interpolation, along with software and data integration problems, have notoriously prevented its realization. AVL systems often archive raw data in order to permit "playback" for investigating incidents; however most historic AVL systems do not include the ability to systematically extract information such as schedule adherence and running time from poll data. The worst case is AVL systems that do not include route and schedule matching; that capability would have to be added for the data to be useful for analysis.

The inability of most AVL systems to deliver data for off-line planning analysis was a major theme of a 1988 conference sponsored by the Canadian Urban Transit Association (see, for example, *(5)*). Through the mid-nineties, this situation continued. A 1998 synthesis of practice with respect to data analysis procedures *(6)* highlights the failure of older AVL systems to provide useful data for off-line analysis. Of the 7 U.S. transit agencies surveyed that had AVL systems, only 3 used AVL data to monitor schedule adherence, and all relied entirely on manual data collection for running time data.

Illustrating AVL system's common lack of orientation to archiving data and an inexpensive way of overcoming that limitation is the recent success of an analyst at Broward County Transit *(7)*. Their AVL system archived incident messages, but not routine poll data. Because incident messages occur only on an exception basis, they could not support most running time and schedule adherence analyses. The poll messages, it turns out, were written to an Oracle database in order to provide a snapshot of the system, but that database was overwritten every two minutes. To archive the poll data, the analyst wrote a program that copies the contents of the Oracle database to a permanent database every two minutes. Now the analyst is developing ways to analyze that data in the permanent database.

In response to the demand for better schedule adherence data and to reduce the demand for voice communication, some AVL suppliers have added a second level of data communication for transmitting event records. “Events” are typically exceptions; however, often timepoint passage is treated as an event. When a bus recognizes that it’s at a timepoint, it sends a message including its own ID and the timepoint ID. These messages, carried on different data channels than the poll messages, give the transit agency a stream of “time-at-location” data that is already matched to route and schedule and is suitable for off-line analysis.

The basic reason for the failure of traditional AVL to provide useful archived data is that they weren’t designed for it because transit agencies didn’t insist on it. Many agencies bought AVL primarily for emergency response; that is most inexpensively done with a simple system that offers no matching to schedule or other operations analysis. Other procurements have called for real-time computer-aided dispatch capabilities, which include schedule matching, but placed little emphasis on off-line analysis capabilities. This failure has to do with fracturing within the transit organization, with AVL procurement often seen as primarily a radio system upgrade run by the operations control department. Departments that would have benefited from off-line analysis either didn’t realize the potential benefits or were unsuccessful in influencing the procurement. “They only ask us to provide the data,” said one AVL vendor; “what they do with it is their business.” Only a small number of transit agencies have had the expertise in house to convert their AVL data stream into a useful database.

1.2.2 APC

Valuable reviews of the history of APCs are found in reports by Levy and Lawrence *(8)*, Boyle *(9)*, and Friedman *(10)*. Unlike AVL, APCs have always been designed with archived data in mind. In spite of the emphasis their name gives to passenger use data, they have also historically been designed to collect and analyze operational performance data. Canadian transit agencies have been particularly active in exploiting APC data. OC Transpo (Ottawa), TTC (Toronto), Winnipeg Transit, Tri-Met, and King County Metro are example agencies that have long benefited from routine reports on passenger loads, running time distribution, and on-time performance.

APC systems include an onboard computer known as the APC analyzer that interprets information received from sensors and converts it into counts of passenger ons and offs. In traditional APC systems, count records are stored on board; there is no connection to the radio system. In older systems, data upload requires manual intervention and may occur weekly; in newer systems, data upload occurs automatically every night using a short-range, high-speed wireless link. In some newer APC systems that are integrated with AVL, count data is not stored at all on board, but is transmitted by radio as an event message.

A problem with many historical APC systems is that the software for data storage and analysis was developed in house or provided by an APC supplier who went out of business. Either way, considerable work was needed to migrate the data system to modern computers. Many of the older APC systems are in some stage of migrating their databases to commercial database platforms on modern operating systems, giving them the capability to develop their own analysis tools and reports.

APC has not yet seen widespread adoption due primarily to its cost and the maintenance burden it adds. Where adopted, counters are typically installed on 10 to 15 percent of the fleet. Equipped buses are rotated around the system to provide data on every route. However, technological advances that will be discussed later may make passenger counters far more common.

1.2.3 Event Recorder

Lying part way between AVL and APC is a system in which events such as door openings and closings are recorded in an onboard computer and uploaded at night. It's an APC without the passenger counter (though passenger counters can be attached); it's an AVL without the radio. If speed events are recorded, it becomes a useful device for investigating incidents. Sometimes passenger counters are installed on a subset of the fleet; if so, passenger counts simply become another kind of data record.

Though not popular in the U.S., event recording has a long history in Europe, beginning with mechanical tachometers (many of which are still in use). For more than 10 years, several Dutch and German transit systems have event recorders on the entire fleet. However, the main purpose there has traditionally been incident investigation. Vendors have typically provided no data analysis capability beyond playback. Research done at the Delft University of Technology has led to software for routine analysis of event recorder data. While it has been developed and used as part of several limited-life projects over the last 20 years, it has been in routine use at a transit agency (Hermes, in Eindhoven) only since 1996.

While event recorders have long been part of the generic "smart bus" blueprint, their adoption in the U.S. has been slow, except as a backup for radio failure. New Jersey Transit began pilot implementation of such a system in 2000. Their plan is to eventually equip the entire fleet with onboard computers for event recording, vehicle health monitoring, onboard system integration (smart bus), and operations supervision (AVL), with passenger counters on some or all of the fleet.

Recently, a new wave of event recorder system installations has begun, driven by Americans with Disabilities Act requirements to provide stop announcements. Because the core onboard system tracks location at the stop level, little additional investment is needed to record stop events; indeed, event recording is an integral part of some products because it is necessary for testing and system development. In at least one case, an agency purchased a stop announcement system that includes event recording, but cannot use the event data because they elected not to purchase a wireless upload-download link. This is another case of the tragic and all too frequent scenario in which data is captured but never seen, only overwritten day after day.

The term "Trip Recorder" is used in related industries (airlines, trucking) to refer to a device storing the latest operations data to aid in accident and security incident investigation – the familiar "black box." It is possible for an event recorder to include that function, although it requires more memory to store records every few seconds than records at every stop.

1.2.4 Standard Vehicle Devices

Transit coaches have had electronic controls and monitoring systems for many years *(11)*. For example, odometers, engine heat sensors, door switches, most destination signs, and most fareboxes operate with digital electronics. AVL, APC, and related systems have sometimes taken input from some of these systems. For example, AVL systems often included engine overheat alarm status in their messages; and both AVL and APC systems have used odometer inputs for dead reckoning. In modern buses, drivetrain system components share data by means of common data bus that can be accessed by a vehicle location data system.

1.3 Technological Advances

Rapid technological advances since about 1995 give recent AVL and APC systems greater capabilities than older systems. The best indicator of what kinds of systems we can expect to see in the future is not the "average" system in use, but the relatively small number of newer systems and older systems with major upgrades.

1.3.1 Core Technologies

Location Determination. Global positioning systems, usually operated with a differential correction for better accuracy, have made location data less expensive and almost maintenance free compared to traditional signpost systems. GPS accuracy has also improved in recent years with the elimination of Selective Availability and improved receivers.

Onboard Computers. Onboard computers have become less expensive, more compact, and more powerful, and have greater capacity for data storage. While older AVL systems used small microprocessors capable of limited functions, newer AVL systems, and all APC systems, invariably include an onboard computer or "Vehicle Logic Unit." The cost of furnishing onboard computers with the higher processing speed and great data storage capacity needed by some applications is falling.

The introduction of general capability onboard computers has led to a new development: AVL and APCs sharing a common location system. This development naturally brings down the cost of the second system, and makes data integration between the two systems easier because they share a common location reference.

Download / Upload. In some older APC systems, data upload from onboard computers involves manual labor using diskettes or a hard link to a laptop. Wireless local area networks (LANs) using a high capacity optical or RF medium have made data upload simpler – it happens automatically when the bus is refueled at the end of the day, or during the overnight as buses are garaged. Uploads (e.g., uploading revised schedules and software) are usually handled by the same system.

Radio Communication. Radio communication has become more efficient. Transit agencies can sometimes split channels, in effect doubling their capacity for data transmission. Some agencies are able to poll frequently enough, and have a long enough message, to transmit passenger counts over the air along with their location and other event data. Where it is available, cellular packet data, in which data is sent in short bursts in unused moments in a cellular telephone network, offers a high capacity medium with nearly no setup cost; however, there is an ongoing cost for transmitting data. Cellular packet data often has a delay of several seconds, which is claimed to be a drawback for some real time applications, but is certainly not a problem for data intended for off-line analysis.

Passenger Counting. New passenger counting technologies have emerged using infrared, video, overhead detection, and improved mat sensors. They are less expensive and more reliable than their predecessors, making APCs less out of reach financially, and making fleetwide penetration more attractive.

1.3.2 "Smart Bus" Design and Vehicle Area Networks

A vehicle area network is a wired network or "data bus" through which devices on a transit coach can communicate by broadcasting messages, obviating the need for direct connections between every pair of devices. Modularity is achieved by adoption of a standard for communication along that data bus. Nearly all suppliers today produce devices that comply with the J1708 family of standards *(12, 13)* published by the Society of Automotive Engineers (SAE).

The J1708 standard provides the backbone for a "smart bus" design, focused on an onboard computer that tracks location with inputs from devices such as GPS receiver, odometer, and door sensors. Other essential parts of the smart bus design are a console serving as interface to the operator and a wireless LAN port for high capacity data upload and download at the garage. By outfitting the onboard computer with sufficient memory, it will serve as an event recorder; if connected to the radio, it will run the onboard side of AVL. Connected to passenger sensors, it can host APC analyzer software, converting sensor data into passenger counts; connected to stop enunciators, it will run the stop announcement system.

To date, the only J1708 networks we know of were installed as a system; however, in principle, once the network is installed, it should be possible to add devices or substitute one J1708 compliant device for another.

System Integration. Using the smart bus architecture, suppliers have begun engaging in cooperative ventures to deliver sets of integrated devices to transit agencies. The most far-reaching is the integration of AVL and APC, taking place for several years at Tri-Met and recently at numerous transit agencies, in which the AVL vendor provides a smart bus system with a single location tracking system and sufficiently powerful onboard computer, and the APC subcontractor provides the passenger sensors and software to convert sensor data to counts. Another example is a stop announcement system vendor who provides the smart bus system, with passenger counters added as a subordinate device.

While it may seem obvious that AVL, APC, and stop announcement systems should share location systems and onboard computers, their integration is only a recent development. (In fact, they are still being implemented independently at some transit agencies that lack the smart bus foundation, with buses having multiple GPS receivers, multiple onboard computers and, in spite of efforts to avoid it, multiple operator interfaces.) In addition to the J1708 standard, integration has been spurred by two developments: the general acceptance GPS as the best basis for determining location, and the lower cost of onboard computers with sufficient processing power and memory to serve multiple functions. Using a smart bus design when procuring an AVL, APC, stop announcement, or event recording system substantially lowers the marginal cost of adding the other functions, and provides flexibility for later procurements.

1.3.3 Related Devices and Applications

Technological advances in devices and systems related to AVL and APC also hold promise for improved access to and quality of operational performance data.

Passenger Information Applications. Providing passengers information on bus location, schedule deviations, time of next bus arrival, and announcement of the next stop is a rapidly developing field. In some cases these systems "hang" off a core AVL system as part of a smart bus design; in other cases, they supply the core location system, which can be shared with other devices. The most common form of passenger information, the destination sign, is still operated manually at most agencies, although improved tracking data and algorithms allows the destination sign to be driven automatically by the location information in the "smart bus" system.

Fare Payment Devices. The traditional electronic farebox has a limited processor and limited data storage capacity, creating at most one record per one-way trip with simply a count of boardings on that trip by fare category and a count of revenue. A new development is the transactional farebox, which produces a time-stamped record for each transaction. If the fareboxes are not networked to a smart bus system, the farebox data stream can be collected in parallel with an AVL or event recorder data stream, and later matched on the basis of time. One drawback is that if the fareboxes aren't networked, their clocks will not be synchronized with the AVL or event recorder clocks.

If manufactured to be J1708 compliant and connected to a vehicle area network, a transactional farebox will broadcast the data composing each transaction so that location- and time-stamped farebox transactions can included in the event recorder's dataset.

Location-correlated farebox transactions – whether correlated by real time location stamping or in post-processing – can serve as a source of passenger use data for systems without passenger counters. While fareboxes cannot be used to track load because they don't register passengers both boarding and alighting, there are methods of estimating load based on the historical symmetry between the boardings pattern in one direction and the alightings pattern in the opposite direction *(14)*.

Transaction data in which a user ID is read from a magnetic strip card or a smart card offer the possibility of tracking linked trips and doing analysis of transfers. Knowing the pattern of where and when a particular farecard was used to enter the system allows one to make a good guess, based on round trip symmetry, of its holder's trip pattern. The viability of this approach has been demonstrated in the New York subway system *(15)* and in the multi-modal Helsinki transit system *(16)*.

As contactless smart cards penetrate the market, and providing that privacy concerns do not block their development, card readers will one day be able to count passengers alighting as well as boarding (at least those carrying smart cards). They will also make transfer data more readily available.

Trip Recorders. Trip recorders are common in the trucking and long-distance coach industry. Typically, they store the last hour's operation data including drivetrain indicators such as speed, braking, and turning, and are used to investigate accidents. Those systems often include location data from GPS, and are therefore candidates for integration with AVL and APC systems. While they are standard on transit coaches in the United Kingdom, we are not aware of their use on U.S. transit coaches.

Security Systems. Camera-based security systems are becoming popular in large U.S. transit systems. Because of the large file sizes associated with image data, image data is typically overwritten every few hours, and used only to investigate incidents. The image data is time stamped, but not (to our knowledge) integrated with a location system.

Scheduling Software. Scheduling software has been virtually universal since at least the 1980's, offering a well-maintained database of the schedule, which is needed for schedule matching for both real-time and off-line AVL data analysis. Some scheduling software vendors have added to their software suites tools for analyzing running time, which can use manually or automatically collected running time data.

Getting schedule data out of the scheduling package database for AVL, APC, and similar applications has been a problem for many years. Agency staff report difficulties that have caused years of delay; those difficulties have been experienced with various scheduling software vendors and with both AVL and APC applications. Scheduling software packages tend to be highly customized according to local scheduling practice, and so a standard interface even for a single vendor has proven elusive.

1.3.4 Integration and Data Standards

Helpful standards have been developed and accepted in some areas related to archived AVL-APC data, while development is still needed in others.

Onboard Communication Standards. Two SAE standards, J1708 and J1587, have been defined that relate specifically to data communication protocols between devices on board public transportation vehicles. Vendors of most onboard devices will comply with the SAE standards. This standard allows an operator to sign in to one device and have that sign in information broadcast over the network to other devices, enabling a single sign-in, which, as mentioned earlier, is an important key to data quality. Devices connected to the network can broadcast events such as passenger counts, door openings and closings, and lift use, and a central onboard computer can log those events and assemble messages sent in by radio.

Farebox manufacturers have historically been reluctant to allow their machines to communicate with other onboard devices, citing the need to prevent fraud by keeping revenue-related information secure. Nevertheless, the J1708 standards cover farebox data and can be applied. Limited integration schemes, such as sharing a common operator log-in, have been applied at a few transit agencies. Recent products advertise J1708 compatibility.

Transit Communications Interface Protocols (TCIP). The TCIP project, started in 1996, is a standards development effort sponsored by the U.S. DOT's Joint Program Office for Intelligent Transportation Systems. Its mission has been to define the data elements and message sets that can be specified as an open data interface for transit data interchange activities. Phase 1 was completed in 1999, and established a transit ITS data interface "Framework" and eight "Business Area Object Standards." Phase 2, completed in June 2001, built on the work of Phase 1 by developing the transaction sets, application profiles, and guidebooks required to test and implement TCIP.

Some of the pertinent TCIP developments include:

- The definition of Automatic Vehicle Location objects including for example: compass bearing parameter, current time, current date, trip distance, position, velocity vector, total vehicle distance, milepost identification, etc.
- The definition of conformance groups that consist of a list of objects required to support a specific function. Performance groups were defined for dead reckoning, triangulation, global positioning system, etc.
- The development of a "Standard on Spatial Representation Objects," including a "Data Dictionary" and the definition of "Message Objects Set."
- The development of a "Standard on On-Board Objects," also including a "Data Dictionary" and the definition of "Message Objects Set."

TCIP object definitions include most schedule features, which are important for matching a vehicle and data collected on it to the route and trip on which it's operating. Some AVL and APC vendors have adopted the TCIP standard for their schedule data. However, TCIP has not yet been widely adopted by transit agencies as part of their product specifications and RFP's. Moreover, there are aspects of schedules that are not covered in TCIP, such as when a bus is simultaneously discharging passengers from an inbound trip and picking up passengers for an outbound trip. Among the people working to improve the TCIP standard are those involved with AVL and APC data.

Location Referencing Guidebook (LRG) Feasibility Study. The goal of this project funded jointly by FTA and the Transit Standards Consortium is to develop more effective practices in the exchange and use of spatial data. The first Technical Memorandum, published in July 2001, is entitled "Definition of Scope of Spatial Data Applications in Transit." It defines the scope of spatial data applications in transit, their relationship to location referencing, and significant barriers to the use of location referenced data.

A second Technical Memorandum will describe existing and emerging spatial data interoperability standards and issues related to how transit may use those standards.

FTA National Transit GIS Initiative. The FTA initiated in the mid 1990's a National Transit Geographic Information System to develop an inventory of public transit assets in the U.S. Although the effort focused primarily at a high national level, it was also designed to encourage more use of GIS tools by transit systems. As a result, the FTA published in January 1996 a report entitled "FTA National Transit GIS: Data Standards, Guidelines, and Recommended Practices." This effort was to some extent a precursor to the LRG project discussed above.

Of potential relevance to this study are the sections in 1996 report related to: a) Transit Feature Conversion Practices, in particular the design of bus stop and route databases, and b) Data Exchange Specifications, intended to facilitate the importing of transit-related databases into GIS programs. However, it appears that the recommended practices and specifications proposed in that report have not found their way into practice.

1.4 Route and Schedule Matching Issues

Successful matching of data captured on a vehicle to route and schedule depends on both good algorithms and good data. Good data includes both data captured during operation and enterprise data such as schedule and stop location data.

Modern AVL systems with a real-time component generally match to route and schedule in real-time, using algorithms that are generally proprietary and customized. APC systems without a real-time component generally provide proprietary, customized matching software executed off-line. To our knowledge, King County Metro is the only transit agency that has developed its own matching software for AVL data. They took over the matching software development when the AVL vendor defaulted. NJ Transit is working closely with its APC-AVL vendor to develop its matching algorithms.

1.4.1 Bus Stop and Timepoint Data

Matching location data to a stop or timepoint is a powerful approach for data analysis, since stops and timepoints are standard locations. Changes in door switch status (open / closed) usually signal that a bus is at a stop. The matching process demands a standard stop and / or timepoint location database against which field location data is compared. Transit agencies have encountered many problems developing and maintaining this stop location database.

Without data matching as a driving application, there has been little historical need for a transit agency to have an accurate stop database. Many agencies have no stop database, since "stops," i.e., the sidewalk space and signs, are owned by the municipalities, and since routes and schedules are detailed only to the timepoint level. Those with stop databases only had to be accurate enough for operators and maintenance personnel to locate the stop.

NJ Transit's experience matching AVL-APC data to scheduled patterns and trips emphasizes the importance of good base map stop location data. On patterns for which at least 90% of the reference locations are coded to within 300 feet of actual, NJ Transit's matching algorithm is able to match 81% of the trips to a scheduled trip and pattern, in contrast to a 65% matching rate overall.

Initially, stop matching will almost certainly require an intense amount of work preparing and correcting the stop database. (In the near future, GPS-based location systems may be able to automatically create base maps of stops and patterns, easing that burden.) Sometimes the agency takes that responsibility; sometimes it is given to an AVL or APC system vendor. A good geocoded base map maintained by a metropolitan planning agency can be of great value. However, geocoding based on street locations is often not accurate enough; field GPS measurements at the stop are needed. Equally important is ongoing work updating the stop location file for both temporary and permanent route changes. Some large agencies report changing 5% of their 10,000 stops each year – that's 500 stops!

Of all the applications requiring stop matching, stop announcements are the most demanding, because they can't rely on post-processing corrections and because matching errors will be visible (audible) to the public. A stop location file updated for stop announcements will serve well for archived data analysis.

Further complicating bus stop location is the fact that a transit agency can have four or more definitions of stop location *(17).* They include intersection or landmark (e.g., Third and Main); intersection quadrant (e.g., Third St. eastbound at Main, near-side); nominal coordinates along an ideal route, usually following the roadway centerline; and coordinates of the point on the curb closest to the bus stop sign. Add to that the complication of determining coordinates – are they measured using the same location system that will be used for bus location, or from a base map, introducing a possible calibration problem – and it is easy to have errors in the stop location database that will complicate matching.

Another complicating factor is that different enterprise systems – scheduling (to the extent that a bus stop is synonymous with a timepoint), facilities, transportation, passenger information systems such as on-line trip planners, and (at the municipal level) traffic each uses "bus stop" for a different function, and may use different definitions that involve multiple locations.

While AVL-APC systems need accurate bus stop locations, the location data they supply can be used to improve the base map. As mentioned earlier, one new APC vendor includes a learning mode function explicitly designed for mapping stop location. Another APC vendor recommends comparing average GPS coordinates at observations of a stop to the GPS coordinates in the base map, and using the average observed to update the base map if the discrepancy is significant and the observations are consistent.

1.4.2 Route and Schedule Data

Route and schedule definitions are generally imported from the scheduling system, a practice that makes sense because of the great care put into the schedule database because of its critical role in operations and payroll. The ability or inability of the scheduling system to generate the kind of route and schedule files needed by the AVL or APC system is a major factor in route tracking and matching success. In our research, the need for a standard interface is echoed by many transit agency and vendor representatives.

One problem stems from the fact that in the route definition, routes are frequently defined as a series of timepoints, not stops, while the APC or similar system is trying to match data to stops and to use stop locations for tracking. Another problem is that the scheduling system often excludes detail needed by the tracking and matching software concerning the path taken by buses during pull-ins, pull-outs, and deadheads. Making up for the deficiencies in the schedule files can be tedious and costly, has held up some projects for years. With schedules redone every 2 to 4 months, an automated process is clearly necessary.

If route changes are not immediately shared with relevant databases, matching problems will result. In the Eindhoven event recorder system, the current schedule is downloaded to onboard computers every day at pull-out to ensure that the schedule against which the operation is matched is current. Obviously, this kind of data sharing requires schedule definitions from the schedule system that are compatible with the onboard system and the matching software.

1.4.3 Getting Valid Identifiers

Matching is a lot easier when the algorithm doing the matching knows what run its bus is supposed to be following. It's then a matter of verifying that the bus is on the expected track. More difficult is determining where the bus is when there is either no run ID, or the bus is clearly following a different route or run.

The first issue, then, is getting operators to sign in. Sign-in is standard with AVL systems connected to the radio. Operators have to sign in to use the radio; if they don't, their bus might be called in for repair, with the operator being disciplined for failing to sign in. APC systems not tied to the radio present a problem. Many agencies are reluctant to force operators to sign in again to another system when they are already signing in to the radio, farebox, and destination sign. When an additional sign-in is required, compliance and validity will often be problems. System integration can avoid the need for an additional sign-in, as at King County Metro, where the sign-in to the AVL/radio is transmitted to the mostly independent APC system. If the APC system can't be integrated to the radio, and agencies don't want to force another sign-in on operators, they have to match their data without run information, as is the case at NJ Transit.

When operators sign in, getting *valid* identifiers for run and for operator is still a challenge. Best results occur when the location system that is recording data is also being use to drive a real-time application such as stop announcements, computer-aided dispatching, or radio communication. Invalid or unmatchable identifiers can be recognized, reported, and corrected in real time. By waiting until post-processing to guess the run, one loses the chance to get valuable information from operators and dispatchers.

Metro Transit (Minneapolis) eases the burden on operators while improving accuracy by providing automated sign-in, communicated by wireless link during pull-out, based on vehicle-block assignments made during the overnight. In their new AVL-APC system, operators will be asked to verify and correct their sign-in information.

As the value of archived data becomes more accepted in the agency, one can expect that dispatchers and supervisors will take more care to ensure the correctness of run ID's. Their cooperation is especially necessary when they order a bus to shift to another run, or to follow an unscheduled itinerary.

1.4.4 Tracking, Matching, and End-of-Line Issues

The AVL-APC vendor usually provides tracking (real-time) and matching (post-processing) algorithms. Vendors claim that their algorithms are largely successful; after all, most buses follow their assigned routes. However, tracking and matching errors are still frequent.

Tracking and matching algorithms rely on the quality of the reference data (e.g., stop locations), sign-in data, and captured data. A general rule suggested by NJ Transit's project manager is that with more and better data, you can make a better match. Agencies have found it valuable to make timepoint and stop records whether a bus stopped at that location or not in order to help matching.

End-of-line operations can be both complex and unpredictable, making it difficult to identify a trip's start and time. Operators approaching the end of the line with an empty bus may feel free to deviate from the prescribed route – after all, it won't affect any passengers. Detected door openings, exits, and entries at the end of the line may simply be the operator adjusting a mirror. A vehicle jockeying for position in a layover area may be mistaken for an early departure. Some agencies report treating first and last segment running time data with some skepticism. End-of-line issues also complicate passenger count analysis, making it hard to correctly attribute counts to the trip that is ending versus the trip that is starting. In business practice, sometimes passenger boardings at one stop should be attributed to one trip or route, and the alightings at the same stop attributed to another. Some trips may begin and end with non-zero passenger loads.

Agencies have used various means to deal with these issues. To prevent false alarms, King County Metro recently implemented a change in their tracking algorithm, ignoring bus movements in layover areas that occur more than 3 minutes before scheduled departure time. Agencies that have only timepoint level (not stop level) data recording sometimes use real or virtual signposts located not at the terminal but a few minute's travel from the terminal. This arrangement resolves some of the matching issues, but complicates running time and schedule adherence calculations. (Stop-level data collection accomplishes the same thing more accurately and with fewer complications.) NJ Transit uses engine shutdown and startup records to help identify the end of the line.

Splitting the trip that is ending from the next trip beginning can be complicated by the fact that, in practice, buses sometimes serve the end of one trip and the beginning of the next trip at the same time, such as when a bus circulating through downtown drops off passengers from an inbound trip while it picks up passengers for an outbound trip. At NJ Transit, matching algorithms have been modified to account for this kind of business practice, attributing passenger boardings that occur in such a zone to the outbound trip while alightings occurring in the same zone are attributed to the inbound trip. Another example of the complexity of modeling real business and passenger behavior include passengers, in hopes of getting a seat, boarding a bus going outbound and staying on as it turns around and returns inbound.

CHAPTER 2

AUTOMATED DATA ANAYLSIS: PRACTICE AND POTENTIAL

The first three sections of this chapter answer the questions, how might archived AVL-APC data be used to improve transit management and performance, and what kinds of data are needed to support the various analyses? Included is a list of analysis and decision-making tools that are either in use or that might be used, based on a survey of transit management functions. The latter part of the chapter deals with practice in archived data analysis.

2.1 Key Dimensions of Automatically Collected Operations Data

To help what AVL-APC data might be useful to support transit management, it is convenient to first analyze four key dimensions of the core data involved: level of spatial and temporal detail, complete vs. exception data, sample size, and data quality.

2.1.1 Levels of Spatial and Temporal Detail: A Hierarchy

An analysis of existing and conceived systems for gathering location-related operations data reveals a hierarchy with five levels of spatial and temporal detail labeled A (least detail) to E, summarized in Table 4. Detail levels A and B typically involve data recording by transmitting the data over the air to a central computer; levels C-E usually involve onboard data storage to overcome the limitations of radio channel capacity, with data "milked" (uploaded to a central computer) each night.

Detail level A represents the least detail: infrequent event-independent location records. This level is representative of many older AVL systems, in which information is captured only about the location of the bus when it is polled. The polling interval, depending on system design, is typically 40 to 120 s, though intervals as low as 16 s and as great as 240 s have been implemented. For special purposes, the polling rate can be increased on some buses at the expense of others. Determining the moment at which the bus passed a particular location (e.g., a timepoint) requires interpolation, causing approximation errors that can be as large as half the polling cycle. This level of detail is the simplest to implement in an AVL system because it requires no onboard vehicle tracking intelligence.

Detail level B includes timepoint records. An onboard computer knows when it reaches the timepoint location, and either records it in an onboard computer, or transmits it over the air as a timepoint message. Knowing the onboard location is best done by the onboard computer tracking vehicle location, either using GPS or dead reckoning. With this tracking capability, users can select and change locations of interest. Most new AVL systems with GPS have this capability. King County Metro has implemented this capability in a signpost system by having the central computer tell the bus, about two polls before arriving at a timepoint, what odometer reading will indicate arrival at the coming timepoint. A less flexible way to obtain level B data is to install wayside transmitters at locations of interest which will trigger a record being made on a passing bus. However, this arrangement makes it hard to change specified locations, and the failure of a transmitter means loss of information at that point.

Detail level C involves a record for each stop. It is standard with APC applications and event recorders tied to stop announcement systems, and rare with AVL. Typically, data are temporarily accumulated in a small number of registers, and a record is closed out and written when the door closes. The record includes either the moment arriving at or departing each stop, or both. Other items often recorded include passenger on and off counts and notice of lift use. The system is also usually configured to create a record for special events such as passing a signpost, operator login, and exceeding a preset period of inactivity (e.g., 2 minutes).

TABLE 4: Levels of Spatial and Temporal Detail for Data Capture

Level	Description	Event-Independent Records	Event Records	Between-Stop Performance Data
A	AVL without real-time tracking	infrequent (typically 60 to 120 s)	-	-
B	AVL with real-time tracking	infrequent (typically 60 to 120 s)	each timepoint	-
C	APC or event recorder	-	each stop	
D	event recorder with between-stop summaries	-	each stop and between-stop events	recorded events and summaries
E	event recorder / trip recorder	very frequent (every second)	all types	all events, full speed profile

Detail level C is usually implemented either onboard data storage, milked each night by a high-speed wireless link. Some newer radio-based systems send stop-level APC records over the air rather than record them onboard. Because only APC-equipped buses, representing 10 to 15 percent of the fleet, send stop event messages, the impact of stop event messages on radio traffic is not too severe.

In one Level C configuration, the onboard computer does not match vehicle location to the route; matching is done by the central computer when the data is processed. With this configuration, no information will be reported for bus stops that are passed without the door opening, unless a signpost is located there. In the more common configuration that lends itself to on-line applications such as AVL, stop annunciation, and next-stop arrival time information, the onboard computer continually matches vehicle location to the route. It can then make a record even for stops that are passed without the door opening by having stops defined as virtual signposts. Users can also define dummy stops for locations that are not actual stops in order to get data on passage moment at those locations. Important such locations can include garage entrances, timepoints that fall between stops, municipal boundaries, and layover and relief points.

In detail level D, in addition to stop-level information as in level C, data on each interstop segment is also recorded. Onboard data storage becomes essential as the amount of information recorded increases. In the configuration used at Tri-Met and in Eindhoven, the onboard computer records data in temporary buffers and includes summary measures (e.g., maximum speed, time spent at crawl speed) in the stop record. NJ Transit's configuration has the additional flexibility that routine interstop events (e.g., change in heading or speed passing a threshold) trigger their own records. This level of detail goes beyond traditional APC by recording operational data, and beyond traditional AVL by including stop-level detail and onboard data storage. Onboard storage requirements are still limited since only a few summary measures are recorded for each segment.

Detail level E involves near-continuous recording (e.g., a record every second or every few seconds) including, at a minimum, time, location, and door status (open / closed). This level provides the trip recorder capability of the familiar "black box" and allows the user to analyze and summarize almost any measure of performance without having to specify and configure the system in advance. However, level E requires more onboard storage capacity and a higher short-range data transfer capacity for the milking operation. NJ Transit uses this level of detail for special investigations (its system allows the user to specify how often a record is to be written). At least one APC supplier reports taking this approach in new installations. Conceivably, over-the-air AVL systems with frequent polling (perhaps with frequent polling limited to a few target routes at a time) can approach this detail level. In every known application, the polling interval is too large to qualify as a trip recorder; however, it would certainly be possible to target a sample of buses for very frequent polling, as is often done when a bus's emergency alarm is set.

2.1.2 Complete vs. Exception Data

A second, related dimension is whether data is captured routinely, or only when a key data item is out of range. Several AVL systems installed in the early and mid-1990's used exception reporting, driven partly by technical limitations but mostly by the preference of controllers (dispatchers) to focus on service that was in need of remedial action. With respect to schedule adherence, a bus was counted as an exception when outside a specified tolerance such as 1 minute early to 5 minutes late. A bus with a true alarm status or other non-standard condition would also qualify as an exception. In those systems, non-exception buses simply reported "I'm OK" instead of giving their location in order to reduce message length and shorten the polling cycle.

It is hopeless to try to estimate headways or running time from exception data, unless one is willing to live with errors as large as the range of the on-time window (e.g., 6 minutes for a tolerance of 1 minute early to 5 minutes late). While schedule deviation can be analyzed in terms of percent within the on-time window, further analysis (e.g., how many departures are within 3 minutes of schedule) is limited.

Researchers at Morgan State University tested the feasibility of using exception data to analyze running time and schedule adherence on MTA (Baltimore) bus routes *(18)*. Because they had records only on buses that were outside the on-time window, they focused on next-segment running time for buses that arrived at a timepoint early or late. If the bus reached the next timepoint "on time," the researchers had to guess when within the on-time window the bus arrived. Also, because the system archived only exception messages, it was also impossible to gauge the amount of missing data.

In an interview, the Morgan State researchers described how they had been requested by the MTA to systematize the way they transformed the raw data into a complete record that would support analyses of running time and schedule adherence. The researchers' response was that the transformation involved large approximations and was an extremely tedious process that took several months and that could not be made routine. The MTA has since discontinued attempts to analyze their AVL data.

Technological advances have essentially eliminated the pressure to limit data collection to exceptions. Therefore, exception data is not considered a viable option in this review. As an example of how industry has moved away from this trend, CTA's AVL system, specified in 1993 and purchased in 1996 to provide location data only on buses that were off schedule, ordered a modification in 2002 to provide location data on all buses.

2.1.3 Fleet Penetration and Sample Size

A third dimension of automatically collected data is the sample size, which depends on fleet penetration and data recovery rate. The main choices with respect to fleet penetration are whether the entire fleet is instrumented, which is standard for AVL and for event recorders, or only a small subset of the fleet, typically 10 to 15 percent, which is standard for APCs. The average sampling rate is the product of the fleet penetration and data recovery rates. For example, if 10% of the fleet is equipped, and data is recovered from 70% of the trips operated by instrumented vehicles, scheduled trips will be observed, on average, 7% of the number of times they are operated. In a 3-month period containing 65 weekdays, 13 Saturdays, and 13 Sundays, that would be about 4.5 observations per scheduled weekday trip, and just under 1 observation of each Saturday and Sunday scheduled trip.

An average sampling rate can mask significant variations across the system. When fleet penetration is small, logistical difficulties coupled with the vagaries of data recovery failure often result in some scheduled trips being sampled well more than the average number of times, and others perhaps going completely unobserved. Management of the rotation of the instrumented vehicles, then, becomes another important factor in determining whether needed data will be available.

Because leader-follower or headway analysis requires valid observations of consecutive pairs of trips, the number of valid headways one can expect to recover is proportional to the square of the data recovery rate and the correct assignment rate. For example, if the data recovery rate is 70% and if a request to instrument a given route results in 90% of the trips on that route having an instrumented bus, one can expect to observe only $(0.7)^2(0.9)^2 = 40\%$ of the day's headways on that route.

A 1998 survey *(6)* found net recovery rates for APCs ranged from 25 to 75 percent, with newer systems having better recovery rates. In 1993, COTA (Columbus) reported that with 11.9 percent of the fleet APC equipped, they netted on average 5 usable samples per assignment each quarter *(19).* That represents a 6.1 percent sampling rate, for a net recovery rate of about 50 percent. Only a small part of that loss was due to mechanical failures, as mechanical reliability was well above 90 percent. More often, data recovery failure stems from failure to match a recorded trip to the schedule or from passenger counts not balancing.

In another example, drawn from our case studies, an audit of King County Metro's fall 1998 sign-up (a four month period) found on average 6 valid APC observations per weekday scheduled trip, for a sampling rate of 7.5 percent. With about 15 percent of the fleet instrumented, that represents a data recovery rate of 50%. Coverage across the schedule was variable. They recovered at least one valid observation of 97% of their scheduled weekday trips, and at least 3 valid observations for 83% of their scheduled weekday trips. On the weekend, 85% of scheduled trips had valid data. In the last year, Metro reports that the recovery rate has improved to the 60 to 70 percent range.

Less is known about data recovery rates for AVL. King County Metro recovers AVL data from about 80 percent of its scheduled trips. However, with the entire fleet instrumented, data recovery rates are not so important with AVL unless there is systematic data loss in particular regions or times.

In general, a small sample size is sufficient to reliably estimate the mean of a quantity with low variation such as running time on a segment, or demand on a particular scheduled trip. A larger sample size is needed to examine variability or extreme values, such as the 90-percentile running time or load. A very large sample size is needed to accurately estimate proportions, such as proportion of departures that are on time *(20).*

Analyses that aggregate scheduled trips into periods, or that aggregate routes, have the advantage of a larger sample size than analyses of individual scheduled trips. That is a reason that many customary analysis methods, developed in a limited-data environment, concentrate on the period rather than the trip.

A large sampling rate allows more timely analysis of data. Over a long enough period of time, even a small sampling rate will yield a large number of observations. However, for management to react promptly to demand and performance changes, or measure the impact of operation changes, analysts need recent data. Therefore, tools used in the active management of a dynamic system will benefit from the high sampling rate that follows when the entire fleet is equipped.

2.1.4 *Data Quality Control*

As data is gathered in real time, there are inevitably errors – errors in attributing it to the right trip or route, in passenger counts, in location. Without sufficient quality control, users may find so many errors that they come to mistrust the data, and perhaps stop using it, defeating the purpose of gathering the data.

With good quality control and post-processing, many errors to be corrected (e.g., trip or route ID's corrected, passenger counts balanced), and other others can be flagged as out of range or inconsistent so that unreliable data will either be omitted from the historic database or entered with a flag by which it can be excluded from data analyses. Corrections can also be semi-automatic, as at Houston Metro where a program displays inconsistencies in route-run ID between AVL and payroll data, and asks the person running the program whether to change the AVL's route-run ID.

2.2 Related Data Items

Analysis opportunities also depend on what other data items are captured in the AVL-APC data stream, and on related databases that might be integrated with the AVL-APC data.

2.2.1 *Related Data Items in the AVL-APC Data Stream*

Other potentially valuable data items that can be captured by AVL-APC and other onboard devices that can integrated into the AVL-APC data stream include:

- *Door open time*. While most systems that make stop-level records (level C and higher) both the door open and close moment, some do not. Door open time is assumed to be part of level of detail D.
- *Stop and start moment*. With stop-level records (level C and higher), additional insight to dwell time and holding can be obtained by recording when the bus stopped and when it began moving again. If the operator of an early bus holds at a stop with the door closed, start moment is a better measure of schedule adherence than door close time. Similar information is obtained by recording when a bus enters and exits a defined zone around a stop. Several AVL systems use the zone concept, either based on signposts (using signal strength thresholds) or on virtual zones defined in terms of GPS coordinates or odometer distance.
- *Passenger entry / exit moment* (for level D). Typically, APCs provide only a summary record of counts. Time stamping each passenger entry and exit provides more insight into dwell time and holding.
- *Off-route location*. Unlike odometer-signpost systems, GPS-based systems offer location information when the bus is off-route, including turnarounds, deadheads, pull-ins and pullouts, detours, and unscheduled departures from route.
- *Event codes.* Events can include mechanical alarms and flags (e.g., engine overheated, lift in use), driver-actuated security alarms, and driver-entered event codes (e.g., to indicate lift use) add to the richness of the database. Event codes to indicate pass-ups are often the only way of capturing something that, to passengers, is a prominent aspect of service quality. Most AVL systems include event code data.
- *Control messages* from the control center to the bus operator. Communication from the control center to the bus is less often in the form of data codes (as opposed to voice). Codes that might be used and recorded include standard messages such as "wait for transfer" or "running hot (early)." More important, but hard to capture in a database, are commands to deviate from schedule service, e.g., to turn back, express, go out of service, or switch to another route. To the extent

messages are transmitted over a data channel (as opposed to a voice channel), they are usually part of an AVL data stream.

- *Traffic control communications.* On a route with signal priority for buses, messages can be sent between the bus and traffic controller (e.g., "I'm a Route 16 bus and I'm 40 s late." "Got your message. You've been assigned priority level 4."). This requires cooperation with the traffic control system, and is not yet known to occur.

- *Farebox transactions*. It is customary for farebox data to be stored separately from AVL-APC data, both on the bus and in archive databases. However, it is certainly feasible for the data archive that contains automatic location data to also include farebox data if the farebox is connected to a vehicle-area network broadcasts transaction data as it occurs. Detail level C would only include a count of the number of transactions of each type at each stop, while detail level D could include records of each transaction.

- *Annunciator & Destination sign*. Records of changes to the annunciator and destination sign messages allow the transit agency to monitor whether information given to passengers matches actual vehicle location and route, and can help confirm changes in route and direction. To our knowledge, even when the annunciator and destination sign are driven by the AVL system, records of their activity are not made in existing systems.

2.2.2 Related Databases

Many analyses using AVL-APC data will benefit from links to other databases. They include:

- *Schedule (including routes and patterns).* Schedule data is needed to identify the route, pattern, and scheduled trip that was operated. Without proper identification, analysis is either impossible or severely complicated. On-time performance and running time deviations are measured against the schedule. It is important to recognize that while the entire schedule changes only a few times a year, minor changes can occur any time to deal with detours, special events, and so forth, and so a method of updating the schedule is vital *(21)*. While the schedule database used for scheduling may only need to know the current routes and schedules, analysis of archived data must always be tied to the route and schedule then in force.

- *GIS.* Stop and intersection inventories are necessary to identify stop and intersection locations. GIS can be valuable for displaying data; e.g., schedule adherence statistics along a route could be displayed on a route map. More complex is using a map-based user interface in an analysis, e.g., analyzing passenger activity or service quality in a specified area. Some planning analyses also use background demographic and land use information.

- *Payroll.* Links to payroll facilitate a comparison between what a driver was scheduled and / or paid to do and what he or she did.

- *Farebox* data, as mentioned earlier, is usually stored by the farebox system as a separate data stream and in a separate database. In traditional systems, only summary data (summarized by trip, hour, or day) is recorded; some newer systems provide transaction data. If there is no APC data, farebox data is often a primary source of passenger boarding counts. Farebox data can provide a check on the AVL-APC data. Some fare payment systems, both magnetic stripe and smart card based, allow one to track individual users or the route previously used. That permits one to identify transfer points (useful for stop matching) and to track linked trips.

- *Maintenance*. AVL-APC data on speed and other aspects of vehicle performance can be used to track demands on the vehicle more precisely than an odometer. Those demands can then be related to vehicle maintenance needs and history.
- *Weather, Incidents, and Special Events*. Users of archived data should be able to know of any unusual weather, roadway incidents, or special events. They could be used to flag data for exclusion in many analyses, or may be the basis of special analyses of how those events affected operations.
- *Customer Satisfaction, Census, Customer Complaints.* APC/AVL data can offer valuable perspectives on the causes of customer satisfaction and customer complaint problems. When combined with the census' demographic and journey-to-work data, the service quality measures available through AVL-APC can help identify inadequately served markets.

2.3 Analysis and Decision Support Tools and Their Data Needs

Data should not be collected for its own sake, but to support analysis and decision tools to improve transit management and performance. Muller and Furth *(1)* highlight the role played by both real-time and archived operational data in the performance improvement cycle. Analysis of archived data helps improve the service plan, while real-time data helps control operations so that they better follow the service plan.

This section examines an array of analysis and decision support tools that can benefit from operational performance data that can be gathered by AVL and APC systems, identifying for each analysis or tool the data items, level of detail, sample size, and related databases needed to support each tool. This review touches all relevant functions of the transit enterprise and includes both tools that are used in practice as well as tools that might be used to improve transit management and performance. A summary of this section is provided in Table 5.

2.3.1 General Service and Contract Compliance Monitoring

General service monitoring, of interest to both upper management and oversight boards, simply tries to answer the question, did the service operate as scheduled? The need to monitor performance of contracted operators is driving many European transit systems to install automated data collection and analysis systems, whether by AVL, APC, or event recorder. The two key analyses at this level are to identify **missed trips**, and to identify the percentage of trips with **large deviations from schedule**.

AVL is not reliable in itself for monitoring missed trips, because missing data does not always indicate a missed trip. However, in combination with other data sources such as payroll and farebox, it may approach or exceed the reliability of analysis based on manual supervision, and certainly provides a check on manual monitoring. APCs are unsuitable for the task if installed on only a sample of the fleet. AVL can be more effective than manual methods at recognizing when a trip is so late (e.g., 15 minutes late) that it is classified, from a customer service perspective, as missed.

Level of detail A can be acceptable for monitoring schedule adherence for self-operated service because interpolation errors inherent in estimating schedule deviations at timepoints will neither be either so large nor so biased as to mask any significant trend in systemwide or route-level schedule adherence. Incident codes and control messages are valuable if either operators or controllers indicate when or why a trip is missed or is very late. A link to the schedule database is necessary to match operated service to scheduled service.

TABLE 5: Decision Support Tools and Analyses and Their Data Needs

Function	Tool / Analysis and [Usage Code*]	Detail Level Needed	Additional Items Needed *(italics = optional)*	External Data Needed *(italics = optional)*
General service monitoring, including contract compliance	- Missed trips [1] - Schedule adherence [4]	A or B	*incident codes, control messages*	schedule
Targeted Investigations - Customer service (complaints) - security / legal (incidents, accidents) - Operator performance	Trip investigation at gross level (was it there? was it off-route?) [4]	A	*off-route, incident codes, control messages*	*schedule, payroll*
	Trip investigation: early, late, overcrowded? [3]	C	door open & close; passenger counts for overcrowding	
	Trip investigation: speed, acceleration [2]	D or E	maximum speed; *records every 2 s or more to measure accel, decel rates; GPS altitude*	
Scheduling and Monitoring Running Time	Route and segment running time analysis (mean and distribution) [4]	B		
	Suggesting running time based on percentiles [3]	B		
	Selecting homogeneous running time periods [3]	B		
	Suggesting half cycle time based on percentiles [2]	B		
	Running time analysis net of holding time [2]	C	door open time, *stop - start time, incident codes, control messages, on-off counts*	schedule
	Speed and traffic delay [2]	D	door open time, *stop-start time*	*schedule*
	Unsafe operations monitoring [0]	D or E	maximum speed; *records every 2 s or more to measure accel, decel rates*	
	Relating running time to weather, roadway incidents, and special events [1]	B		weather, roadway incident data, special event data

TABLE 5, continued

Function	Tool / Analysis and [Usage Code*]	Detail Level Needed	Additional Items Needed (*italics = optional*)	External Data Needed *(italics = optional)*
Schedule Adherence and Connection Protection **(service and operational quality)**	Percent early, late by timepoint [4]	B (timepoint-level) or C (stop level)	door open time, *stop - start time, incident codes, control messages*	schedule
	Distribution of schedule deviation at a timepoint [3]	B or C		
	Graphical display of schedule deviation distribution along a route [2]	B or C		
	Experienced lateness and earliness [1]	C	stop-start time, *passenger counts*	*farebox transactions with linked trip data*
	Connection protection [1]	C	door open time or stop - start time, control messages	*farebox transactions with linked trip data*
Headway Analysis **(service and operational quality)**	Headway deviations (mean and distribution by timepoint) [3]	B (timepoint-level) or C (stop level); all buses reporting	door open time, *stop - start time, incident codes, control messages*	schedule
	Impact of headway variability on passenger waiting time for random passenger arrivals [1]	C	door open time, *stop - start time*, on-off counts, incident codes	*farebox*
	Plot successive trajectories (bunching analysis) [2]	C	door open time, *stop - start time, incident codes, control messages*	schedule
Demand Analysis	Load profile (mean ons, offs, and load by stop along a route; also passenger-miles) [4]	C	on-off data or (for estimated alightings, load, and pass-mi) farebox transactions	
	Load variations [3]	C		
	Analysis of trip maximum loads and max load points [1]	C		
	Time-dependent demand and load analysis, and suggesting trip start times to achieve load targets [1]	C	door open time, on-off counts or (for estimated load) farebox transactions	
	Analyze overload, lift, bicycle, and other events by stop and time [3]	C	incident codes and/or farebox transactions	
	Transfer and linked trip analysis [1]	C	farebox transactions with card ID's	farebox

Function	Tool / Analysis and [Usage Code*]	Detail Level Needed	Additional Items Needed *(italics = optional)*	External Data Needed *(italics = optional)*
Geographic and Planning Analysis	Geocoding stops and other points of interest [2]	C		GIS
	Mapping bus path through shopping centers, new subdivisions, etc. [3]	E		
	Comparing measured vs. nominal stop locations [1]	C		
	Relate on-off data to demand rates in traffic analysis zones and to geographic database [1]	C		GIS, regional travel demand model database
	Relate service quality data to geographic database [1]	B or C		GIS, schedule
Utilities	Monitoring system failures [4]	A	system diagnostics	
Other Operations Analysis	Operator performance (schedule adherence, on-time start, running time, headway maintenance) [1]	B (timepoint level) or C (stop level)	door open time, *stop - start time, incident codes, control messages, on-off counts*	schedule, *farebox data*
	Dwell time analysis [2]	C	door open time, stop-start time, passenger entry-exit moment, on-off counts, *farebox transactions, incident codes*	
	Layover and pull-in / pull-out analysis [0]	B	door open time, *incident codes, control messages, off-route*	schedule
	Control effectiveness: any service quality monitoring or service analysis, related to control messages	*as required for each analysis*	incident codes and control messages	
	- Before / after study - Special event / weather analysis	*as required by the type of analysis*	*as required by the type of analysis*	*as required by the type of analysis*

TABLE 5, continued

Function	Tool / Analysis and [Usage Code*]	Detail Level Needed	Additional Items Needed *(italics = optional)*	External Data Needed *(italics = optional)*
Passenger Information Monitoring	Prediction accuracy (match announced stop or predicted arrival time with actual) [1]	C	door open time, annunciator	schedule, GIS
	Accuracy of route data in destination sign and farebox [0]	A	destination sign, farebox	schedule
Payroll	Verify sign-in data [2]	A		schedule, payroll
	Examine operator's duty when there's an overtime claim [2]	A	off-route, *incident codes, control messages*	schedule, payroll
Maintenance Management	Analyze maintenance incidents [0]	D	incident codes, *control messages, on-off counts, GPS altitude, vehicle health indicators*	*maintenance, altitude*
	Monitoring vehicle demands [0]	D	*on-off counts, GPS altitude, vehicle health indicators*	maintenance, *GIS*
	Analyze failure trends [0]	D	incident codes, *control messages, on-off counts, GPS altitude, vehicle health indicators*	maintenance, *GIS*
Strategic Planning	Trends analysis [2]	*as required by the type of analysis*	*as required by the type of analysis*	*as required by the type of analysis*

* Usage codes: 4 = used commonly by agencies with AVL-APC data; 3 = used by some agencies with AVL-APC data; 2 = used by only a few agencies with AVL-APC data; 1 = used experimentally or ad hoc; 0 = not used

For monitoring contracted services, especially if there are incentives or penalties for missed trips and on-time performance, the interpolations necessitated by level A detail may present too much of a challenge, especially if the polling interval is long. Level of detail B data or better is more appropriate.

Many cities use their AVL and APC systems to track systemwide and route-level schedule adherence. Windsor, ON has an interesting executive information system that provides summary reports intended for upper management. In Eindhoven, the Netherlands, a report is produced nightly summarizing performance on the previous day, available by password on the worldwide web.

2.3.2 Targeted Investigations: Customer Service, Security, Legality

Many transit agencies use their AVL data archive to **investigate customer complaints, incidents or accidents, and cases of suspected operator misbehavior**. This need is often met using "replay" capability in systems oriented primarily to real-time applications, in which data from the subject trip is reviewed as if it were happening in real time. If stop or timepoint records are stored in a database, this function can be met using general query capability.

Level of detail A is adequate to answer "was the bus in question really there?" Several transit agencies report being able to disprove an accident claim by referring back to archived AVL data. Level A can also identify whether a bus was seriously early or late. However, since being early is often a matter of only 1 or 2 minutes, level B or C is generally needed to verify complaints about early buses. Ideally, the location data should indicate departure time, or at least when the doors closed. To investigate a complaint about an overcrowded bus, passenger counts are ideal. Without passenger counts but with stop level data, one could at least check for an unusually long gap between buses.

Off-route location capability, part of many cities' AVL systems, is valuable for investigating **off-route incidents**. A record of incident codes and control messages enable investigators to see how the operator and controllers responded to an incident. A link to the schedule database is necessary for investigating schedule deviation complaints and off-route incidents. A link to the payroll database is valuable for verifying the identity of an operator in question.

If an investigation deals with the **bus speed** or **excessive acceleration or deceleration**, level D or E is needed. Level D, if configured to record the bus' maximum speed between stops as at Tri-Met, can help verify whether a bus was speeding. With Level E and frequent records (every 2 s or more), one can review speed and acceleration at any moment, including the moment of a crash or injury. This is a natural extension of hub tachometers, which have been widely used for decades in British and other European bus systems for investigating speed-related incidents. Altitude data, a GPS component that is often omitted in AVL systems, could be used to infer grade, which might have a bearing on some accident analyses.

2.3.3 Scheduling and Monitoring Running Time

Scheduling departments are interested in updating scheduled running time based on observed running time. Level of detail B or higher is preferred, since running times are usually scheduled between timepoints. For **scheduling at the stop level** – a logical step transit agencies are beginning to take as part of providing better passenger information expectations grow – detail level C or higher is needed. Transit agencies applying **schedule-based signal priority** need schedules at the stop or intersection level, requiring detail level C or D.

Most users of archived AVL-APC data have a means of analyzing running time, displaying **mean and the distribution of running time** for either segments or routes.

While many transit agencies make schedules based on mean running time, some prefer to use percentile values, for which automated data gathering and analysis is a boon. Tri-Met uses low percentile (40 percentile) values to bias the schedule toward late rather than early operation, while Eindhoven uses high percentile (85-percentile) values to promote operator compliance with holding by offering them an attainable schedule. More commonly, as at OC Transpo, schedule makers choose running time subjectively based on the distribution of observed running time.

Some agencies, including King County Metro and Eindhoven, have software tools that **suggest scheduled running time,** in combination with a parallel tool for **establishing scheduling periods**, i.e. periods of homogeneous running time for which a common running time can be assigned. Users supply a criterion such as "find as long a period as possible for which a common scheduled running time is within 2 minutes of the 75-percentile observed running time for at least 90% of the scheduled trips in that period."

At King County Metro, schedules are built beginning at the segment level, so running time analysis is done at the segment level. In Eindhoven, schedules are first made at the route level, with running time later allocated over segments, so analysis is first done at the route level, followed by another tool that suggest segment running times.

Tri-Met has begun to experiment with using percentile analysis of running times as a basis for **establishing half-cycle time** (running time plus recovery time in a single direction). Because a primary purpose of recovery time is to limit the probability of a trip starting late due to a late arrival, it stands to reason that should be based, at least in part, on an extreme value such as 90- or 95-percentile running time, rather than simply set as a percentage of running time. That way routes with greater running time variability will have more recovery time, while those that operate with less variability have less.

Ideally, scheduling tools should use **net running time**, which excludes holding time (when an operator intentionally holds because he or she is ahead of schedule). **Determining holding time** is tricky, and is done by only a few agencies with AVL-APC data. At Eindhoven and at NJ Transit, it is done using stop-level data, schedule matching (because operators should hold only when ahead of schedule), door open and close times, and departure times. If there is a long gap between door close time and departure time and the bus is ahead of schedule, the gap is treated as holding. Because in good weather holding may occur with the doors open, NJ Transit's system is being upgraded to provide a record of how many passengers boarded and alighted every few seconds (the main onboard computer frequently queries the onboard passenger count analyzer while doors are open and writes records), allowing them to designate periods of inactivity while the bus is ahead of schedule as holding time. Eindhoven's algorithm, which lacks passenger detection data, looks for unusually long dwell times (times with the door open) when the bus is ahead of schedule. Additional data that might help identify holding time include incident codes that might help explain long dwells due to lift use or a fare dispute. Another version of holding is still harder to detect – "killing time" en route to avoid being early. Data and algorithms that would help **detect "killing time**" would be valuable and are being developed at NJ Transit.

Speed and traffic delay analysis is helpful for identifying locations in which there may be a need for infrastructure or traffic improvements such as signal retiming, turning restrictions, signal priority, or physical priority. It is also helpful for monitoring the impact of any change in infrastructure, traffic, and operator behavior. Both related to running time net of dwell time, and therefore require door open and close time. Several transit agency APC data analysis suites include speed analysis. Delay can be estimated in two ways. One is to subtract from running time observed dwell time and an "ideal" running time, which might be running time during a period of light traffic. To our knowledge, no transit agency does that kind of delay analysis. The second, done by Eindhoven, attempts to directly measure delay as time spent below crawl speed with the doors closed. Eindhoven's data analysis system shows delay distribution (mean, 85-percentile, maximum) by segment along a route.

Detailed running time analysis can be used to **monitor operations for unsafe practices** such as excessive speeds and accelerations. This is a natural extension of the trip recorder function, but looking for unsafe practices routinely. An analysis matched to the schedule might indicate where excessive speed and acceleration are correlated to operators being behind schedule; that might be an indication that scheduled running times should be liberalized. However, such an analysis risks a “Big Brother” reaction from operators, even if it is well intentioned and not used punitively.

There can be many reasons to study the **impacts of weather, roadway incidents, and special events on running time** and schedule adherence. Most challenging is relating to roadway incidents, which have an important geographic and time dimension whose extent is often unknown to the transit agency. Developing an interface with a traffic management center database may be worthwhile.

For tools involving estimating mean running time, speed, or delay for a given period, data from 10 to 15 percent of the fleet will be adequate since variability is relatively small and data are aggregated over several trips in a period. However, analysis and scheduling tools that examine running time variability or percentile values call for full fleet coverage.

2.3.4 Schedule Adherence and Connection Protection

To passengers, schedule adherence is a matter of service quality. From the service provider perspective, schedule adherence reflects the quality of the service plan (the schedule) and of operations control. Each aspect – service quality, service planning, and operations control – calls for its own focus.

The most common analysis, a gross measure of **percentage of timepoint departures that are early / on-time / late**, is adequate for monitoring service quality. For the service provider, it’s a combined measure of the quality of the service plan and of operational control. The disadvantage of such a combined measure is that if schedule adherence is poor, one can’t always tell if the problem is a bad schedule or poor control in keeping operations on schedule.

Estimating the proportion of trips that are on time requires a sample of at least 200 observations. For general system monitoring, a 10% sample is adequate, because data from many periods and routes can be aggregated. For example, STM (Montreal) uses an APC sample to evaluate operations control at each garage, tying manager evaluations to on-time performance. However, analysis at the route-period level calls for data from the entire fleet.

Extensive AVL-APC data can support better, more targeted tools for analyzing schedule adherence. First, analysis can go beyond defined early / late thresholds, providing a **distribution of schedule deviation.** Different analysts may be interested in different thresholds. For example, 3 minutes late might be the threshold for an analysis of “good” service; 5 minutes late the threshold for “adequate” service; and 7 minutes late the threshold for alerting dispatchers. **Standard deviation of schedule deviation,** a summary of how broad is the schedule deviation distribution, is an indicator of how unpredictable and out-of-control an operation is; it is part of Eindhoven’s daily service quality report.

A **graph of the distribution of schedule deviation along a route** is a standard tool used in Eindhoven to improve scheduling. Clear patterns often emerge that suggest immediate scheduling improvements. For example, if operators routinely depart late but still arrive on time at the end of the line, there is too much time in the schedule – the operators know it and start late to avoid running hot. If service alternates between being early at some points on the route and late at others, the distribution of running time between segments is off.

A more refined measure of service quality as perceived by passengers can be obtained by weighting schedule deviations by passenger on-off counts to estimate **experienced lateness and earliness**. Often the best schedule adherence is when passenger demand is lightest, so that passengers as a whole perceive service as being worse than a simple "percentage on-time" statistic might suggest. Such an analysis should recognize that being early is of no consequence to passengers getting off, but can have serious consequences to those getting on. (Experienced earliness for passengers getting on could be translated into lateness by assuming that a bus's early arrival will result in some passengers waiting a full headway.) If passenger counts aren't available, historic counts or location-matched farebox data could be used to weight schedule deviations. In some ways, historic counts might be preferable since an observation of nobody boarding an early bus might not mean that nobody was inconvenienced – the bus may have had nobody waiting for it because it arrived so early! Distributions of experienced earliness and lateness (how many passengers experienced a bus arriving *x* minutes early and *y* minutes late) may also be helpful. To our knowledge, such analyses are not part of any routing data analysis program.

Passengers are particularly interested in whether they can make their connections. Arriving 4 minutes late is not a problem if the time allowed for the transfer is 5 minutes, but it could be a big problem if the allowed time was only 3 minutes. However, if the departing trip is held – again, the convergence of schedule planning and operations control – it's another matter still. To our knowledge, standard tools for analyzing connections have not yet been developed. **Connection protection analysis tools** need accurate arrival and departure time data, integration with the schedule, and a capacity to do analyses across routes. A more accurate tool will incorporate passenger demand data: ons and offs at a minimum, but ideally transfer volumes. Data on control messages such as requests for holding will permit tools that **analyze operational control**.

2.3.5 Headway Analysis

On routes with small headways, **headway regularity** is important to passengers because of its impact on waiting time and crowding. It is important to the service provider because of how crowding tends to slow operations, and because a good deal of operations control is focused on keeping headways regular. Standard summary measures, used at several transit agencies with AVL-APC data, are standard deviation, average absolute deviation from scheduled headway, and distribution of headways. (Mean headway deviation, which is usually around zero because positive deviations are balanced by negative deviations, is of little value.)

Headway data allows one to estimate **passenger waiting time**, assuming random passenger arrivals (or any other specified arrival process). Stop-level data will permit a more accurate estimate than timepoint data. Estimates will be still more accurate if they use boarding data to know how many passengers arrived during each headway and to average waiting time over many stops. Records of pass-ups make the estimate still more accurate; without them, one must either assume no pass-ups, or make rough assumptions. To our knowledge, such an analysis has not yet been developed for AVL-APC data, but it has been applied at the MBTA (Boston) using automated data from rapid transit lines. After experimenting with various measures of the impact of headway variation on passenger waiting time *(22),* they settled on two indirect but simpler measures: percentage of passengers waiting more than one scheduled headway (calculated assuming uniform passenger arrivals and no pass-ups), and percentage of headways greater than 1.5 times the scheduled headway *(23).*

A **graph of successive trajectories,** an analysis used at a few transit agencies, is valuable for giving insight into bunching. Eindhoven's analysis package uses a different color for each vehicle, permitting one to follow a single vehicle to look for vehicle-specific effects or operator-specific trends. It also shows the scheduled as well as actual trajectory. One way to develop and test better control policies is to carefully investigate trajectories.

For such an examination, being able to view all relevant data – ons and offs, lift use, control messages – would be helpful.

Headway analysis requires data from successive buses. More than any other analysis, it calls for a high rate of data recovery, since one lost trip means two lost headways. Doing headway analysis with a sample of instrumented buses poses the logistical challenge of getting all the buses operating on a route to be instrumented buses. With level B location data, headway can be monitored at timepoints. Level C location data plus door open times or stop-start times is preferred, so it allows headways to be analyzed at stops, and because headways can change substantially between timepoints. A link to the schedule database is needed to compare headways with the schedule.

2.3.6 Demand Analysis

Average passenger ons and offs by stop and its derivatives, **load by segment** and **passenger-miles,** are necessary inputs for many route and schedule planning tools, including headway determination; designing short turns, zonal service, express service, and alternating deadheading; evaluating stop location changes or stop amenity additions; and National Transit Database reporting. These tools require on-off counts and stop-level records, and are part of most APC systems.

Sampling methods for NTD passenger-miles reporting that take advantage of APCs are needed. Only a few agencies with APCs have NTD sampling plans tailored to their capabilities (custom sampling plans are permitted by the FTA if certified by a qualified statistician). Sampling methods for NTD passenger-miles reporting that take advantage of APCs are needed. Only a few agencies with APCs have NTD sampling plans tailored to their capabilities (custom sampling plans are permitted by the FTA if certified by a qualified statistician). Many are still using FTA Circular 2710.1A, because it remains the only standard sanctioned sampling plan. However, it is extremely inefficient for a transit agency with APCs to use this plan because it involves using only a tiny fraction of the APC data collected, and presents unnecessary logistical challenges in getting instrumented buses onto randomly selected trips. It would be much easier on transit agencies, and yield much more accurate estimates, to use data collected from the entire APC subfleet, which in many transit agencies amounts to tens and hundreds of thousands of trips a year.

This plan, developed with manual sampling in mind, involves 549 one-way trips sampled at random. Meeting the plan is a primary constraint on how APC-instrumented buses are assigned at several agencies, presenting logistical challenges. It would be much easier on transit agencies, and yield much more accurate estimates, to use data collected from the entire APC subfleet, which in many transit agencies amounts to tens and hundreds of thousands of trips a year.

Analysis of **load variability** is important for scheduling and for evaluating operational control, and is part of a few agencies' suite of analysis tools. On routes with a demand-based headway, the main goal is often to prevent extreme crowding from occurring more than infrequently; this suggests that headways should be based on extreme values of load (e.g., 90-percentile maximum loads) rather than on average loads at the peak point. Directly relating headway decisions to observations of extreme load focuses attention on improving operational control to prevent overcrowding.

Analysis of **crowding** usually involves a comparison against vehicle size or seats and the loading standard for the relevant period and type of service. Besides load at the point of greatest overall passenger volume, customer-centered measures of crowding should be calculated such as maximum load (regardless where it occurs), amount of time spent at various levels of crowding, and number of standees and passenger-minutes spent standing. Some APC users analyze measures of this kind. Analyses of crowding should ideally be correlated with pass-up records and with headway so that transit management can see to what degree crowding problems might be solved with better headway control.

Time-dependent demand analysis is a desirable tool for determining trip start times that will lead to balanced loads. Using passenger activity data combined with measured headway data, and averaging over many days, one can derive the passenger arrival rate as a function of time. Combining these arrival rates in small (e.g., 1-minute) time slices using a reference frame that moves at the speed of a bus allows one to predict more accurately the passenger load on a trip based on its assigned start time and the start time of its leader. Such a tool, currently under development at the request of the transit agency of the Hague, requires stop-level records with on-off counts (or farebox transactions, for a rougher estimate of passenger load), and that all the buses serving a route be equipped.

When incident codes or the farebox is used to register overloads and special uses, combining that data with location data at detail level C can be helpful for analyzing **overloads, lift use, bicycle use, and other special events.** This feature is part of many AVL and APC systems. However, it is difficult to enforce operator-activated event codes (as opposed to automatic codes from networked devices) unless those codes are being transmitted in real-time by radio and subject to immediate supervision.

If fare media include unique ID's, as is the case with both magnetic and smart card media, combining farebox transactions with location data at detail level C permits **transfer activity** and **linked trip analyses**. Linked trip analysis is especially important in Canada, where linked trips are the standard measure of transit use. Efforts using both magnetic stripe and smart card media indicate success in inferring a passenger's exit point by tracking where the fare ID is next used to enter the system *(15, 16);* there are still no routine tools, however. Without location data, it is still possible to do a route-to-route analysis, from which location can often be inferred since most route pairs have only one logical transfer point. This approach is being used experimentally at the Chicago Transit Authority.

2.3.7 Geographic and Planning Analysis

Equipped vehicles can be used to **geocode stops and other points of interest**. In one deliberate version of this tool, reported to be part of a new Israeli APC system, a transit analyst with a laptop computer connected to the onboard console rides along as the bus follows its route, stopping at each stop or point of interest while the transit analyst keys in stop identifiers at the right moment, creating a file of accurate stop locations. A less intrusive version has the bus making frequent records in order to **map a bus path**, e.g., through a shopping center parking lot, new subdivision, or detour.

Routine **comparison of measured vs. nominal stop locations** is helpful for detecting errors in the stop location database, or places where service differs from its official plan, such as buses stopping away from the official bus stop because the stop is inaccessible due to construction, or has somehow "migrated" based on passenger convenience. At least one vendor recommends this type of study during the installation phase; we don't know of any regular reports with such comparisons. Variation in observed stop location from bus to bus can also be an indication of problems in the matching algorithm, in GPS or wayside signals at that location, or in equipment.

Planning studies beyond the route level generally involve a geographic database for answering questions like what is the likely demand for a new routing. There is usually a wide gulf between route-level data (on-off counts) and geographic demand data (travel demand and modal preference from zone to zone). Studies that update geographic travel demand data using on-off data are usually specialized and done only occasionally as part of large study. More routine methods are needed to **relate on-off data to demand rates in traffic analysis zones and to geographic databases** so that geographic demand analyses benefit from accurate and timely APC data.

Likewise, there is a need to **relate service quality data to geographic databases**. This would allow analyses of variations in service quality in different part of the service area, and could be used in travel demand models to explain and predict transit mode share.

2.3.8 Utilities

Analysis suites for AVL-APC data generally include utilities that allow managers to check for faulty components of the data collection system. Examples are identifying signposts that weren't heard from; identifying areas with poor GIS reception; identifying buses whose odometers vary from the distance covered according to their route; and identifying buses with high APC count discrepancies.

2.3.9 Other Operations Analysis

The wealth of data that an automated location system affords enables the transit agency to analyze its operations to find ways of improving its service or efficiency. The example analyses presented in this section will probably not be routine, but may have a place as part of an agency's research into improving its operations.

Published *(24)* and unpublished studies indicate that much of the variance in running time and schedule adherence can be explained by operator behavior. An **analysis of performance by operator** is a valuable tool for training operators and for experimenting with different methods of supervision and control. To account for the bus-bunching phenomenon, an operator's performance must always account for the position of its leader. Performance elements can include schedule adherence, running time, layover time and on-time dispatch, headway maintenance and bunching, and more. To our knowledge, only specialized studies have been done using AVL-APC data at Tri-Met.

Correlations between data items will be valuable for examining how operators adjust their running time when late or early, or when they have a heavy load, or when their leader is running hot. Do operators that are beginning to run early intentionally slow down, and do operators that are getting behind speed up? Being able to focus on individual operators may reveal patterns or early or late departures. Agencies may not want to use the AVL data directly to discipline operators, but it can certainly be used to help dispatchers and supervisors better target their efforts and gauge their effectiveness. Operator performance can be analyzed as a function of experience in general and experience on a route. Uncovering systematic differences between operators, whether due to experience or any other factor, can be valuable in schedule planning and in planning methods of supervision to prevent bunching.

Analyzing operator performance must be done with careful respect for operator acceptance and safety. If used for discipline, there is a danger of sabotage. Still worse, safety can be compromised if operators are punished for getting behind schedule. That does not necessarily mean that operators should not be given feedback on their performance. Experience with automated data collected in Rotterdam's tram operation shows that operators may enjoy getting a written record of their performance – for the first time operators had written evidence to show their family what a good job they were doing in staying on schedule.

This analysis can be done using timepoint or stop-level records. Incident codes and control messages are useful for omitting or flagging trips with unusual circumstances or special instructions. On-off counts and farebox data can allow the analysis to account for passenger demand and crowding.

Dwell time analysis, in which door open time is correlated with on, off, and load data, allows one to better understand the dwell process and to examine the impact of relevant changes, such as changes in fares, changes in fare payment method or equipment, introduction of low floor buses, changes that affect crowding levels, and traffic-related measures that change the time it takes for a bus to berth, its ability to approach the curb, and the delay it encounters returning to traffic. NJ Transit has developed such a tool. It requires stop level data with door open and close times and on-off counts. Bus stopping and starting times allow analysis of initial and final delays. Farebox transaction data is helpful for indicating fare type. Incident codes can be helpful to flag lift or ramp use, fare disputes, or other problems that might be registered.

Layover analysis is valuable to determine the degree to which on-time dispatch problems are due to insufficient recovery time versus lack of control and discipline at the start of a route or at pull-out. Level B or better location data is sufficient, but only if the data permit one to determine when a trip has ended and when the following trip begins. Often layover points allow the operator some freedom in where bus is driven and where it is parked, making it more difficult to match vehicle location to a schedule than when a bus is on route. An operator's leaving and reentering the bus creates a dilemma for the passenger counting system – are we seeing the first boarding passenger of the new trip, or is it just the operator getting back on a few minutes before the trip begins? These questions could be better answered by using improved algorithms and by capturing more data, such as length of time the bus is standing, passing dummy timepoints in the layover area, engine on/off status, and whether the operator's seat is vacant. To analyze pull-outs and pull-ins (on time? related as they should be to first trip start time and last trip end time?), off-route location data will be needed.

Tools that analyze the causes of bunching are needed to help guide efforts to prevent lateness and bunching.. How much of bunching is due to starting late? To a leader that is running early? To a random surge in demand? To congestion on the bus slowing down the boarding and alighting process?

Ad hoc analysis capability is valuable for analyzing **control effectiveness.** This capability would allow analysts to see whether the control actions they took or messages they sent had the desired effect. This capability involves the analysis tools described earlier (e.g., graphing trajectories, schedule adherence), plus access to incident codes and control messages. For example, when Tri-Met implemented a policy for short-turning in an attempt to improve headway regularity in the p.m. peak, they used their APC database to examine its effectiveness.

Ad hoc analysis capability in the areas described above are valuable for **before / after studies** to evaluate the impact of changes in schedule, control, or infrastructure. Likewise, it permits analysis for **special events** and **special weather conditions** by allowing analysts to focus on particular times and places of interest.

2.3.10 Passenger Information Verification

Information supplied to passengers regarding operations relies primarily on the schedule and, for real-time information, on real-time AVL data. Archived data with location detail level C, plus door open time, can be helpful in **checking accuracy of predicted arrival times and next stop announcements** and in developing improved prediction algorithms. Estimating time till a vehicle arrives at a downstream stop is related to, but different from, determining a new scheduled running time. Various analyses could be helpful, including those that account for the running time of previous buses, the relative speed of the subject bus, the schedule deviation, and the load and headway of the subject bus (to predict dwell time at intervening stops). Valuable supplementary data include incident codes and control messages (for excluding certain trips, or analyzing special situations) and on-off counts. To our knowledge, only specialized analyses have been done (e.g., *(25)*).

Another valuable analysis would be **monitoring route data in the farebox and destination sign**. If captured in the AVL-APC data stream, they could be analyzed with location data to see if they are correct and change where they should change. To our knowledge, such an analysis is not yet being done.

2.3.11 Payroll Verification

Archived location data at any level of detail can be helpful for verifying a driver's actual duty against the schedule and payroll databases. The location data can help resolve payroll disputes. Likewise, investigating operator duties with overtime claims can be helpful to verify, for example, whether an operator made an unscheduled stop during pull-in. Such analyses have already been discussed under the heading Targeted Investigations.

Using AVL or APC data to discipline operators always carries the risk of controversy (Big Brother) and sabotage. Some agencies report using their archived data only to detect patterns of abuse, which are then used to direct supervisors in their traditional role.

2.3.12 Maintenance Management

If maintenance incidents are registered, then location data with the relevant incident codes will permit an **analysis of maintenance incidents**. Location data at detail level A will permit correlation with location-specific factors. Detail level D will permit correlation with operational performance measures such as maximum speed. GPS altitude data, permitting one to infer grade, may be useful. On-off counts permit correlation with load. Including control messages in the database can be useful if there are messages that relate to the incident. A link to the maintenance database can also be useful to verify whether maintenance incidents are being properly registered both onboard and in the garage, and to identify causal factors related to a bus' maintenance history.

Trip recorder-type functions of AVL-APC data should permit the calculation of **measures of vehicle demand** that may be more effective than mileage in maintenance management. For example, detail level D can be configured to count acceleration / deceleration cycles, and detail level E would permit a study of speed profiles (e.g., rapid acceleration) that might provide insight when matched to a vehicle's maintenance history. A still more sophisticated analysis could use data on passenger load and grade (from GIS, gyroscope, or GPS altitude data) to better model demands placed on the drivetrain, steering, and braking system. Records of **door and wheelchair lift use** might also be useful for maintenance.

Analysts will want *ad hoc* analysis tools to relate these more detailed measures of vehicle use to various items in the vehicle's maintenance history in order to establish relationships, design improved operation and maintenance strategies, and to evaluate changes to operations and maintenance. Ideally, tools might be developed to **analyze trends** and to **predict and prevent failures.** For just this reason, NJ Transit plans to enhance its onboard system to monitor a variety of mechanical health indicators as well as grade (using GPS altitude). Measures they already analyze with data currently available include engine hours, engine idle hours, accel-decel cycles, time spent at various speeds, and miles traveled.

2.3.13 Strategic Planning

For almost any item of interest, it is desirable to analyze how it may have changed quarter by quarter or year by year. Analysts will want to know trends in demand, in running time on particular routes, in schedule adherence, in lift use, and so on.

While every transit agency does trends analysis, it is usually done outside the AVL-APC database and analysis suite using exported results. One challenge to doing trends analysis within an AVL-APC database is that schedules and routes change, complicating comparisons to previous years or periods.

2.4 Data Analysis Practice

2.4.1 Analysis Software

Software used to analyze archived AVL-APC data have come from five different sources: the AVL-APC vendor, a scheduling software supplier, a third party with a standard product, a custom software developer, and in house. Each arrangement has its own advantages and drawbacks.

Software developed in house. Most analysis of archived AVL-APC data uses home-grown software tools. This arrangement has worked well for some agencies, allowing them the flexibility to adapt to their particular needs and enterprise databases, and ensuring that tool development is closely tied to need and likely use.

Since the mid-1990s, self-developed database and reporting software for AVL-APC data has used commercial off-the-shelf (COTS) database platforms on PC networks. COTS platforms have the advantage of being less expensive and benefit from regular upgrades, necessary in this age of technological advance. Coding for standard and *ad hoc* reports is prepared either in a database query language or using report-generating software such as Crystal Reports and Brio. Analyses that demand more complex calculation are often performed with spreadsheets or statistical analysis packages, using database software to retrieve and export the data for analysis.

Tri-Met is a leading example of an agency using archived AVL-APC data analysis. Their integrated AVL-APC system provides stop-level records that, for the 65 percent of the fleet, include passenger counts. The data is stored and managed in an Oracle database. Using a query language, selected data (e.g., by route, direction, times, dates) can be extracted. Extracted tables are then imported to SAS for statistical analysis; scripts for standard analyses are stored and reused. Sometimes results are imported to Microsoft Access for easier formatting.

King County Metro, with separate AVL and APC system, uses multiple databases and applications. Its AVL data is stored in an Informix databases. For schedule deviation analysis, scripts coded in Microsoft Access provide a friendly user interface for selecting AVL route, direction, time, date range, and so forth. The analyses themselves were programmed in query language and are performed by Informix, which produces output in the form of Excel tables and graphs. Analysts may do further manipulations of the Excel tables. For running time analysis, a query language program runs every two weeks on the AVL Informix database, extracting data that is then input to their scheduling package that includes an add-on software product for running time analysis. Raw APC data is kept in an Oracle database. Using programs prepared in the Focus query language, summary records are created and exported to a Microsoft Access database, which has been programmed to offer a friendly user interface and nice reports. There are also standard reports done using query language from the original databases, as well as *ad hoc* reports using query language.

Metro Transit of Minneapolis, a third example, did running time analysis of data from its now obsolete AVL system using macros written in Microsoft Excel, once the analyst had extracted the data of interest from the database. In its new AVL system, now in implementation, they are working with the AVL vendor to define analysis and reporting needs; they plan to share responsibility for development of analysis software.

A final pair of examples is New Jersey Transit and Broward County Transit, whose APC / event recorder and AVL systems (respectively) are operational and expanding. They are using COTS database platforms for data management and COTS report generating software Brio and Crystal Reports for analysis.

Unfortunately, developing one's own database and reporting software demands resources and expertise that many transit agencies lack. Also unfortunately, tools developed at one agency are usually not transferable to another.

Software supplied by equipment vendors. Software supplied by APC vendors provide useful reports including on / off / load profiles, running time distributions, and on-time performance. However, it usually lacks flexible query capabilities.

Historically, AVL vendors provided software related to real-time applications only; for archived data analysis, their job ended when they handed the transit agency the data. Often the only archived analysis tool is the ability to play back the AVL data stream. Some AVL suppliers include a genuine database and analysis function, but tend to offer only elementary analyses such as on-time performance percentages and reports on how often various event codes were transmitted. For two of our case study agencies, AVL vendors are developing more comprehensive database and analysis capabilities as part of their procurement contracts.

Software that is coupled to onboard equipment limits the flexibility to add other onboard equipment, or to replace aging equipment with equipment from a different vendor. There is a chance that the vendor will go out of business, or will not keep improving the software. A note of caution comes from reviewing experience with farebox data. While the major electronic farebox vendors also supply software for analyzing farebox data, most larger U.S. agencies who rely on farebox data for estimating ridership and analyzing revenue have found that they had to export the data to a database developed in house on which they could run their own reports because the vendor's software didn't provide the flexibility they needed.

Software supplied by scheduling system vendors. Analysis programs offered by scheduling system vendors focus on analyzing running time data to suggest scheduled running times. For example, one vendor's software analyzes segment level data (timepoint to timepoint), and provides some useful tools for choosing scheduled segment running times. Because scheduled running times are tied to the scheduling system, it is suggested that scheduled running times can be semi-automatically entered into the scheduling system database. It was not intended for analyzing route-level running times (e.g., to look at extreme values to see how much recovery time is needed). Ironically, it does its analysis without reference to scheduled departure and arrival times or headways, and therefore cannot analyze schedule or headway adherence.

Software coupled to the scheduling system has many of the same disadvantages as software coupled to an equipment supplier. One case study agency that uses such a tool for running time analysis has to use its own database and software tools for other analyses and *ad hoc* queries.

Third party software. In the Netherlands, Delft University of Technology's Transportation Engineering Laboratory has developed the database and reporting software TriTAPT (Trip Time Analysis in Public Transport) for automated location and passenger count data. Various editions have been applied over the last 20 years to several Dutch transit agencies; the current edition is being used in Eindhoven and in the Hague. It features many useful single-route reports; excellent graphical representations, including proportional scaling to represent distance and time intervals; attention to distributions and extreme values as well as mean values; a professional user interface that allows users to select days and times to be included in an analysis; edit capability that allows an analyst to suppress outliers; and practical tools for suggesting scheduled route times. It has been applied with data gathered using a variety of automated data collection equipment, including APCs, event recorders, and AVL systems of different makes. Interfaces have been developed to scheduling system databases. It uses a custom database to speed processing, but includes an export and import utility so that data tables can be transferred to and from text files.

In Germany, the Hannover transit system Uestra developed their own database and reporting software for AVL data; a related spin-off company has recently commercialized it. An experimental transit planning-GIS system developed by the University of Montreal has been used with APC data at several Canadian transit agencies.

However, it relates to passenger counts and route planning, with no running time and schedule analysis.

In the U.S., to our knowledge, the first application of third party software for analyzing AVL data is underway at the Chicago Transit Authority for use with a new smart bus system that features stop announcements and event recording on all buses and passenger counting on a sample of the fleet. Called RideCheck Plus, it includes some GIS capabilities including links with demographic data and mapping capabilities. It has previously been used for analyzing ride check data and APC data.

Third party software for analyzing archived AVL-APC data has the advantage of not being tied to any particular brand of equipment or scheduling system. As a stand-alone product, it is likely to continually improve, unless the product is discontinued. It offers the benefits of standardization and replication. A major disadvantage of third party software in the U.S. is that transit agencies' funding mechanisms often forbid them to buy software "just" for data analysis, although such a purpose can often be justified within the context of a major AVL or APC system procurement as was the case at CTA.

Research-Oriented Analysis. Several specialized AVL analyses by university research teams have been reported in the literature, including University of Michigan analysis of raw Ann Arbor Transit Authority AVL data *(26)* and the earlier-mentioned analysis of MTA data done at Morgan State University. In both of these cases, the specialized processing required to analyze these datasets left them inaccessible to staff analysts. In contrast, Tri-Met's APC database, developed in-house, supports analyses by both staff analysts and researchers from Portland State University *(27)*.

2.4.2 Example Analysis Reports

Both TCRP (5) and CUTA (6) have published syntheses of practice that include numerous samples of reports generated from archived AVL and APC data. This report also contains a large number of example analysis reports in its case studies.

Most common are demand, schedule adherence, and running time analyses. Some of the best practices evident in those reports include:

- allowing user-requested dates and periods of the day (or trips)
- showing the sample size and the distribution, not just the mean, of items such as running time
- plotting trajectories of sequential trips
- distinguishing operators to allow the analyst to look for operator-specific trends
- automatic generation of performance summaries for high-level management distributed daily on the worldwide web
- showing correlation of load with headway or schedule deviation
- reports that go beyond echoing or summarizing data to support a specific management function

An example of the latter is a set of three reports used in Eindhoven to support schedule building. One analyzes historic trip time data to identify periods of the day with relatively homogeneous travel times. Another uses historic data for a given period of the day to recommend segment running times based on a user-selected statistical criterion (e.g., x-percentile running time, with the user selecting x). A third shows what travel time would be if intersection delay were eliminated, an analysis aimed at suggesting a schedule for routes when signal priority at intersections is installed.

2.4.3 Database Software

Virtually all the North American agencies reviewed use COTS database platforms for their AVL-APC data archives. They have the advantage of low cost, manufacturer support, and standard query languages. Transit agencies often have expertise in a database system, and prefer that their AVL-APC data use the same one.

A disadvantage of COTS database platforms and reporting software is that they can be slow when a lot of data is involved. More powerful report generating tools (available at 3-10 times the cost of their low-end counterparts) help overcome this problem by periodically pre-staging the data most likely to be used in reports and analyses. Response speed for large data sets can also be reduced by use of special data structures optimized for fast data retrieval.

2.4.4 Data Storage at Several Levels of Summarization

AVL-APC data is archived at several different levels of summarization.

The first level, the lowest level, is the raw data itself. Nearly all agencies archive their raw data streams in case they are needed for incident investigation, to investigate a problem that occurred in later processing, or as a test data stream for new methods. However, the raw data is rarely used.

The second level has records mostly at the stop-level or timepoint-level. This level includes all event records; events in addition to stop and timepoint events include sign-on changes, and other events as the system is configured. Most analyses are based on this level.

Higher levels of summarization at the trip and route-direction-period level are used at one agency (King County Metro) to reduce storage and speed processing for older APC data. Once data is a year old, it is little used except for trend analysis.

2.5 Integration with GIS

It is ironic that AVL-APC databases full of location data are not usually part of a geographic information system (GIS). The fact is that most common analyses are based on the route network (routes, stops, timepoints), which, once defined in the schedule file, can be analyzed without further reference to their geographic locations. Simple and common GIS uses at the front and back end of AVL-APC data analysis include using GIS base maps as inputs to tracking and matching algorithms, and use map formats for reports. Examples of the latter include a route map showing on-off activity by bubbles at stops, or a route map with colors indicating speed on each segment.

Integrated with GIS, AVL-APC data can permit interesting and improved GIS analysis. For example, by integration APC on/off data into a GIS, one could predict demand on a proposed route much better than using demographic data. With better GIS integration, one could query for the on-time performance on a given highway, incorporating all routes using that highway. Integrating transit speed and reliability data into GIS can permit more accurate demand model predictions. We are still a long way from having the models and interfaces needed to automatically incorporate AVL and APC data into the geographic travel time and transportation demand files used in long range planning by transit agencies and MPOs. A large part of the reason for the weak relationship between them is that demand models are concerned with linked trips, while APCs capture only ons and offs. Demand models would most benefit from an integration of APC counts and location-stamped farebox data that includes transfer information.

The University of Montreal has developed an experimental transit-GIS system used in three Canadian transit agencies. A few vendors, including a scheduling software vendor and a 3rd party vendor of ride check and APC analysis software, include GIS analysis of existing ridership in their analysis suites.

CHAPTER 3

FINDINGS AND GUIDANCE

This chapter summarizes findings and offers guidance for moving toward productive utilization of AVL-APC data. Findings and guidance are offered in four general areas:

- System Design and Data Capture
- Analysis and Decision Support Tools
- Quality and Integration of Other Data Sources
- Organizational Issues

3.1 System Design and Data Capture Issues

3.1.1 Stop Level or Timepoint Level Detail?

There are several differences between designing for stop level data and for timepoint level data. Those differences make this question – stop-level or timepoint-level data capture – a key decision point in system design. Because in practice schedule analysis is generally performed at the timepoint level, timepoint level data is adequate for traditional scheduling and schedule adherence applications. This logic has driven most new AVL systems in the U.S. to provide archivable data at the timepoint level, and not at the stop level. In contrast, in European practice, every stop is a timepoint, and so location systems there naturally provide stop level detail. Also in contrast, APC and stop announcement systems, which are inherently oriented toward the stop level, provide stop level detail.

One difference between a design for timepoint-level versus stop-level data relates to onboard system architecture. Stop location recognition generally involves door sensors; timepoint recognition does not. Recognition of timepoints (and of stops when a bus does not stop) is based, for GPS-based systems, on entering or departing a zone, usually circular, about the point's nominal coordinates. Including door sensors in the system architecture has important implications for data capture, as information on when doors open and close are useful for several analyses. And while accuracy of timepoint recognition unaided by door sensors is usually adequate, accuracy is better when door sensor data is available. The difference is particularly important when variations might cause some buses to be registered as passing a timepoint as they approach a stop, and others as they leave the stop. Such inconsistency introduces random errors on the order of half a minute to data used for schedule adherence and running time analysis.

A second difference is in system capacity. One might imagine that by simply designating every stop a timepoint, a timepoint-based system could be converted to one that captures stop-level data. However, bus routes typically have 10 to 20 times as many stops as timepoint. Many AVL systems that rely on real-time radio transmission for data recording, especially those in larger metropolitan areas where radio channels are scarcer and have to be shared among more buses, do not have the capacity to handle messages for every stop. There are also APC systems that transmit stop records over the air; however, they involve only a small fraction of the fleet, and may run up against capacity limitations if such a system were expanded fleetwide. In contrast, AVL and APC systems that record data in an onboard computer have no such capacity limitations.

While timepoint data is all that is needed for traditional scheduling and schedule adherence applications, there are several reasons to consider stop-level data:

Accuracy. Time / location accuracy will be improved because door sensor data help determine when a bus arrives and departs at the stop closest to the timepoint.

- *End-of-line problems with timepoint data.* Agencies with only timepoint records generally find data for the first and last segments of the line to be unreliable because of unpredictable operator behavior at the end of the line. With stop data, operations near the end of the line are better tracked and far less information about the final segment is lost.
- *Holding and net running time.* When determining needed running time from data on actual running time, it is helpful to identify and subtract time operators spend holding because they were ahead of schedule. Identifying holding time requires records of when a bus arrives at and departs from any stop at which an operator may hold, which can include stops that are not timepoints.
- *Better analysis.* Analysis of headways, estimated passenger waiting time, and other performance and service quality measures are much rougher if done at the timepoint rather than the stop level. To the extent that transfers occur at stops that are not timepoints, stop level data is needed to analyze connection protection.
- *Stop-level schedules.* The possibility of getting stop level data raises the question, why don't American transit agencies follow their European counterparts and schedule bus service for each stop? Stops are where passengers wait to meet buses, and passengers want to know when to expect the bus to arrive. The enthusiastic reception the public has given to real-time next arrival prediction systems underscores the public's desire for stop level schedules. While agencies may still prefer to do schedule-building in terms of timepoints rather than stops, stop level data is still needed to accurately allocate segment running time between timepoints, rather than assuming a constant average speed. Not providing the public with stop-based schedules is probably a good example of practice being driven by the historic lack of data – practice that should change as stop-level data becomes available.
- *Support signal priority.* While traffic signal priority is becoming more accepted and implemented across the country, most traffic and transit agencies prefer that priority be given only to buses that are behind schedule *(28).* While it is possible to condition priority on schedule deviation at the latest timepoint, it is more accurate to use schedule deviation at the latest stop, since priority can make a bus that was late at one timepoint become early before it reaches the next. Developing stop-level schedules and measuring schedule deviation at the stop level for effective schedule-based transit signal priority requires stop-level data.
- *Future integration with passenger counters and other onboard devices.* A location system configured to make stop-level records can readily be expanded to add passenger counts to its stop records if sometime in the future the transit agency decides to install passenger counters on its buses. Onboard data recording also allows the system to make records of data from other devices, including the farebox, destination sign, lift, door sensors, and engine and drivetrain sensors.

In conclusion, the choice of timepoint versus stop level data is a critical decision issue for AVL system design. Timepoint-level data is sufficient for traditional scheduling and schedule adherence practice. However, stop-level data gives a transit agency more capabilities. It allows an agency to improve its scheduling, operations control, and service quality monitoring. It is an important component to transit signal priority. And it is necessary to develop accurate stop-level schedules, a public information enhancement that is likely to improve ridership and customer satisfaction.

3.1.2 Location-at-Time vs. Time-at-Location

AVL systems are capable of providing time-at-location data – i.e., a record of when a bus passed a predetermined point of interest such as a stop or timepoint – or location-at-time data, i.e., a record of where a bus was at a given time. Time-at-location data generally requires real-time tracking of bus location against a route profile. Location-at-time data can be obtained without tracking, and is typically provided by polling schemes.

For management use of archived data, time-at-location data is strongly preferred to location-at-time data. Almost all time-oriented applications of archived AVL-APC data (schedule adherence, running time, schedule deviation) are based on arrival or departure time from locations of interest.

Time-at-location can be estimated in post-processing from location-at-time by interpolation. The accuracy of the interpolation depends on the polling cycle. Maximum likely interpolation error is about half the polling cycle, so that, for example, if buses are polled every 60 s, the maximum likely interpolation error is 30 s. Therefore, if the polling cycle is short, reasonable accuracy can be attained. At least one vendor of next arrival systems predicts bus arrival time at each stop based on location-at-time (poll) data, and records predicted arrival times, which can be used in analysis as estimates of actual arrival time. The vendor claims that a test shows that most predictions to be quite accurate. With a short polling cycle and careful monitoring and tuning of the arrival prediction process, estimation errors may indeed be small enough to support running time, schedule adherence, and similar analyses. More testing of such an approach is needed, however, for it to win the confidence of managers and other data users.

3.1.3 Between-Stop Records: Detail or Summary?

Data on what happens between stops is of less importance than stop data. Many systems with onboard computers are capable of making frequent records of speed, location, heading, and so forth. "Frequent" could be as often as every second. They can also make records when triggered by events such as crossing a speed or acceleration threshold. What data on performance between stops is worth capturing? And should full detail be kept in the analysis database, or is a segment level summary sufficient?

For general analysis in the analysis database, stop level summaries appear to be sufficient. Example summary measures are maximum speed (practiced at Tri-Met), number of seconds spent below crawl speed (practiced in Eindhoven), and number of accel-decel cycles (being developed at NJ Transit).

Full detail is most useful for incident investigation, suggesting that full detail be stored in a lower priority archive (perhaps less accessible, or with a more limited life). Full detail can also be used to investigate the path taken by a bus in order to map a route through a shopping mall parking lot, for example, or for research.

3.1.4 Is Exception Data Sufficient?

While exception data can be used for some real-time applications, it cannot support most operations analysis methods. New systems should avoid being restricted to exception data if they hope to do analysis of the archived data. This is not to detract from the value of exception analysis and reporting, which can be done both in real time and off-line using a complete stream of data as opposed to exception data only.

3.1.5 Onboard Integration with Unified Location Capability and Interfaces

The smart bus design, with a central computer tracking location and supplying location information to other devices that need it, is clearly superior to stand-alone systems that each determine their own location. A shared location system will usually be as accurate and intelligent as the most demanding application (which often is stop announcements), providing other applications with more accurate location information than they might otherwise have. Maintenance and improvement are simpler as well.

Using the SAE J1708 standard for on-vehicle networking makes product integration easier and less expensive. It makes systems expandable as devices are added or upgraded so as to become networkable.

Integration to allow a single interface with the operator and a single link to the central computer system also has many benefits. A single operator interface simplifies the operator's task and makes getting valid sign-in data more likely. Sharing a single interface to the central computer – usually a high-speed link as part of a wireless local area network based at the garage – is cost-effective and allows easy data upload and download and software upgrades.

Transit agencies have been successful in specifying J1708 networks and compliance in their procurements, and suppliers have been quick to announce J1708 compatibility.

3.1.6 Real-Time vs. Off-Line ID Verification and Tracking

Matching observed location data to the schedule is much easier when the bus has valid sign-in data. System design should therefore maximize the chances of getting valid sign-in data. Features that increase the chances of getting good sign-in data include:

- Integrated smart-bus design with a single operator interface.
- Range and validity checks during sign-in.
- Automatic sign-in based on daily vehicle-block assignments (as practiced at Metro Transit, Minneapolis) or operator dispatch, or using an operator-held smart card to convey operator ID and assignment information to the bus.
- Verifying and updating sign-in data during operation allows sign-in errors to be corrected far more easily than after the fact. For example, if bus location is tracked in real-time, discrepancies between entered run ID and what's actually being run can be noted. Operators can be asked to correct the sign-in information or explain the discrepancy. With AVL-CAD, discrepancies can be automatically brought to the attention of supervisors for correction.

Verification during post-processing can automatically correct some errors and facilitate matching. Even if errors can't easily be corrected, post-processing verification can be a deterrent to future sign-in failures, depending on how seriously sign-in failures are treated by management. In practice, sign-in failure is often not taken seriously when the failure doesn't directly affect service, as with farebox sign-in. Sign-in failures that directly affect operations, e.g., preventing the stop announcement or radio system from working, are apt to be taken more seriously.

3.1.7 Onboard Data Recording vs. Over-the-Air Transmission

Transmitting data by radio to a central computer is necessary for real time AVL applications, and can also be used to record data used in off-line analysis. If the data is coming in to the control center anyway, why not just record it there? Radio-to-central-computer has been used effectively for recording stop-level APC data from a sample of the fleet and for recording timepoint data from the entire fleet (sometimes both applications in the same radio system).

However, using radio as part of the recording scheme is inherently limiting due to radio channel capacity. Radio system performance depends on bandwidth available, fleet size and fraction of the fleet instrumented, message frequency (from every stop, or only from timepoint), and message length. Limitations tend to be most severe in large cities where radio channels are scarcer and fleet size is greater.

Therefore, before accepting radio-to-central-computer as a means of recording data, a transit agency should consider whether it has sufficient capacity for any envisioned expansion of the data collection system. While the radio may be able to handle timepoint records from every, and/or stop records from a sample of buses, could it handle stop records from every bus should a decision be made to go to stop level reporting? Could its messages be expanded to include door sensor data, or fare collection data, or other data that might be added?

Another alternative, used by next arrival prediction systems in several cities, is to use commercial frequencies such as cellular digital packet data (CDPD). This alternative, where available, overcomes capacity limitations, and reduces the fixed cost of radio infrastructure, but involves a cost for every message transmitted.

Recording data on board entails a cost (falling with technological advances) to equip all the vehicles with sufficient storage capacity, but presents no practical capacity limitations. This arrangement is consistent with the smart bus architecture, with the onboard computer available to control a variety of tasks. For systems with AVL/CAD, this arrangement separates data recording for real-time purposes, which must go by radio and therefore must be limited in frequency of messages and message length, from data recording for off-line analysis, which becomes essentially unlimited in capacity.

Onboard storage is probably necessary in many cities to get fleetwide stop-level data. It is certainly necessary to get speed data at frequent intervals, records of each passenger entry and exit or farebox transaction, and frequent records of speed, location, and heading. These data items are needed to support the trip recorder function, advanced algorithms for route matching and identifying holding time, and some demand analyses.

In conclusion, transmitting data over the air only can support traditional schedule and demand analyses that require timepoint-level operations data and a small APC sample. However, using onboard data storage frees the agency from capacity limitations, enabling more complete data capture, which in turn enables a richer and more accurate set of operations tools. Onboard data storage is also helpful for integrating with other devices in a smart bus architecture.

3.1.8 Capturing Data from Single-Purpose Systems

Against the trend toward onboard integration is an opposite trend toward single purpose, stand-alone systems with location information for passenger information purposes (next stop and next bus information). Their design makes them inexpensive and attractive in some environments, and raises the question of whether location information can be extracted from these systems to support operations analysis. In principle, it should be possible to record and process their operations data for off-line analysis. That possibility is beginning to get the attention of suppliers and their customers. For example, one vendor of next arrival systems is developing reports from archived data for its customers. (This vendor's product differs from standard AVL and APC systems in that data are all transmitted to the vendor's own computers and stored and analyzed there, with communication by internet and commercial wireless links to buses, public display panels, and transit agency planners.)

3.1.9 Is Equipping a Sample of the Fleet Sufficient?

Sample size and timeliness of data are affected by the fraction of the fleet equipped. With radio-based AVL systems, it is the custom to equip the entire fleet for real-time purposes, so archived data benefits from full fleet coverage. With APC, it is the custom to instrument only 10 to 15 percent of the fleet. Tri-Met is an exception, with about 60 percent of its fleet instrumented with APC and all new buses ordered with APC.

Equipment that Counts Passengers. Most analyses that require passenger counts can be done with 10 to 15 percent fleet coverage. A sample of this size is sufficient to estimate mean or total passenger activity, and will therefore support estimates of average peak load, total passenger-miles, and service quality measures tied to the number of passengers affected.

Determining extreme values of passenger load – e.g., the 95-percentile load, or the fraction of trips with load exceeding a threshold – needs 100 percent coverage. Extreme values are of interest on routes that are subject to overcrowding. Traditional practice, influenced by the historic paucity of data, uses average peak load for both scheduling and passenger service quality monitoring. However, on routes that are subject to overcrowding, analysis and decision tools based on extreme values are preferred because they more accurately reflect the impacts of crowding on both operations and passenger satisfaction. Therefore, transit agencies would benefit from having full APC coverage on routes subject to overcrowding.

Equipment that Records Operations Data. Analysis and decision tools using operations (time-location) data are best served with 100 percent fleet coverage. A smaller sample is sufficient to estimate mean values, but important analyses and tools depend more on extreme values. One example is the percentage of trips on a route earlier or later than a certain threshold. Another is 95-percentile running time, which is (or should be) an important determinant of scheduled cycle time (running time plus layover) on routes with significant running time variation. Headway analysis also requires that every bus serving a route be equipped. Full coverage also means that data will be timelier and is more likely to be available to investigate a complaint.

A mix of full and sample coverage can be achieved in several ways. One example is a fleetwide real-time AVL-CAD system supplemented by a sample of passenger counters; another is a fleetwide event-recording system, perhaps tied to a stop announcement system, in which a fraction of the buses also have passenger counters.

As Tri-Met has found, there are advantages to having full fleet coverage of both AVL and APC. With a smart-bus onboard architecture in which location data and route and schedule tracking are already being done by a unified location system, the marginal cost of adding passenger counters is far smaller ($1,000 to $3,000 per bus) than the cost of stand-alone APCs ($6,000 to $10,000 per bus). Of course, with fleetwide coverage there are many more passenger counters to maintain.

One reason to prefer fleetwide coverage is to avoid the logistical problems of rotating an instrumented subfleet. First, there's the need to develop a sampling plan, address special requests for data collection, assign the instrumented buses, and verify assignments. Second, sorting buses to make the correct assignment adds one more layer of complication. Third, with randomly occurring missed assignments, equipment failures, and data recovery problems, it is quite a chore to track which scheduled trips have been missed and make sure an instrumented bus is sent out on the correct assignment. Headway analysis is particularly demanding because of the need to have instrumented buses covering an entire route, which is especially complicated if the agency practices interlining.

Rotating specially equipped buses is also difficult because there are many kinds of special equipment, not just data collection equipment. Some buses have special ergonomic seats, and some operators need those buses. Instrumented buses are bound to age, and there are often political, labor, and management reasons for distributing new and older buses in ways that may conflict with a data collection schedule. Therefore, operations are substantially eased when the entire fleet is equipped.

Another benefit of full coverage is the ability to investigate complaints. Many complaints (the bus was early or wasn't there) can be investigated using operations data; however, overcrowding complaints are best investigated with passenger count data.

In conclusion, full coverage is recommended for time-location data, and something between 10 percent and full coverage for passenger counters, depending on the fraction of routes in the system that are subject to overcrowding. Full coverage with passenger counters carries advantages in operations logistics, timeliness of passenger count data, and ability to investigate complaints of overcrowding, and is therefore warranted in larger transit systems if the system architecture permits the addition of passenger counters at little marginal cost.

3.1.10 Incorporating Control Communications Data

Codes used to communicate by radio from the operator to the control center using a data channel (e.g., to indicate pass-up, lift use, or a missed relief) are generally captured in systems connected to the radio.

However, when supervisors (or controllers or dispatchers) tell an operator to deviate from the scheduled plan, the change is often not captured in the data stream in a form amenable to later analysis. Capturing the change in plan makes tracking easier, allows analysts to flag deviated service for exclusion from certain analyses, and enables one to analyze the effectiveness of the control action. Naturally, supervisors' immediate concern is to control service, not to provide a neat data archive. However, to analyze operations well, and particular to analyze the effectiveness of control actions, information on control decisions should somehow be coded and captured.

3.1.11 Fare Payment - Location Integration

Location-stamped counts or transactions from fareboxes and other fare payment devices such as smart card readers afford several useful analysis tools, including getting demand by stop (of particular value if there is no APC) and, if transactional data with transfer information is available, linked trip origin and destination. However, to date integration of farebox data with location data has been done only in limited experiments. There is somewhat of a chicken-and-egg problem: who wants location-stamped farebox data when there are no tools to analyze it, and who wants to develop the tools when there's no data for them?

Fare transaction data and location data can be integrated onboard, in real time, as well as off-line. There are two ways to integrate farebox and location data on board. One is for the fare device to broadcast its transactions on a J1708 network, where the fare transaction messages can be captured into the AVL-APC data stream. This arrangement has the advantage, especially if there is no APC, of making the fare transaction data available for a host of analyses of archived AVL-APC data. With a J1708 network, it is a simple matter to receive farebox transaction data (although new message definitions may be needed to get transfer information). The greater challenge is getting farebox manufacturers to provide fareboxes that that will broadcast their transaction information on the network.

A second approach to onboard integration is for the location system to transmit location information to the farebox, which is most readily done on a J1708 network. However, it would require substantial modification of the farebox data system to receive and record the location information. Analysis would then have to be done the farebox data stream. A disadvantage of having real-time location stamps in the fare transaction data stream is that location data generally benefits from postprocessing for better route and schedule matching. Therefore, adding fare transaction data to the location data stream, which usually gets the benefit of improvement during postprocessing, seems better than adding location data to the fare transaction data stream.

It is also possible to integrate farebox and location data after the fact using separate farebox and AVL data streams, using time as a key to correlate the streams. The challenges here with respect to system design are to synchronize the systems, and to see that they have common identifiers. This could be accomplished by a common sign-in interface and limited onboard integration in which the onboard computer broadcasts its clock time periodically to the farebox. Another way to synchronize is for fare collection devices to have inexpensive GPS add-ons used to receive GPS-broadcast universal time.

3.1.12 Learning Mode

A "learning mode" facility in which the location system is used to map stop or other locations is valuable. In one variation, implemented by an Israeli APC supplier, a traffic checker or analyst rides along holding a laptop computer connected to the onboard computer, entering codes for stops and turns at appropriate locations which can then be matched to the GPS coordinates from the vehicle location system. The paired data could either be stored as events in the AVL-APC data stream, or in the traffic checker's laptop.

A less deliberate variation is to use routinely gathered location information, together with door opening data and other available indicators, to map the route actually taken and stop locations actually used, and to compare them against the nominal route map in order to update (or even create) the route and base map.

3.1.13 Location Accuracy

Our review of practice suggests that location accuracy *per se,* as opposed to matching error, is not a significant problem. Well-documented problems with GPS signals in natural and man-made canyons can be resolved using odometers and gyroscopes as backup. Transit agencies generally require, and vendors comply with, reasonable accuracy requirements.

3.1.14 Passenger Counter Accuracy

Our review also finds general satisfaction with passenger counter accuracy. Still, transit agencies are looking for advice on how to specify and test for accuracy. Vendors sometimes use terms whose meaning is unclear like "98% accurate."

A great advantage of APCs is that, by counting both ons and offs, they offer a natural check. Large counting errors are almost certain to result in on-off imbalance, and can therefore be easily screened. The accuracy issues are then a) at what level of on-off imbalance should counts be discarded; b) what percentage of counts are imbalanced; and c) of the remaining counts, what is their mean and standard deviation of count error? Accuracy of automated counts can therefore be reduced to two main dimensions: how much data is so inaccurate that it gets thrown out, and how accurate is the remainder. There are also finer dimensions, such as how accurate the counts are on crowded versus uncrowded buses, or under different ambient conditions.

On-off discrepancies can occur for reasons other than mechanical counting error. Early in the morning, a bus might fill at the garage with operators before the counter is turned on; the first trip may then show many alightings, yet no boardings. Likewise at the end of the day, the system may shut down before everyone's gotten off. A mistake in identifying the end of a trip that results in on or off counts being assigned to the wrong trip will also result in on-off imbalance.

Experience with state-of-the-art passenger counters suggests that they may be more accurate than routine manual counts. Testing their accuracy therefore requires either a special effort by teams of well-motivated observers, or videos that can be reviewed and compared with automated counts off-line. One APC vendor has developed such a video system, which it uses internally for testing and analysis and makes available to customers for validation testing.

3.2 Analysis and Decision Support Tools

3.2.1 Growing Analysis and Data Needs

The transit industry is making a transition from being data poor to data rich. Its traditional analysis and decision support tools don't need much data, because they had to make do with little data. Automated data gathering systems will never reach their potential for improving transit performance and management unless the industry's analysis and decision support tools change to taking advantage of automatic data collection.

Service standards, always data driven, should become more sophisticated. For example, agencies use average load standards to prevent overcrowding, because overcrowding slows operations, promotes further bunching, and hurts passenger satisfaction, especially when passengers have to be passed up. However, overcrowding is not an average but an extreme condition, and therefore should be governed by analysis of extreme values such as the 95-percentile load. As an example, an agency might have a standard that average load at the peak point should not exceed 50 passengers. Not that 51 or even 60 passengers is necessarily a problem; what they want is for extreme loads, say loads above 65, to occur no more than a small percent of the time, and they have learned over the years that if the mean is kept below 50, occurrences above 65 will be relatively rare. However, this service standard does not recognize that some routes may have greater load variability than others. Suppose a route, through better scheduling and operational control, comes to have lower variability in peak load. Its average load could be allowed to rise to, say, 55, and occurrences of loads above 65 will still be rare. And suppose another route, due to poor operational control, has loads that routinely fluctuate between 30 and 70, with a mean of 50. It meets the simple service standard, but fails to provide the level of service desired. A load standard that is based on an extreme value such as 95-percentile load or fraction of loads exceeding 65 passengers would correctly reward a transit agency for its efforts to improve operational control.

As a second example, why should service standards relate to load at the route's average peak point, and not to load wherever the maximum load occurs, which is what passengers care about? The answer is data collection convenience – a point checker can quickly measure load at a given point on many trips, while finding the maximum load on each trip requires onboard data collection. With automated data collection, that limitation disappears, and service standards should adjust accordingly.

The same reasoning applies to scheduling running time and recovery time. A lot of scheduling practice is about avoiding extremes – keeping low the probability that a trip arrives so late that it forces the next trip to start late. Logically, that means focusing on an extreme value such as 95-percentile when determining combined running time plus layover, sometimes called half-cycle time. However, standard practice is based on mean running times plus a fixed percentage for recovery time, with subjective adjustments in response to complaints. Traditional practice, again bound by data limitations, doesn't adequately recognize that a route with low running time variability can afford to have a smaller half cycle time, and that a route with greater running time variability may need a longer half cycle time.

Transit agencies like Tri-Met, King County Metro, OC Transpo, and Eindhoven who have good automatically collected operational data are finding more and more uses for it. One vendor representative explained it this way: "We are selling an addiction to data." At first, agencies say they only need a little data to perform their traditional analyses. But once they have it and see what they can do with it, they want more. Eventually, their whole mode of operation changes as they become data dependent – or data driven, to use a more positive term.

3.2.2 Analysis Software

Examples can be seen of analysis and reporting software developed by transit agencies themselves, by equipment or scheduling system vendors, and by third party vendors. At some agencies, AVL-APC data has become systematically accepted through the agency and have made a large impact on operations planning and service quality analysis. At those agencies, scheduled running times are regularly compared with distributions of observed running time, and schedules adjusted as needed; likewise for crowding, with adjustments being made to headways. Planners have ready access to standard reports on schedule adherence and passenger counts. Service analysts use the database to do *ad hoc* studies looking, for example, at running time variability and recovery time need, and the difference in performance between more and less experienced bus operators.

However, development of analysis and reporting tools for archived AVL-APC data is still in its infancy. While some transit agencies have the staff skills and resources needed to develop their own database and analysis tools, most do not. Many of the software tools being used offer only a few analysis reports and lack useful graphical representations and flexible user interfaces. In house tools developed at one transit agency are usually not transferable to another. Tools provided by equipment vendors and scheduling software suppliers are generally limited by the purpose of the parent product, and provide only standard reports, not allowing *ad hoc* query of the database. Many of the analysis tools suggested in Chapter 2 have never been demonstrated.

There is a need for better analysis and reporting software that can be used by multiple agencies and independently of the make of data collection equipment. Transit agencies who have good archived AVL-APC data need better analysis and reporting tools to more effectively use the data, and those who don't have good data need to see what they could do if they had it, so that their next procurement won't err on the side of missing data capture opportunities.

3.2.3 General Desirable Qualities of Analysis Software

Chapter 2 lists many specific analyses based on archived AVL-APC data. Some other desirable qualities of analysis and reporting software are listed below.

Modularity with Respect to Source Data and Final Formatting. Analysis software should be capable of working with any source data stream through a standard interface, so that the tools developed for one agency can work for another, even if the other has APCs instead of AVL, or a different vendor, or timepoint records instead of stop records. The interface should be open in the sense of allowing additional data items, and should be achievable in the sense that common AVL and APC data streams with timepoint or stop records can be fit into the specified format.

At the product end, agencies may differ in how they want to format their final reports. Analysis software may offer a default format for its reports, but it should also export to a text table that users can customize or manipulate as they wish using database, spreadsheet, Internet, and other software.

Open Database. Analysis software should hold the AVL-APC data in an open database so that users can access it to do any *ad hoc* analysis they wish.

Data at Different Levels of Summarization. Because most analyses involve data at the route or segment level, it is probably most efficient for databases to hold data already summarized at the segment and trip level. However, analysts should still have access to data at the finest level at which it was collected for special purpose reports and research.

A Range of Reporting Capability. With the modularity and openness described, analysis software will offer users a range of reporting capability to meet the needs of both large and small transit agencies. Users can choose between fully developed reports, analysis programs whose output they format themselves, and access to a database to program and format *ad hoc* or custom analyses. This range of reporting capabilities suits the range of transit agency sizes and capabilities, in which those with little IT capability can use the canned reports, while those with greater IT capability can do their formatting and analyses.

Multi-Route Analysis Capability. While most analyses of archived AVL-APC involve a single route or pattern, or simply aggregate over all routes, there are also desirable analyses that involve analysis of data from trips on multiple patterns or routes.

For example, demand analysis on a corridor may be concerned with sections of several routes using that corridor, and may be interested only in passengers who got both on and off along the corridor of interest. Another example is headway analysis along a trunk shared by several patterns or several routes. Analysis tools are needed that combine data from relevant patterns and routes to analyze demand or service along a street or in an area.

3.2.4 GIS Integration

Several vendors and transit agencies with in-house GIS expertise have developed reports that map results of AVL-APC analysis. A few are beginning to integrate other geographic data such as road networks, census data, and regional transportation model data at the traffic analysis zone level.

On the other hand, incorporation of AVL-APC data into a GIS database is still at a primitive stage of development. At one level, simply associating data (e.g., passenger counts or schedule deviation) with a location allows one to query the database based on geography in order to determine, for example, daily boardings or schedule adherence in a certain neighborhood. Ideally, GIS capability should allow an analyst and analysis to move back and forth between lat-long coordinates and plain-language descriptions of location. A further step would be to directly incorporate AVL-APC data in a geographic transportation planning database, using on/off counts to update travel demand parameters of traffic analysis zones, and running time data to update network travel times.

Several challenges remain for AVL-APC data to be incorporated into a geographic transportation-planning database. One is that demand models are concerned with linked trips, not simply boardings and alightings. Another is that demand models are concerned with the geographic location of a trip maker's ultimate origin and destination, not first and last stop. Better data capture (such as farebox transactions with location stamps in order to get transfer information) and better models are needed for AVL-APC data to relate better to transportation planning-GIS databases.

3.3 Quality and Integration of Other Data Sources

3.3.1 Data and Interface Standards

The SAE J1708 family of standards for networking onboard devices has been generally accepted in the industry. Most vendors readily comply with them, although reluctance has been reported on the part of some farebox vendors. As buyers, transit agencies can insist on compliance with ITS standards when purhasing equipment and software.

TCIP standards for schedule data definitions have been established, are still evolving to correct deficiencies. They have been voluntarily adopted by some vendors, although they are not yet widely used by transit agencies as part of their product specifications. Because schedule data is such an important part of an AVL-APC data archive, further development of this standard will be important for interoperability of both onboard systems and analysis software.

An ongoing project, funded by FTA and the Transit Standards Consortium, to develop a Location Referencing Guidebook is expected to help guide and standardize definitions of geographic objects such as bus stops.

3.3.2 Stop Locations

Transit agencies first implementing automatic vehicle location will almost certainly need to correct their stop location database, because automatic applications do not forgive errors and omissions the way manual data collection can. Some agencies and vendors have used dedicated crews to field map all stop locations using mobile GPS units. Buses themselves can be configured to be those mobile units. One vendor explicitly includes the mapping function in its APC system, allowing the bus itself to serve as the GPS probe. This function requires a crew member in addition to the bus operator and requires that the bus used for mapping stop at all stop locations.

One transit agency reports that starting with a well-calibrated GIS base map based on aerial photography can reduce the burden of field mapping. Field mapping is then needed only to make corrections when consistent discrepancies between the base map and automatic location data is observed. Discrepancies are inevitable, and a valuable feature of AVL-APC data processing is to recognize and alert analysts to them.

Bus stop location is an ambiguous term that can have several meanings. Data transfer and sharing between applications using bus stop location data require attention to the definitions used and the level of accuracy in each database. The ongoing FTA-sponsored effort to standardize location referencing in transit applications should help in this regard.

3.3.3 Schedule Interface

Transit agencies report that the interface between the scheduling system's route and schedule database and the schedule files of AVL and APC systems has often been an implementation obstacle. Getting the AVL-APC vendor and the scheduling package vendor to agree on and implement the interface has been a source of considerable delay in some projects. Some AVL-APC vendors report that they eventually "figured out" particular scheduling package databases. In spite of TCIP standardization efforts, the highly customized nature of schedule databases, due largely to varying scheduling practices at different transit agencies, still makes it difficult to get all the needed route schedule data out of the schedule package database in the desired format. Transit agencies are well advised to support further development of schedule-related TCIP standards and to insist on compliance with them from vendors of schedule-related products.

Automated data collection may require more complete schedule data, and force a change in scheduling practice, than what agencies are accustomed to. For example, at one agency the AVL system required departure and arrival times for deadhead segments, pull-outs, and pull-ins, items that were not explicitly part of the transit agency's scheduling practice. The transit agency had to decide whether to change its scheduling practice or to get the AVL vendor to adapt its system to the agency's scheduling practice.

3.3.4 Fare Collection Data Stream

Little effort has been made to date to integrate fare collection archived data with AVL or APC archived data. A research project recently conducted at MIT for the CTA found several obstacles to integrating the data sources. One is that the farebox and AVL clocks were not synchronized. Another is that the fare transaction data contained a lot of unexplained anomalies. For example, one would expect that people transferring from one route to another (the farebox transaction data includes route transferred from for some passengers) would board the second bus where the two routes intersect, providing a means of identifying a stop; however, the data show such transfers occurring at multiple stops. Because farebox data is not usually analyzed at this level of detail, its quality in many of these respects (e.g., accurate clock, accurate transfer data) hasn't mattered. Improving the quality of farebox data will challenge efforts to integrate it with AVL-APC data.

3.4 Organizational Issues

3.4.1 Raising the Profile of Archived Data in Procurement Decisions

A reason that many AVL systems have failed to deliver their potential in terms of useful archived data is that those who specified and designed the systems either did not emphasize the importance of archived data, or more likely, did not recognize important differences between the needs of real-time data and those of archived data *(21)*. Time and again, procurements have focused on real-time applications, with the implicit expectation that archived data analysis would somehow happen. Some of the lack of appreciation of the character and value of archived data has been on the part of vendors whose primary product is real-time information; and some has been on the part of transit agency staff who managed procurements. There is a need for transit agency staff who are involved in system procurement to better understand how system design affects what data are captured, what the data quality will be, and what off-line analyses it will be able to support. At the same time, there is a need for decision makers to appreciate the importance of archived data acquisition and analysis for improving system management and performance.

3.4.2 Making Control and Supervision Better Support Data Quality

Control and supervision has traditionally been concerned about performance, not about data collection. In fact, to the extent that operations control orders buses to deviate from their schedule plan, they make schedule matching more difficult.

At the same time, there is a growing awareness that collecting and later analyzing data can also contribute to improving performance. There is an opportunity to improve data quality by changing control and supervision practices. Examples include detecting and correcting invalid sign-on data, providing data to the AVL-APC system when a bus is deviated from its schedule, and standardizing codes for control messages.

3.4.3 Staffing and Skill Needs

To date, transit agencies making good use of AVL-APC data have been able to do so only because of the strong set of staff skills they have been able to employ. The lack of available data analysis tools has meant that agencies have needed the skills to develop their own database and analysis tools. The only known exception is the transit agency in Eindhoven, where the Delft University of Technology provided the primary IT skills needed. In the future, however, with the development of third party analysis software and increasingly relevant and accepted standards for data definitions and interfaces, the need for IT expertise in order to exploit archived AVL-APC data should decline. However, agencies will still need the staff and expertise to analyze the data.

Managing data quality takes considerable staff effort. At agencies with good AVL and APC data, one or two staff members are usually devoted to overseeing that the systems deliver the data they should. Identifying and correcting accuracy problems sometimes takes considerable IT expertise, especially if matching algorithms or data objects have to be changed. With AVL systems, matching, accuracy, and data capture issues are often just as much a problem for real-time applications as for archived data, so that little additional work is needed for archived data itself.

3.4.4 Managing an Equipped Subfleet

Where APCs are used, it is common practice equip only 10 to 15 percent of the fleet. The logistical issues of managing an equipped subfleet are significant. A program manager is needed to make assignments of instrumented buses, and to check whether assignments were made. Effort is needed on the part of garage supervisors to stage the instrumented buses properly, and by transportation supervisors to make sure they are used where assigned. Studies that require that all the buses operating on a route or along a trunk be instrumented at the same time (e.g., for a headway analysis) are particularly demanding; a few missed assignments and the data collection effort will have to be repeated another day. If so many instrumented buses are needed in one place that inter-garage transfers are needed, that presents another layer of complication.

Aging of the equipped subfleet also poses problems. At first, the instrumented buses are new, and there may be demands to assign the new buses in ways that conflict with a data collection program. As they age, restrictions on using them on certain routes or runs can develop. For example, they may not have the ergonomic seats needed by some operators, or the low floor required on some routes. They may have maintenance problems with wheelchair lifts or other bus systems.

Partly for this reason, one agency (Tri-Met) is well on its way to equipping its entire fleet with APCs. When passenger counters can be integrated into a smart bus system that already includes automatic vehicle location and schedule matching, the marginal cost of adding passenger counters drops dramatically.

3.4.5 Avoiding Labor Opposition

Suspicion of “the spy in the cab” or “big brother” is natural. There is always the danger of sabotage if transit operators resent the way a system monitors them, especially if they believe that the system is unfair or inaccurate.

For the most part, transit agencies that have adopted AVL and APC have avoided incapacitating labor opposition. Agencies have generally been successful at communicating the security benefits of AVL, which builds operator support. APCs generally don’t engender opposition, partly because their name suggests they’re only counting passengers. Some agencies intentionally avoid directly challenging an operator with AVL-APC data. However, some report using the data to identify patterns of abuse or poor performance and alerting supervisors where and when problems are likely to occur.

CHAPTER 4

OPPORTUNITIES FOR FURTHER RESEARCH AND DEVELOPMENT

4.1 Opportunities Related to System Design and Data Capture

4.1.1 Matching Software for Location-at-Time Data Streams

There are three types of AVL systems in use that provide location-at-time data that might benefit from the development of matching algorithms that use interpolation to produce estimated time-at-location records. One is older AVL systems designed primarily for emergency response. Another is newer AVL systems that provide a stream of frequent poll messages containing location-at-time in addition to less frequent timepoint messages. Normally, the frequent location-at-time data is discarded for off-line analysis. Finally, some next arrival systems capture only location-at-time data. In all of these cases, converting the location-at-time data into stop records with estimated time of arrival or departure at each stop would allow stop-level analysis of running time and headway.

The value of investing in this effort is questionable, however. The market for retrofits to legacy AVL systems is shrinking as AVL systems are replaced. In the case of next arrival systems, predicting arrival time at stops is their core business. Therefore, they poll buses frequently, and pay attention to their arrival prediction algorithms. They already report close agreement between their predicted arrival time and actual arrival time, and their predicted arrival time may well be are more accurate than simple interpolation. And in the case of newer systems that mix time-at-location messages with timepoint message, if users seriously want stop records, they are probably better served by modifying the system to create onboard stop records (e.g., by adding door sensors and enhancing the storage capacity of their onboard computers) than by using algorithms to create stop records that would include estimated arrival time, but would lack door open and departure time.

4.1.2 Better End-of-Line Matching Algorithms

Agencies need better algorithms to determine arrival and departure time at terminals. If running time data from the first and last segment of a trip is not reliable, it becomes impossible to determine trip running time, and important component of schedule planning. This problem is most severe with AVL systems that make timepoint records but not stop records. However, even with stop records or similar frequent location records, finding patterns in operator behavior at ends of lines can help reduce the amount of data that is suspect or discarded.

4.1.3 Passenger Counter Accuracy Definitions and Guidelines

Transit agencies want better definitions of APC count accuracy and guidelines for specifying and testing accuracy. This would entail a review of definitions used by vendors and APC users; a review of imbalance tolerances used in screening; and a review of specification and testing procedures. It would involve statistical analysis of APC counts and comparison counts, and a mathematical analysis of accuracy definitions and test procedures.

A paper on this subject by a member of the research team was recently presented and is soon due to be published *(29)*.

4.1.4 Capture of Transfer Information

Capturing information on passenger transfers and linked trips, whether by means of fareboxes, smart card readers, or other devices, would enable additional useful analyses. Linked trip information is particularly important in Canada, where linked trips are the primary measure of passengers. Developments in that direction are already underway or in planning at several transit agencies.

4.2 Opportunities Related to Analysis and Decision Support Tools

4.2.1 Develop Better Reports and Reporting Software

For archived AVL-APC data to achieve its potential impact on transit management and performance, agencies need better software tools for analysis and reporting, as reported in the previous chapters. Many of the tools suggested in Chapter 2 have not yet been programmed and demonstrated, and others need to be demonstrated more in order become mainstream.

Ultimately, software development should be undertaken by the private sector, where it is most likely to be updated to the latest hardware and management practices. At the present, the primary research need is (a) to develop a catalog of useful reports, including those in use as well as new ones, and including sufficient explanation to allow others to program them, and (b) to demonstrate those reports using real AVL-APC data. Ideally, transit agencies should be provided with prototype software that will allow them to see the value of these reports with their own data.

4.2.2 Advancing Practice in Service Design, Monitoring, and Control

Transit service standards, scheduling and planning methods, and control practices need to be modified to take advantage of the transition from a data-poor to a data-rich environment. For example, there is an opportunity for agencies to provide better customer information, including reliable stop-level schedules; to measure and monitor crowding the way passengers feel it (looking at extremes, not means, and not restricted to a peak volume point) and to adjust schedules accordingly; to set recovery times according to each route's running time variability rather than applying a standard fraction of scheduled running time; to set running time based on reliability criteria that consider the probability of late arrivals, early departures, and operator compliance; and to develop schedules that are appropriate to signal priority where it is applied.

While such changes must be made by transit agencies themselves, research and development of new analysis reports (see previous section) should spur this development.

4.2.3 Analysis Software Definitions and Interfaces

To promote modularity between analysis tools and data collection devices – so that analysis software developed for one agency can be used at another that may use different data collection hardware – standard data element definitions are needed for the types of data found in an AVL-APC database. Interfaces to and from analysis software should be developed using national standards where available and tested, probably in case studies, to see what inconsistencies or ambiguities there may be that might prevent a standard interface, and to suggest ways to correct these deficiencies.

4.2.4 National Transit Database Sampling Using APCs

There is a need for standard sampling plans for estimating passenger-miles that take advantage of the large sample size afforded by APCs and that interfere as little as possible with the broader data collection program of a transit agency that has APCs.

4.2.5 Integration with Geographic and Transportation Planning Databases

Further development is needed to demonstrate linking AVL-APC events and analysis results to locations and integrating them into a geographic database with normal GIS query and mapping capabilities.

Research and development are needed for incorporating AVL-APC data (on-off counts, running times, schedule deviations) into transportation analysis databases, using passenger counts to update demand parameters and performance data to update travel time parameters.

4.2.6 Speed Profile Summary Measures

The stop level appears to be the most important level for analysis. For data stored at this level, useful and standard summaries of the speed profile on interstop segments are needed. Summaries that are being used in practice include maximum speed and time spent below crawl speed. Other summary measures have been proposed, such as number of accel-decel cycles, time spent above cruise speed, and number and intensity of acceleration or deceleration events. Research is needed to identify and test (including testing their usefulness) such measures.

4.2.7 Simulation Tool Development

Detailed AVL and APC data provide some of the basis needed to develop an accurate simulation of transit operations. Such a simulation tool would be valuable for testing new schedules, new methods of control, and impacts of changes such as introducing low floor buses, changing fares and fare payment methods, consolidating stops, and applying signal priority. This simulation tool could also be used to determine measures of effectiveness, such as travel time, fuel efficiency, speed, delay, and cost.

4.3 Opportunities Related to Other Data Sources

4.3.1 Using AVL-APC to Update the Base Map

While many transit agencies informally use their AVL-APC data to update stop or timepoint locations by noting consistent discrepancies between observed location and expected location, formalizing this process would likely result in more accurate and more easily created and updated base maps.

4.3.2 Schedule Interface Standards

Transit agencies and AVL-APC vendors and integrators are clearly interested in more complete standards for the schedule data interface. As mentioned earlier, TCIP standards for schedule objects are being refined by a group whose members includes people concerned with archived AVL and APC data.

4.3.3 Matching AVL to Fare Collection Data

There is a need to demonstrate the value of integrating fare collection data with AVL-APC data. Technologically, we know that it is possible to integrate the data, either by combining separate data archives, or by having fare transactions recorded as events in the AVL-APC system. However, until there is a demonstration of such a joint archive yielding valuable scheduling, planning, or revenue information, it is unlikely that such a system will be specified and realized.

4.4 Opportunities Related to Organizational Issues

4.4.1 Documenting the Benefits of Archived Data

Transit agency staff and consultants who recognize the value and distinct needs of archived AVL-APC data need ammunition to justify the investment. Case histories are needed that document operating cost savings, service improvements, and ridership gains that stem from the use of archived AVL-APC data. As such reports are written, the U.S.

DOT's ITS office already has an aggressive program for publishing and summarizing them.

4.4.2 *Modifying Control Practice to Support Archived Data Quality*

As described in the previous chapter, the era of automated data collection should lead to changes in control and supervision practice, including coding control actions and putting them into the data stream, and changes that help improve data quality. Research is needed to recommend and document such changes and their impacts.

4.4.3 *Guidelines for Rotating an APC Subfleet*

With APCs becoming more common, there is a need for guidance on managing an equipped subfleet, including guidance on data collection programs using APCs that will complement the released TCRP report *Performance Measurement Guidebook (30)*.

REFERENCES

1. Muller, Th.H.J. and P.G. Furth. "Integrating Bus Service Planning with Analysis, Operational Control, and Performance Monitoring." Proceedings of the Intelligent Transportation Society of America Annual Meeting, Boston, 2000.
2. Casey, R.F. *Advanced Public Transportation Systems Deployment in the United States: Update, January 1999.* Report FTA-MA-26-7007-99-1, U.S. D.O.T., 1999.
3. Casey, R.F. *et. al. Advanced Public Transportation Systems: The State of the Art: Update, 2000.* Report FTA-MA-26-7007-00-1, U.S. D.O.T., 2000.
4. Okunieff, P.E. *AVL Systems for Bus Transit.* Synthesis of Transit Practice 24, Transportation Research Board, 1997.
5. Furth, P.G. "Using Automatic Vehicle Monitoring Systems to Derive Transit Planning Data." Canadian Urban Transit Association: Proceedings of the International Conference on Automatic Vehicle Location Systems, Ottawa, pp. 189-200, 1988.
6. Furth, P.G. *Data Analysis for Bus Planning and Monitoring*. Synthesis of Transit Practice 34, Transportation Research Board, 2000.
7. Cevallos, F. "Using dBASE to collect AVL Data." *The dBase Developer's Bulletin16,* July, 2002.
8. Levy, D. and Lawrence, L. *The Use of Automatic Vehicle Location for Planning and Management Information.* STRP Report 4, Canadian Urban Transit Association, 1991.
9. Boyle, D. *Passenger Counting Technologies and Procedures.* Synthesis of Transit Practice 29, Transportation Research Board, 1998.
10. Friedman, T.W. "The Evolution of Automatic Passenger Counters." Proceedings, Transit Planning Applications Conference, 1993.
11. Schiavone, J.J. *Understanding and Applying Advanced On-Board Bus Electronics.* TCRP Report 43, Transportation Research Board, 1999.
12. *Surface Vehicle Recommended Practice for Serial Data Communications between Microcomputer Systems in Heavy-Duty Vehicle Applications.* Standard SAE 1708, Society of Automotive Engineers, 1993.
13. *Surface Vehicle Recommended Practice for Electronic Data Interchange between Microcomputer Systems in Heavy-Duty Vehicle Applications.* Standard SAE 1587, Society of Automotive Engineers, 1996.
14. Navick, D.S. and Furth, P.G. "Using Location-Stamped Farebox Data to Estimate Passenger-Miles, O-D Patterns, and Loads." To appear in *Transportation Research Record,* 2002.
15. Barry, J.J., R. Newhouser, A. Rahbee, and S. Syeda. "Origin and Destination Estimation in New York City with Automated Fare System Data." *Transportation Research Record* 1817, pp. 183-187, 2002.
16. Lehtonen, M. and Kulmala, R. "The Benefits of A Pilot Implementation of Public Transport Signal Priorities and Real-Time Passenger Information." To appear in *Transportation Research Record,* 2002.
17. Kemp, J. "Automatic Passenger Counting and Data Management System Project: Overview, Current Status, Lessons Learned." Presented at Orbital Users Group meeting, 2001.
18. Lee, Y.-J. *et.al.*. "Effect of Automatic Vehicle Location on Schedule Adherence for Mass Transit Administration Bus System." *Transportation Research Record* 1760, pp. 81-90, 2001.
19. Barry, V. "Ongoing Use of Automatic Passenger Counters In Route Planning." Proceedings, Transit Planning Applications Conference, 1993.
20. Furth, P.G., J.P. Attanucci, I. Burns, and N.H. Wilson. *Transit Data Collection Design Manual.* Report DOT-I-85-38, U.S. Department of Transportation, 1985.

21. Kemp, J. “Archived Data Services: Lessons Learned – Things You Didn’t Have to Think About When It Was Just AVL.” Presented at Transportation Research Board Annual Meeting, 2002.

22. Wilson, N.H. *et. al.* “Service-quality monitoring for high-frequency transit lines.” *Transportation Research Record* 1349, pp. 3-11, 1992

23. “Service Delivery Policy.” Massachusetts Bay Transportation Authority, 1996.

24. Strathman, J.G., Kimpel, T.J., and Dueker, K.J. “Evaluation of Transit Operations: Data Applications of Tri-Met’s Automated Bus Dispatching System.” To appear in *Transportation Research Record,* 2002.

25. Dailey, D.J. and Maclean, S.D. “Transit Vehicle Arrival Prediction: Algorithm and Large-Scale Implementation.” *Transportation Research Record* 1771, pp. 46-51, 2001.

26. *Evaluation of the Advanced Operating System of the Ann Arbor Transit Authority: Transfer and On-Time Performance Study: Before and After AOS Implementation.* Electronic report available EDL# 13147, ITS Office, U.S. D.O.T.

27. Strathman, J.G. *et. al.* “Service Reliability Impacts of Computer-Aided Dispatching and Automatic Vehicle Location Technology: A Tri-Met Case Study.” *Transportation Quarterly* 54:3, pp. 85-102, 2000.

28. Furth, P.G. and Th.H.J. Muller. “Conditional Bus Priority at Signalized Intersections: Better Service Quality with Less Traffic Disruption.” *Transportation Research Record 1731*, pp. 23-30, 2000.

29. Kimpel, T.J., J.G. Strathman, D. Griffin, S. Callas, and R.L. Gerhart. “Automatic Passenger Counter Evaluation: Implications for National Transit Database Reporting.” Presented at Transportation Research Board Annual Meeting, 2003, and forthcoming in *Transportation Research Record*, 2003.

30. Ryus, Paul. *A Guidebook for Developing a Transit Performance-Measurement System.* TCRP Report 88, Transportation Research Board, 2003.

APPENDIX A: Tri-Met's Experience With Automatic Passenger Counter and Automatic Vehicle Location Systems*

The Tri-County Metropolitan Transportation District of Oregon (Tri-Met) is one of about 30 metropolitan transit agencies that have deployed both automatic vehicle location (AVL) and automatic passenger counter (APC) systems (Casey, 1999). These technologies are important components of the agency's new automated bus dispatching system (BDS). The AVL and APC systems at Tri-Met recover comprehensive operations and passenger activity data at the bus stop level that is archived for later analysis. The agency has gained a reputation as an industry leader in the areas of data archiving and the application of archived data to performance monitoring and analysis.

Prior to the implementation of AVL and APC technology, Tri-Met, like others in the transit industry, relied on costly manual data recovery. During lean budget years in the early 1980s, the agency began using stand-alone APCs instead of ride checkers, and recovered very little operating data beyond what was required for Federal Transit Administration reporting. With its AVL and APC systems, Tri-Met is now automatically collecting and archiving over 500,000 stop and event data records per day. Offline analysis of this data supports a wide array of agency activities. The transformation from scarce to abundant data has contributed to a variety of functional changes that have both enhanced service quality and improved efficiency.

This report examines factors related to the development of Tri-Met's BDS that have contributed to its success in data archiving and analysis. The organization of the report includes a description of the BDS, a review of the types of analyses that have been supported by archived data, a discussion of factors that influenced decisions made on the design of the system, reflections on unrealized expectations and missed opportunities, and concluding observations.

TRI-MET'S AUTOMATED BUS DISPATCH SYSTEM

Tri-Met's BDS became fully operational in 1998. Its main features are as follows (see Figure 1):

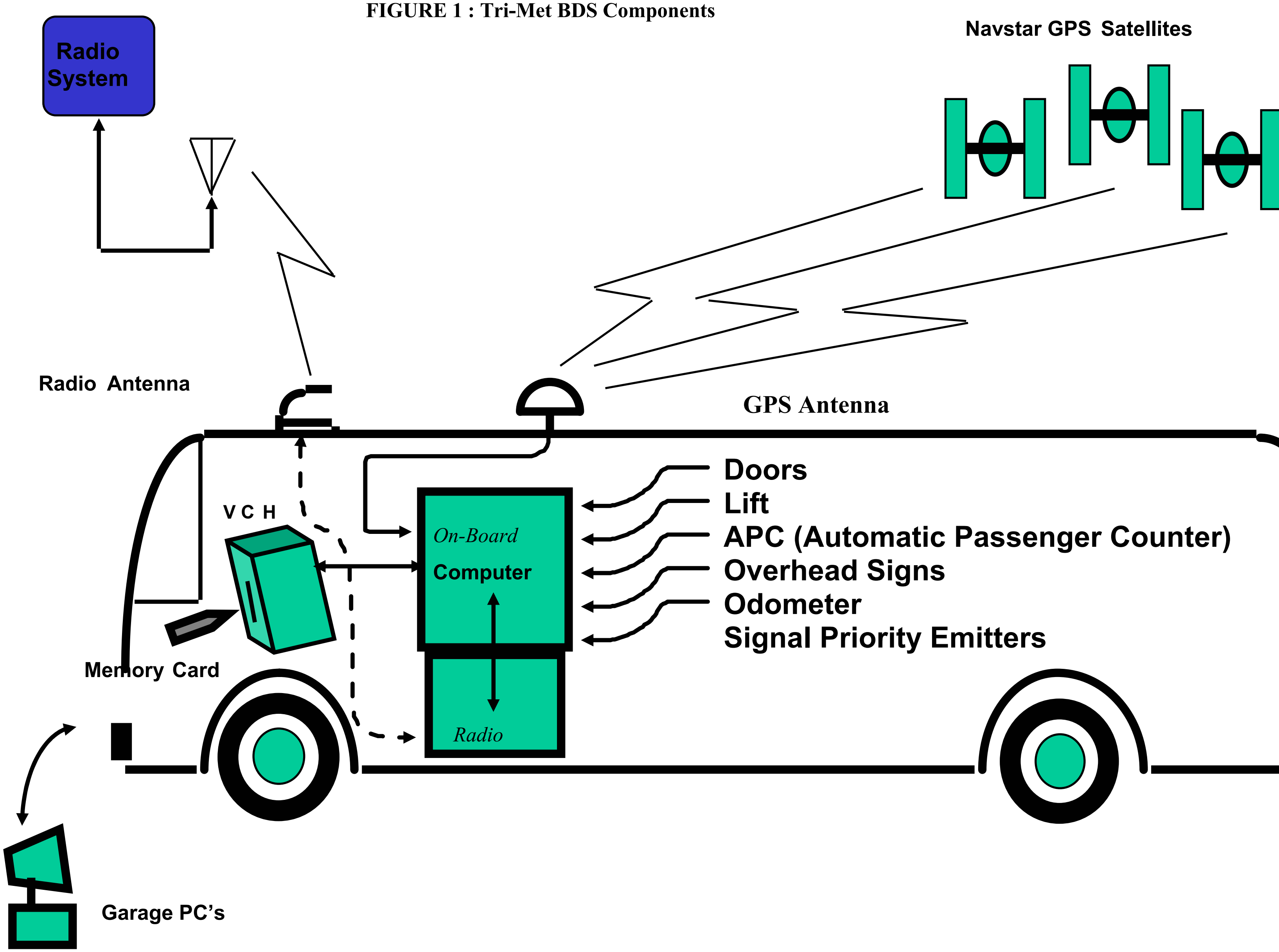
- Orbital Sciences Corporation's Intelligent Fleet Management System, which includes Automatic Vehicle Location (AVL) using a satellite-based global positioning system (GPS);
- Voice and data communication within a pre-existing 450 MHz Motorola mobile radio system;
- On-board computer for temporary data storage, a vehicle control head displaying schedule adherence to operators, detection and reporting of schedule and route deviations to dispatchers, and two-way, pre-programmed digital messaging between operators and dispatchers;
- Infrared beam-type Automatic Passenger Counters (APCs) installed on approximately 70% of the bus fleet and all new bus acquisitions;
- New dispatching center containing nine CAD/AVL consoles.

A PCMCIA memory card containing the route schedule is inserted in the vehicle control head when an operator logs on at the beginning of each block. A vehicle's current status is related to its scheduled status to determine schedule deviation in real time, which is displayed on the control head screen (see Figure 2). When schedule deviations exceed predetermined values, an exception report is automatically transmitted to the dispatch center. Exception

---

* This case study was prepared by James G. Strathman of Portland State University. Tri-Met staff interviewed include Rick Gerhart, Ken Turner, Steve Callas, and Jon Lutterman. Their insights are gratefully acknowledged. Tom Kimpel and Pete Fielding also provided many helpful comments on an earlier draft.

FIGURE 1 : Tri-Met BDS Components



reports are also transmitted when vehicles deviate from their routes. Exception reports are listed on the dispatcher's CAD screen along with other attention requests (e.g., mechanical problems, traffic and on-board incidents, delays, etc.) that are transmitted by vehicle operators from the control head keypad. In addition to schedule status, the control head screen displays freeform text messages from dispatchers to operators. The AVL system also contains a covert microphone and silent alarm key, providing enhanced security.

Data generated by the BDS is written automatically on the memory card. A data record is written at each bus stop, producing approximately 500,000 stop records per day. In addition, a data record is written at each location where an operator-initiated text message is generated or a schedule exception, producing about 25,000 event records per day. Stop records contain the following information: route number, direction, trip number, date, vehicle number, operator ID, bus stop ID, stop arrival time, stop departure time, boardings, alightings and passenger load (on APC-equipped vehicles), door opening, lift usage, stop dwell time, maximum speed since the prior stop, longitude, and latitude. An example is shown in Table 1.

When vehicles return to the garages at the end of each day the data are transferred from the memory card to a PC and then uploaded to the network. A post processing operation then matches the stop data records to the schedule database. About 97 percent of the stop records are successfully matched to the schedule database. The matched data is organized in an Oracle 8.1 database and stored on an eight processor Sun 4501 computer with 500 gigabytes RAID, 8 gigabytes RAM, and 1 gigabit Ethernet connection. Presently, the BDS data warehouse contains more than 120 million stop and event records, along with the related schedules. Summary tables have been designed to report at the stop, time point and trip levels.

APPLICATIONS OF AVL-APC DATA AT TRI-MET

Archived AVL and APC data support a variety of regular reporting activities at Tri-Met. The data are also used to support analysis of specific needs and problems that relate to both short and long term planning, scheduling, and operations. Tri-Met has dedicated one staff position to reporting and analysis of AVL and APC data. In addition, the agency has engaged Portland State University researchers in a series of projects to evaluate the effects of its BDS on service performance and to help identify potential applications of the data. The projects have been jointly sponsored by the USDOT Region 10 University Transportation Center.

Detailed quarterly performance reports represent the most prominent use of AVL and APC data at Tri-Met. The report provides performance data at the trip, route, and system levels. An example of the trip level report for the Winter 2001 quarter is shown in Table 2. The table covers afternoon and evening outbound trips for the 4-Fessenden, an interlined route extending north from downtown Portland. The columns in the table report the scheduled trip departure time, train number, and departure location. This is followed by six indicators related to passenger activity: average boarding rides per trip, average maximum load per trip, the average maximum load factor (average maximum load divided by seating capacity * 100), the percentage of trips exceeding 130% of seating capacity (a service standard), the number of pass-up events reported by operator-keyed messages, and the number of trips generating valid APC data. In any given quarter there are between 60 and 70 scheduled trips. In this example, valid APC data were recovered on about 75% of the trips.

The columns on the right side of Table 2 report indicators of operating performance: scheduled running time, median actual running time, running time ratio (median divided by scheduled running time * 100), the observed running time coefficient of variation, median speed, headway adherence (the percentage of stops in which the actual headway ranges from 50% to 150% of the scheduled headway), the scheduled recovery time, the median actual recovery time, and the recovery ratio (median divided by actual recovery time * 100). The final three columns in the table report stop level on-time performance, where early is represented by departures more than one minute ahead of schedule and late is represent by departures more than five minutes behind schedule.

Trip level data is also aggregated in the quarterly performance report to provide summary indicators. Table 3 is an example of weekday performance reporting by service period for the 4-Fessenden. Several new indicators are reported in Table 3, including (in the top half of the table) boarding rides per revenue hour and the percentage of trips with passenger loads exceeding seating capacity due to deviations in headways. This latter measure is based on a regression relating passenger loads to headway deviations.

In the bottom half of Table 3 are three indicators of excess passenger waiting time associated with deviations from scheduled headways. These measures are based on work by Wilson et al. (1992), and assume a constant rate of passenger arrivals at bus stops. The three measures shown in the table are average excess wait times per passenger

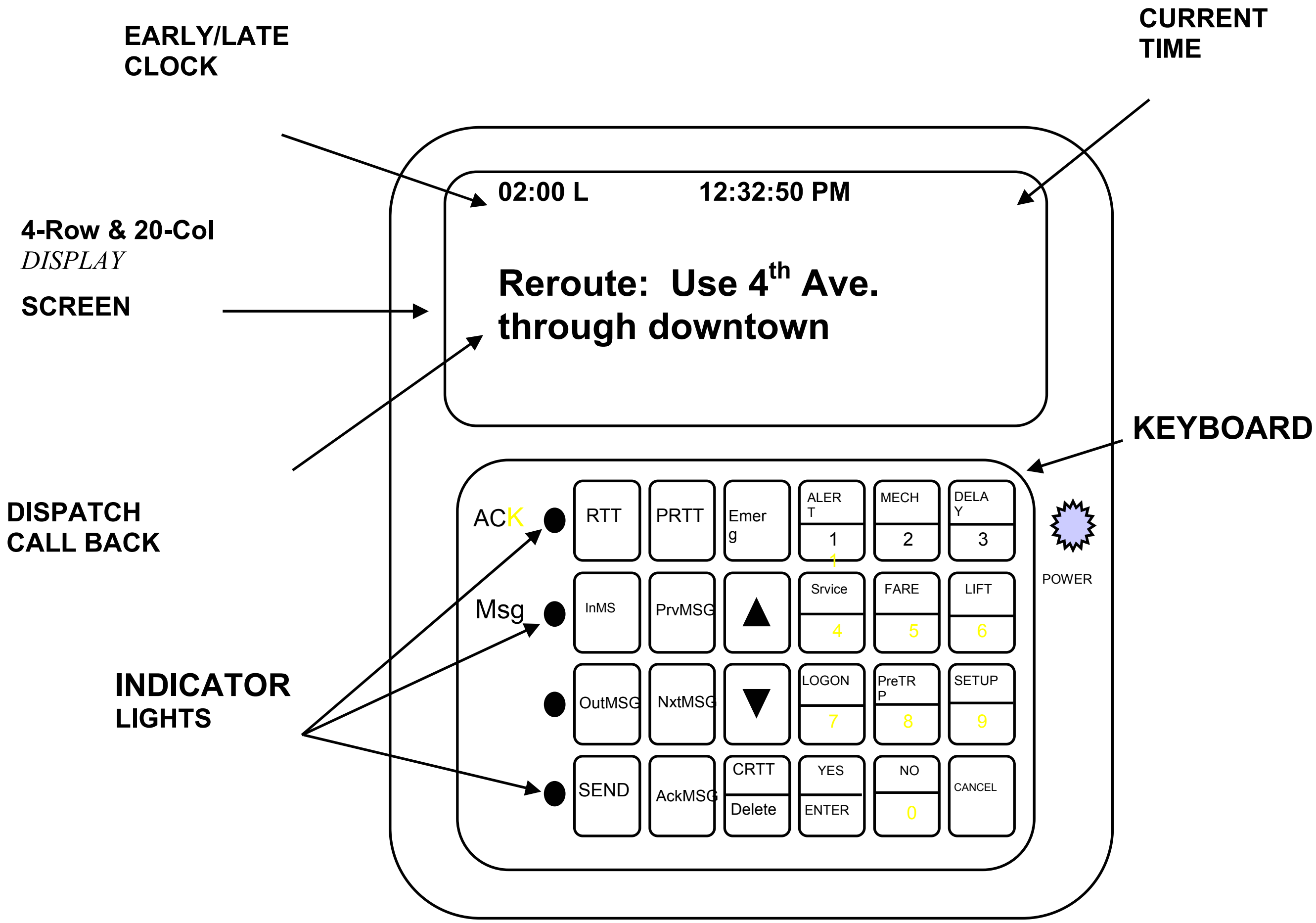


FIGURE 2: Vehicle Control Head

TABLE 1: BDS Data Records

CardRep - Notepad

File Edit Search Help

BUS= 2134 TRAIN= 1401 DATE=20000313

STOP ID	STOP NAME		ARRIVE	LEAVE	ON	OFF	MI	FT N	FT E	D	L	DWEL	SPD	LOD
9302	5TH & HOYT	(OP)	16:25:02	16:25:04	0	0	0			0	0		17	2
0	unscheduled stop		16:25:06	16:25:08	0	0	0	-868	28	1	0	1	3	2
9222	5TH & EVERETT	(NS)	16:25:08	16:25:52	2	0	1			0	0		18	4
9303	5TH & COUCH	(NS)	16:26:04	16:26:40	8	0	1	-10	6	1	0	21	20	12
7631	5TH & PINE	(OP)	16:27:14	16:28:2				2	5	1	0	45	19	15
	Scheduled times --------->		16:28:00	16:28:0										
7635	5TH & STARK	(FS)	16:28:30	16:29:5				-42	2	1	0	17	17	18
7585	5TH & ALDER	(FS)	16:29:58	16:31:2				8	14	1	0	24	16	22
7645	5TH & YAMHILL	(FS)	16:31:28	16:31:4						0	0		15	22
7633	5TH & SALMON	(FS)	16:31:50	16:33:2				143	80	1	0	34	20	35
	Scheduled times --------->		16:33:00	16:33:0										
3639	MADISON & 4TH	(FS)	16:33:46	16:35:2				-14	59	1	0	22	15	43
	time=16:36:02; Fare Evasion.													
3635	MADISON & 1ST	(NS)	16:36:10	16:36:1						0	0		24	48
2641	HAWTHORNE BRIDGE & EAST	(AT)	16:37:16	16:37:3				28	60	1	0	7	30	46
0	unscheduled stop		16:38:50	16:39:0				47	1119	1	0	5	29	45
2594	HAWTHORNE & 6TH	(NS)	16:39:02	16:39:0						0	0		15	45
2597	HAWTHORNE & 9TH	(NS)	16:40:04	16:40:2				10	23	1	0	6	24	45
2599	HAWTHORNE & 12TH	(NS)	16:41:28	16:42:0				-2	8	1	0	18	23	43
	Scheduled times --------->		16:39:00	16:39:0										
2595	HAWTHORNE & MAPLE	(FS)	16:42:10	16:42:1						0	0		31	43
2603	HAWTHORNE & 16TH	(FS)	16:42:22	16:42:5				8	-12	1	0	9	29	38
2596	HAWTHORNE & POPLAR	(NS)	16:42:58	16:43:2				23	-100	1	0	4	21	37
2607	HAWTHORNE & 20TH	(FS)	16:43:50	16:44:1						0	0		22	36
2608	HAWTHORNE & 23RD	(NS)	16:44:20	16:44:2						0	0		31	36
2612	HAWTHORNE & 25TH	(FS)	16:44:30	16:44:5				6	31	1	0	5	32	34
2614	HAWTHORNE & 28TH	(OP)	16:45:12	16:45:4				12	29	1	0	15	21	32
2615	HAWTHORNE & 30TH	(NS)	16:45:54	16:46:0						0	0		28	32
2617	HAWTHORNE & 32ND PL	(FS)	16:46:16	16:46:4				1	-3	1	0	6	29	29
2620	HAWTHORNE & 34TH	(FS)	16:46:50	16:47:2				2	32	1	0	11	19	27
2623	HAWTHORNE & 37TH	(FS)	16:48:00	16:48:2				1	-85	1	0	5	26	25
2625	HAWTHORNE & 39TH	(NS)	16:49:16	16:50:3				11	-80	1	0	12	22	17

TABLE 2: Trip Level Performance Report

Route: 004 - Fessenden | **Weekdays** | **December 3, 2000 to March 3, 2001**

Direction Outbound to Lombard & Burlington

Start Time	Train	Start Location	Avg. Boarding Rides	Avg. Max Load	Max Load Factor	Percent Over Capacity	# of Pass Ups	APC Obs.	Sched. Run Time	Median Run Time	Run Time Ratio	Run Time CV	Median Speed	Headway Adher-ence	Sched. Recovery	Median Recovery	Recovery Ratio	On Time	Early	Late
									h:mm:ss	h:mm:ss			mph		mm:ss	mm:ss				
1:37 PM	450	SW 6th & Salmon	62	31	79%	0%	4	53	0:54:00	0:55:37	103%	9%	11.9	92%	32:00	26:44	84%	46%	4%	50%
1:52 PM	451	SW 6th & Salmon	59	32	82%	8%	1	53	0:54:00	0:53:12	99%	9%	12.4	91%	27:00	26:31	98%	72%	7%	21%
2:07 PM	435	SW 6th & Salmon	68	39	101%	8%		53	0:54:00	0:55:04	102%	8%	12.0	91%	22:00	19:43	90%	71%	6%	23%
2:22 PM	441	SW 6th & Salmon	68	34	86%	4%	2	47	0:56:00	0:56:20	101%	7%	11.7	98%	21:00	20:21	97%	75%	7%	17%
2:37 PM	438	SW 6th & Salmon	77	38	98%	2%	4	53	0:56:00	0:55:44	100%	8%	11.8	95%	22:00	21:02	96%	76%	5%	19%
2:48 PM	434	SW 6th & Salmon	76	39	101%	9%	3	45	0:56:00	0:58:20	104%	8%	11.3	80%	24:00	18:16	76%	52%	3%	46%
3:00 PM	442	SW 6th & Salmon	68	35	89%	4%	8	54	0:56:00	1:00:26	108%	7%	10.9	84%	18:00	14:00	78%	72%	6%	22%
3:08 PM	443	SW 6th & Salmon	66	38	98%	8%	1	53	0:56:00	0:56:35	101%	7%	11.7	89%	17:00	16:20	96%	79%	6%	16%
3:15 PM	453	SW 6th & Salmon	55	33	84%	2%		57	0:56:00	0:55:45	100%	5%	11.8	85%	20:00	19:05	95%	82%	5%	13%
3:22 PM	449	SW 6th & Salmon	62	35	89%	2%		60	0:56:00	0:55:48	100%	6%	11.8	62%	26:00	25:16	97%	65%	4%	31%
3:33 PM	446	SW 6th & Salmon	73	42	109%	4%		50	1:00:00	0:56:24	94%	6%	11.7	86%	21:00	24:22	116%	71%	19%	10%
3:47 PM	454	SW 6th & Salmon	78	39	91%	8%	3	13	1:00:00	0:59:54	100%	4%	11.0	94%	19:00	19:52	105%	74%	18%	8%
4:01 PM	456	SW 6th & Salmon	98	50	117%	19%	3	37	1:00:00	1:00:10	100%	6%	11.0	89%	17:00	15:20	90%	69%	3%	29%
4:13 PM	448	SW 6th & Salmon	76	44	113%	9%	3	56	1:00:00	1:00:28	101%	8%	10.9	71%	20:00	16:53	84%	52%	3%	45%
4:26 PM	436	SW 6th & Salmon	76	45	114%	6%	2	48	0:57:00	0:57:34	101%	8%	11.5	80%	26:00	23:18	90%	64%	3%	33%
4:38 PM	439	SW 6th & Salmon	71	48	124%	34%	2	50	0:57:00	0:57:54	102%	8%	11.4	57%	29:00	21:02	73%	31%	2%	67%
4:51 PM	440	SW 6th & Salmon	65	36	92%	0%		55	0:57:00	0:54:52	96%	4%	12.0	48%				71%	6%	24%
5:03 PM	445	SW 6th & Salmon	72	48	124%	25%	6	51	0:57:00	0:55:52	98%	8%	11.8	82%	19:00	18:49	99%	62%	12%	26%
5:14 PM	444	SW 6th & Salmon	65	44	114%	15%	3	53	0:57:00	1:02:44	110%	8%	10.5	82%	26:00	18:25	71%	52%	2%	46%
5:27 PM	452	SW 6th & Salmon	61	41	106%	9%		46	0:57:00	0:56:04	98%	12%	11.8	86%	28:00	28:28	102%	76%	6%	18%
5:39 PM	447	SW 6th & Salmon	70	39	100%	2%		58	0:53:00	0:52:57	100%	4%	12.5	63%				41%	3%	56%
5:54 PM	455	SW 6th & Salmon	55	38	96%	6%	2	48	0:53:00	0:53:14	100%	7%	12.4	80%	20:00	19:18	96%	78%	6%	16%
6:07 PM	450	SW 6th & Salmon	56	36	93%	0%	3	53	0:53:00	0:54:11	102%	8%	12.2	83%	22:00	17:48	81%	52%	6%	43%
6:20 PM	435	SW 6th & Salmon	42	34	88%	0%		53	0:50:00	0:50:28	101%	7%	13.1	81%	27:00	27:08	100%	84%	6%	10%
6:33 PM	451	SW 6th & Salmon	49	32	82%	0%	1	52	0:50:00	0:51:05	102%	6%	12.9	95%	29:00	27:18	94%	77%	7%	16%
6:45 PM	434	SW 6th & Salmon	46	28	71%	0%		45	0:50:00	0:48:54	98%	4%	13.5	82%				59%	1%	39%
6:56 PM	458	SW 6th & Salmon	34	24	61%	0%	1	53	0:50:00	0:48:11	96%	5%	13.7	78%	21:00	22:25	107%	85%	7%	8%
7:11 PM	443	SW 6th & Salmon	46	28	71%	0%		52	0:48:00	0:49:06	102%	8%	13.4	86%	23:00	21:06	92%	68%	5%	27%
7:26 PM	453	SW 6th & Salmon	44	32	83%	0%		58	0:48:00	0:50:40	106%	6%	13.0	77%	29:00	20:27	71%	33%	1%	66%
7:41 PM	448	SW 6th & Salmon	33	21	53%	0%	1	56	0:48:00	0:47:01	98%	4%	14.0	75%				69%	4%	28%
7:56 PM	446	SW 6th & Salmon	33	21	53%	0%		50	0:48:00	0:44:46	93%	6%	14.7	89%				85%	10%	5%
8:11 PM	439	SW 6th & Salmon	40	28	72%	0%		48	0:48:00	0:48:42	101%	7%	13.6	89%	16:00	14:26	90%	75%	3%	22%
8:26 PM	436	SW 6th & Salmon	40	26	66%	0%		46	0:45:00	0:45:46	102%	4%	14.4	69%				35%	3%	61%
8:41 PM	444	SW 6th & Salmon	30	23	58%	0%		54	0:45:00	0:49:05	109%	9%	13.5	82%	19:00	13:08	69%	53%	1%	46%
8:56 PM	445	SW 6th & Salmon	31	21	53%	0%		50	0:41:00	0:42:15	103%	5%	15.6	87%				54%	3%	43%
9:10 PM	455	SW 6th & Salmon	32	25	65%	0%		50	0:41:00	0:43:46	107%	9%	15.1	93%	24:00	20:40	86%	63%	1%	36%

* Indicates variability exceeds +/- 20% at the 80% confidence level.

TABLE 3: Summary Performance Report

	Trips	Boarding Rides	Rides/ Rev.Hr	Max Load	Load Factor	% Over Capacity	% Due to Headway	# of Passups
Outbound								
Early AM	7	130	27.1	15	38%	0%	0%	0
AM Peak	10	342	42.3	22	57%	1%	0%	0
Midday	31	1,670	61.0	31	79%	2%	44%	32
PM Peak	10	707	74.7	43	110%	13%	46%	21
Night	22	806	48.0	26	67%	0%	0%	6
Total by Direction	80	3,655	55.0	29	73%	3%		59

	On Time	Early	Late	Sched. Headway	Headway Adhere.	Excess Wait Time	Wait Time Per Trip	Total Wait Time
Outbound				*hours/min.*		*min./sec.*	*hours/min.*	*hours/min.*
Early AM	84%	5%	11%	0:16	93%	00:36	0:11	1:17
AM Peak	66%	4%	30%	0:12	87%	00:39	0:21	3:39
Midday	74%	6%	20%	0:14	90%	00:34	0:30	15:49
PM Peak	59%	5%	36%	0:13	74%	01:09	1:21	13:32
Night	69%	5%	26%	0:21	89%	00:51	0:30	11:19
Total by Direction	70%	5%	24%				0:34	45:39

and per trip, and excess wait time over the quarter. The wait time measures are intended to reflect service quality from the passengers' perspectives.

The quarterly performance report serves a variety of functions at Tri-Met. For example, service planners sort the data on boarding rides per revenue hour to identify instances where service is greatly underutilized, and focus their efforts on changes that will either improve utilization or reduce the amount of service provided. At the other end, they identify instances of heavily utilized service and consider options such as adding trips or improving schedule and headway adherence. Schedule writers evaluate observed running and recovery times to determine whether adjustments in the schedule are needed. The operations staff assess schedule and headway adherence to identify routes (and, after further analysis, locations) where field supervisors can better focus their control efforts.

APC data are also compiled to produce a quarterly passenger census report by stop. Prior to the extensive deployment of APC's in the bus fleet, passenger census data was collected manually every five years at a cost of about $250,000. Table 4 provides an example of quarterly passenger census reporting for selected bus stops on the 4-Fessenden. The columns in the table report the stop location and ID, average daily boardings and alightings, and lift activity. This data is useful in identifying opportunities for stop consolidation or the need for additional stops.

Analysis of archived AVL data is also undertaken in response to specific issues that arise related to service design, delivery, and use. For example, customer complaints about lateness or excessive speed can be checked against the data. Patterns of fare evasions (as reported by operator event messages) can be mapped to target locations for enhanced fare inspections. Patterns of early departures or excessive layovers can be identified to aid field supervisors in focusing on problem operators, while patterns of high on-time performance can be used to identify meritorious recognition. Data are also used in evaluating special projects. For example, Tri-Met has initiated a project ("Transit Tracker") to provide real time bus arrival information to passengers on selected routes. Predicted arrival times are posted on the web as well as on variable message boards installed at a limited number of stops. Arrival predictions are based on the scheduled running time from the current bus location, with updates provided on an approximate 90 second interval. Archived AVL data are being used to evaluate the accuracy of arrival predictions. In another example, an analysis of dwell and running time data has been undertaken to determine whether a fleet of recently-acquired low floor buses reduces boarding, alighting and lift times compared to standard buses. In another case, the additional running time associated with a route deviation was estimated. Another example concerns an evaluation of the reduction in running time means and variances due to signal prioritization on selected routes. A final example is a project to redefine service standards for vehicle capacity utilization focusing more generally on morning and evening peak hour periods rather than individual trips.

To date, the primary applications of archived AVL and APC data at Tri-Met have been dedicated to improving service quality and ensuring maximum effective utilization of resources. However, there is increasing interest in using this technology directly to reduce costs. Two examples of cost-motivated opportunities are the yard mapping of buses and time-slipping. Presently, the daily yard mapping process at the agency's three bus garages is the responsibility of three employees. The locational referencing of vehicles with AVL may be sufficiently accurate to generate yard maps automatically, which would yield a savings of about $150,000 annually. Secondly, time slips for operators logging overtime are currently filled out and processed manually. Alternatively, overtime could be calculated and forwarded automatically when operators log off the BDS.

The collaboration between Tri-Met and local university researchers has yielded a series of studies of the effects of the BDS on operations and scheduling. Initial efforts dealt with defining performance measures and collecting baseline data prior to the implementation of the BDS (Strathman et al., 1999). This was followed by an evaluation of the initial performance impacts of the new system (Strathman et al., 2000), which found a 9% improvement in on-time performance, an 18% reduction in running time variation, a 3% reduction in mean running time, and a 4% reduction in headway variation. It was estimated that these improvements produced a savings in passenger waiting and in-vehicle travel times valued at $3.5 million annually. It should be noted that these improvements occurred without substantive changes in practices related to service design and delivery. It was posited that they were the consequence of providing enhanced information to dispatchers and operators.

Another study evaluated efforts to improve bus spacing along the downtown bus mall (Strathman et al., 2001). In this study, dispatchers alerted field supervisors of impending delays. Field supervisors responded by imposing control actions (holding, short-turning, and switching trip assignments). It was found that these actions resulted in a 16% reduction in passenger load variation, which was maintained through the maximum load points of the study routes. Another study sought to improve Transit Tracker arrival time predictions by estimating delays related to lift

TABLE 4: APC Passenger Census Report

Tri-Met Passenger Census - Winter 2000/2001
Weekday All-Day Ons and Offs by Route and Stop

Route: 004 - Fessenden **Outbound to Lombard & Burlington**

Stop Location	*Location ID*	*Ons*	*Offs*	*Total*	*Monthly Lifts**
6TH / SALMON	7789	250	0	250	23
6TH / YAMHILL	7807	153	183	336	21
6TH / ALDER	7747	250	129	379	15
6TH / STARK	7797	129	78	207	10
6TH / PINE	7787	93	65	158	5
6TH / COUCH	7758	96	52	148	9
6TH / EVERETT	9298	83	64	147	9
EVERETT / 4TH	9546	15	23	38	1
EVERETT / 2ND	1612	27	16	43	1
ROSE QUARTER TC / B1 OUTBOUND	1097	618	73	691	48
WILLIAMS / WHEELER	9547	7	5	12	0
WILLIAMS / BROADWAY	6357	25	17	42	3
WILLIAMS / TILLAMOOK	6370	9	25	34	1
WILLIAMS / THOMPSON	6369	17	31	48	2
WILLIAMS / RUSSELL	6354	56	55	111	36
WILLIAMS / GRAHAM	6362	24	70	94	25

activity and bridge closures (a number of east side routes serving downtown Portland must pass over draw or lift bridges on the Willamette River) (Dueker et al., 2001).

The extensive passenger data produced by the large-scale deployment of APCs provided the focus for several other Tri-Met/PSU projects. In the first project, passenger data were combined with service attributes and demographic variables in route segment corridors (using GIS to apportion data from the American Community Survey) to estimate a transit utilization model (Kimpel et al., 2000). In the second project, the validity of APC boarding, alighting and load counts was assessed relative to counterpart data obtained from on-board cameras (Kimpel et al., 2002). A framework for sampling archived passenger data was also developed consistent with NTD and internal reporting requirements. Although a sampling approach based on a limited number of APC-equipped buses had been approved by FTA a decade earlier, Tri-Met continued to rely on manual passenger counts due to difficulties in assigning APC-equipped buses to sampled trips. Extensive deployment of APCs has now mitigated bus assignment problems. Automating NTD data recovery is expected to save Tri-Met about $50,000 per year.

The precision and confidence levels required for NTD reporting can be met with a relatively small sample of bus trips. Given the extensive deployment of APCs, the NTD sampling framework was extended to provide monthly passenger boardings by route at higher levels of precision and confidence. Previously, monthly route-level boarding estimates were derived from revenues, but the growing mix of fare payment alternatives made those figures increasingly uncertain.

The final Tri-Met/PSU project focused on schedule efficiency and operator-related effects on running time variation (Strathman et al., 2002). Archived AVL data were used to construct running time distributions by route and time period. From these distributions, typical running, recovery and layover times were calculated and compared to standards recommended by Levinson (1991), as well as Tri-Met's own scheduling standards and conditions identified in the agency's labor agreement. It was found that 81 of the agency's 104 routes had excessive running, recovery, and layover times, while the scheduled times in the other 23 were inadequate. It was estimated that adjusting the schedules would potentially yield a $7 million per year savings in operating cost. The contribution of operators to running time variability was assessed with a fixed effects model. After controlling for various operational causes of running time variation, the model estimated the variation attributable to operator-specific differences. The results indicated that operator differences account for the majority of running time variation, and that running time is inversely related to operator experience. This suggests that "bus bunching" problems experienced in peak periods can be partly attributed to the mixing of operators with varying experience on a given route. The runcutting process is an important determinant of the mix of operator experience in its delineation of work assignments among part-time, split shift, and straight shift operators. Runcuts may thus generate a mix of work assignments that is optimal in terms of cost-effectiveness, but sub-optimal in terms of headway maintenance.

THE APC-AVL CHRONOLOGY AT TRI-MET

The chronology of APC and AVL planning, deployment and use at Tri-Met covers a 20 year period. A 1980 visit to OC Transit in Ottawa, Canada, left Tri-Met staff impressed with the potential of APCs. Ten units were acquired and installed in the bus fleet in 1982. However, the initial experience with APCs was plagued by equipment malfunctions and data problems. More than 80% of trip-level assignments of APC-equipped buses either failed to return any data or generated data records that were screened out due to inconsistencies in passenger counts, time/distance tolerances, and trip assignment information.

While it was clear that APCs were not initially delivering on their potential at Tri-Met, a series of severe budget cuts in the early 1980s spared them from elimination. A 1984 budget cut led to a lay-off of Tri-Met's entire staff of traffic checkers. Passenger counting for FTA Section 15 reporting was contracted out and, for several years, virtually no additional data were collected for planning and performance monitoring. At that time, a commitment to "force the APCs to work" was made, and a person was appointed the task as his only responsibility. His efforts ranged from hardware maintenance to programming, data processing, and reporting.

A major challenge with the APCs involved trip and time point referencing the passenger data. The bases for referencing the data stream were time clock and odometer readings. Operator radio log ins were used to identify work assignments, trips were identified by time breaks in the data associated with scheduled layovers, and time point location was determined from odometer readings. Post processing of the data included a comparison of total daily boardings and alightings, and day records were screened when the difference between the two exceeded 10% of the total passenger activity. The resulting data reports on time point level passenger activity and on-time performance served as input for two in-house developed programs, Interactive Schedule Maintenance (ISM) and Schedule Writers Analysis Package (SWAP), which were used to monitor and adjust bus schedules.

By 1986, the number of APC-equipped buses had grown to 50 (about 10% of the fleet) and passenger data collected from these units was used for Section 15 reporting for the first time. Continuing problems with equipment malfunctions and screening of data, coupled with difficulties in assigning APC-equipped buses to routes selected for data recovery, led to uncertainty about the compatibility of the resulting sample data with the recommended Section 15 sampling procedure (UMTA, 1978). A subsequent review of the system led to the development of a cluster-based sampling approach (Strathman and Hopper, 1991), which was approved by FTA for Section 15 reporting.

Tri-Met's involvement with AVL technology can be traced to the late 1980s. A feasibility study assessing signpost and Loran-C systems was completed in 1988. However, subsequent internal budget proposals for a stand-alone AVL system with a projected cost of about $3 million failed to gain support for five consecutive years. In the early 1990s, an operations control plan (Tri-Met, 1991) identified a critical need to replace the agency's bus dispatch computer. An AVL component was recommended as part of the upgrading of the dispatching system.

The proposed replacement of the bus dispatch computer was approved by Tri-Met management, and tentative approval, subject to cost, was also given to specifying AVL capability in the request for proposals (RFP). A Bus Dispatch System (BDS) RFP was issued in late 1993, and generated 5 qualified responses. The proposal by Orbital Sciences Corporation was selected in mid-1994. Orbital's bid of $6.5 million included $.5 million for adding AVL functionality to the BDS. A contract was awarded in mid-1994. The shake down of the BDS began in 1997, and the transition to full operation was completed by early-1998.

KEYS TO SUCCESS

Based on its experience with APCs, Tri-Met concluded that it was important to take a "hands on" project management approach in designing and implementing the BDS. The main reason given for this assessment was that the transit industry market for APTS technologies was fairly small and, as a result, vendors typically lacked both experience and key insights related to transit operations and planning/analysis environments.

Probably the most critical organizational factor contributing to success was that the person selected to manage the BDS project had diverse and substantial operations experience, including service as an operator, trainer, and data analyst for FTA Section 15 reporting. Management of the BDS project from the development of the RFP through final implementation was his only responsibility. He was familiar with Tri-Met's experience in archiving APC data, and was aware of the value of this data to service planners, schedule writers, and operations managers. He thus assembled a project team that represented dispatching, scheduling, information systems, maintenance, operations analysis, and service planning. Because dispatchers were primarily interested in real time applications of the BDS, it was the presence of the other interests on the project team that ensured substantial attention to data recovery and archiving issues in the design of the system.

Another important consideration was that the vendor selected for the project was strongly motivated to deliver a successful product. The Tri-Met project was the company's first AVL job in the transit industry, and the client's satisfaction was seen as important for subsequent work with other transit properties. In turn, there was a talented resident staff of programmers, database specialists and data analysts at Tri-Met who worked closely with the vendor to ensure success.

Tri-Met brought valuable prior experience with APCs to the BDS project. It was recognized that an APC-AVL interface would provide a much-improved basis for locational referencing of the APC data, which would allow passenger activity reporting to be upgraded from the time point to the stop level. It was determined that an AVL-APC interface producing stop-level records would require on-board data storage, given capacity constraints on radio transmission. The decision to stop-reference the AVL-APC data thus shifted attention away from an archiving framework based on data generated from a "location-at time" polling cycle.

Synergistic economic considerations are also worth noting. Tri-Met might not have acquired an AVL system had it been linked to the necessary upgrade of its bus dispatch computer system, which substantially reduced AVL's cost by compared to a stand-alone system. Similarly, with on-board location referenced data storage already in place to serve the AVL system, the cost of adding APCs declined from about $5,000 to $1,000 per vehicle.

RETROSPECTIVE CONSIDERATIONS

Overall, Tri-Met managers are pleased with the performance of their AVL/APC system. The data recovered and archived by these systems has allowed them to conduct detailed analysis and reporting of bus operating performance and passenger activity, which is increasingly contributing to improvements in dispatching, operations control, service planning and scheduling. At the same time, potential opportunities were untapped in the design and implementation of the AVL/APC systems. In addition, there were expectations that were not realized. Untapped opportunities and unrealized expectations include the following:

- Instrumentation for monitoring major vehicle components was not included in the AVL system specifications. A recent review of maintenance-related digital messages between operators and dispatchers found a frequency of about 35 per day, or about 6% of vehicles in peak service. Also, records show that in a recent month 235 vehicles (35% of the vehicle fleet, with an average age of 7.5 years) required a road call. Maintenance needs thus involve considerable expense and result in service disruption. Instrumentation of vehicles would have provided a means for creating an automated maintenance log, which may have provided a means for detecting some problems before they deteriorated to the point of failure. Maintenance staff did not actively participate in the development of the BDS specifications. This may have been the result of experiences with the pre-existing Motorola radio system, which was capable of monitoring and reporting engine temperature and oil pressure. The radio system produced numerous "false positives" (e.g., reporting low oil pressure because the engine had been shut down during layover), and the monitoring function was disconnected. Tri-Met is now exploring the addition of drive train monitoring hardware that would provide operating data to the AVL system that could be reported to maintenance at the end of each day.
- Tri-Met has installed electronic registering fareboxes in their bus fleet. However, these fareboxes are not connected to AVL/APC system. Farebox, AVL, and APC data were not integrated for several reasons. First, a farebox interface with the BDS was considered to be too expensive. Second, an interface would have added to the complexity of the system, and there was a desire to limit potential complications while allowing for future expansion. Tri-Met is now considering a shift to smart cards, and an integrated system yielding origin-destination data would be very useful in service planning and marketing. A number of buses also include cameras that are not AVL-integrated. Discs containing digital video images must be physically removed. The images are time referenced, but not location referenced.
- It was expected that the new BDS would not only yield more effective dispatching, but also result in an extension of the responsibility of dispatchers into the area of operations control (Tri-Met, 1991). The monitors displaying real time vehicle location, schedule and route deviation reports, and prioritized digital messages from operators have added a substantial amount of information to the dispatching environment. Dispatchers state that their workload has grown to deal with the increase in information and, consequently, operations control responsibility takes a low priority. Operations control remains the responsibility of field supervisors, but these individuals still lack access to the real time information available to dispatchers. To overcome this shortcoming, Tri-Met is planning to acquire mobile data terminals for field supervisors' vehicles. There is also an interest in elevating the importance of operations control among the various responsibilities vested in field supervision.
- It was intended that the data warehouse structure would allow for data to be conveniently imported into a geographic information system (GIS) to facilitate mapping and spatial analysis of operations at the stop, segment, and route levels. The data warehouse structure allows for AVL and APC data to be queried using structured query language for purposes of operations planning and performance monitoring. While all AVL and APC data are archived, only the most current data are available in disaggregate form. Disaggregate data from previous time periods are archived on removable storage media and can be reconstructed if the need arises. The limiting factor for permanent storage of data on the database server is storage capacity.
- The integration of Tri-Met BDS data with a GIS has been steadily evolving over time. A distinction should be made between visualization and mapping capabilities, and spatial analysis applications. At present, the capabilities for undertaking project-specific spatial analysis are excellent due to the disaggregate nature of the data. One of the principal advantages of data for spatial analysis is that information is collected at the bus stop level, allowing for aggregation to higher levels. The data can be assigned to point locations such as stops, time points, and transfer points or linear features such as routes and route segments. The data can be further summarized according to logical divisions such as block, trip, direction, and time period. In most instances, it is first necessary to summarize information on the database side prior to linking it with corresponding geographic features in a GIS. The signal prioritization project at Tri-Met made extensive use of BDS data within a GIS in the early project planning stages for locating priority intersections and analyzing the potential impacts of stop location (near side/far side) on bus performance. Another use of GIS for spatial analysis involved allocating socioeconomic data to transit service areas

in order to create spatial variables for use in econometric models. This technique was applied in a recent study involving an analysis of passenger demand (Kimpel et al., 2000).

- One area that needs increased attention concerns the use of GIS as a visualization tool to aid in decision making. BDS information is often displayed on maps. For the most part, this information tends to be descriptive in nature. As an example, service planners might display passenger boardings and alightings at bus stops in order to determine candidates for consolidation. The visualization of bus performance information is presently underexploited. There are a number of summary reports generated at Tri-Met on a regular basis such as route performance reports and passenger censuses. It would be advantageous to be able to visually display information on performance measures at alternative levels of aggregation as they vary over time and space.

Some of the issues discussed above, including integration of fareboxes and drive train monitoring, were the result of an overall intent to minimize the complexity of the BDS. Tri-Met's goal was to design and implement a system that was relatively simple and straightforward, while allowing for future expandability. Limiting the complexity of the system helped to insure that it worked as intended, in contrast with other experiences in the transit industry where adoption of new technology has been problematic (Hall, 1980). Presently, four years into successful implementation of the BDS, Tri-Met is exploring a variety of upgrades and extensions to address the needs and opportunities discussed above.

CONCLUSIONS

Tri-Met has over 20 years of experience working with APC technology and data, and about 4 years of experience with AVL technology. It appears that much of the agency's recent success with data archiving and analysis can be traced to its previous efforts to obtain useful data from APCs. It was recognized that problems associated with identifying trips and referencing of passenger activity to time points in the APC data stream could be remedied by integrating APC and AVL technologies. A consequence of APC-AVL integration was the decision to define data records on the basis of "time-at-location" rather than "location-at-time," which could only be achieved through on-board data storage.

The role of the project manager was vital in ensuring that the experience with APCs was reflected in the design of the data recovery and warehousing structure of the BDS. AVL functionality was not a central feature of the BDS project, and data recovery and warehousing were not priority concerns for the dispatchers that the BDS was designed primarily to serve; their interest in AVL was essentially limited to real time use. Having a project manager who was experienced in transit operations and in analyzing operating data was a key to success.

Given the fairly rapid evolution of AVL technology from signpost to satellite-based systems, Tri-Met clearly benefited from being a relatively recent adopter. Although the agency could have developed a data recovery and archiving system with the earlier generation technology (following the experience of King County Metro, for example), the near-exact locational referencing of the current technology was ideally suited to its needs. In contrast, Tri-Met was a fairly early APC adopter, but this technology has not changed as much over time. Thus Tri-Met's experience with APC systems can be seen as an important element in explaining the success they have had with data recovery and archiving in the BDS environment.

Archived AVL-APC data support a variety of regular performance reporting functions and specific evaluation needs at Tri-Met. The agency employs analysts who bring a high level of skill and initiative to these tasks. There are also several experienced programmers in residence, whose primary responsibility has been to maintain aging (soon to be replaced) scheduling software, who can also be called on to support AVL-APC data analysis. The existence of a talented staff can be traced to the data-poor and financially constrained era of the early 1980s, as can the collaboration between the agency and local university researchers.

The wealth of operating data recovered by the BDS has resulted in more comprehensive and valid performance reports. Automating the data recovery process has also shortened the time span and greatly reduced the costs associated with analysis of operations issues. It is unlikely that much of the analysis discussed in this report would have been undertaken in a manual data recovery environment.

As described in previous sections, the implementation of the BDS and its AVL/APC components has yielded tangible improvements in service quality and measurable savings in operating and administrative costs at Tri-Met. A number of potential benefits, however, have not yet been realized. The operations control environment and practices have changed only slightly from the pre-BDS era. In the scheduling arena, where substantial efficiency savings are potentially achievable, archived AVL-APC data are still used in an ad hoc fashion. This may change with the implementation of new scheduling software. While the agency's agreement with operators prevents the use of archived data for disciplinary purposes, its use in spotlighting or rewarding meritorious performance has gone untapped. Given the ability to comprehensively recover data on fairly unambiguous measures of performance, there is an opportunity to introduce incentives that could subsequently improve service quality and reward those who are contributors.

Studying the implementation of advanced transportation technology in a specific setting can identify how that technology leads to changes in practices and improvements in performance. For some of the documented changes it is also possible to monetize the corresponding benefits at this scale of analysis. However, many of the beneficial effects of new technology on performance are subtle and indirect, and are likely to be only vaguely perceived or missed entirely in case study investigations. It is also important to recognize that the changes following implementation of advanced technology may not necessarily be beneficial in every respect. Capturing overall effects resulting from the implementation of advanced technology requires analysis at the industry scale, focusing on such indicators as total factor productivity and costs. In this respect, Gillen et al. (2001), and Gillen and Haynes (2001) provide good examples of industry-level assessment of the economic consequences of advanced technology.

It is difficult to assess the effects of the BDS experience on intra-organizational performance at Tri-Met. Fielding (1987: 181) observed that "(t)ransit organizations are … hierarchically structured with little attention given to information sharing at the second and third levels of management." With respect to BDS design and implementation at Tri-Met, it was recognized at the project development stage that a variety of units in the organization stood to benefit from archived AVL and APC data, including dispatching/operations control, scheduling, planning, finance, marketing, and maintenance. Following implementation, all of these units are, to varying degrees, benefiting from analysis of archived data. The benefits are most evident in the areas of planning and scheduling. Units that were less active in the initial phases of design, implementation, and analysis have now become more active contributors to possible extensions of the BDS, including farebox/smart card integration, addition of mobile terminals to field supervisors' vehicles, and addition of drive train monitoring devices. At the second level of management, there is growing utilization of the ability to directly measure factors closely associated with service quality, effectiveness, and efficiency. In addition to measurement, the ability to evaluate the causes and consequences of quality, effectiveness, and efficiency problems is steadily evolving. These developments are providing senior managers with a much-improved understanding of strategic options in their efforts to advance agency performance.

REFERENCES

1. Casey, R. 1999. Advanced Public Transportation Systems Deployment in the United States. Report No. DOT-VNTSC-FTA-99-1. Volpe National Transportation Systems Center, Cambridge MA.

2. Dueker, K., T. Kimpel, J. Strathman, R. Gerhart, K. Turner and S. Callas. 2001. "Development of a Statistical Algorithm for Real-Time Prediction of Transit Vehicle Arrival Times Under Adverse Conditions." Final Report. Center for Urban Studies, Portland State University.

3. Fielding, G. 1987. *Managing Public Transit Strategically: A Comprehensive Approach to Strengthening Service and Monitoring Performance*. San Francisco, CA: Jossey-Bass Publishers.

4. Gillen, D., E. Chang and D. Johnson. 2001. "Productivity Benefits and Cost Efficiencies From Intelligent Transportation Systems Applications to Public Transit: Evaluation of Advanced Vehicle Location," *Transportation Research Record* 1747, 89-96.

5. Gillen, D., and M. Haynes. 2001. Measuring Aggregate Productivity Benefits From Intelligent Transportation System Applications: The California Experience," *Transportation Research Record 1774,* 52-59.

6. Hall, P. 1980. *Great Planning Disasters*. Berkeley and Los Angeles, CA: University of California Press.

7. Kimpel, T., J. Strathman, D. Griffin, S. Callas and R. Gerhart. 2002. "Automatic Passenger Counter Evaluation: Implications for National Transit Database Reporting," Center for Urban Studies, April.

8. Kimpel, T., J. Strathman, K. Dueker, D. Griffin, R. Gerhart, S. Callas and K. Turner. 2000. "Analysis of Passenger Demand and Transit Service Reliability," Center for Urban Studies, June.

9. Levinson, H.S. 1991. *Synthesis of Transit Practice 15: Supervision Strategies for Improved Reliability of Bus Routes*, National Cooperative Transit Research and Development Program, Transportation Research Board, National Research Council, Washington, DC.

10. Strathman, J., K. Dueker, T. Kimpel, R. Gerhart, and S. Callas. 2002. "Evaluation of Transit Operations: Data Applications of Tri-Met's Automated Bus Dispatching System," Transportation, 29, 321-345.

11. Strathman, J., K. Dueker, T. Kimpel, R. Gerhart, K. Turner, S. Callas and D. Griffin. 2001. "Bus Transit Operations Control: Review and an Experiment Involving Tri-Met's Automated Bus Dispatching System." *Journal of Public Transportation*, 4(1): 1-26.

12. Strathman, J. K. Dueker, T. Kimpel, R. Gerhart, K. Turner, S. Callas. 1999 "Automated Bus Dispatching, Operations Control, and Service Reliability: Baseline Analysis," *Transportation Research Record* 1666, 28-36.

13. Strathman, J. and J. Hopper. 1993. "Empirical Analysis of Bus Transit On-Time Performance," Transportation *Research-A,* 27A:2, 93-100.

14. Strathman, J. and J. Hopper. 1991. "An Evaluation of Automatic Passenger Counters: Validation, Sampling and Statistical Inference," *Transportation Research Record* 1308, 69-77.

15. Strathman, J., K. Dueker, T. Kimpel, R. Gerhart, K. Turner, P. Taylor, S. Callas and D. Griffin. 2000 "Service Reliability Impacts of Computer-Aided Dispatching and Automatic Vehicle Location Technology: A Tri-Met Case Study," *Transportation Quarterly*, 54:3, 85-102.

16. Tri-Met. 1991. *Operations Control Plan for the Tri-County Metropolitan Transportation District of Oregon*. Portland, OR: Tri-Met.

17. Tri-Met. 2000. *Tri-Met Stop Level Passenger Census*. Portland, OR: Tri-Met.

18. Urban Mass Transportation Administration. 1978. "Sampling Procedures for Obtaining Fixed Route Bus Operating Data Required Under the Section 15 Reporting System." Circular 2710.1. Washington, DC: U.S. Department of Transportation.

19. Wilson, N., D. Nelson, A. Palmere, T. Grayson and C. Cederquist. 1992. "Service Quality Monitoring for High Frequency Transit Lines." Paper No. 920781, Transportation Research Board, 71st Annual Meeting, Washington, DC.

APPENDIX B: New Jersey Transit's Developing APC and Archived Data User Service*

A contract for what is expected to become a large scale APC implementation at New Jersey Transit was let in 1998. In the pilot implementation, begun in 2000, eight buses operating out of the Ironbound Garage in Newark are equipped with passenger counters and an event recording system that logs, with time and location stamps, operational events such as door open, door close, turns, fare register activity, engine start and shutdown, speed threshold transition, electrical condition and coach temperature. Data are automatically uploaded nightly, matched to the schedule, and loaded into a database. Planning and management reports have been developed for this database that permit both automatic and user-requested reporting. The database can also be queried *ad hoc*.

While the number of vehicles equipped to date is small, the data system architecture will accommodate expansion to the entire bus and rail fleet. Expansion to one of NJT's light rail systems is imminent. The on-board architecture is likewise designed to allow modular expansion as more intelligent devices and data interfaces are added to vehicles.

NJT is working with a company that has traditionally been an AVL vendor. This is unusual since NJT's system does not include real-time location reporting by radio. However, it uses the same location system used in AVL, complemented by passenger sensors. The development of an archived data system in cooperation with a vendor whose business has traditionally dealt with real-time data has underscored the differences between providing real-time data and providing archived data for planning and management.

Special Features of NJT

NJT is an extremely diverse transit system. Its statewide operations include urban bus service, long distance express bus service to New York and Philadelphia, other rural long distance bus service, two light rail systems, commuter rail, and paratransit. Including contracted services, it operates about 2100 buses out of 27 garages located around the state. It has to coordinate ITS architecture with several metropolitan planning organizations, three state DOT's, and several toll collecting agencies. As part of the US DOT's Model Deployment Initiative for Advanced Traveler Information Systems (ATIS) in the New York metropolitan area, NJT's route and schedule data is maintained and published by a private contractor.

NJT's enormous service area militated strongly for GPS as its location technology. Its large number of garages made manual methods of retrieving data (e.g., collecting disks from on-board computers) out of the question. Its diverse fare structure, and the fact that rural buses use distance-based fares and provide change, makes its farebox system requirements unique, creating some unique opportunities.

HISTORY OF NJT'S AUTOMATED DATA PROGRAM

Experience with Signpost-Based AVL

In the early 1980's, NJT had a signpost-based AVL system installed as part of a new radio system. The entire state was covered with fewer than 500 signposts, most centered on the City of Newark, NJ. That system was not designed for supplying archived data. It is used mostly for managing urban headways and logging pullouts.

Getting an AVL Vendor to Do APC

NJT began planning for APC in 1983. In 1994, they sponsored a demonstration project with 6 instrumented buses. The demonstration project was followed by a technology assessment report in 1996. During this process, the vision for the system expanded beyond passenger counting to include event logging and travel time analysis, and an

---

* This report was prepared by Peter G. Furth of Northeastern University. Agency staff interviewed were Jim Kemp, Manager of Data Coordination and Integration and Project Manager for the APC / archived data service project; Kevin Landrigan, Director of Schedules and Data Integration, Bus Service Planning Department; Doug Cattan, Director of Bus Systems Development; and Glenn Newman, Manager of Geographic Information Systems.

expandable "smart bus" on-board architecture. Based on the technology assessment, the procurement specifications came to look more like a request for proposals than a standard bid specification, because NJT wanted to give bidders the freedom to use any technology to meet functional needs. In particular, the spec was written in a way to permit an AVL vendor to compete using its location system with the addition of passenger counting equipment, even though the specification was not for AVL *per se*. The bid was won by Raytheon TMS, which was subsequently bought by Orbital Sciences, an AVL vendor that at the time had just become the North American distributor of Iris passenger counters.

To a large extent, NJT's story is one of a transit agency and an AVL vendor learning together what are the differences between AVL and an archived data system, and making that archived data system work. NJT project manager Jim Kemp focused on lessons learned about those differences in a presentation given recently at the Transportation Research Board Annual Meeting (January, 2002). He reminded his listeners several times that supporting an archived data service presents a whole new set of data challenges, "because the data's not just for AVL, anymore." Those lessons learned are incorporated in this case study.

Limited Implementation

Notice to Proceed, issued in 1998, envisioned retrofit of 140 buses and installation in 100 new buses to be ordered soon. However, budget priorities quickly shifted funding to acquisition of additional new buses. A pilot implementation on 8 buses in the Ironbound Garage in Newark (about 10 percent of the garage's transit fleet) began in 2000. At first, the instrumented buses were rotated among the routes served out of that garage, providing a test of the system's breadth. Experience with 10 percent sampling led staff to conclude that with that small a sample, automated data collection was sufficient for quantitative performance measurement, and to identify gross problems, but was insufficient to design solutions.

More recently, the eight instrumented buses were concentrated primarily on one route, Line 62, a long route that serves Newark Airport and therefore has a dedicated fleet of 17 buses furnished with luggage racks. With 8 of those 17 buses gathering data automatically, planners began to get data heretofore unknown: repeated samples of every trip in the schedule – and moreover, of consecutive trips on the same day. This level of data has given them insight and confidence enough to generate solutions, i.e., change route headways and schedules, and has led to the conviction that equipping the entire fleet is worthwhile. Line 62 now serves as a good testbed for development of reports and other applications using archived data. Because Line 62 has 20 different "branches" or service variations, it is also a good test of algorithms that try to match the event log to the actual route taken.

The next expansion of the data gathering system will be to the (Newark) City Subway, one of NJT's light rail lines. A complication of this line is that it includes segments of underground trackage; and that the subway sections are of course impervious to GPS signals. To overcome this limitation without changing the system configuration, several GPS repeaters will be used to rebroadcast GPS signals in the tunnel. The wide, low floor doors of the light rail vehicles posed a challenge to passenger counters, but tests indicate that at least two vendors have counters that will work.

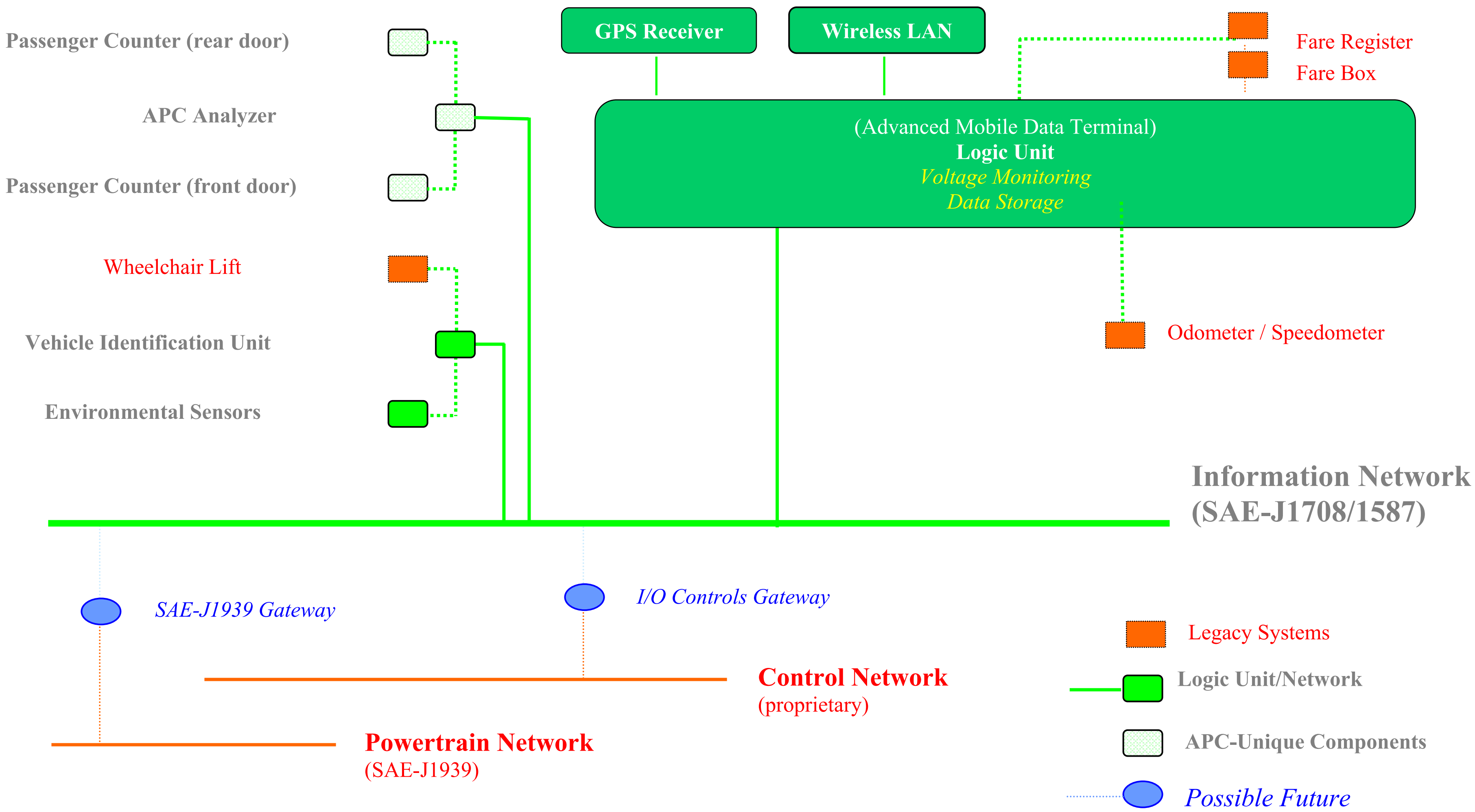
In the meantime, all new bus purchases continue to specify inclusion of an integrated GPS antenna/receiver with an RS232 serial port, but not wiring for an on-board network or other data-related equipment.

INTELLIGENT BUS APPROACH

Direct Connections and Network Connections

The heart of the data collection system is an on-board computer, called the vehicle logic unit, which serves as the event recorder. It identifies and stores events, stamping each one with time and location (GPS coordinates); it provides the network clock, and it monitors voltage. Other on-board devices are connected either by direct connection or by means of an SAE J1708 vehicle area network (VAN) that has been installed in the instrumented buses. The on-board architecture is illustrated in Figure 1.

FIGURE 1: Current On Board Architecture



Devices that are directly connected to the vehicle logic unit include, a GPS unit, a wireless LAN transmitter / receiver used for downloading and uploading data, the electronic fare register, and the odometer / speedometer. The logic unit also includes an integral operator console (data terminal), which is not used in the current implementation. Devices connected by means of the J1708 network include a vehicle identification unit and a dedicated on-board passenger count processor called the APC analyzer. The vehicle identification unit is connected to the doors, the wheelchair lift, the engine, and environmental sensors (coach temperature). The APC analyzer is connected to passenger counting sensors and door sensors.

The J1708 network is versatile because it will allow additional devices (for example, stop enunciators) to be connected as they become installed. The on-board architecture was designed to be expandable as other smart bus features, including digital radio that will enable real-time AVL, are added, as illustrated in Figure 2. In that future scenario, integration relies almost exclusively on the J1708 network, rather than direct connections to the vehicle logic unit. This architecture simplifies wiring and permits modular software and device development because it allows data sharing between devices without having to go through the vehicle logic unit.

Each night during servicing, data stored in the event recorder is broadcast via a standard IEEE 802.11 high speed wireless local area network (W-LAN). The data is received by a computer installed in the garage, stored, and forwarded by wireline to headquarters. There, the data is processed and reduced, and loaded into the central database. The raw data is also saved. The wireless LAN is also used to upload software and data updates to the vehicle logic unit.

Farebox Integration Experience

In the early 1980's NJT became the first transit agency in the world with networkable fareboxes, which feature an RS485 serial port to connect to a spotter display. In 1999, NJT experimented with connecting those fareboxes to the event recorder computer. Their farebox vendor made software modifications to activate an additional RS485 output channel, which was then connected to an input channel on the event recorder computer. The fareboxes were programmed to strip off and transmit operator sign-on information and certain transaction data. The experiment was successful, but is currently turned off (by reverting to the unmodified farebox software).

Farebox integration is still contemplated as the data collection system develops, though the project manager recognizes that there will inevitably be discrepancies between counts as recorded by the farebox computer and the passenger count sensors, which may require reconciliation. Another possible arrangement is to integrate the farebox data stream with the event recorder data stream off-line, using common tags such as vehicle ID, time, and events such as door openings. This arrangement, too, would require some software modifications and on-board device integration because their farebox data is not currently time-stamped or connected to the doors, and they would need to ensure a common clock.

EVENTS CAPTURED

Every event listed in Table 1 generates a record with a location and time stamp. Captured events include door open and close, acceleration and deceleration across two speed thresholds, and start and end of turn. The system manager can set the speed and turn detection thresholds. The system also generates periodic records that include coach temperature and electrical status. The system manager can set the period (typically 30-120 seconds) and the threshold for electrical discrepancy identification. By setting the period to a small value, near-continuous records of location can be obtained (since the periodic records also include time and location stamps). For example, the period was once set to 6 s in order to map a bus's route through a shopping center.

The APC analyzer stores signals from its sensors in a buffer, accumulating counts of passengers entering and leaving, and then broadcasts the counts when triggered by a door close event. Likewise, the logic unit stores GPS location every second in its buffer. When a sufficient change in direction (heading) is detected, a record is made which includes the time at which the turn actually began or ended.

ARCHIVED DATA SERVICE APPROACH

NJT's data service approach begins with the concept of the data warehouse in recognition of the many types of data and databases that must relate to each other for effective data analysis, illustrated in Figure 3.

FIGURE 2: Possible Future Onboard Architecture

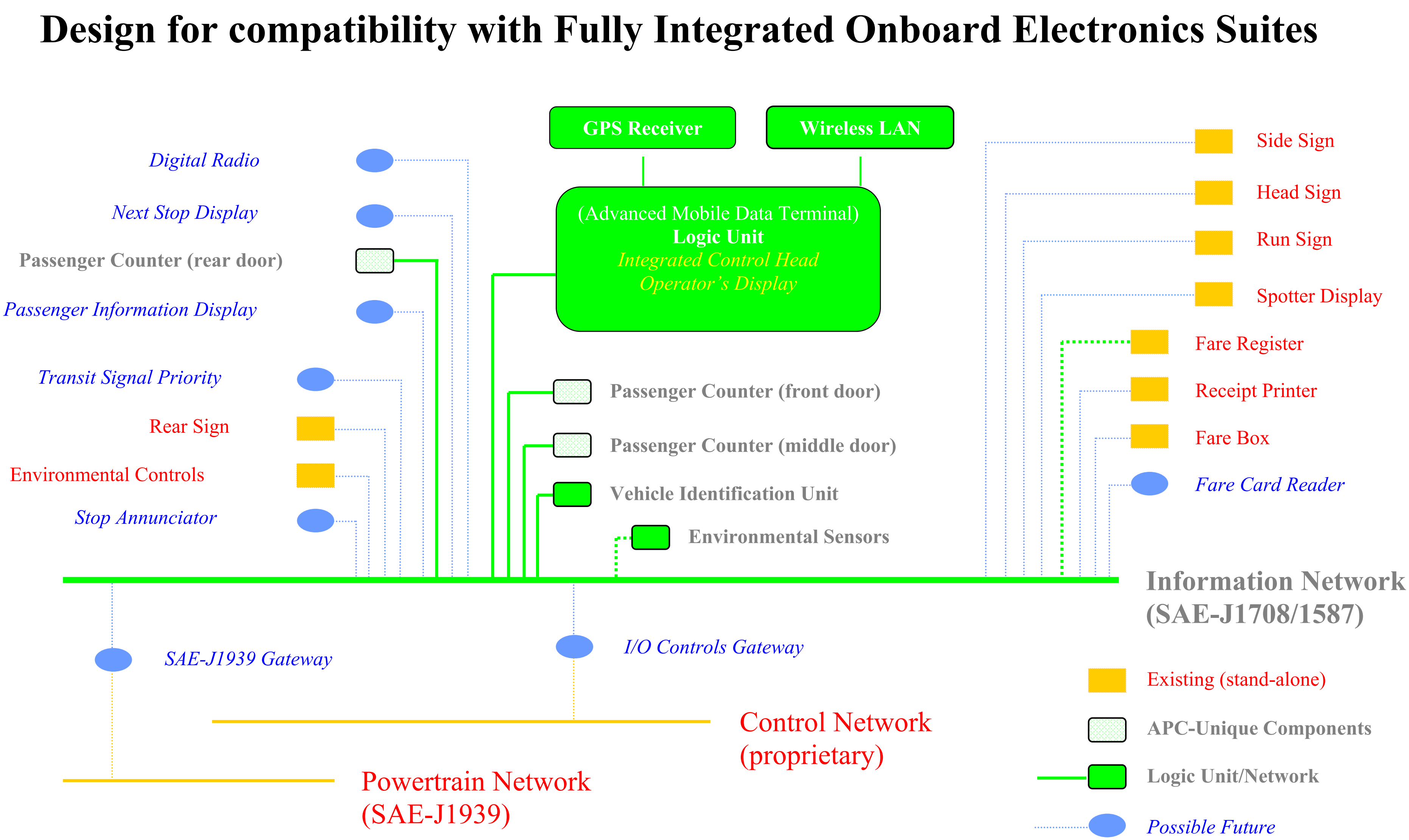
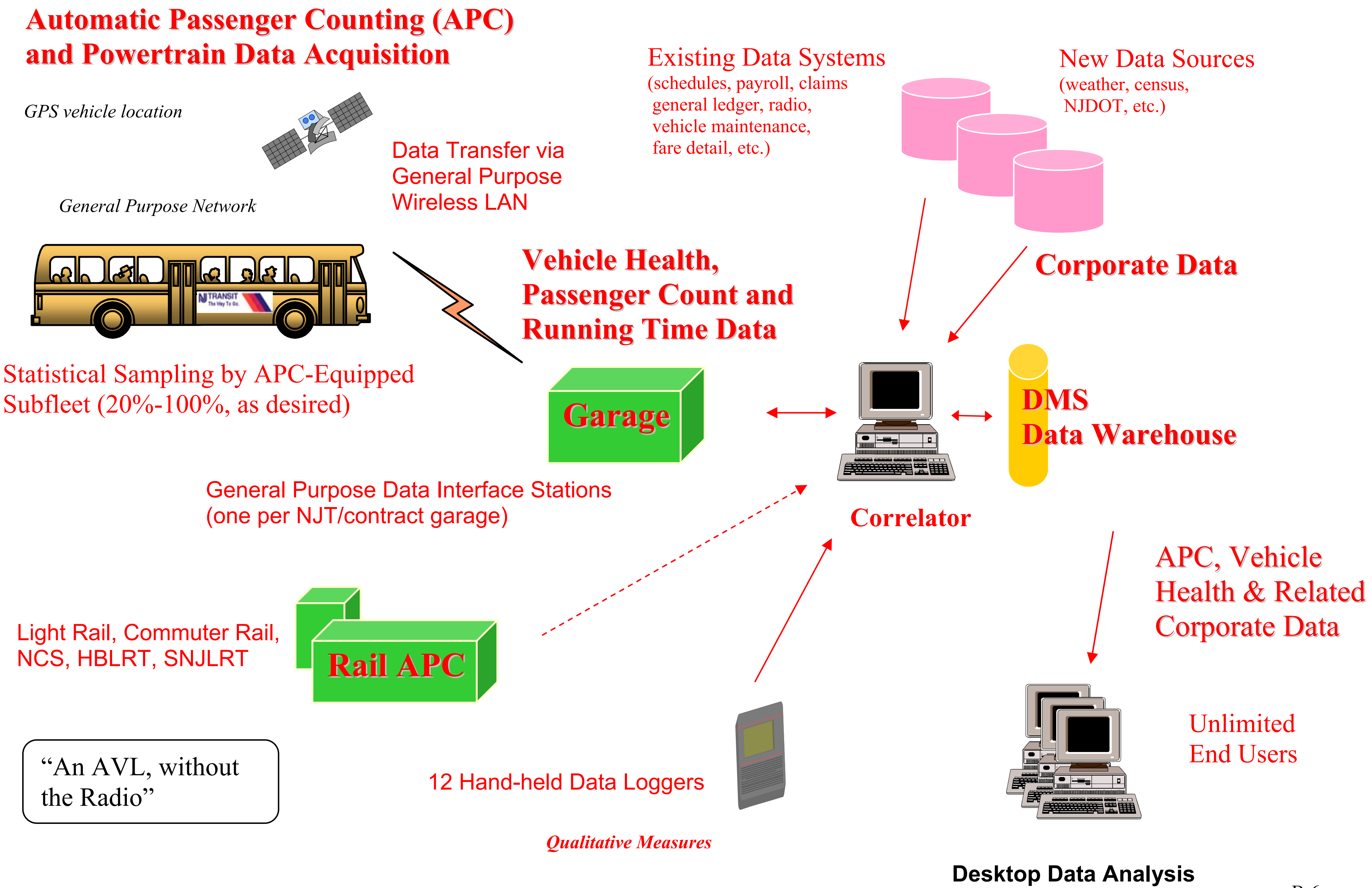


FIGURE 3: Overall System Architecture



Event data from the automated data gathering system is stored in databases at three levels of detail. The primary level used for analysis is "correlated data," i.e., data that has been matched to a particular trip, with detail on each stop and each interstop segment. In addition, the raw event data is maintained for various reasons (testing new matching algorithms, investigating incidents, backup in case matching has to be redone). In addition, pre-aggregated trip level summaries are stored in a separate table that allows faster retrieval.

Other sources of operations data in the warehouse concept will include farebox data, data gathered by field checkers using hand-held devices, and automatic data collection systems on NJT's various rail systems. These sources will be complemented by other corporate data systems, including schedule, payroll, and vehicle maintenance, and by external data sources including weather, census, and NJDOT highway data.

Implementation of the data warehouse concept is still in its infancy. Nevertheless, the concept underscores several important points about archived data use, including the following:

- Operations data must be correlated to the schedule in effect at the time the data was collected, so a full history of schedule data is needed.
- Links to other corporate data systems such as payroll should be facilitated by the use of common tags such as time and operator ID. Care must be taken to ensure the integrity of those tags, since operators routinely relieve each other on a given bus, and frequently perform service for which they were not scheduled (as directed by supervisors in response to incidents or absences).
- External data sources such as weather are valuable inputs for data analysis. The data system should be configured to relate to those data sources in the database in which they come, rather than require that those data be manually input to the operations data database. Few of these external sources are likely to have data tags relevant to transit, however. So it will likely also be necessary to develop the means to relate much of this external data by geographical location rather than through common tags.

DATABASE AND REPORTS

Data matched to trips and stops is held in an Oracle database, which can be queried using SQL. Standard reports are developed in house by the project manager using a data manipulation and report-generating product called Brio Enterprise. The project manager reports that the additional power and ease of use that come with a higher end report generating package is well worth the extra expense. He finds that it takes from 1 to 16 hours to develop a report, depending on the complexity.

They have also created some map reports using Thematic Mapping within MapInfo.

Some example reports are included as Figures 4-6. Figure 4 is a single trip analysis report, paralleling a ride check report, with an emphasis on tracking running time and passenger load. It was the first report in demand, since it replicates the report planners were used to from manual counts. Figure 5 is page 1 of a summary of trips that violate a performance standard in running time, standing time, standees, or on-time performance. It averages performance over all samples of a given scheduled trip (e.g., the Route 62 10:25 am weekday trip inbound on pattern 1). Figure 6 provides a performance summary by route / direction / period, including statistics on peak load, standees, passenger-miles, running time, standing time, and schedule adherence. In both Figures 5 and 6, performance statistics that don't meet the standard are highlighted.

INTEREST IN VEHICLE MAINTENANCE

Regarding future development of the automated data collection system, NJT management has expressed an interest in getting data that can help them with vehicle maintenance. Some of the expressed needs have to do with real-time data – given the transit agency's enormous service area, they would like to be able to use GPS to locate a crippled bus, and mechanical sensor data to help diagnose the problem (can the bus continue, or should it be taken out of service?). Like large commercial fleets, they'd like the ability to remotely interrogate the vehicle for diagnosis.

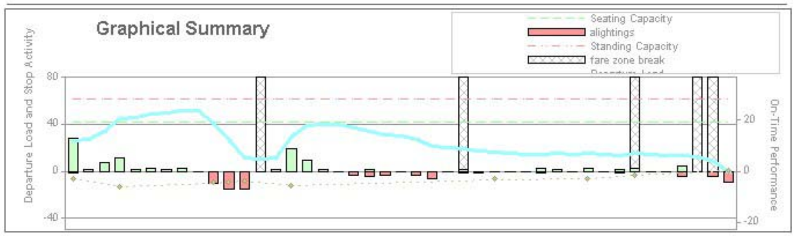
Interest in using archived data to help with vehicle maintenance is strong, but less well formed. One idea is to use data gathered from the event logger to complement the traditional measures of vehicle use – miles, engine hours and idle hours. For example, the event logger can also count acceleration / deceleration cycles, note instances of heavy engine demand (e.g., a bus full of passengers accelerating up a hill), and track data from mechanical sensors such as engine

FIGURE 4: Single Trip Analysis Report

APC Single Trip Analysis Report

Route 62 **Run** *(n/a)* **10:45 AM** **Outbound Trip** **Saturday, Jun 29, 2002**

Bus 1844 (MetroD) **Operator #** *(n/a)* **Saturday Schedule** **Jul 02 pick** **PENNLANE to WDBRCTML**



Trip Summary:

Origin Arrival Time: 10:43 AM Pre-trip Layover: 17.1 min

98 Passengers Max Load: 52 Scheduled End Time: 12:30 PM Trip Length: 28.9 miles 634.3 Psqr Miles

Scheduled Running Time: 105 min; Actual: 102 mir Max Passenger Standing Time: 12 min

Trip Detail:

Actual Time	Correlated Stop Description	Ons	Offs	Departure Load	Scheduled Time	Fare Zone
10:47 AM	PENN STATION BUS LANES	28	2	26	10:45 AM	P-1
10:51 AM	MULBERRY ST. & COMMERCE ST.	1	0	27		P-1
10:53 AM	COMMERCE ST. & BROAD ST.	7	0	34		P-1
10:55 AM	BROAD ST. & BRANFORD PL.	11	0	45	10:49 AM	P-1
10:58 AM	BROAD ST. & WEST KINNEY ST.	1	0	46		P-1
10:59 AM	BROAD ST. & PENNINGTON ST.	3	0	49		P-1
11:00 AM	BROAD ST. & SOUTH ST.	1	0	50		P-1
11:00 AM	BROAD ST. & PARKHURST ST.	2	0	52		P-1
11:05 AM	unidentified stop	0	1	51		P-1
11:08 AM	NEWARK AIRPORT TERMINAL A	0	10	41	11:04 AM	P-1
11:11 AM	NWK AIRPORT TERMINAL B	0	15	26	11:07 AM	P-1
11:13 AM	NWK AIRPORT TERMINAL C	0	15	11	11:10 AM	P-1
11:20 AM	RT 1& 9 & NORTH AVE.	0	1	10		P-2
11:24 AM	EAST JERSEY ST. & CATHERINE ST.	1	0	11		P-2
11:30 AM	BROAD ST. & W JERSEY ST.	19	0	30	11:25 AM	P-2
11:34 AM	RAHWAY AVE. & BROAD ST.	9	0	39		P-2
11:36 AM	RAHWAY AVE. & DE HART PL.	1	0	40		P-2
11:38 AM	RAHWAY AVE. & ELMORA AVE.	0	0	40		P-2
11:39 AM	SAINT GEORGES AVE. & HAGEL AVE.	0	3	37		P-2
11:40 AM	SAINT GEORGES AVE. & PARK AVE.	1	4	34		P-2
11:41 AM	SAINT GEORGES AVE. & THOMPSON AVE.	0	3	31		P-2
11:42 AM	SAINT GEORGES AVE. & DRAKE AVE.	0	1	30		P-2
11:42 AM	unidentified stop	0	3	27		P-2
11:43 AM	SAINT GEORGES AVE. & CHANDLER AVE.	0	6	21		P-2
11:44 AM	SAINT GEORGES AVE. & WARREN ST.	0	1	20		P-2
11:45 AM	SAINT GEORGES AVE. & VICTORY ST.	1	2	19		P-3
11:46 AM	SAINT GEORGES AVE. & CRESCENT AVE.	0	2	17		P-3
11:48 AM	SAINT GEORGES AVE. & WOOD AVE.	0	1	16	11:45 AM	P-3
11:51 AM	SAINT GEORGES AVE. & STILES ST.	0	1	15		P-3
11:52 AM	SAINT GEORGES AVE. & APGAR TERRACE	0	1	14		P-3
11:53 AM	WEST SCOTT AVE. & ST. GEORGES AVE.	2	2	14		P-3
11:57 AM	ELIZABETH AVE. & WEST GRAND AVE.	1	0	15		P-3
11:58 AM	IRVING ST. & ELIZABETH AVE.	0	1	14		P-3

APC Single Trip Analysis Report

Actual Time	Correlated Stop Description	Ons	Offs	Departure Load	Scheduled Time	Fare Zone
11:59 AM	IRVING ST. & BROAD ST.	2	1	15	11:56 AM	P-3
12:02 PM	EAST HAZELWOOD AVE. & LEESVILLE AVE.	0	1	14		P-3
12:03 PM	EAST HAZELWOOD AVE. & WHELAN PL.	1	2	13		P-3
12:09 PM	ROOSEVELT AVE. & HARRISON AVE.	2	0	15	12:08 PM	P-4
12:11 PM	ROOSEVELT AVE. & HAYWARD AVE.	0	1	14		P-4
12:13 PM	PERSHING AVE. & ROMANOWSKI ST.	0	1	13		P-4
12:16 PM	ROOSEVELT AVE. & TERMINAL AVE.	4	4	13		P-4
12:24 PM	MAIN ST. & ELEANOR PL.	0	1	12		P-5
12:30 PM	WOODBRIDGE CTR. & WOODBRIDGE TERR	0	4	8		P-4
12:29 PM	WOODBRIDGE CENTER	0	9	-1	12:30 PM	P-4

FIGURE 5: Trip Exception Report

Route Summary Report

11/09/02 - 11/15/02

89 samples 70 trips

Trip Exception Summary

Atis Route Id	Service Type	Direction	Pattern Desc	Trip	number of samples	average running time discrepancy	avg max standing time	avg max standees	percent of timepoints "on time"
62	Saturday	Inbound	NAPTTRMA to PENNLANE	9:50 PM	1	2.7	0.0	0	**0%**
			SMTHDAVD to PENNLANE	9:50 AM	1	6.7	0.0	0	**33%**
		Outbound	PENNLANE to SMTHDAVD	5:25 AM	1	-3.1	**14.0**	20	100%
	Sunday	Inbound	BRODJRSY to PENNLANE	7:00 AM	1	-4.3	0.0	0	**0%**
		Outbound	PENNLANE to BRODJRSY	10:00 PM	1	2.2	5.0	1	**0%**
			PENNLANE to SGEOWOOD	8:00 PM	1	-5.8	0.0	0	**13%**
	Weekday	Inbound	BRODJRSY to PENNLANE	3:46 PM	1	**18.3**	0.0	0	**0%**
				3:46 PM	1	**15.6**	0.0	0	**0%**
			MTROPARK to PENNLANE	4:45 AM	1	1.6	**14.0**	16	**0%**
			NAPTTRMA to PENNLANE	3:30 AM	1	-4.2	0.0	0	**0%**
				11:20 PM	1	4.4	9.0	4	80%
				3:30 AM	1	-5.9	0.0	0	**0%**
				11:20 PM	1	-4.4	0.0	0	**0%**
			SMTHDAVD to PENNLANE	8:49 AM	1	**20.6**	**18.0**	7	**0%**
				11:49 AM	1	2.0	0.0	0	**0%**
				6:30 AM	1	**14.7**	**11.0**	6	50%
				12:49 PM	1	1.3	0.0	0	**44%**
			WDBRCTML to PENNLANE	11:32 AM	1	**19.5**	**15.0**	7	**0%**
				12:32 PM	1	**10.9**	**14.0**	7	**30%**
		Outbound	PENNLANE to BRODJRSY	4:50 PM	1	8.6	0.0	0	**0%**
				11:20 AM	1	-0.5	10.0	7	**0%**
				11:20 AM	1	-8.3	8.0	9	83%
			PENNLANE to NAPTTRMC	4:00 AM	1	-4.0	0.0	0	**0%**
			PENNLANE to SGEOWOOD	1:30 PM	1	5.1	9.0	6	**0%**
			PENNLANE to SMTHDAVD	6:35 PM	1	**21.4**	0.0	0	**0%**
			PENNLANE to WDBRCTML	2:05 PM	1	-1.8	4.0	3	**40%**
				10:05 AM	1	**10.5**	0.0	0	**36%**

FIGURE 6: Route Summary Report

Route Summary Report

11/09/02 - 11/15/02

89 samples 70 trips

Atis Route Id	Service Type	Direction	Time Of Day	number of samples	average max load per trip	max observed load	average productivity psgr mi per hr	average running time discrepancy	average standees per trip	average standing time per trip	percent of timepoints "on time"
62	Weekday	Inbound	Early Morning	9	26	58	294	-2.0	2	1.6	68%
			Morning Peak	4	**45**	52	360	5.5	6	7.5	69%
			Late Morning	1	**49**	49	350	**20.6**	7	**18.0**	**0%**
			Mid-Day	4	31	49	204	7.3	2	3.8	**38%**
			Early Afternoon	4	34	49	186	3.9	2	5.5	**48%**
			Afternoon Peak	4	30	40	240	**16.1**	0	0.0	**33%**
			Evening	2	18	20	192	-1.4	0	0.0	**44%**
			Late Night	9	33	46	261	-1.6	0	1.0	90%
		Outbound	Early Morning	7	36	42	462	-4.6	0	0.0	88%
			Morning Peak	6	26	31	235	2.2	0	0.0	86%
			Late Morning	4	37	51	315	-0.1	4	4.5	**48%**
			Mid-Day	7	**44**	50	445	0.3	4	5.7	76%
			Early Afternoon	3	32	45	286	-4.2	1	1.3	73%
			Afternoon Peak	1	11	11	106	8.6	0	0.0	**0%**
			Evening	2	27	34	315	**14.0**	0	0.0	**38%**
			Late Night	2	32	35	304	-2.6	0	0.0	87%
	Saturday	Inbound	Early Morning	2	31	33	369	5.7	0	0.0	85%
			Morning Peak	2	38	41	345	9.1	0	0.0	78%
			Late Morning	1	41	41	392	6.7	0	0.0	**33%**
			Mid-Day	2	23	29	166	-1.6	0	0.0	100%
			Late Night	1	25	25	386	2.7	0	0.0	**0%**
		Outbound	Early Morning	2	**45**	62	466	-4.3	10	7.0	100%
			Morning Peak	1	32	32	290	4.3	0	0.0	80%
			Late Morning	1	**46**	46	312	-7.6	4	10.0	100%
			Mid-Day	1	**58**	58	453	-7.7	16	10.0	83%
	Sunday	Inbound	Morning Peak	1	27	27	270	-4.3	0	0.0	**0%**
			Late Morning	1	25	25	280	-13.0	0	0.0	100%
			Early Afternoon	1	24	24	217	5.4	0	7.0	100%
		Outbound	Early Morning	1	32	32	469	-2.1	0	0.0	100%
			Mid-Day	1	**45**	45	399	4.9	3	7.0	100%
			Evening	1	31	31	431	-5.8	0	0.0	**13%**
			Late Night	1	**43**	43	633	2.2	1	5.0	**0%**

temperature. Finer data such as this might improve the agency's ability to diagnose problems as (or before) they appear, to predict failures, and to optimize maintenance schedules.

GEOGRAPHIC AND SCHEDULE DATA ISSUES

Like any other transit agency, NJT knew roughly where its routes and stops were and what its schedule was. However, they found that an archived data service required far greater accuracy in their database than had been needed for either operations or AVL. They also found differences that had to be resolved between the simple data model that had first been proposed and NJT's actual business model.

Schedule and Geographic Databases

NJT uses a private contractor to maintain transit route and schedule information for its Automated Traveler Information System (ATIS). Whenever NJT changes a stop or schedule, it sends an updated schedule file to the ATIS contractor. The ATIS contractor then produces an integrated extract of revised pattern, stop and schedule files for NJT's internal use. That extract is, in practical terms, the source of route and schedule data for the archived data service. A pattern is modeled in the ATIS extract simply as a sequence of stops, as opposed to a sequence of links going from one timepoint to another as would be found in a scheduling database. There can be many patterns on a given route, each representing a variation of the route.

NJT also has a GIS group that maintains its own database of stop locations, patterns and routes. The agency's GIS is built on a base map provided by a vendor, Navtec. In this base map, a link is a roadway link between intersections.

While the GIS group was working with a mobile GPS unit to survey the actual locations of bus stops, the ATIS contractor, under a more significant time constraint, geocoded stop locations from whatever simple descriptions were available such as "Third and Elm." Experience trying to match location data as determined by the GPS units of equipped buses to stops in the ATIS extract showed that geocoding was too inaccurate for automatic passenger counting, however. It also led to occasional gross errors; for example, a data entry error in one digit once misplaced a stop by 40 miles; another time, the geocoding software found the right intersection ("Third and Elm") but in the wrong town.

In the GIS, a pattern is defined as a sequence of timepoint intervals, where a timepoint interval is a sequence of stops common to all routes using those stops. Therefore, every route branching point forces the start of a new timepoint interval. Planning for the eventual merging of the ATIS and GIS databases is underway. Currently, about 95 percent of NJT's stop locations have been accurately field measured for the GIS. Since the process of updating stop locations (in response to real route and stop changes) is never ending, this percentage can be expected to remain relatively constant for the foreseeable future.

Geographic Model Subtleties

In this section, we describe some of the time-space data model issues that NJT has had to deal with. The business model that the company follows is not as simple as data modelers might like it to be. It is usually not possible to alter business practices to comply with a simple data model; rather, the data model must adapt to the complexities of the business model.

Sequence of Stop Patterns, Timepoints, or Road Segments?

Is a route (or a route branch) a sequence of stops, a sequence of timepoints, a sequence of timepoint intervals (a timepoint interval being a sequence of stops with a timepoint at each end), or a sequence of road segments? In the schedule database, a route is a sequence of timepoints; in the ATIS extract, a sequence of stops; in the GIS route definition, a sequence of timepoint intervals; in the underlying GIS map, a sequence of road segments; and in actual service, all four.

In the event log, stop locations are noted if a bus actually stops – whether or not it opens its doors. If a bus does not stop, time at a stop or a timepoint has to be interpolated from straddling event records.

Current practice is to automatically match event data to stops in the APC database. Applying the schedule data, e.g., comparing a vehicle's departure time with the schedule, therefore requires matching a timepoint to a stop. This presents two common complications. First, timepoints may not be located at bus stops. Timepoints have been selected as good places to control a route's operation, and sometimes the best place to control the operation is not at a stop. NJT's ATIS database deals with this problem by providing the ability to add dummy stops to represent timepoint locations that are not actual bus stops. Second, when a stop is used a timepoint, the timepoint database will typically use a single point for an intersection, while stops in the two directions of a route are in different locations. Therefore, several stops may be associated with any single timepoint. The nominal coordinates for the timepoint itself may be either calculated as the average of those of its associated stops, or simply copied from an associated landmark or intersection. NJT's matching process identifies timepoints in the context of the patterns on which they occur.

Before APC data was available, there was little need to pay attention to stops between timepoints. Therefore, stop lists had many inaccuracies that had gone undetected for years. Sometimes the correct set of intermediate stops were listed, but not in the correct order.

NJT's data model does not explicitly account for the roads along which buses operate. Analyzing all the bus service along a particular roadway segment – perhaps a trunk shared by several routes – is not facilitated by the current data model. The GIS group is currently working on a project to merge congestion data, provided by NJDOT, with transit demand to estimate the amount of person-delay due to congestion. They currently assume bus loads, but would prefer to be able to get bus loads directly from the APC database. Doing so will require links between road segments and route segments.

Likewise, analyzing the quality of bus service in an area – e.g. a particular town or neighborhood – demands data links beyond a simple collection of routes. NJT's GIS group is just beginning to do geography-based analyses using the APC/event data. The goal for the future is to have GIS scripting prepared for commonly needed geographical analyses so that the analyses will be available to users without GIS expertise.

Bus Stop Issues

Buses often stop at locations that are not known bus stops. Some of those locations may be actual bus stops that the stop database should learn about.

At terminals and other major stops, a "bus stop" may actually encompass an array of stop locations.

Buses can "stop" more than once at a given stop. For example, they can close their doors, creep forward a bit, and then open their doors again.

Two Trips at Once

On many routes, especially those from South Jersey suburbs to Philadelphia, buses serve, in effect, two trips at once. While passing through downtown Philadelphia, they are discharging passengers from the inbound trips at the same time they are boarding passengers for the outbound trip. Any route ending in a loop can exhibit this pattern; even routes without loop ends can exhibit it if passengers are allowed to stay on board during the turnaround. Passengers can board near the end of a line heading outbound, remain on board during the turnaround, and then continue in the opposite direction. (Why? To get a seat, get protection from the weather, or simply escape boredom.) A data model that doesn't recognize this possibility can result in strange ridership figures.

Two Buses Serving the Same Trip

Sometimes, due to high demand, a platoon of buses is dispatched to serve what in the schedule is a single trip. This is especially common in light rail operations, where two or more vehicles may be joined together into a train.

MATCHING TO ROUTE / SCHEDULE

NJT had their APC vendor develop software, called a "correlator," to match the logged operations data to scheduled trips. It is intentionally designed to not rely on operator input, but instead to determine route and trip by recognizing operation patterns. Based on location, it matches a stop event to a known stop (or set of candidate stops, if several stops are close to the logged location). It then matches the next stop, and looks for patterns that include both those stops. The process continues until it has narrowed the patterns down to a single route / branch. Then it looks for a trip time in the schedule on that route / branch that matches the logged time.

If, once the correlator has determined the route / branch, the correlator finds that subsequent stops do not match the selected route / branch, or if there is no scheduled trip on the selected route / branch around the logged time of the trip, or if the correlated portion of the trip is insufficient for analysis purposes, the trip is discarded by the reporting software as unmatched. Currently, about 50 percent of the logged trips go unmatched. NJT hopes to substantially reduce this figure in future generations of the matching algorithms. One anticipated enhancement is to allow the correlator to back up and try a different route if it finds subsequent stops not matching the selected route. Another is to match entire vehicle blocks at once rather than matching trip by trip. That should be effective for the roughly 80 percent of vehicles that follow their assigned block. For those who don't, trip-level matching can be performed as is done today.

INTEGRITY AND CLEANSING OF AVL / APC DATA

The project manager emphasizes that data integrity is much more important with archived data than with real-time AVL. The last line of defense is the strict screening process done by the reporting software, in which up to 50 percent of the data currently collected is filtered out. (As stated earlier, they aim to reduce this quantity using better matching algorithms.)

Hardware problems are easy to detect. Every night, four email messages are automatically generated to report on the data upload and correlation processes. If the upload didn't take the usual amount of time, or didn't contain data from the usual number of vehicles, it's obvious that there is a problem. The system automatically generates an email message to the project manager, vendor, and maintenance contractor if an equipped bus doesn't upload data after more than three days. Problems with the on-board computer – for example, registers that occasionally reset themselves – are readily detected in the data stream. If the passenger counting sensors go bad, it shows up immediately as an imbalance between ons and offs. Current screening requires ons and offs to be within 5 passengers or 10 percent of total ons per trip, and within 5 percent over the day. (Typical variance over the day is about 1 percent.) The allowed trip-level variance is route-specific, and can be increased for routes in which passengers carry over from one trip to the next.

NJT found that routes sometimes had a negative load near the start of the day. This is explained by a bus being loaded with employees at the garage before being turned on (the passenger counter turns on when the bus does), whose alighting after the system is turned on causes an apparent negative load. The reporting software will reset such initial negative loads to zero rather than discard the data.

Another standard test is for agreement between GPS displacement and odometer distance data, and between GPS and clock data. Another is to count the number of periodic records (records triggered by a user-set period, normally 2 minutes); it should correspond to the amount of time the system was on.

ORGANIZATIONAL ISSUES

Working With an AVL Vendor

NJT's approach has been to rely on its primary vendor to develop most of the components of its data system. This was the vendor's first large project dealing with APC's and archived data. Furthermore, within either organization, few people, if any, understood the promise of automated data and the demands that automated processing of data would put on other corporate data bases, including stop lists, route definitions, and schedules. Naturally, that led to some growing pains. Table 2 provides an extensive list of lessons learned in the words of the project manager.

In-House Capabilities and Support

NJT's project has advanced in spite of limited organizational support. The project plan developed by the project manager calls for part-time input from many staff members plus the following full-time positions: project manager, project engineer, database administrator, system engineer(s), technician(s), and data analyst. In the pilot implementation, there has been less than one dedicated full-time position: one person who has served as project manager, analyst, technician and database administrator, while also bearing responsibility for other projects.

NJT's GIS group plays a vital support role in supporting the stop and route database. Every pick, they spend three weeks entering stop and route changes into the GIS system. It is standard procedure to field-locate stops with GPS using vendors located in different parts of the state. The GIS manager estimates that for all the time they spend coding real stop and route changes, they spend nearly half as much making corrections as the stop and route database continues to step up from its pre-APC level of (in)accuracy. They are also beginning to develop GIS applications using the archived passenger count and operations data.

NJT's Bus Operations group was instrumental in supporting the physical installation of the system. They have remained cognizant of system maintenance needs, but have thus far left the actual maintenance effort to an outside contractor. Skeptical at first with respect to bottom line benefits, they have since become some of the system's strongest advocates.

NJT's IS department coordinated the interface with the ATIS contractor, but has otherwise mostly stayed uninvolved, with the vendor providing most IS support to the project. However, as user demand builds, and as staff begin the see the project less as a "device" project and more as a "data systems" project, it is possible that as the project grows, the IS department will take over the data system hardware support, with the vendor providing support for software (though that could also move in-house, leaving the vendor responsible only for software upgrades). It is also possible that the entire operation could be web-enabled and outsourced to the vendor. The optimal mechanism for ongoing support has yet to be determined.

Selling the System

One challenge with any technology project is selling it internally. The system provides strategic information that promises to revolutionize the way transit plans and manages its services, but the likely impact depends on many factors, both internal and external, and is difficult to quantify. Upper management is looking for savings the system will generate. To a degree, the system was therefore "sold" to upper management in part for its ability to reduce data collection costs for passenger-miles estimation, a National Transit Database (NTD) requirement. They estimate that 50% of the time of their 17-person checker staff is devoted to NTD sampling. However, eliminating the need for manual sampling on buses would require system implementation at all 27 garages. Pilot implementation in a single bus garage simply cannot impact NTD sampling requirements much. Part of the reason light rail was chosen for the next stage of implementation is that NTD manual sampling needs for that segment of the business can be eliminated with less investment in light rail than in bus.

Planners who have begun to use archived data recognize that its benefits go far beyond NTD passenger-miles sampling. The intensive passenger count data available on Line 62, covering weekend as well as weekday (a rare luxury in the world of manual counts) allowed planners to realize that there was overcrowding during parts of the weekend, and nearly no use in other periods. They shifted service from the underutilized periods to those in need of more. To the planners involved, the archived data had proven very valuable – it had allowed them to alleviate crowding at no net cost. However, because these changes were done in a fixed budget environment (the environment that operations planning usually finds itself in), it was hard to show management a clear, direct savings since the net cost of operations remained unchanged.

NJT's Bus Operations group was skeptical at first with respect to bottom line benefits, but now, having seen actual reports and having been given the opportunity to consider and shape the future capabilities of the onboard equipment, they have since become some of the system's strongest advocates.

TABLE 1: Captured Events

Every event record includes location and time stamp

- system initialization
- engine start
- fare register log on*
- acceleration above "crawl" speed *(user sets crawl speed threshold)*
- deceleration below "crawl" speed
- deceleration below "zero" speed (user sets "zero" speed threshold)
- acceleration above "zero" speed
- start of turn (user sets turn angle threshold)
- end of turn
- door open
- door close (includes passenger on and off counts)
- fare register transaction*
- fare register zone change/end of trip*
- periodic temperature/electrical record (forces a record every x seconds, where x is user-set)
- data compression
- data download
- entry into diagnostic mode
- exit from diagnostic mode
- engine shut off
- system shutdown

Table 2: Lessons Learned

- Timely communication with stakeholders is more important than progress toward objectives.
- Stakeholders must be given the opportunity to make original contributions during project development – start with a blank page, not a strawman specification.
- Provisions for adequate technical and administrative staff support should be addressed prior to RFP release.
- Project Managers should be encouraged to immediately suspend projects that appear to be going out of control. Project sponsors need to understand that this may be a necessary part of getting the job done in a way that works.
- Requests for Proposals (RFP's) should specify and prioritize measurable results, how results and performance will be measured, the specifics of the environment in which the system is to operate, and any desired project controls. They should not specify deliverables that imply a specific design. Only a few pages are needed.
- Proposals should specify how expected results will be achieved, what they will cost, how they will be delivered, how the delivery process will be controlled, and the type, quantity and timeliness of project support that will be expected from the customer.
- Following selection, suppliers should work with customers to identify and prioritize testable requirements, and how they will be tested, in preparation for issuance of a revised proposal and confirmed specification.
- Customers should review the supplier's support requirements and confirm the identity and availability of appropriate project support personnel before finalizing the contract.
- Conformed specifications for products, services, service delivery, results to be achieved, and required support – with revised estimates of budget and schedule – should be developed and validated before notice to proceed.
- Customers typically don't know how to express requirements in a meaningful way. It is the supplier's responsibility to ensure a thorough understanding of the customer's real requirements and the environment in which these requirements will have to be met.
- Suppliers may not know how to find out how much the customer hasn't told them. It is the customer's responsibility to identify as many special cases and specific uses as possible, to thoroughly test the supplier's understanding of requirements and the environment prior to issuing a contract for system development.
- Consider hiring a specialist to assist in defining testable requirements during development of the contract specifications.
- Have the developer draft the requirements document; this provides a way to assure understanding, and helps ensure the developers will know how to implement.
- Have the users validate requirements by identifying how they will be tested; knowing how they'll be tested removes ambiguity.
- Have the developer break requirements down into manageable pieces that can be implemented as distinct releases; each release will require its own documentation, hand-off review and acceptance test.
- An early deliverable should be a Top-Level Architecture addressing all functionality of all planned releases; top-level documentation should include confirmed requirements, required outputs, data model, rules for data integrity, and an acceptance test plan tied to requirements.
- Change control and traceability of requirements, through all stages of development, testing and acceptance, is an absolute necessity; a project without change control and traceability is by definition "out of control."
- A responsible supplier will seek to delay Notice to Proceed until a confirmed specification has been validated by the customer.

- Risk management planning should address measures to minimize the impact of losses of critical project personnel in each phase of systems development; these may include:
    - formally validating requirements;
    - delivering complete documentation with each release;
    - frequent team meetings;
    - comprehensive records of test data, defects and system change requests.
- The risk management plan should include contingencies for customers' inability to make timely delivery of required project support.
- The vendor should perform a system test against all relevant requirements prior to delivery of each release; the customer should review test documents and duplicate tests as necessary to confirm results for formal acceptance following delivery.
- A formal acceptance test plan should be prepared for each release. It should address normal cases, boundary cases, exception cases, and known errors. It should include:
    - test cases, with pre-requisites and set up;
    - inputs (at raw data level);
    - test steps;
    - expected results;
    - expected duration to run;
    - traceability (every test related to a confirmed requirement; every requirement thoroughly tested).
- Contracts should specify procedures for formal requirements validation, design reviews, requirements traceability, change control and system verification.
- Contracts should specify break points at the completion of requirements definition, top level design and detailed design, with formal review and acceptance of respective interim documentation as a prerequisite to continued development.
- Contracts should include clauses automatically stopping work if customer feedback from requirements validation or design review is late; the supplier must not be pushed to develop products in the absence of validated requirements.
- Contracts should specify that suppliers will be fully compensated for delays or rework resulting from customers' failure to provide the agreed upon level of project support.
- CPFF Contracts should specify a schedule for fixed fee and retainage to be invoiced only at completion of associated milestones, and should provide for discounting of these amounts in proportion to any specified functionality not delivered.
- CPFF Contracts should specify requirements for cost tracking, budget projections and performance reporting that clearly distinguish original work from correction of defective products previously developed.
- Contracts should specify reductions of fee (profit) for failure to adhere to procedures and schedules per the agreed service delivery specifications.
- Contracts should specify that suppliers are solely responsible for the impacts of the loss of project personnel due to management decisions (i.e. anything other than illness, death, war, childbirth, spousal relocation, or natural disaster); customers should be compensated for costs and delays for which their suppliers are responsible.
- IN SUMMARY:
    - DO include everyone from the very start;
    - DO make sure the vendor understands your business model;

- DO build project controls directly into the contract;
- DO perform formal design reviews and verification tests;
- DO establish a formal software change control process;
- DO keep everyone continuously informed;
- DO slow the project down of you think you need to; and
- DON'T succumb to pressures to focus on development at the expense of communications with stakeholders

(source: Jim Kemp, New Jersey Transit)

APPENDIX C: AVL- and APC-Related Data Systems at King County Metro*

HISTORY AND MAIN FEATURES

Metro of King County, the largest transit agency in the Seattle area, was the first large scale user of APC's in the U.S., beginning in 1982. About 180 buses, 15% of Metro's bus fleet, are equipped with passenger counters. They use signpost / odometer to track vehicle location. Development of APC data processing software has all been in-house over a period that has seen migration of the primary processing and reporting from mainframe to desktop computers.

Metro also acquired an AVL-CAD system that began operation in 1992. It also uses signpost / odometer, but with the new system came a whole new set of signposts. When the vendor defaulted in 1993, Metro took possession of the software code and has since developed all the AVL software for both real-time and historical analysis.

The AVL and APC systems have some degree of on-board integration. Metro developed an on-board interface that allows the APC on-board computer to get information from the AVL on-board computer including operator sign-on, the AVL's signpost information (obviating the need to maintain the old APC system signposts), odometer counter, and the AVL clock. However, AVL and APC data streams and databases are completely separate.

Since 2000, Metro has had a useful database of AVL timepoint data. A desktop application produces tables of schedule deviation data that can be exported to a spreadsheet for further analysis. Running time analysis is done using software supplied by Metro's scheduling system running from a database loaded with extracts from the primary AVL database. Periodic and special reports from the timepoint database are developed by request by a data analyst using Structured Query Language.

Metro also has a desktop application for its APC data. They maintain a desktop APC database, created from their more detailed APC data archive using the Focus query language. For trips, the desktop database stores both individual and summary records (mean, maximum, minimum); for timepoint segments and bus stops, it stores only summary records. Several reports are available using a simple interface.

Metro has developed and implemented an advanced Computer Aided Dispatch system. Considerable staff time and effort has been spent on monitoring and improving the reliability of the AVL location data upon which it depends. Archived AVL and APC data benefit in two ways from their connection to the real-time CAD system. First, they benefit from the accumulated efforts of the past to improve location data reliability. Second, the fact that operator sign-in data is monitored by dispatchers ("coordinators") makes the incidence of bad sign-in data much less frequent.

Metro's AVL system has a special feature for determining time at timepoint without having to interpolate between polls. As part of the regular polling routine, the control center predicts the bus odometer reading at the next timepoint and sends that prediction to each bus. When the bus reaches that point, it records the time, and sends that time back as part of the next poll message. While the APC system does not share that timepoint-finding mechanism, it records the time at which buses arrive and depart from every stop, and many timepoints are, for practical purposes, stops.

Schedule analysts make use of both AVL and APC operating data. Since 2001, when AVL accuracy (as measured by spot checks) improved to about the same level as APC, AVL has been the preferred data source because it is a much larger data source, with data capture on 80% or more of all operated trips, while APC records are captured for only about 8% of operated trips. AVL's larger data capture also means that its dataset is more likely to contain recent data on routes of interest. In addition, AVL data is transferred nightly to the historic archive, while the historic APC archive is updated about once a month.

---

* This case study was prepared by Peter G. Furth of Northeastern University. Agency staff interviewed were Tom Friedman, Project/Program Manager; Darrel Riley, Information Systems Analyst; Brad Kittredge, AVL Data Analyst; Walt Miller, Radio Maintenance Lead; Steven Masumoto, Senior Schedule Planner; Tony Longo, APC Data Coordinator; Dave Rynerson, Senior Research Associate, Transit Scheduling & Service Development; and Mike Berman, Geographic Information System Program Manager.

.

Metro has a strong emphasis on sharing data across the agency and with the public. It has a well developed program of sharing its schedule data, AVL data, and APC data across the agency so that users in every department can benefit.

This case study first describes how Metro's AVL and APC systems are configured to capture data, with associated issues related to data accuracy. Next, it describes the various databases using AVL, APC, and related data and applications using those databases. It concludes with a section on trip matching issues.

AVL AND APC CONFIGURATION AND DATA CAPTURE ISSUES

AVL Configuration and Data Captured

On-board equipment supporting the AVL system includes a dedicated on-board computer connected to the radio, control head (operator interface with the radio), signpost receiver, and odometer. Operators sign in with their ID and their run ID. The location system uses signpost signals and odometer readings to track vehicles, assuming that they are following their assigned route. Signposts signals are used to anchor location at key points, and odometer readings provide distance traveled along the route from those key points. Location matching versus the route is done at a central computer in the control center. Off-route buses are identified, but not tracked.

To capture time at which buses pass a timepoint without having to interpolate between polls, Metro KC has developed an innovative approach they call "virtual timepoints." About 3 minutes before a bus reaches a timepoint, the central computer calculates what a bus's odometer reading will be when it passes that timepoint, and passes that value to the bus as part of the routine interrogation. The bus stores that value temporarily, and when its odometer reading matches the target, the bus computer records the time.

The AVL radio system uses two data channels that are each capable of polling 6 buses per second at 2400 baud. The practical polling rate for the two channels combined, however, is only about 10 to 12 buses per second. In the peak, with about 1100 buses operating, the polling interval is therefore about 90 to 100 s; midday, with about 600 buses operating, the polling interval is about 60 s. Faster polling (up to 2 s) can be set for individual vehicles, either automatically due to an emergency alarm, or at the service coordinator's discretion.

AVL messages include the routine poll, request to talk, and alarms. The contents of a routine poll are shown in Table 1.

TABLE 1: AVL Routine Poll Contents
odometer reading
odometer reading at last signpost and signpost ID
time at last virtual timepoint and timepoint ID
vehicle identifiers

The AVL-CAD system's features include display of schedule deviation for each bus and average deviation for all the buses serving a route at the time of inquiry. Real-time location of Metro's buses is available to the public on the web using an application developed under the federal ITS Model Deployment Initiative's Smart Trek project.

AVL Missing Data and Data Accuracy Issues

Real-time tracking works for the 80 to 90 percent of the bus trips that follow the run entered by the operator at sign-in. The 90 percent capture rate is regarded as close to the practical limit. The other 10 percent are readily identified because they pass signposts they aren't "supposed" to pass, or because they fail to pass signposts they should pass, according to their entered run number. Of this 10 percent, failure to sign in accounts for about 3 percent. This is a common problem with trippers (buses doing a short assignment). Because sign-in failure is treated as a radio failure and is therefore cause for a bus to be recalled to the garage, few regular run operators fail to sign in. The other 7 percent of bus polls classified as off-route are due to incorrect sign-in information, reassignment by a supervisor, or to a bus actually being off-route.

The central computer checks for odometer accuracy by comparing the difference in readings between signposts with the known distance between signposts. The discrepancy, called "confidence factor" and measured in feet, is recorded in the AVL data stream. However, it is not used to calibrate odometer readings when predicting odometer reading at the upcoming timepoint. Malfunctioning odometers are readily identified, triggering a repair order. They estimate that only about 6 of Metro's 1400 coaches have malfunctioning odometers at any given time.

Metro's location accuracy goal is ±200 feet 95% of the time; it is just now being met. There are less than 300 signposts, in contrast with Metro's 3000 timepoints. The signposts provide good coverage downtown, but in the outer areas accuracy can suffer as timepoints are sometimes several miles from the nearest signpost. Besides odometer inaccuracy, signpost signals are often distorted by environmental factors, especially utility and trolley wires, making it hard to correctly locate the signpost based on its field. The AVL system notes the location of maximum field strength; the APC system, by contrast, notes odometer readings when the field is first picked up and when it is lost, and averages the two. In an imminent upgrade, the APC system will convert to using point of maximum field strength as does AVL.

APC Configuration and Data Captured

Several years ago, Metro staff developed an APC-radio interface, which passes on to the APC system the following items used by AVL: operator sign-in, signpost ID, odometer, and time / date. Otherwise, the APC and AVL systems function independently. APC on-board records are manually uploaded each week.

Primary record types and contents are listed in Table 2 (there are also secondary record types used as diagnostics). The signpost record makes it unnecessary for other records to note odometer reading at the previous signpost; signpost ID is sufficient. The mile record was added to help follow express buses on freeways. The idle record helps with determining when a layover ended. The hour record is a kind of system "heartbeat" that confirms whether the system was on and when it crashed.

TABLE 2: APC Record Types and Contents

Record type	Trigger	Contents in Addition to Standard Items
[standard items in all records]		*odometer reading* *ID of last signpost* *time* *identifiers*
Stop record	door close	door open and close time on and off count
Signpost record	passed signpost	odometer reading at first and last detection of signpost signal
Idle record	stopped 2 minutes with doors closed	door open and close time on and off count
Mile record	traveled 1 mi without making a record	(standard items only)
Hour record	every hour	(standard items only)

APC Missing Data and Data Accuracy Issues

About 15 percent of Metro's fleet is equipped with APC's. Equipped buses are circulated over the entire schedule each sign-up. An audit of the fall 1998 sign-up, lasting four months, found valid data for 97% of scheduled weekday trips, with 83% having at least 3 observations. On the weekend, 85% of scheduled trips had valid data.

With 15 percent of the fleet equipped, one would expect in a perfect world to observe each weekday scheduled trip 15 percent of the days, or 12 days, in a 4-month sign-up period. With about 6 valid observations per weekday scheduled trip, Metro's recovery rate relative to fleet penetration is about 50 percent. Causes for lost observations include logistical issues (getting the equipped buses out), errors in passenger counting, and errors in trip matching.

A full-time APC Data Coordinator is responsible for ensuring the integrity of the APC data. He checks the data every week, and occasionally rides buses to check data accuracy and investigate problems. In the filter now used, a vehicle's APC data is rejected if its off total is no more than 7% below or 15% above its on total for the entire day. The asymmetric filter is a concession to systematic counting bias. There is currently no test for on-off balance at the trip level, though Metro staff would like to see it added. Data are also rejected if the difference in odometer reading between signposts varies too much from the actual distance between signposts. Data are also rejected if the trip cannot be matched to a route / trip in the schedule. Matching error can be due to inconsistency between the operator sign-in and what was actually operated, as discussed earlier. There are also particular matching issues discussed in a later section of this case study.

Given that running time and on-time performance can be evaluated using either AVL or APC data, the fact that APC observes only a small sample of operating trips gives AVL a strong advantage. In addition, differences in how the two systems record location, in spite of their shared use of signposts, odometers, and sign-in information, affect their accuracy and usefulness. AVL makes a record at each timepoint, while APC makes a record at each stop and signpost. Time at locations where a record is not made must be estimated by interpolation.

In Metro's business model, schedule adherence and running time are based on time at timepoints. In contrast, no measure of significance is based on time at stops or signposts. Therefore, for analyzing operations data, AVL data is preferred to APC data because it measures timepoint arrivals directly and is a much larger sample.

Advantages of Integrating AVL with APC

In addition to the obvious cost and maintenance advantages of having the APC share AVL devices, APC data quality benefits from the on-line check provided by the AVL system. Through the real-time components of the AVL system, sign-in information is to some extent verified, so that APC matching benefits from having correct run numbers most of the time. Their common clock makes the data sources compatible.

Control Center Report Log

Service coordinators in the control center who communicate by radio with bus operators make logs of significant events and dispatch actions. There are codes for standard events and actions. Example events, which operators are required to report by radio, include "very late" (either 15 minutes late, or later than the upcoming recovery time plus 5 minutes), pass-by, and wheelchair pass-by. Example actions are trip cancelled and trip short-turned.

It is generally accepted that coordinators reliably record control actions and significant events as they are reported. It is hard to validate, of course, how complete is the record of events because it depends on operators calling them in. However, operators tend to be conscientious about calling in for extreme lateness or overloads, because if they don't and a complaint is lodged, they may face discipline for failing to call in the event. Event records provide a degree of redundancy with AVL and APC data, and are useful as a check on the AVL and APC data. Also, because AVL and especially APC data are not captured on every vehicle, the control center log is a valuable source for incident and complaint investigations and other research.

On-Board AVL Log – An Unused Data Source

As a diagnostic tool, the AVL vendor provided that the AVL/radio on-board computer would log events including reaching a timepoint, passing a signpost, and operator issuing a request to talk. These event records are only written in the on-board computer, and must be uploaded manually.

Metro has considered this data source, which is still used occasionally to investigate maintenance issues (e.g., is a signpost being picked up accurately)? However, with the data processing enhancements that have made poll data more reliable and the lack of similar development for event log data, the AVL on-board event log is not considered a necessary source of bus operations data.

DATABASES AND APPLICATIONS

Multiple Data Sources

While Metro's AVL and APC systems share some on-board and wayside components, their data streams and databases are independent. Considering other sources of ridership and operational data – supervisors, operators, and farebox – Metro analysts often have several data sources for a given measure of interest. Table 3 is a guide prepared by one of Metro's planners comparing and contrasting attributes of the various data sources.

TABLE 3: Attributes of Data Sources for Ridership and Operational Performance Analysis

Attributes	APC	AVL	Supervisor	Driver/Conductor	Farebox
Mode:					
Regular Service	x	x	x	x	x
Custom Bus	x	x		x	x
School Service	x	x		x	x
DART				x	
Streetcar				x	
Special Event			x	x	x
Level of Detail:					
Bus Zone	x				
Time Point	x	x	x		
TPI	x	x			
Trip	x	x		Varies	
Block	x	x		x	x
Route	x	x		x	
Ridership:					
Ons	x			Varies	Partial inbound
Offs	x				Partial outbound
Load	x		x	Varies	
Seats	x				
Psgr.Miles	x				
Lift Use					x
Type of Fare					x
Schedule Performance:					
Arr./Lv. Time	x	x	x		
Running Time	x	x			
Data Access:					
Desktop	x	x	Limited	Limited	
Printouts	x	x	x	x	x
Download	x	Limited	x	x	x
Pitfalls:					
Early bias		x			
Time Point Estimation	x	x			
First/last Time Point	x	x			
Under-counting	x				x
Sample:					
Average Size	5-6%	75%	Low; variable	Low; variable	80-90%
>1 Obs.M-F,SAT,SUN	94,74,82%				
Lag time	2-3 months	one day	one week	one month	one month
History	one decade	2 years	two decades	Varies	eight years
Sample Irregularities:					
Re-routed service	x	x			
Weekend sample	Low				
Problem Trips	x	x			
Problem Routes	x				
Missing Data Types:					
Paper Transfers					x
Flash Passes					x
Passes not swiped					x
Ride Free Area					x

DART = Dallas Area Rapid Transit
TPI = Time Point Interval

(source: Dave Rynerson, Metro of King County)

When offered multiple data sources, analysts tend to gravitate toward the one they find the most reliable. For operational measures such as on-time performance and running time, the preferred source for several years was APC. However, once AVL accuracy became as good as APC (in 2001) and readily available, AVL has become the first choice because of its far greater sample size and because it's more recent. APC is consulted less often, but is still used regularly, e.g. if the AVL data doesn't give a consistent story, or is inconsistent with supervisors' observations.

The focus on a single preferred data source is no accident, as Metro has made it an institutional priority to make AVL a reliable source of operational performance data and APC a reliable source of ridership data. The same attention has not, for example, been given to measuring operational performance using APC, or ridership using the farebox. (However, attention is now being focused on the farebox due to a new revenue sharing scheme involving five transit agencies around Puget Sound.)

Making Schedule Data Accessible

Metro's AVL and APC data analysis applications rely on schedule data. One application, the running time analysis program ATP, is an add-on to Metro's scheduling package, HASTUS, and therefore interacts directly with the HASTUS database. For the other applications, as well as for other uses such as the control center's real-time functions, Metro has developed a program that extracts HASTUS schedule data every two weeks and merges it with GIS data to create an Oracle database, called Distribution Database, that is distributed across the agency using an intranet.

AVL Databases and Applications

In the control center, the Data Acquisition and Control System, developed in-house to control radio communications for both real-time applications and data archiving, directs a record to be written of event messages, including passing a new signpost, or passing a timepoint, and operator-initiated events such as sign-on. Routine polls are distributed to real-time applications such as the control center and web-based public information displays BusView and MyBus, but are not recorded to an archive. Nightly, an extract of the day's records is filtered and rewritten to a historic database in Informix. The historic database stores AVL data in one table and operator initiated events (e.g., overload, Request To Talk) in another. Only records of timepoint passage (e.g., bus x passed timepoint y at time t) are saved in the AVL table. AVL records do not include schedule data; analyses versus the schedule (e.g., schedule adherence) are done on the fly using schedule data maintained in separate tables in the AVL database.

Desktop Reports

The AVL Data Access Project (ADAP) succeeded in 2000 in making reports using AVL data available with a user-friendly report request format. The report request interface is written in Microsoft Access; it passes parameters to the Informix database, which processes the queries and returns results in the form of Excel tables and graphs. It offers two reports, for which the user selects the range of dates and hours:

- *Timepoint Arrival.* For a selected timepoint, it creates a table showing deviation from scheduled arrival time for every trip passing the timepoint, with columns repeating for each selected day. If multiple routes use the timepoint, results can be sorted by time within route, or simply by time to get a picture of multi-route activity at the timepoint. The only summary statistic is average deviation for each scheduled trip. The table is exported to Excel for further analysis.
- *Block Summary.* For a selected block, it creates a table of schedule deviations at each timepoint in the block. There is a column for each selected day, plus columns identifying the timepoints and their scheduled time. The only summary statistic given is average deviation for each timepoint. The table is exported into Excel to allow further analysis.

These reports are used by scheduler planners, service quality supervisors, operations base chiefs, the customer assistance staff, and by departments that investigate incidents. The wide acceptance and usage of AVL data speaks to the success of Metro's efforts related to historic AVL data.

The application also includes a running time analysis that has fallen out of use since the ATP software (described later) has become available for this function. The application also includes management reports on radio system performance.

Custom Reports Using SQL

Involved reports take too long to generate to make them suitable for desktop access using Microsoft Access. Metro's staff includes a data analyst proficient in Structured Query Language programming who develops reports as requested by managers, schedule planners, and the supervisory board. There are two regularly used reports:

- *Schedule Adherence Problem Locations.* This report identifies the timepoint / period of day combinations with the worst on-time performance. Field supervisors in the Service Quality Department use this information to help them focus their supervision the next month.
- *On-Time Percentage by Trip.* This report, produced once for each sign-up, is a huge table showing percentage of arrivals early / on-time / late for every scheduled trip at every timepoint for all the weekdays in the 4-month period. It includes no summary statistics, except a single fleetwide percentage early / on-time / late. The table is imported into a spreadsheet for further analysis.

Another report in development will report schedule adherence by operator. One weakness of this analysis is that it relies on operator sign-in data to know operator ID. The current timekeeping / payroll system does not permit a view of operator ID – vehicle ID correspondence for each day, so there is no way of verifying operator ID, and the extent of misentered operator ID's is unknown. Data analysts look forward to an upgrade to the payroll system that will allow verification and correction of operator ID using historical data, or better yet, real-time coordinated entry or verification.

Running Time Analysis Application ATP

The Schedule Planning Department also maintains a database of AVL timepoint data, configured to interface with the running time analysis application ATP, an add-on that came with their scheduling software package, HASTUS. Every two weeks, a program extracts and formats the most recent batch of records from the historic AVL database and transmits them to Schedule Planning, where they are loaded into the ATP database. The ATP database contains records for timepoint segments (timepoint intervals), with the following fields: timepoint ID's, time at first timepoint, elapsed time to reach second timepoint, route and run ID, and a code indicating the data source (AVL).

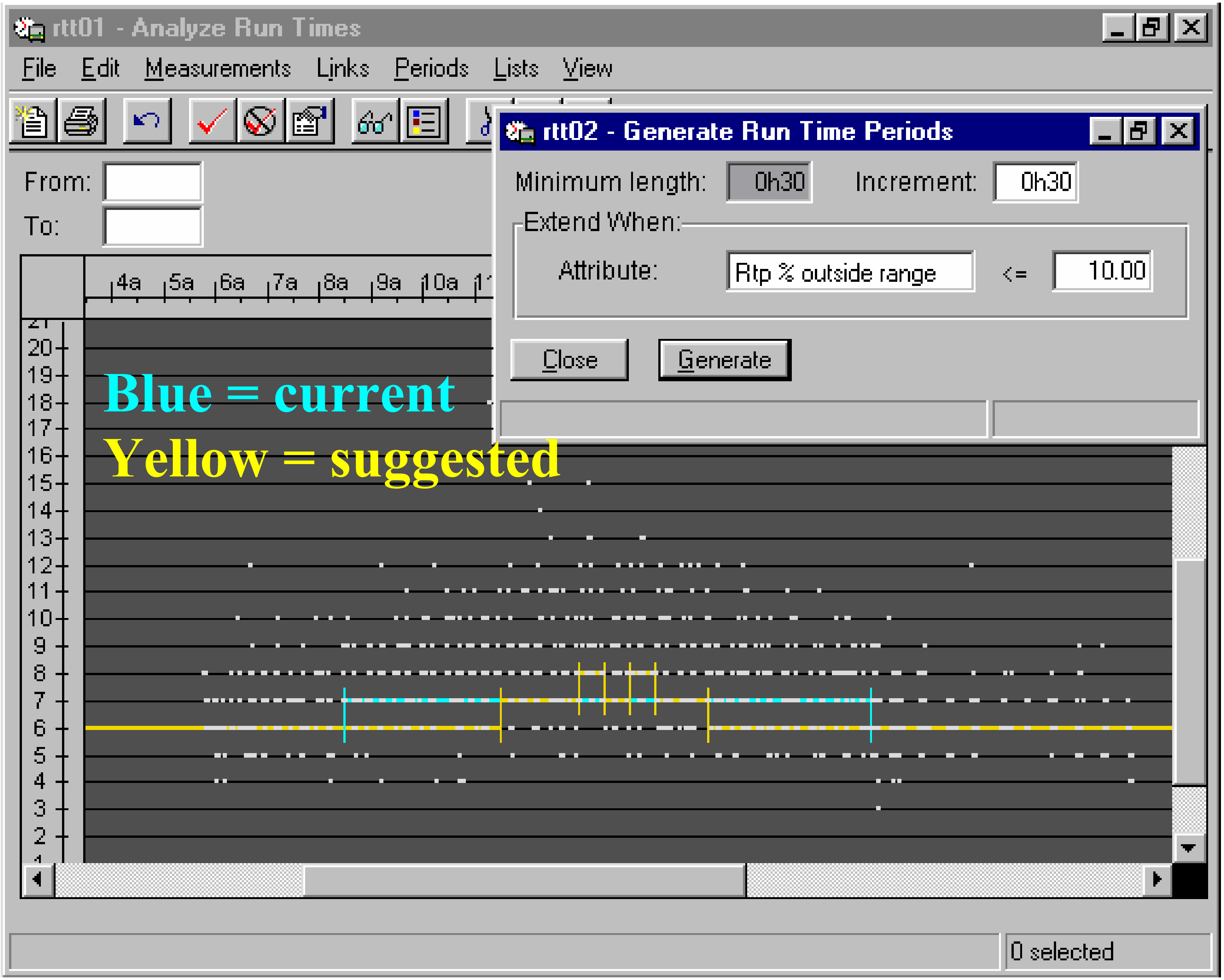
ATP's analyses and reports include the following, illustrated in Figure 1:

- *Distribution of segment running times.* The report produces a scatterplot of segment running times versus clock time. With it, one can see both random variation and systematic variation over the day. Summary measures, by user-selected time increment (e.g., 60 minutes), include mean and standard deviation of running time and percentage of observations outside a user-specified range. The report also produces the data in tabular form, allowing further analysis using a spreadsheet. Normally, the report is run for a single timepoint segment; however, users can specify a sequence of two or more timepoint segments to analyze running times on observed trips that covered that sequence of segments. The report is not set up to readily examine an entire route; to do that, the user has to enter the complete sequence of timepoint segments that make up the route.
- *Suggest new running time periods and running times for a timepoint segment.* The user specifies a tolerance range, e.g., [-1 minute to +3 minutes], an outside-tolerance percentage allowed, e.g., 20%, and a time increment, e.g., 60 minutes. The program then divides the day into periods of one or more time increments and suggests for each a running time for which mean observed running times in that period lie within the specified range with the specified level of exceptions. The greater the range and the exception percentage allowed, the longer the schedule periods will be.

ATP analysis does not involve matching to scheduled trips; it only compares observed running times between timepoints within a user-specified time window against scheduled running time, taken directly from the HASTUS database. However, ATP does not offer schedule adherence analysis, or the ability to analyze observations of a particular scheduled trip (e.g., the weekday 7:12 trip).

ATP analyzes timepoint segments, not routes. When a timepoint segment or series of segments is served by more than one route, there is no easy way, short of listing blocks to be included, of separating observations of one route from another. This structure is consistent with the scheduling software's logic and Metro's practice of making the timepoint segment the primary unit of running time, with route running time being simply the sum of its constituent segments. This approach, which contrasts with first setting running time for a route and then allocating time across segments, has the advantage of consistency when multiple routes share a segment. It can also be an impediment, such as when it desirable to allow one route's running time between a pair of timepoints to differ from other routes, e.g., because the subject route follows a different routing. To accomplish this, Metro uses in the schedule database multiple timepoints to represent essentially the same point.

Figure 1: ATP Analysis Showing Observed, Scheduled, and Suggested Segment Running Time by Time of Day



The ATP database has been active for over two years. They have decided to keep data only for the most recent four sign-ups (16 months), amounting to about 22 million records. Metro sees no need to analyze older running times (though they could, using APC data, analyze running times for the last 15 years).

APC Databases and Applications

Primary APC Database: APC Access

The APC data stream is processed weekly and summarized both monthly and at the end of each sign-up. The raw archive is stored as an Oracle database, which can be queried using the Focus query language. However, most APC data users use a Microsoft Access database called APC Access containing summary records created using Focus from the raw APC archive. The APC Access database contains records at three levels: bus stop (internally called a “zone”), segment (“timepoint interval”), and trip. Only at the segment and trip levels are records kept for each observation. Aggregate records, showing minimum, maximum, and average values for a given sign-up, are kept at all three levels of detail. The number of records for the spring 1999 sign-up, is shown in Table 4. As mentioned earlier, in a typical sign-up, 97% of weekday trips are recovered at least once, and 83% are recovered at least 3 times. In analysis, a minimum of three observations is preferred to minimize the influence of outliers.

The database package used has the advantage of a user-friendly desktop interface, but it becomes slow with a large numbers of records. That is part of the reason that individual bus stop observations are not kept in the database. The database is available agency-wide using an intranet.

APC Access includes schedule adherence and ridership reports. Reports are all in tabular form. Schedule adherence reports include:

- *Timepoint Running Times.* This report compares a bus’s observed running times, averaged over a user-specified date range, by timepoint segment against scheduled running time. It examines running time both including and excluding dwell time at the stop at the first timepoint, if there is a stop there.
- *Schedule Adherence by Block / Timepoint / Route.* These two reports give average, earliest, and latest deviation from schedule for each scheduled trip of either a block or a route at all the key timepoints on that block or route, or at a selected timepoint for all blocks and routes serving that timepoint. Both arrival and departure times are compared with schedule, if there is a stop at the timepoint. It also compares scheduled and actual running time on the timepoint segment ending at the subject timepoint.

TABLE 4: Number of APC Records by Type, Spring Sign-up, 1999

	Individual Observations	Aggregated (min, max, average)	Notes
Bus stop	4 to 5 million	87,134	Individual observations are not saved beyond the sign-up. Aggregated records, by route and time period, are saved for weekday only
Timepoint interval (segment)	506,00	157,338	all days; included deadheads
Scheduled trip	75,433	23,885	all week; excludes deadheads; excludes passengers traveling only in downtown “Ride Free Area”

- *Deadhead Summary.* This reports compares actual to scheduled running time and arrival time for deadhead routings.

Ridership reports include:

- *Bus Stop Activity Summary.* For each selected bus stop (one can select individual stops or a series of stops on a route), it shows average ons, offs, approaching load, and seating capacity for all the trips observed at that stop by period of the day (AM, midday, PM, evening, night). If the stop is served by multiple routes, results are shown for each route and for all routes combined. Also included are number of scheduled trips observed, number of observations, and number of times the bus actually stopped. The report includes no information on variation between trips or on correlation with headway. In one format of this report stops are listed in order along a route; however, that format requires so much staff time to sort the groups of records for the variants of each route into a useful order that it is available only for the Fall sign-up each year. Another variation of this report covers park-ride lots, bus tunnel stations, and transit centers.

- *Timepoint Screenline.* This report combines ridership and operational data, showing load and schedule adherence for individual trips by route past one or more timepoints.
- *Timepoint Load Profile.* This report shows, in tabular form, ons, offs, and load by timepoint segment. It aggregates over all observations of a scheduled trip, giving mean, minimum, and maximum of each measure. Because some timepoint segments may have been observed more often than others, the mean on-off-load data are not necessarily consistent.
- *Summarized Time Point Loads.* This report displays total daily ons, offs, and load by timepoint along a route.
- *Trip Ridership.* This report shows mean, minimum, and maximum ons, offs, and load for each scheduled trip by route, aggregating over all the observations in a specified period. It also calculates riders per hour and average load factor.

APC Data in a Geographic Information System

Metro has developed a GIS application called GIS Toolbox that incorporates schedule data with geographic data. It is used to maintain the route and stop database, and includes many useful analyses of schedule data including analysis of terminal congestion for layover points used by multiple routes and measures of geographic accessibility.

The GIS Toolbox also includes some analyses that use APC timepoint and trip summary data, extracted from the APC Access database into the GIST database. Gaps in the APC data are populated using average values. GIST's APC-related applications include:

- preparing load profiles along a route map (the greater the load, the heavier the line along the route)
- displaying ridership data and vehicle volumes for a GIS link by clicking on the link

Other Planning Applications of APC Data

The APC analysis program includes automated procedures to feed ridership summary data to other applications used in the planning and schedule planning departments. One is a Change Management Program (CHAMP) that tracks rationale and budgetary impacts of service changes. Another is the Subarea Allocation Model (SAM), used in mid-range (6-year) planning, that analyzes ridership and cost in three divisions of Metro's service area. A third is the Transit Resource Allocation Model (TRAM), which does route performance evaluation using cost and ridership measures. TRAM maintains a historic Paradox database of route summary data including route ridership by period of the day.

Control Center Log Database

As mentioned earlier, operator-initiated events transmitted over a radio data channel are recorded in a separate table in the Informix database that holds the AVL timepoint data. Those event records are called Coordinator Service Records (CSR's). For now, at least, no logical links have been developed between Coordinator Service Records and AVL data that might show, for example, relationships between non-extreme lateness and operator-reported overloads.

The CSR database, which goes back 12 years, is available across the agency and is used for a variety of research purposes such as investigating complaints and incidents. Service coordinators use it to review and learn from past scenarios that required control actions. Schedulers also use it to research routes under review; however, as the AVL and APC databases, which report routine as well as extreme lateness and load in full detail, have become more reliable, the importance of the CSR database for scheduling has diminished.

MATCHING ISSUES

AVL and APC data streams use different procedures to match the data to route and schedule. For AVL, matching is done in real time by the Data Acquisition and Control System (DACS) software to support the AVL's real-time functions. It relies on the run numbered entered by the operator, and confirms that the bus is following the scheduled run. If a valid run number is not entered, no matching is done. If a valid run number is entered but the bus is not where it is expected to be, DACS continues throughout the day trying to match the bus to the entered run number, taking note of signposts bounded by a early / late time range as possible locations to get the bus back on-route.

APC data matching is done after the fact. A matching algorithm starts with daily dispatch number and is verified by run number entered by the operator. Parsing is based on the sequence of signposts passed, pull-in and pull-out times, and idle times during the day. If a matching or scaling error is found, only data through the last valid trip in the block is admitted. New software, currently being written, will allow the matching to try to get back on track (similar to AVL tracking) so that subsequent trips, if operated correctly, could also be admitted to the database.

Metro has spent considerable effort dealing with the difficult issue of end-of-line matching. While it might seem ideal to have a signpost at the start and end point of each trip, whose passing would be a positive indication of trip start and trip end, variable operator behavior at the end of line makes it impossible on many routes to establish a clear start or end point. Metro has dealt with this issue in several ways.

They locate signposts near, but not too near, the end of the line. Using a signpost right at the start of a the route can give lots of meaningless readings as buses move around during their layover period; better to move the signpost downstream of the start enough so that buses only pass it when they have really begun their trip.

Of course, there's a tradeoff – moving the signpost in from the ends of the route makes signpost signals more meaningful, but it makes it more difficult to determine end-to-end running time. Several analysts who use AVL and APC data recommend caution when interpreting running time or schedule adherence data for the first or last segment of a route.

The AVL and APC matching algorithms look for idle times, even if they don't occur at an expected layover point, and tries to match them to a trip endpoint. The AVL route matching algorithm ignores odometer changes that occur between the end of one trip and two minutes before the scheduled start of the next trip, under the assumption that operators rarely begin more than two minutes early. Odometer changes during the layover period are much more likely to be operators jostling their positions at the terminal than movements along a route. The system actually monitors for a lack of motion within the end of the trip to stamp the end of a trip, and start of motion within a specified time to start the next trip. Motion after the trip end is recorded but not assumed to have route importance. This algorithm enhancement, put in place about two years ago, both simplified and improved Metro's route matching. APC will soon have similar logic.

CONTINUAL IMPROVEMENT IN DATA QUALITY

Use of the AVL/APC data is widespread and is deliberately propagated to a wide set of users inside the agency as well as to the public. Due to the control given by owning the source code for the AVL software, King County Metro has gone from capturing 50% of timepoint crossings 4 years ago to almost 90% recently, with small further increases expected soon due to process improvements. The AVL system was able to develop techniques to validate its own data quality, which has proven invaluable in determining where to concentrate attention on making improvements. Once the issues of tracking were understood by the AVL staff, the basic logic in the AVL has remained constant over three years, with minor enhancements approximately every 6 months.

APC is also seeing data capture increases as analysts understand how to leverage AVL data better to assist APC uses. APC is undergoing a major rewrite to be able to leverage the ideas used in the AVL system to assist APC logic. A recent major enhancements to the AVL system was to rewrite the reporting section just to provide data that replicated APC data format so that the APC system could get location enhancements by doing analysis of AVL data. The overlap has been very beneficial in increasing data quality for both systems. Use of the ATP module along with ADAP has allowed the analysis staff to spend less time doing schedule analysis and more time developing reports for other departments, for instance creating reports for Service Quality.

A final current issue is compliance with state and federal guidelines governing data retention that apply to AVL, APC, and radio data. King County is in the process of making sure that they have implemented the state statutes for data retention. Other transit agencies that hope to implement data storage for their systems should be aware that they may be subject to state or federal guidelines.

APPENDIX D: The Chicago Transit Authority's Experience with Acquiring and Analyzing Automated Passenger and Operations Data*

The CTA has long been interested in AVL and APC technology. Planning efforts to acquire an APC system, first begun in the mid-1980s, were repeatedly renewed but then faltered. In 1993, they issued an RFP for an AVL system whose implementation phase, begun in 1996, is due to end in 2002. However, for various reasons, they still find themselves unable to match AVL data routinely to the schedule, crippling most opportunities for data analysis both on-line and off-line. Through this process, the CTA has learned valuable lessons about planning, management, and procurement procedures for automatic data systems. Those lessons are being applied in a new procurement, contracted in 2002, for APC's as part of an automatic stop announcement system.

Experimental analyses using AVL data have confirmed the technical feasibility of obtaining and analyzing useful archived data on operations. Nevertheless, in spite of a recent enhancement that makes it possible to obtain location data fleetwide with 40 to 70 s polling, there are no plans to archive data other than operator-initiated exceptions.

In the stop announcement system now under contract, passenger counters with on-board data recording were requested on 15 percent of the fleet. However, due to the "smart bus" design, the on-board infrastructure will permit event recording on all buses, whether or not they are equipped with passenger counters, so that while passenger counts will be limited to a 15% sample, operations data will be collected on every bus, every day. They are also considering a policy of working toward having APC's fleetwide, because the "smart bus" design makes the marginal cost of adding passenger counters relatively small.

SYSTEM DESCRIPTIONS

AVL System

The CTA's AVL system, specified in 1993 and under contract since 1996 to Orbital Sciences, includes two levels. The first, implemented virtually fleetwide, is called the Bus Emergency Communications System. The original system provides data only when there is an operator-initiated event. An enhancement, due for delivery in November 2002, provides location data on all buses with 40 to 70 s polling. The second level, called the Bus Service Management System, was designed as a pilot project for one garage, equipping 250 buses that would serve two of CTA's busiest routes. These buses will know their schedules and track themselves against the schedule. This level has been demonstrated, but is not fully operational because it was not possible to keep the 250 instrumented buses in one garage.

With the Bus Emergency Communications System, buses have a mobile data terminal (on-board computer, display, and keypad) connected to the radio, GPS receiver, odometer, and silent alarm. Buses instrumented for the Bus Service Management System also have wireless (infrared) modems for uploading schedule data. The AVL system uses GPS to determine location. While the on-board configuration includes a connection to the odometer for dead-reckoning backup, that facility is not being used.

Data archiving in both the Bus Emergency Communications System and the Bus Service Management System is limited to exceptions. Routine analysis of that data is limited to *ad hoc* needs such as reviewing histories for incident investigations. There have also been some successful experiments with generating, archiving, and analyzing frequent (2 minute or less) location data from the AVL system.

Bus Emergency Communications System

The Bus Emergency Communications System is active on 1600 of the CTA's 1800 buses. It polls buses every 15 s to see whether the operator has initiated an event; example events include sign-in, request to talk, extreme lateness, and

---

* This case study was prepared by Peter G. Furth of Northeastern University. Agency staff interviewed were Kevin O'Malley, Manager of Data Services, Planning Department; Tom Pluger, Network Systems Manager, Control Center; Paul Gross, General Manager, Data Services and Technology Development; Eric Wile, summer intern and MIT graduate student; and Angela Moore, Project Manager, Bus Operations and former summer intern and MIT graduate student. Art Scanlon of Clever Devices and Richard Ruskin of RSM Services also contributed valuable information.

silent alarm. If an event has been actuated, buses return the message with their location status (GPS coordinates, speed, and heading) and identifiers. If no event has been initiated, the return message does not include location data. This system does not include matching data received to the schedule.

Event messages are passed through the radio system's "distribution tier," making them accessible outside the radio system. In the current design, the event messages are both sent to the control center and archived in a database. The archived data is used to review histories as part of incident investigations.

A system enhancement, due for implementation in November 2002, will be for "no event" messages to also include location status variables. This change is expected to increase the polling interval to 40 to 70 s; however, it means receiving positive location data from all buses, instead of just exceptions. Like the event messages, these messages will be accessible through the radio system's distribution tier. In the control center, they will be used to update a location register for each vehicle. However, while archiving the location data from "no event" messages is technically feasible, there is no plan within the CTA to archive it. Whether software to match the location data to the schedule will be developed remains to be seen.

Bus Service Management System

A second part of the AVL system, the Bus Service Management System, involves buses that know their schedules and track themselves against their schedule. The on-board computer automatically creates an "off-schedule" event when the bus passes a timepoint outside a specified tolerance (typically, 1 min early to 5 min late). These "off-schedule" events are processed like other event messages in the Bus Emergency Communication System.

While this kind of exception reporting limits the burden on the radio system and on dispatchers, it makes running time analysis impossible, and hides from schedule adherence analysis any schedule deviations that fall within the specified threshold.

The Bus Service Management System has faced two significant obstacles to implementation, one of them fatal. The fatal obstacle was the idea of reserving a set of buses for two routes. When a large order of low-floor buses arrived, the public relations aspect of how these buses were distributed took priority over data system integrity. Many of those buses were put into the garage in which the BSMS buses were housed, forcing BSMS buses to be distributed to other garages across the fleet. The other garages do not have the equipment needed to transmit schedule information to the BSMS buses; and with only a fraction of the buses on the target route equipped with BSMS, the system's value for controlling the line is lost.

Another significant obstacle has been interfacing with the schedule database. Nevertheless, the difference between the schedule format now provided by the scheduling system and the format required by BSMS is minor enough that a route's schedule file as retrieved from the scheduling system can be easily modifed by control center staff to satisfy the BSMS interface format; that was done for the BSMS system's acceptance testing. Still, the fact that the process of getting the schedule into the desired format is not automated has frustrated many staff memebers.

Experimental Data Gathering and Analysis

Two students from the Massachusetts Institute of Technology, as part of a research and summer internship program sponsored by the CTA, have demonstrated the ability to acquire location data from the AVL system and analyze it off-line. One of them, now a CTA employee, captured the data in on-board floppy disks; the other developed a way to capture location data using the radio polling system.

A feature of the buses instrumented for the Bus Service Management System is that one can insert a floppy disk into the Mobile Data Terminal and have it make a record every 4 s including location status (GPS coordinates, speed, and heading), door status, and identifiers. (This feature was not provided with data collection in mind; it was intended for system diagnosis and as a possible means of downloading schedule data.) Getting data on a route's operation for a single day required first getting BSMS instrumented buses assigned to the runs that serve that route; second, having a researcher insert a disk each bus's Mobile Data Terminal, and then retrieve the disks at the end of the day. This is an all-day process, as buses enter and leave service on a route all through the day. In this manner, data was collected for all the buses operating on two routes for a few days each.

The data was analyzed by several student researchers in an *ad hoc* manner using spreadsheets and custom programs. They examined the distribution of running time by segment and for the route, by time of day. Another analysis combined the event data with transactional boardings data from the farebox system to examine the relationship between boardings and running time – to what extent does a higher load slow down a bus? Unfortunately, fare transaction records do not include a time stamp, so matching the boardings data to particular stops and trips was a tedious process, relying in part on data from farebox transfers to establish the location of transfer stops.

Another interesting analysis done with this dataset was to aggregate it to coarser levels to see how much information is lost as the polling interval increases. Trajectories were plotted using the original 4-second data, as well as using data from only every 70 seconds or so, and data from every 4 minutes. (The 70 s interval corresponds to expectations of the Bus Service Management System working with 250 buses; 4 minutes was the expected poll interval if the entire fleet reported location.) With 4-second detail, one can see dwell times and intersection delays clearly. At 70 s, dwell times are lost, and only the worst intersection delays are visible; however, the overall shape of the trajectory, and in particular the bunching phenomenon, is clear. At 4 minute resolution, bunching is still obvious, but it becomes impossible to detect where or why it began.

Another student researcher obtained location data using the radio system. A feature of the AVL system, generally not used, is that individual vehicles can be tagged for data collection. They will then be polled for their GPS coordinates every 60 or 120 s, depending on radio channel capacity. To get data on a particular line, the researcher prepared, from the schedule, a file of run numbers serving that line. One program he created processes messages from the Bus Emergency Communication System, creating a database of log-in messages that include run number (entered by the operator) and the vehicle number. Another program queries the log-in database every minute searching for vehicles working the runs of interest; it then tags those vehicles for data collection. The poll messages from the tagged vehicles are then fed into a database. To avoid overburdening the radio system, the number of tagged vehicles never exceeded 100.

The database of poll data from the tagged vehicles was analyzed both in real-time and off-line. An example real-time analysis was to plot vehicle location on a map, with an indication of schedule deviation. This was done for the bus route serving CTA's North Racine facility, and was made available by intranet to CTA employees working there. This report became very popular – employees would check it to know when to leave the office to catch the bus. (This experiment turned out to be a clever way of building institutional support for automated data gathering!)

Off-line analyses done using the poll data included plotting trajectories, finding the distribution of running time, and analyzing schedule deviation. Running time analysis was of particular interest to the scheduling department. Schedule deviation analysis was of interest to operations supervisors. However, determining schedule deviation at route endpoints (e.g., when are buses leaving early) was complicated by three things. First, there is the general end-of-line identification problem that all automated systems deal with – when does a trip end and its layover, or the next trip, begin? Second, GPS signals in the downtown, where many routes have their endpoint, are often unclear due to canyon and multipath effects. This problem would be lessened if the AVL system used odometer information, which it cannot use until the route / schedule data issue has been solved. The third complication is the need to interpolate between polls, which, particularly with 2 minute polls, can introduce considerable error.

The experiment of tagging buses for location polling was stopped in June, 2002, about 7 weeks after it began, because the computer program managing the radio system began crashing about every 2 weeks. Until now, no causal relationship between the extra polling and the crashing has been established.

Recognizing the value of routinely collected location data was established, a change order for the Bus Emergency Communication System, due for delivery in November, 2002, will make location data part of every message returned from the buses. However, as mentioned earlier, that change anticipates using the frequent location data for real-time applications only; there are no plans for archiving the data.

Smart Bus System: Stop Announcements, Event Recording, Passenger Counting

When the CTA was hit with a court order to provide automatic stop announcements, management decided to turn it into an opportunity to meet archived operations data needs as well. They asked for a "smart bus" design that would allow APC's and other devices to connect, sharing the location and route matching facility that would be part of the stop announcement system. They also changed their procurement style from a rigidly specified design to functional specifications and an invitation for proposed features.

The system now under implementation follows the smart bus concept, with an on-board computer and an SAE J1708/J1587 vehicle area network for digital communication between the computer and on-board devices. Integrated devices include GPS receiver, odometer / speedometer, and door sensors as part of the location system; stop enunciator and destination sign driven automatically by the on-board computer; an operator console for sign-in and display to operator; and a high speed port for data upload and download at the garage. (The door sensors will be connected to the on-board computer directly rather than through the J1708 network; latency in network communication, sometimes as much as 0.5 s, can delay stop announcements enough to bother some passengers.) Other on-board devices can readily be linked to the network. A subset of buses, 15% of the fleet, will have passenger counters, which have their own processor that will broadcast over the network a J1708/J1587 message whenever the doors close, indicating the number of ons and offs at the stop. The prime contractor is Clever Devices, Inc.

The on-board computer includes the event recording function. Events to be recorded in the new system, with time and location stamp, include bus stopping at a stop, door opening, door closing with passenger on / off counts, bus leaving a stop and sign-in changes. Event data will be uploaded nightly at the garage and then transferred to a central Oracle database.

The smart bus system vendor had the option of sharing the AVL system's GPS receiver and operator console; however, they chose to install their own instead.

Data Analysis Using Third Party Software

Analysis and reporting will be done using third party reporting software, Ridecheck Plus Software, a product of RSM Services. RideCheck Plus includes over 60 standard reports and graphs developed originally for manual surveys and ride checks and which can be used equally well with APC data (see examples at www.rsmservices.com). For any given report, users can select the "filter," i.e. the date range, time range, and so forth, to be included in the report. Several reports relate to operations data. One, illustrated in Figure 1, shows the difference between observed and scheduled running time, by timepoint segment and by trip for a given route. Results are presented in a table, formatted to mirror a schedule table with a row for each scheduled trip and a column for each timepoint. The entry for a timepoint refers to the segment beginning at that timepoint. "Observed" running time is the average of all observations of the selected trip in the specified date range. Other reports show percent early / on-time / late by timepoint and route-wide.

Because the data will be stored in an Oracle database, CTA staff will be able to query it for *ad hoc* needs. They could also develop their own reports, although there are no development plans at this time since the focus now is on getting useful reports from Ridecheck Plus. As part of the contract, RSM will develop a few additional reports requested by the CTA.

LESSONS LEARNED

Unanticipated Data Needs

As people throughout the CTA become more accustomed to getting data using modern information technology, new data needs that had never been considered in earlier AVL-APC planning have arisen. Examples include:

FIGURE 1: Average Observed Minus Schedule Running Time Report, By Trip and Segment

Metropolitan Transit Authority — *Ridecheck Plus*

Running Time Crosstab — *2000 August (Weekday)*

3: University/Highland & Oak Park - Eastbound — *41 Trips*

				Differential Running Time, minutes (greater or less than scheduled time)											
Trip	Start Time	Block	Branch	2955 VALLE YOUNK H	150 31ST UNIVE N	2957 WALMA FRONT E	2743 UNIVE 31ST E	844 COTTA MLK P E	2973 WALNU 6TH E	332 6TH UNIVE N	27 11TH MADIS N	1643 HARDI SCHOO H	2224 PARKF MALL W	1642 HARDI HILLS E	3145 PARKF LAY O H
1	5:43a	301	EB-03/4	-1		1	0	0	0	0	0				
3	5:49a	31	EB-05								3				
5	6:07a	35	EB-12	0	0	0	0	0	0	0	0	0			
7	6:09a	302	EB-05								-9				
9	6:26a	303	EB-07	0		0	0	0	0	0	0	0			
11	6:41a	34	EB-01	1		-2	2	-1	-2	1	0				
13	6:53a	33	EB-05								0				
15	6:57a	31	EB-02	0	-3	1	-2	3	-1	0	0				
17	7:13a	302	EB-02	-2	3	3	-3	2	0	-3	3				
19	7:40a	304	EB-01	2		0	1	-2	2	0	-1				
21	8:04a	33	EB-01	-1		-1	0	0	-1	1	2				
23	8:19a	35	EB-01	0		0	0	0	0	0	0				
25	8:49a	34	EB-01	-2		1	-1	0	0	1	4				
27	9:09a		EB-03/4	-1		2	0	6	-4	-1	6				
29	9:53a	33	EB-01	-5		3	-1	2	1	0	0				
31	10:08a	11	EB-03/4	-4		-3	4	6	-1	-1	0				
33	10:53a	34	EB-01	-3		3	-1	-1	0	0	0				
35	11:08a	32	EB-03/4	-1		1	0	0	0	0	0				
37	11:53a	33	EB-01	-2		2	-1	1	-1	0	1				
39	12:08p	35	EB-03/4	0		0	0	0	0	0	0				
41	12:53p	34	EB-01	-2		1	0	2	-0.5	1	-1.5				
43	1:08p	31	EB-03/4	-4		3	0	5	-7	2	-4				

08-Jun-01

Source: RSM Services, Inc.

- *Legal Investigation of Disruption Impact.* CTA operations sometimes suffer disruptions due to incidents for which another party is to blame, e.g., a grade crossing accident caused by a motorist trying to sneak around a closed gate. The legal department would like a review of operations and passenger count data to estimate the cost of (a) operations changes and (b) lost revenue due to the incident to support a claim against the party at fault.
- *Security Investigations.* Heightened security consciousness since the September 11 attack on America has prompted numerous requests regarding operations, particularly during disruptions and for the September 11 evacuation of the city itself.
- *Headway-Based Ridership Analysis for Competing Routes*. As part of its Bus Rapid Transit program development, the CTA operates a limited stop route, X49, overlaid on local route 49. These routes compete for passengers whose mode of ridership is to simply use the first bus that arrives, whether on the 49 or X49. Understanding the ridership pattern of the X49, then, requires not just ridership counts on the X49, but boarding counts tied to location and stop-level arrival time on all the buses on both routes in a given period.

The lesson learned is clear: it is impossible to specify in advance all the analyses one will want to do with automatically gathered data. The data system must be open enough to allow future development of analysis tools.

Evolution of APC from Orphan to Piggyback

Until 1998 or so, APC was stuck as a "small market" item. Because the market for APC was small, there were few vendors. Due to low sales volume and limited competition, prices were high. Some large companies like Westinghouse with the resources to develop major technological advances ventured into the market; however, when the market did not quickly grow, they stopped investing in development and soon left the market altogether. Meanwhile, high prices, questions about vendor longevity, and relative lack of technological innovation kept the market small, and kept the CTA reluctant to purchase.

The price of APC's essentially mandated that only a sample of the fleet be instrumented. While the kind of data gathered from a sample of vehicles was attractive to the planning department, it found little support in other CTA departments. In contrast, automatic data gathered fleetwide was attractive not only for planning, but also for many other aspects of management, including scheduling, operations control, and incident investigation.

Now the advent of GPS and smart bus open architecture designs has made it possible for APC's to piggyback, at a relatively low marginal cost, on the location and data transfer systems justified to meet another, more pressing agency need – in the CTA's case, automatic stop announcements. While the current contract calls for instrumenting only a subset of vehicles with APC's, their small marginal cost has the agency considering including passenger counters in all new vehicle acquisitions, much like Tri-Met, until the entire fleet is equipped.

Hard vs. Functional and Open Specification

Past specifications the CTA wrote for APC's tended to specify details of the design, much like specifying the PA system for a bus. The result was often that nobody could meet the spec, or that a company could meet the spec only by developing something that had never been made – a risky proposition for both the vendor and their client.

The approach taken in their most recent procurement tended more to be a functional and an open specification. One person described the approach as "going shopping." While the request for proposals (RFP) did specify some aspects of the design, and specified certain functions as necessary, it also described many features that the CTA would find desirable – but not necessarily required – and it asked vendors to propose what else they might add. The CTA realized that in a field in which they had no experience, they were bound to not know exactly what they wanted, and a vendor with experience with other transit agencies might have a product that would benefit the CTA. Another reason for avoiding a hard spec is that for many desirable but non-critical features of the system, the CTA had little basis for estimating the cost – would it be a trivial add-on, or a major addition? Unlike the past, the market had grown to the point that there were several vendors with off-the-shelf products; and while those products have the same basic function, they still differ in many respects. To reduce both cost and risk, the CTA was interested in meeting their needs as much as possible with off-the-shelf products, and did not want a specification that legally obligated them to reject vendors that didn't meet non-critical aspects of a design specification.

This approach was rewarded in the procurement process. There were several responsive bids from companies with proven products.

The capture and analysis of archived operational data tended to play a minor role in procurement process. The specification called for an open on-board architecture and an open database, with a strong recommendation that vendors use enterprise software platforms (Oracle for general database, ESRI products for GIS), but otherwise left it up to the vendor to propose what data items to capture, how the data would be processed, and how it would be analyzed. The winning proposal included as a subcontract a third party analysis software vendor, RSM Services, Inc., whose software product RideCheck Plus will be used for analyzing the APC data.

One drawback to the open specification is that much of the detailed specification is shifted to the negotiation and contract writing phase. Negotiations over what items to include and their cost took considerable time and effort. In the end, however, the process led to a design that was both economical and that better meets client needs than if the CTA had specified the entire design from the start.

Focusing Responsibility for the Schedule Interface

In the RFP for the stop announcement / smart bus system, the CTA wanted to avoid again being the middleman between a location system vendor and the scheduling system vendor. The RFP assigned complete responsibility for the schedule interface to the stop announcement contractor. The RFP instructed bidders to not only budget their own time to work on the schedule interface, but also to budget funds to pay the scheduling system contractor for their time working on the schedule interface.

Making Schedule Matching Part of Core Operations

Traditional AVL and APC systems have a high tolerance for incorrect route and schedule data. Even if the schedule data is incorrect, or if an operator fails to enter the correct route / run ID, a bus can still operate, count passengers, and radio in its location. In planning for their new system, the CTA recognized the genius of using the location system to drive automated stop announcements: the bus would have to know what route it was following, or stop announcements, audible to all, would be wrong. To reinforce the reliance of core operations on correct route as well as location data, the new system is designed to automatically change the destination sign as well. This way, most discrepancies between run numbers as stored in the on-board computer and run numbers actually being operated will demand the immediate attention of operators and supervisors. This will ensure that most of the archived data will be correctly matched to begin with. It also converts schedule matching from an off-line processing problem to an operations problem.

Procedures are now being conceived for how to update route / run when a bus deviates from its scheduled run. Management admits that is hard to specify procedures in advance because, with the present lack of automated data collection, they don't really know enough about how operations deviate from the schedule to specify the correct procedures for notifying the system. The CTA anticipates an evolving learning process as the new data offers insights into current operation and control procedures and serves as a basis for improved operation and control procedures.

A REMAINING PROBLEM: DISTRIBUTING A SAMPLE OF INSTRUMENTED BUSES

In both the design of its Bus Service Management System (part of the existing AVL system) and its coming APC system, only a subset of the fleet will be instrumented. Considering the differing needs of passenger count data, operation data, and the very real logistical issues involved, management is looking for guidance on how these buses should best be distributed over the course of a scheduling period.

LONG TERM PROSPECT: TWO FLEETWIDE DATA SOURCES?

With the expansion of the AVL system's location reporting expected in November, and with the smart bus system's event recorder function coming on line in the next year, the CTA will have two independent, fleetwide sources of location data. For the near term, the trip recorder data will be used in archived applications because it is matched to route and schedule. However, the desire for real-time applications could lead to the development of accurately matched AVL data. In that case, archiving that AVL data would give the CTA interesting opportunities, by merging and comparing AVL and trip recorder data, to improve the accuracy and enhance analyses of both its AVL and event recorder systems.

APPENDIX E: The Société de Transport de Montréal's Experience with APC Data*

HISTORICAL DEVELOPMENT OF AUTOMATIC PASSENGER COUNTING SYSTEM

The Société de Transport de Montreal (STM) is the major transit operator in the region of Montreal, Quebec, operating a fleet of over 1,600 buses as well as a subway network. The STM has a long history of using sophisticated methodologies and computerized tools for activities like scheduling and planning. They were for example, among the original implementers of tools such as the HASTUS scheduling systems and the MADITUC network analysis model, and have had long standing cooperative arrangements with transportation research centers and researchers at the Ecole Polytechnique and the University of Montreal. This interest in using sophisticated methodologies and tools has placed a prime focus within the organization on the collection of data and on the quality of data.

The STM first became interested in using Automatic Passenger Counting technology to gather service planning-related data in the early 1990's. They identified a variety of benefits that would result from using an automated technology for the collections of data, thereby enhancing service planning, scheduling, and management activities [STM, 1993]. The benefits identified included:

- Access to an increased volume of information concerning passenger loads, load profiles, and comparison of passenger loads to service standards.
- Reduction of the time between data collection and the production of standard reports.
- Availability of information on a daily, weekly, monthly or board period basis.
- An enhanced quality of information on which to base the location of stops or shelters, and route restructuring.
- Reports that allow a comparison of established schedules with real operating conditions, and a more accurate modification of schedules to reflect real arrival and departure times.
- The production of performance indicators on a regular basis, and for any given period (daily, weekly, monthly, board period).

Having identified the desirability of an APC, they conducted from 1991-1993 a demonstration of the two alternative technologies prevailing at the time: infra-red beams and contact mats on stairs. On the basis of this demonstration, an APC system, called SCAD, was installed on 175 buses in 1995. Each bus was equipped with pressure sensitive mats on the steps, an on-board computer for storing of data, an infrared transmitter for downloading data from the bus on a daily basis to the computer in each garage, and a radio antenna to detect wayside radio signposts. Distance is measured from the vehicle odometer, with periodic resetting using the radio signals from the wayside signposts. There are 160 signposts scattered across the STM's service area. The SCAD system collects data on boardings and alightings at each stop, engines on/off, wayside signposts identification numbers, "idle" operation, etc. Data is downloaded automatically at pull-in, and processed overnight. Initial reports of total passenger loads and schedule adherence at timepoints by run are available on the computer network within 48 hours, with other reports generally available one week later. The precision of passenger counts is around 95%, and the SCAD system has become a key source of information for the STM. It has clearly become institutionalized within the management framework of the STM (see discussion below).

The STM started introducing low-floor buses in the late 1990's. This immediately created a problem for the SCAD system, which used the logic of sequential movement up or down steps in order to determine the direction of passenger movement (boarding vs. alighting). As the number of low-floor buses increased with each successive bus order, and the STM had adopted a policy to move to entirely accessible routes, it became obvious that it would no longer be feasible to ensure a full sampling of all runs, using the current technology. The STM therefore first attempted to adapt the existing treadle mat-based technology to the low-floor bus context, but the results could not meet STM's accuracy requirements.

* This case study was prepared by Brendon J. Hemily, with input from Michel Thérer of STM.

As a result, the STM undertook a demonstration project in 1999 to test alternative technological approaches, including a vision recognition system, an active infrared detector system (that detects movement through the sequential reflection of successive infrared beams) and a passive infrared system (that detects human body heat). Having predefined accuracy requirements (see below), the different technological approaches were compared in actual service. The demonstration found that infrared-based passenger counting, active or passive, fulfilled the STM's accuracy requirements, achieving the following results:

- 2.4% - 2.7% difference in average passenger loads (automatic vs. manual counts)
- 6.9% - 7.8% error in total counts (automatic vs. manual counts)
- No discernible bias between errors in counting boardings vs. alightings

In addition, the internal business case analysis estimated that the new system would result in economic savings of Can$1.36 million (US$ 890,000) over the life of the project as compared to reverting back to a manual based data collection system, and would result in a 36% internal rate of return. As a result, a Request for Proposals was issued in late 2000 to acquire a new **low-floor bus SCAD system**. One interesting aspect of the new system was that it was to be designed to complement rather than replace, at least during a three year transition period, the existing SCAD system. Although this requirement made more complex the design of the new system, STM management felt that a phased approach was critical; it was important to keep the current system and reports in place because they had been institutionalized, and involved several departments. A phased approach would also help to maximize the useful life of the existing SCAD equipment, and defray the initial capital outlay, allowing for phased replacement expenditures over time.

As a result, the low-floor SCAD project was developed in two phases. Phase 1 (15 months) would involve:

- equipping 50 new low-floor buses with infrared passenger counters, GPS receivers, and on-board computers, connected using the SAE J1708 Vehicle Area Network standard,
- installing automatic data downloading at all 7 garages,
- data processing software that builds the interfaces with STM's computerized scheduling, bus stop inventory, and vehicle assignment systems, carries out the validation, correction, identification, and reporting of data as specified by the STM, and inserts the Low-floor bus SCAD data into the existing SCAD database,
- mobile diagnostic equipment, and
- the supplier's standard report generating software.

Phase 2 would allow for installation of the SCAD equipment in an additional 120 buses over the next two years.

Phase 1 was initiated in 2001. The passenger counting technology acquired incorporates in fact a combined approach using both passive and active infrared technologies. The passive system identifies people as opposed to inert objects, while the active technology detects movement. Two other technological changes have been implemented at the same time. First, the new system uses GPS location technology, supplemented by odometer readings in cases of lost GPS signal. The existing signposts will continue to be used for the existing SCAD buses. Eventually, once a full complement of low-floor SCAD-equipped buses is acquired (12% to 15% of fleet), and the old buses are phased out, the entire system may be entirely converted over to the new system. In the mean time, both signposts and GPS will be used in parallel on the respective fleets, and the data will be integrated through software interfaces and integrated databases. Second, the infrared downloading is to be replaced on the new system by a wireless LAN network, allowing the eventual elimination of the current PC computers in the refueling bays.

In addition, the STM is reviewing the standard report generating software provided by the supplier in order to evaluate potential application of other reports and possible changes to existing business practices.

Major Reports

Among the many reports produced by the SCAD system, there exist five that have become standard management tools. These are:

- *Passenger Load / Running Time Reports (Figures 1 and 2).* These two related reports, available within 48 hours of data collection, provide an initial summary of the data collected. The Passenger Load Report includes scheduled and real departure times, arrival times and passenger loads by timepoint. The Running Time Report includes scheduled and actual terminus departure times, and running and wait times, for the total run and between timepoints.

- *Schedule Adherence Report (Figure 3).* This report provides the percentage distribution of schedule adherence by route and timepoint, broken down by percentage of on-time observations and percentages (by number of minutes) that are early or late. It also calculates the percentage of observations within the acceptable criteria for on-time performance defined as falling between –1 to +3 minutes of schedule.
- *Ridership Profile Report (Figure 4).* This report provides the detailed passenger activity by stop, measuring boardings and alightings, and percentage of each at each stop. This report is the basis for selecting stops for deploying bus shelters.
- *Service Proposal Report (Figure 5).* This critical report, produced at the end of each board period, for each route, reports for each quarter hour segment, the service standards for maximum load and headway, and the maximum load point. More importantly, the report compares the current frequency and passenger loads, by quarter hour segments, to the frequencies and projected passenger loads proposed for the next board period. It also provides the distribution of bus stops according to passenger loads and passenger load service standards. This report is the key basis for discussion between the Service Planning and Operations Departments on proposed adjustments to frequencies and running times for the next board period.

ISSUES

Institutionalization of Data Collection and Information Utilization

One of the more interesting aspects of the STM's experience with automatic passenger counting over the years has been the degree to which the data collection process and utilization has been "institutionalized" within the organization. In contrast to many transit systems using APC's, where Service Planning management and staff must constantly struggle with management and staff in the Operations Department to ensure that APC-equipped buses are deployed as much as possible according to the sampling plan, passenger counting has been effectively woven into the fabric of the STM's organizational standard operating procedures. This is the result of several factors.

First, over the course of ten years of experience, a well established routine process has been put in place. This involves the development of a detailed sampling plan prior to each of the five board periods. Every route is to be sampled four times, all day, on weekdays, and two to four times on weekends, per board period. The sampling plan is sent to the Operations Chief responsible for each route for validation prior to the board period. The agreed upon plan is then communicated daily to the Bus Starter via the RACOM dispatch system.

Data collected is downloaded every night and a first set of validation procedures are conducted automatically. These include the following routines:

- Vehicle Assignment: initial match of raw data to real runs, as obtained from RACOM
- Validation: initial match of trip using the identified sequence of radio signposts, to determine if the bus actually made the assigned trip
- Matching: matches data to bus stop inventory data
- Load-Leveling: corrects systematic passenger counting errors (negative passenger loads, or non-zero passenger loads where inappropriate)

Various criteria are used to reject suspect data including: +/- 10% for inter-signpost distance; over +/- 20 minutes from scheduled departure from garage; over +/- 10 passengers at terminals. However rejected data is not discarded and is analyzed a posteriori.

Processed data is posted within 48 hours to database in the form of the Passenger Load / Running Time Reports (see above) which includes scheduled and real departure times, arrival times and passenger loads by timepoint. Each Operations Chief has the responsibility of verifying the collected data, and has the ability to temporarily set aside data that he/she believes is invalid for reasons that must be justified. The sampling plan is later adjusted to replace any identified runs where problems were encountered.

Fulfillment of the sampling plan is also monitored through a standard report on a weekly basis, that compares planned and actual percentage of runs for data collection (by A.M. peak, P.M. peak, and totally), by garage and by day. This report is sent to each Garage Superintendent, and problems are discussed.

Beyond the well established process for data collection described above, STM adopted in 1992 a strong corporate strategy, focusing on customer-orientation and cost-effectiveness, that has had several implications for the institutionalization of the SCAD passenger counting system.

First, this customer-orientation strategy places great emphasis on measuring customer satisfaction, and on the performance of the system with respect to key variables related to customer satisfaction. Data collection on schedule adherence and ridership have become strategic tools, and the SCAD APC system is the key source of this data for the organization. Senior management supports the system and uses the SCAD data for monitoring system performance.

Perhaps most interesting in the institutionalization of the APC system at the STM are some of the unique and standard uses made of the data for management incentives purposes. In order to encourage customer-orientation throughout the organization, the STM has created a competitive system between garages : an incentive system has been designed to encourage garage superintendents and operations chiefs to achieve system objectives, including on-time performance, and the SCAD system is one of the principle tools used for measuring performance with respect to corporate objectives, and the co-related distribution of rewards and bonuses.

As a result, ensuring the completeness and validity of the data have become not only vague corporate objectives, but personal ones for Operations Management as well, creating a number of behavioral responses. Sampling plans, prepared by Service Planning are carefully reviewed by Operations Management. The initial Passenger Load / Schedule Adherence Reports, produced within 48 hours, are carefully examined by the Operations Chiefs, who can request the exclusion of suspect data for valid reasons they must justify. The validity and accuracy of the data is constantly being challenged by Operations Management (e.g. Operations Chiefs and Garage Superintendents) which forces Service Planning staff to constantly be vigilant as to its accuracy. A comparison to manual counts by checkers is conducted once a year to ensure validity. Furthermore, fulfillment of the sampling plan becomes critical; any systematic variations or problems are an important issue to resolve, and the Sampling Plan Fulfillment Report serves as one of the bases for performance evaluation of Starters.

Finally, the data produced by the SCAD system leads to the production of the critical Service Proposal Report described above. This report, produced at the end of the Board period, is the Service Planning Department's proposal of how service should be adjusted, in terms of frequencies by route and time period, to reflect a comparison of actual passenger loadings to service standards, and actual running times to scheduled ones, based on the data collected by the SCAD system. It becomes the basis of discussion between Service Planning and Operations staff and management, and is taken seriously by both sides: any changes in service design may affect schedule adherence, which in turn could affect performance and rewards.

Through the production of this report on a regular basis, the STM's SCAD automatic passenger counting system has been used to create a systematic framework for incremental decision making to tailor the design of the service produced to match the real conditions of ridership demand and operating conditions as they develop.

Measurement of APC System Accuracy

In the course of its experience with APC's, the STM has spent considerable effort defining the measurement of error and of system accuracy [STM, 2000]. Its requirements in terms of system precision, are defined as follows, based on the following maximum measurements of counting errors on valid trips:

- 10% on the measurement on the Error on passenger counts
- 5% on the measurement on the Average error for counts of boardings
- 5% on the measurement on the Average error for counts of alightings
- 5% on the measurement on the Differential for average passenger load

where the above terms are defined as follows:

Error on passenger counts = $(\sum\varepsilon / \sum MC) * 100\%$

$\sum\varepsilon$: sum of the differences (in absolute numbers) between manual counts (by checkers) and automatic counts (using the APC) at front and rear doors for all stops

$\sum MC$: sum of manual counts (in absolute numbers) at front and rear doors for all stops

Average error for counts of boardings = $(\sum\sigma b / \sum CB) * 100\%$

$\sum\sigma b$: sum of the differences (in absolute numbers) between counts for boardings, obtained manually and automatically, at front and rear doors for all stops
$\sum CB$: sum of the manual boarding counts (in absolute numbers), at front and rear doors for all stops

Average error for counts of alightings = $(\sum\sigma a / \sum CA) * 100\%$
$\sum\sigma a$: sum of the differences (in absolute numbers)
$\sum CA$: sum of the manual alighting counts (in absolute numbers), at front and rear doors for all stops

Differential for average passenger load = $(\sum\sigma l / \sum CL) * 100\%$
$\sum\sigma l$: sum of the differences (in absolute numbers) between the estimated passenger load (using the APC) and the real passenger load (from checkers) for all stops
$\sum CL$: sum of the real passenger load for all stops

"Valid" trips

To be valid, a trip must be "balanced," i.e. the difference between boardings and alightings should be less than 5 in absolute numbers. Valid trips must represent at least 85% of all observed trips. In addition:

- The above calculations must be based on the raw data obtained from the passenger counters, and should not be adjusted.
- In order to avoid any bias created through the manual counts by checkers, which is the basis for the above measurement of precision, trips for which manual counts show an absolute difference greater than 3 between boardings and alightings will not be used in the evaluation.
- Automatic transfer of on-board data must be successfully completed daily for 95% of data collected.

CONCLUSIONS

The STM case study illustrates a transit system that has now been involved in the development and use of an APC system for a decade, and where through the combination of corporate strategy, business processes, and management incentives, the APC system has become a critical corporate source of data, and the key for a systematic approach to service improvement, based on continuous measurement of the evolution of demand and operating conditions. This has occurred despite the limitations of the older technology, the focused effort on a small number of reports, and the relatively simple format for these reports. STM staff believes that an even greater use of the new APC system can be made by using more state-of-the-art reports.

REFERENCES

1. Société de Transport de Montréal (formerly STCUM). Systèmes de collecte automatique de données-rapport final (Automatic Passenger Counting Systems-Final report), 1993.
2. Société de Transport de Montréal. Projet SCAD plancher surbaissé; Rapport d'essais (Low-Floor Bus APC System-Test Results), 2000.

FIGURE 1: Trip-Level Arrival Time (Hr. arr.) and Load (Ch.) at Timepoints

R_CHARG1.FRX

Rapport sur les charges

(Trié par heure réelle)

Ligne: 506 Newman Voie Reservee

Période: 2002/04/01 au 2002/04/15

Type jour: Semaine Direction: Est Heures: 04:30 à 12:00

Heure départ réelle	Heure départ prévue	Numéro tournée	Date sortie	S00801(2801051) Terminus Stcum Monette/lafleur Hr.arr.	Ch.	S010 (2801344) 90 Airlie Hr.arr.	Ch.	S012 (2801584) Newman Dollard Hr.arr.	Ch.	S014 (2802064) Newman Lapierre Hr.arr.	Ch.	S062 (2802500) Newman Angrignon Hr.arr.	Ch.
6:02:24	6:03:00	12	2002/04/08	5:54:46	1	6:06:34	18	6:09:16	29	6:13:34	35	6:15:44	35
6:13:49	6:11:01	22	2002/04/09	6:13:49	1	6:18:49	14	6:21:11	28	6:28:17	37	6:34:00	42
6:18:39	6:18:00	38	2002/04/12	6:11:43	5	6:22:35	12	6:25:07	16	6:29:50	18	6:33:42	15
6:29:55	6:30:00	6	2002/04/02	6:26:45	1	6:33:43	16	6:36:21	25	6:41:15	40	6:45:07	42
6:36:26	6:36:00	17	2002/04/02	6:26:48	7	6:40:12	15	6:43:52	29	6:49:26	51	6:52:44	49
6:44:50	6:42:01	41	2002/04/12	6:32:22		6:48:44	19	6:51:16	28	6:55:48	45	6:59:40	47
6:47:17	6:47:01	9	2002/04/11	6:46:05	3	6:51:41	12	6:54:25	25	7:00:17	45	7:04:11	46
6:51:14	6:52:00	27	2002/04/04	6:51:14	1	6:55:54	21	6:58:24	24	7:03:48	44	7:08:28	48
6:57:40	6:57:00	54	2002/04/02	6:57:40	1	7:01:41	3	7:05:05	5	7:10:49	19	7:13:59	21
7:01:01	7:14:01	25	2002/04/11	7:01:01		7:26:04		7:28:27		7:30:38		7:32:49	
7:07:25	7:06:00	31	2002/04/02	7:07:19	2	7:10:41	11	7:13:27	15	7:18:23	28	7:23:18	30
7:18:25	7:18:01	42	2002/04/12	7:15:29		7:23:11	24	7:25:35	23	7:32:27	39	7:36:05	42
7:19:14	7:18:01	42	2002/04/02	7:19:14	1	7:23:11	17	7:25:41	21	7:32:39	29	7:36:15	36
7:19:25	7:18:01	42	2002/04/11	7:19:17	1	7:23:33	36	7:27:21	45	7:31:21	45	7:35:44	45
7:23:58	7:21:00	80	2002/04/02	7:15:40	1	7:28:52	18	7:31:44	35	7:37:14	48	7:41:44	50
7:24:32	7:25:01	56	2002/04/05	7:23:20	3	7:28:54	13	7:32:00	18	7:36:40	37	7:39:54	38
7:30:08	7:29:01	27	2002/04/04	7:30:08	1	7:34:44	33	7:37:24	43	7:43:21	73	7:45:31	73
7:32:59	7:33:00	7	2002/04/04	7:32:59	1	7:36:14	7	7:39:20	6	7:44:20	29	7:47:32	38
7:41:54	7:41:01	46	2002/04/11	7:41:36	9	7:45:30	18	7:47:48	18	7:54:24	38	7:58:06	43
7:44:01	7:45:01	40	2002/04/02	7:36:37	5	7:48:53	9	7:51:33	22	7:56:27	23	8:00:27	22
7:49:38	7:52:01	45	2002/04/12	7:46:48	1	7:56:46	8	8:00:02	26	8:04:32	48	8:09:00	54
8:02:11	8:02:00	54	2002/04/02	7:58:05		8:05:01	2	8:08:23	6	8:12:34	18	8:16:06	29
8:08:26	8:07:00	80	2002/04/02	8:01:42	3	8:13:40	33	8:16:38	40	8:22:10	71	8:26:25	66
8:12:04	8:12:01	10	2002/04/11	8:11:40	1	8:16:54	22	8:19:22	40	8:24:10	60	8:29:12	62
8:17:34	8:18:00	56	2002/04/05	8:12:42	6	8:22:40	24	8:25:24	38	8:31:08	63	8:34:34	67
8:26:05	8:24:00	23	2002/04/11	8:26:05	1	8:30:47	29	8:34:43	33	8:40:17	43	8:45:43	44
8:32:06	8:32:01	46	2002/04/11	8:27:58	2	8:36:28	15	8:39:24	22	8:44:50	34	8:47:59	31
8:38:49	8:40:00	40	2002/04/02	8:28:49	3	8:43:25	17	8:46:11	22	8:52:05	28	8:56:36	29
8:50:28	8:50:01	2	2002/04/05	8:38:45	6	8:54:12	22	8:57:28	31	9:04:16	50	9:08:03	51
			*Moyenne:		2		17		25		39		41

Produit le: 2002/05/03

FIGURE 2: Running Time (Temps) and Holding Time (Attente) for a Route, Both Overall (Total) and by Segment

Rapport sur les temps de parcours

(Trié par heure réelle)

R_TEMPS1.FRX

Ligne: 103 MONKLAND

Période: 2002/04/01 au 2002/04/15

Type jour: Semaine **Direction:** Est **Heures:** 04:30 à 12:00

Heure départ réelle	Heure départ prévue	Numéro tournée	Date sortie	Total		G048 (107403) Grand Boulevard Monkland		M23801(126991) Metro Villa-maria	
				Temps	Attente	Temps	Attente	Temps	Attente
6:16:33	6:15:01	8	2002/04/12	0:15:43	0:00:23	0:09:52		0:05:51	0:00:23
6:31:40	6:30:00	36	2002/04/15	0:15:45	0:00:43	0:10:26	0:00:11	0:05:19	0:00:32
6:45:41	6:45:00	54	2002/04/10	0:17:06	0:00:42	0:11:12	0:00:08	0:05:54	0:00:35
7:31:11	7:31:00	52	2002/04/11	0:12:55	0:00:36	0:05:57	0:00:10	0:06:58	0:00:26
7:47:00	7:46:00	21	2002/04/05	0:12:00	0:01:32	0:04:49		0:07:10	0:01:32
7:47:14	7:47:01	12	2002/04/04	0:20:48	0:02:10	0:12:50	0:00:18	0:07:58	0:01:52
8:02:38	8:03:01	23	2002/04/15	0:22:56	0:01:25	0:13:00	0:00:16	0:09:56	0:01:08
8:13:05	8:17:01	6	2002/04/15	0:20:36	0:01:00	0:12:56	0:00:12	0:07:40	0:00:48
8:17:35	8:17:01	54	2002/04/10	0:13:13	0:00:35	0:03:47	0:00:06	0:09:26	0:00:29
8:48:24	8:47:00	12	2002/04/04	0:17:04	0:00:47	0:10:16	0:00:10	0:06:48	0:00:36
9:15:04	9:15:01	6	2002/04/15	0:15:58	0:00:40	0:10:34	0:00:10	0:05:24	0:00:29
9:59:15	9:59:00	35	2002/04/15	0:19:47	0:04:26	0:13:17	0:00:35	0:06:30	0:03:52
10:10:21	10:10:00	6	2002/04/15	0:18:24	0:01:30	0:12:22	0:00:12	0:06:02	0:01:18
10:22:42	10:21:00	37	2002/04/15	0:19:32	0:01:12	0:12:52	0:00:11	0:06:40	0:01:00
10:43:25	10:43:01	36	2002/04/15	0:17:06	0:07:02	0:11:04		0:06:02	0:07:02
10:54:23	10:54:01	35	2002/04/15	0:18:32	0:00:54	0:12:49		0:05:43	0:00:54
11:06:05	11:05:01	6	2002/04/15	0:19:45	0:00:44	0:13:49	0:00:10	0:05:55	0:00:34
11:15:49	11:16:00	37	2002/04/15	0:16:58	0:00:52	0:12:18	0:00:12	0:04:40	0:00:40
11:27:16	11:27:00	71	2002/04/15	0:23:28	0:01:33	0:17:16	0:00:54	0:06:12	0:00:40
11:37:11	11:38:00	36	2002/04/15	0:16:49	0:02:53	0:10:36		0:06:13	0:02:53
11:50:49	11:49:00	35	2002/04/15	0:16:41	0:00:44	0:11:25	0:00:06	0:05:16	0:00:38
			*Moyenne:	0:17:40	0:01:33	0:11:07	0:00:12	0:06:33	0:01:21

R_LIGNE2.FRX

Rapport sommaire par ligne et endroit (Trié par ligne/endroit)

Centre: 58000 - Legendre

Période: 2002/04 au 2002/04

Type jour:Semaine Direction:Tous Heures:Tous

Ligne	Dir.	Endroit	Nb. obs.	************* % en avance ************ -5	-4	-3	-2	-1	0	************* % en retard ************ +1	+2	+3	+4	+5	Performance (-1 à +3)
69	E	B000 grenet / de serres	130	0.7	0.7	0.7	0.0	2.3	27.6	55.3	10.0	0.7	0.0	1.5	96.1
		B034 salaberry / de l'acadie	130	2.3	0.0	0.0	7.6	26.1	30.7	22.3	6.1	3.0	0.7	0.7	88.4
		M280 station henri-bourassa	130	0.7	0.0	0.0	0.0	8.4	32.3	35.3	15.3	4.6	2.3	0.7	96.1
		H048 h-bourassa /st-michel	130	1.5	0.0	0.7	3.0	6.1	23.0	19.2	20.7	8.4	6.1	10.7	77.6
		N017 lacordaire /renoir	130	1.5	0.0	0.7	0.7	3.8	4.6	13.8	8.4	19.2	10.0	36.9	50.0
	O	N060 college m.-victorin	94	0.0	0.0	0.0	0.0	4.2	39.3	42.5	5.3	4.2	1.0	3.1	95.7
		N001 h.-bourassa /gariepy	94	0.0	0.0	0.0	2.1	10.6	34.0	18.0	15.9	8.5	1.0	9.5	87.2
		H048 h-bourassa /st-michel	94	0.0	0.0	1.0	1.0	19.1	28.7	23.4	8.5	8.5	2.1	7.4	88.2
		M280 station henri-bourassa	94	0.0	0.0	1.0	6.3	12.7	34.0	15.9	11.7	8.5	4.2	5.3	82.9
		B034 salaberry / de l'acadie	94	0.0	0.0	2.1	5.3	8.5	20.2	18.0	18.0	6.3	6.3	14.8	71.2
Sommaire ligne 69			**1 120**	**0.8**	**0.0**	**0.6**	**2.5**	**10.0**	**26.8**	**26.8**	**12.0**	**7.2**	**3.4**	**9.2**	**83.1**
Sommaire Legendre			**1 120**	**0.8**	**0.0**	**0.6**	**2.5**	**10.0**	**26.8**	**26.8**	**12.0**	**7.2**	**3.4**	**9.2**	**83.1**

FIGURE 4: Boarding and Alighting Distribution by Line and Location

SCAD_ACHAL_04DISMD

Distribution de l'achalandage

Période de liste: 01S

Ligne: 24 SHERBROOKE

Réelles et estimées Type service: Semaine Direction: Toutes Heures:

No. ligne	Type service	Direction	Trajet	Arrêt	Montant Absolu	Montant %	Descendant Absolu	Descendant %
24	Semaine	Est	B M238E106	1 METRO VILLA-MARIA /	1 278	12.94	3	0.03
				2 DECARIE / DUQUETTE	42	0.43	4	0.04
				3 DECARIE / NOTRE-DAME-DE-GRACE	175	1.77	108	1.09
				4 DECARIE / COTE ST-ANTOINE	177	1.79	135	1.37
				5 DECARIE / SHERBROOKE	377	3.82	73	0.74
				6 SHERBROOKE / VENDOME, DE	309	3.12	52	0.52
				7 SHERBROOKE / CLAREMONT	375	3.80	89	0.90
				8 SHERBROOKE / GROSVENOR	561	5.67	167	1.69
				9 SHERBROOKE / LANDSDOWNE	141	1.43	104	1.05
				10 SHERBROOKE / STRATHCONA	201	2.03	67	0.68
				11 SHERBROOKE / METCALFE	128	1.30	82	0.83
				12 SHERBROOKE / REDFERN	176	1.78	120	1.22
				13 SHERBROOKE / CLARKE	211	2.14	167	1.69
				14 SHERBROOKE / GREENE	164	1.66	167	1.69
				15 SHERBROOKE / WOOD	67	0.68	59	0.60
				16 SHERBROOKE / ATWATER	256	2.59	528	5.34
				17 SHERBROOKE / CHOMEDEY	62	0.63	51	0.51
				18 SHERBROOKE / ST MARC	159	1.61	172	1.74
				19 SHERBROOKE / ST MATHIEU	340	3.44	198	2.00
				20 SHERBROOKE / GUY	173	1.75	320	3.24
				21 SHERBROOKE / BISHOP	237	2.40	232	2.34
				22 SHERBROOKE / MONTAGNE, DE LA	185	1.87	202	2.04
				23 SHERBROOKE / PEEL	275	2.78	462	4.66
				24 SHERBROOKE / MCGILL COLLEGE	335	3.39	489	4.94
				25 SHERBROOKE / UNIVERSITY	270	2.74	357	3.61
				26 SHERBROOKE / AYLMER	181	1.83	125	1.26
				27 SHERBROOKE / BLEURY, DE	285	2.88	404	4.08
				28 SHERBROOKE / JEANNE-MANCE	55	0.56	80	0.81
				29 SHERBROOKE / ST URBAIN	163	1.65	113	1.15
				30 SHERBROOKE / ST LAURENT	110	1.12	371	3.75
				31 SHERBROOKE / HOTEL-DE-VILLE, DE	92	0.93	130	1.31
				32 SHERBROOKE / ST DENIS	95	0.96	192	1.94
				33 BERRI / CHERRIER	1 752	17.73	605	6.11
				34 CHERRIER / ST HUBERT	35	0.35	20	0.20
				35 CHERRIER / PARC-LAFONTAINE, DU	74	0.75	261	2.64
				36 SHERBROOKE / PANET	76	0.77	355	3.58
				37 SHERBROOKE / PLESSIS	20	0.20	125	1.26
				38 SHERBROOKE / HOPITAL	49	0.49	286	2.89
				39 SHERBROOKE / PAPINEAU	55	0.55	420	4.24
				40 SHERBROOKE / DORION	32	0.33	248	2.50
				41 SHERBROOKE / LORIMIER, DE	33	0.33	297	3.00
				42 SHERBROOKE / PARTHENAIS	22	0.23	359	3.62
				43 SHERBROOKE / FULLUM	13	0.13	189	1.91
				44 SHERBROOKE / IBERVILLE, D'	12	0.12	233	2.35
				45 SHERBROOKE / FRONTENAC	7	0.07	294	2.97
				46 RACHEL / MONTGOMERY	47	0.48	210	2.13

Produit le: 2002/04/09 Page 1

Mise en garde: *Les résultats de calculs sont arrondis à l'unité*

FIGURE 5: Service Proposal and Load Analysis

Applicable Standard (Norme) for Load (Ch.) and Headway (Int.);
Current (Reél) vs. Proposed Number of Departures, Headway (Fréq.), and
Corresponding Load (Charge) by 30-min Period;
Stops (Arrets) and Time (Temps) Spent at Four Levels of Loading (Charge) by
Scheduled Trip

R_SERV2.FRX

Réels et estimés

Proposition de service

Ligne: 100 CREMAZIE   Liste: 01S 2001-08-27 au 2001-12-09

Type jour: Semaine   Direction: Ouest   Trajet: M276H079 Station Cremazie - 4505 Hickmore (Mtee De Liesse)

Quart d'heure	Tour	Heure départ	Temps parcours	Nb. éch.	Arrêt charge maximum	Norme Ch. / Int.	********* Réel ******** Départs	Fréq.	Charge	******** Proposé ******* Départs	Fréq.	Charge	*** Répartition des arrêts selon la charge *** Charge <=45 Arrêts	Temps	45 < Charge <= 55 Arrêts	Temps	55 < Charge <= 65 Arrêts	Temps	Charge > 65 Arrêts	Temps	Total Temps
05:15	5B	05:15	30	4	129617 Cremazie / Acadie, De L'	45 / 40	3		25	2	15.0	31	39	20:53							20:53
	2E	05:23	30	7	129617 Cremazie / Acadie, De L'	45 / 40		8	9				39	18:03							18:03
05:30	6C	05:38	30	3	104602 Cremazie / Acadie, De L'	45 / 40		15	27				39	22:49							22:49
05:45	33	05:45	30	9	1905590 Decarie / Opp 290	45 / 40	5	7	33	5	6.0	42	39	22:14							22:14
	36	05:51	31	3	1901889 Cote De Liesse(nord) / Opp	45 / 40		6	16				39	20:40							20:40
	37	05:57	32	2	1901889 Cote De Liesse(nord) / Opp	45 / 40		6	37				39	20:48							20:48
06:00	1	06:03	33	5	1901889 Cote De Liesse(nord) / Opp	45 / 40		6	67				14	6:43	11	7:21	13	11:23	1	1:03	26:30
	8	06:09	33	7	104602 Cremazie / Acadie, De L'	45 / 40		6	55				9	6:41	30	18:28					25:09
06:15	E 44	06:15	33		1901889 Cote De Liesse(nord) / Opp	45 / 40	5	6	53	6	5.0	33	22		17						
	9	06:21	33	6	1905962 Cote De Liesse(nord) / 5850	45 / 40		6	54				16	8:55	23	16:40					25:35
	8E	06:27	33			45 / 40		6													
06:30	6C	06:33	33	3	1901889 Cote De Liesse(nord) / Opp	55 / 31		6	34				39	29:29							29:29
	12	06:39	33	5	129587 Cremazie / Henri-julien	55 / 31		6	57				17	14:03	18	10:13	4	3:01			27:17
06:45	14	06:45	33	5	104602 Cremazie / Acadie, De L'	55 / 31	5	6	35	4	7.5	49	39	25:03							25:03
	E 7C	06:51	33		104602 Cremazie / Acadie, De L'	55 / 31		6	29				39								
	16	06:57	33	3	1901889 Cote De Liesse(nord) / Opp	55 / 31		6	62				15	9:41	14	11:06	10	9:32			30:19
07:00	18	07:03	33	8	105262 Cremazie / St Laurent	65 / 31		6	33				39	26:12							26:12
	E 1	07:09	33		104602 Cremazie / Acadie, De L'	65 / 31		6	38				39								
07:15	20	07:15	33	7	1901889 Cote De Liesse(nord) / Opp	65 / 31	6	6	41	5	6.0	59	39	28:15							28:15
	9	07:20	33	2	105912 Cremazie / Berri	65 / 31		5	56				30	20:03	6	3:20	3	0:52			24:15
	8	07:26	33	4	129617 Cremazie / Acadie, De L'	65 / 31		6	59				25	19:20	6	4:37	8	4:30			28:27
07:30	2	07:32	33	3	105262 Cremazie / St Laurent	65 / 31		6	54				30	25:07	9	6:27					31:34
	12	07:38	33	4	105262 Cremazie / St Laurent	65 / 31		6	52				30	23:10	9	5:30					28:40
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]				[illegible]	[illegible]							30:57

APPENDIX F: OC Transpo: A Pioneer in APC Use for Service Improvement*

HISTORICAL DEVELOPMENT OF APC AND AVL/C SYSTEMS

APC System Development

OC Transpo operates a 930 bus fleet and a diesel LRT line in the city of Ottawa, Ontario. OC Transpo was one of the earliest implementers of APC technology in North America. Much of the following historical information is derived from the sources listed in the References, and complemented by information obtained during the interviews with OC Transpo staff.

The first generation APC system was developed and implemented starting in 1975. The original system configuration used the successive breaking of two infrared light beams to detect passengers and determine the direction of passenger movement. Data collected was interpreted by an on-board microprocessor, stored in memory, and initially retrieved manually using a portable cassette unit for upload to a mainframe computer. Distance was measured using the odometer. By 1979, the APC system was installed on a fleet of 33 APC buses (approximately 5% of the fleet).

From 1979 through to 1984, the system expanded to a fleet of 66 APC equipped buses spread out over three different bus types (sizes) at four different garages. Data retrieval was successfully upgraded to diskettes in 1981, and then to an automatic radio data transmission system in 1984. Although the system was in full production and collected large volumes of information, its utility was severely hampered by the cumbersome reporting sub-system, which only allowed for end-of-booking data reporting and included no ad-hoc information retrieval.

In 1987, a major custom-designed upgrade was made to the APC software system by completely redesigning and replacing it. Apart from relatively minor hardware enhancements, the bulk of the upgrade consisted of improvements to the daily data processing, reporting, and data access system. Problems of obtaining the right type of data, when needed and in the proper format were central in the new design.

Following the upgrade, each bus is equipped with: infrared light beams covering all boarding and alighting channels; an on-board microprocessor to interpret the beam breakages and to store passenger activity and other logs describing the bus progress by time and distance; a microwave receiver; and a radio control module which enables transfer of data from the bus to the central computer. On-street, 35 microwave signposts are strategically placed on traffic signals or utility poles to help locate the buses accurately. The system now includes 80 instrumented buses (9% of the bus fleet at the time), based proportionately on the fleet mix of 18 meter articulated, 12 meter standard, and 9 meter medium coaches.

The data is transferred at night automatically from the buses to a central computer through the radio system, and a number of automatic procedures take place to sort and validate the downloaded data overnight.

The largest part of the new APC system involved a total re-design and re-write of the software processing and reporting system. The new system enabled the following improvements:

- Incorporate multiple samples of the same run
- Retain all valid run data from a bus daily assignment, even if the assignment was interrupted or changed part way through the day
- Inclusion of bus stop data in the trip schedule itineraries and reporting to the bus stop level
- Incorporate ability to capture and report time utilization data
- Reports can be generated at any time during a board period, and users are given total flexibility to specify, for any report, the date range (e.g., March 14-28) and the time period for analysis (e.g., 07:00-09:00), as well as the specification of route, route direction, route section or geographical area. Geographical areas are created through the bus stop inventory by identifying the geographic area to which each stop belongs.

---

* This case study was prepared by Brendon J. Hemily, with input from Joel Koffman of OC Transpo.

- A whole new array of reports was developed, as discussed in the next section.

Data matching and quality was enhanced through a whole set of new tools and procedures. They included:

- *Hardware Diagnostic Capability.* A new Diagnostic Module provides a tool for the user to pinpoint precisely the source of any on-bus problem, while the Passenger Counting Module monitors certain component elements and appends a special diagnostic log at the end of each transmission of data from a bus. This log records the number of beam breakages, the time the beam was broken, the number of door openings, the number of odometer pulses received, and the number of signposts passed. The log is then used by the data processing unit to produce a daily detailed diagnostic report that identifies suspected malfunctions.
- *Automatic Generation of Bus-to-Run Assignments at Start of Day.* The APC Maintainer uses a new program to automatically produce the bus-to-run assignments and store them in a data file, which is then automatically input into the central database at the beginning of each day, and flagged on the Bus Starter's daily file. The generation of the assignments is totally under the user's control. Choices include a random sample; all runs on a route; or specific bus numbers chosen for specific bus runs.
- *Automatic Nightly Processing and Checks.* First, starting at 18:00, the Automatic Radio Transmission System starts polling each bus in sequence to determine if it has returned to the garage. If it has, its data is downloaded via radio. Once received, the Raw Data Processor, which has already received from the central computer all bus-to-run assignments made by Bus Starters for that day, initiates a number of steps. It changes relative time to absolute time, splits the data into bus runs using the bus-to-run assignment data, and initializes distance and passenger count data for each run. When this is complete for all runs from a bus, the processed data are passed to a run validation procedure that verifies the data from each run by checking the following items against pre-set user-defined tolerance criteria:
    - Actual versus scheduled pull-out time?
    - Actual versus scheduled pull-in time?
    - Total versus scheduled run length?
    - Number and sequence of signposts passed?
    - Difference between total ons and offs?

Validated data is split into trips and matched to individual stops. These procedures are completed by 04:00 and generate a report for the APC Maintainer that outlines the following status conditions:

- Buses successful/not successful in data transmission
- Data passing/not passing validation
- Data passing/not passing matching.

The APC maintainer then reviews the reports, verifies cause of data breakdowns, and accepts validated data up to the point of any breakdowns. He then posts all validated data to the 'good' data file.

These procedures were extremely advanced at the time they were introduced, and helped tremendously to reduce the amount of lost data, reduce the time spent diagnosing problems, and helped increase the credibility of the data among the many departments relying on it. The quality of the system implemented in the late 1980's has continued to serve OC Transpo well during the following 15 years. With the arrival of wide-door low-floor buses, a study was undertaken in 1999 to evaluate available APC technology that could be used in low-floor buses and on their new diesel LRT line. Starting in 2001, all new buses added to the APC bus fleet will be equipped with overhead passenger detection technology that is capable of counting simultaneous passenger movements through the wider doors. As of November 2002, 12 articulated buses and their new diesel LRT line are operational with this new equipment.

The APC system has now been in place at OC Transpo for 24 years and is well entrenched. The sampling plan is well followed (90%), and a report monitoring fulfillment of the sampling plan is used to evaluate the performance of Bus Starters. Data from the system is extensively used by planning, scheduling, operations, and management staff.

Automatic Vehicle Location and Control (AVL/C) System

In the early 1990's, OC Transpo explored and eventually implemented an AVL/C system. However, OC Transpo chose to adopt a very different technological approach for their AVL/C system from those common at the time. The system uses inexpensive Surface Acoustic Wave (SAW) tags on the buses that are read by 26 wayside readers spread throughout the region as they pass by, and this information is then transmitted to the central computer at OC Transpo. Using the knowledge of bus passage times, an algorithm interpolates estimated location and schedule adherence of each bus. Location and schedule adherence is re-calibrated with the next passage in front of a wayside reader. This approach was adopted because it offered, at the time, a less expensive approach to vehicle location and schedule monitoring, without the requirement to invest in expensive on-board computers, integrated data/voice communications and data communications channels. It has provided an adequate approach to vehicle monitoring, but is limited by the large expense of the wayside readers. At this point, with the substantial decrease in cost of on-board equipment and the increased reliability of GPS location, OC Transpo is exploring the potential of re-designing its AVL system.

The technology adopted by OC Transpo, although effective for real-time control, somewhat narrows its use for off-line data use; with positive location determination occurring only at the 26 wayside readers systemwide, it provides far less detail than the APC system, although, unlike APC, it provides fleetwide coverage. The only extensively used report from the AVL/C system pertains to missed trips, runs, and low-floor bus assignments.

APC: Data Capture and Reduction for Classifying Vehicle Time Utilization

Measuring Vehicle Time Utilization

OC Transpo was one of the first transit systems to use the APC data gathering system to measure in detail the time utilization of buses. This is achieved through the recording of three different types of data records:

1. *Passenger Log.* In addition to counting the passenger activity by door at a bus stop, this log also records
    - the amount of total time at the stop (stopped time) measured from odometer pulse preceding wheel stop to odometer pulse following wheel starts, and
    - the amount of time spent actively boarding and alighting passengers (dwell time) measured by time from first door opening to time of last passenger beam breakage at a stop.

2. *Idle Log.* Whenever the bus is idle for at least 45 seconds without any passenger activity, a record is issued with the next odometer movement (pulse) that includes the start time of idleness along with the duration of the idle period.
3. *Stop-and-Go Log.* This log is intended to measure the time spent in traffic congestion (i.e. moving very slowly). The use of information from this log provides a relative relationship rather than an absolute measure of very slow movement between routes. The current logic that drives this log is the requirement to encounter a string of five sequential events where events 1, 3, and 5 are ‘idle for less than 45 seconds’ and events 2 and 4 are ‘movement of less than 400 feet’. In essence, it is trapping and reporting upon periods of very low average speed.

Figure 1 illustrates the logic that controls the generation of these three logs.

As a result, the APC system can break vehicle time utilization into the following categories:

- time spent between stops - congestion free
- time spent between stops – in congestion
- time spent between stops – idle time
- time spent at stops – actively boarding/alighting passengers
- time spent at stops – excess time.

It is only recently that other commercially available APC systems have provided equivalent data on vehicle time utilization. This data can also be presented graphically, as shown in Figure 2, to illustrate how much of each type of delay occurs systemwide and by route-period-direction. Interestingly, this report is no longer available at OC Transpo due to plotter obsolescence and relative lack of interest in the report. The fact that this report does not show where on a route the delay occurs may have contributed to its discontinuance.

MAJOR APC REPORTS

The true value of the APC system lies in the report generation programs. The goal of the reporting system is to present data in such a way that it meets the needs of its users: managers need overview reports that provide a gauge of overall service performance, while planners and schedulers require a more in-depth analysis of the service elements. Also, managers for the most part need standard reports at relatively fixed times during the board period or year while planners and schedulers may require data on specific service elements at any given moment. The APC department acts as a Service Bureau generating both regular reports and as ad-hoc reports for users in different departments.

The 'good database' for a board period contains a large volume of data. OC Transpo typically operates about 8000 one-way trips in a weekday and it is quite common to have over 22,000 captured trips in a 3 month board period. The reporting system is designed to respond to specific requests for data by presenting a range of reports, which are controlled by a series of 'data selection' parameters. The data selection criteria include:

- Day Type (Weekday, Saturday, Sunday, Special)
- Date Range (entire booking or from date/to date)
- Time Period (5 standard periods early, AM Peak, Midday, PM Peak, evening, all day, or from time to time)
- Route Class (regular line haul, regular local, regular peak, interline express, interline non-express, all routes)
- Route/Direction
- Route Section (from stop, to stop)
- Geographic Area (list of stops)

This ability to tailor the selection of the information to fulfill the data need is a powerful function of the system.

The data produced by the APC system is organized into two broad categories that respond to the users' varied needs - Management Overview and Detailed Service Statistics. Table 1 lists the statistics available under each category.

TABLE 1: Available APC Statistics at OC Transpo

MANAGEMENT OVERVIEW STATISTICS

Performance

- Overall Service Provision
- Route Performance Exceptions
- Revenue Time Utilization
- Schedule Adherence

Passenger Activity

- Passenger Volumes at Screenlines
- Maximum Passenger Loads
- Passenger Boardings
- Passenger Alightings
- Distribution of Maximum Passenger Loads
- Distribution of Schedule Adherence
- Passenger Activity for Major Centres

DETAILED SERVICE STATISTICS

Performance

- Regular Service Provision
- Interline Service Provision
- Schedule Adherence at Timing Points by Time Period
- Schedule Adherence by Route
- Revenue Time Utilization by Route by Time Period
- Dwell Times by Route by Time Period
- Point Check Report
- Door Usage by Route
- Running Times by Section
- Distribution of Running Times
- Deadhead Running Times
- Route Direction Summary
- Route Rank Report (Revenue/Cost Ratio, Pass Km per Km, Overloading, Schedule Adherence, Excessive Idle Time, Excessive Layover Time, etc.)

Passenger Activity

- Load Profiles by Stop / by Time Period
- Load Profiles by Section / by Time Period
- Load profiles by Route
- Passenger Activity by Stop / by Time Period
- Passenger by Section by Time Period
- Passenger Activity by Area by Time Period
- Passenger Activity by Route
- Passenger Volumes at Screenlines

The APC System data is presented through a series of reports that have been standardized over time. The following provides descriptions of some of the most commonly used reports:

Summary Of Running Times (SORT) [Figure 3]

The SORT report produces summary statistics on running times by section for requested routes and patterns. The scheduled running time, observed mean running time, and number of observations are presented for various sections each hour of the day.

Distribution Of Running Times (DORT) [Figure 4]

The DORT report shows the distribution of running time over each hour of the day; between two specified route-day type combinations. There is also a summary that shows average running times (scheduled and actual) of all the patterns in five periods of the day.

Deadhead Running Times (DHRT) [Figure 5]

The DHRT report lists scheduled deadhead links with their mean scheduled, mean running times, with deviation between mean observed, and standard deviation of observed running time.

Revenue Time Utilization (RMOC) and Running Time Utilization By SECtion (RTUBSEC) [Figures 6 and 7]

The RMOC report is a data table showing, for each observed trip between a given stop pair, total time spent vs. scheduled, with a breakout into five categories of time: moving time, stop-go time, idle time. dwell time, and excess time. The RTUBSEC report shows mean and standard deviation of each running time category for a given each route-day type-time period combination.

Route Direction SUMmary report (RDSUM) [Figure 8]

The RDSUM report (Route Direction SUMmary) presents trip-level operations and passenger use statistics for each scheduled trip for a given route-direction. Every observation is listed, as well as an average for the scheduled trip. Statistics include actual departure time (from the origin terminal), arrival time (at the destination terminal), running time, boardings, maximum load, and maximum load point. Scheduled departure, arrival, and running times are also presented, as are deviations from those scheduled times.

Passenger Activity and Load profile by StoP (PALPSTP) [Figure 9]

The PALBSTP report produces mean, median, minimum, maximum, and standard deviation of passenger boarding, alighting, and load per bus for a sequence of stops.

PoinT CHECK (PTCHECK) [Figure 10]

The PTCHECK report gives a summary of service for a specific location. Ons, offs, load, actual time, and difference from scheduled time are shown for each observation. For any scheduled trip with multiple observations, it also shows averages.

Passenger Volumes At SCReenlines (PVASCR) [Figure 11]

The PVASCR report lists the number of passengers by route crossing a defined screenline by the five major time periods and by 15-minute intervals during the AM and PM peak hours. Screenline volumes are reported only on screenline sections that have trips captured. However, reported volumes are factored up to represent all trips scheduled crossing that screenline section.

Passenger Activity By AREa (PABARE) [Figure 12]

The PABARE report shows passenger boarding and alighting by bus statistics (mean, median, minimum, maximum, standard deviation) in a defined geographic area. Planners define a geographic area as a set of bus stops and direction of service.

Overall Service Provision (OSP) [Figure 13]

The OSP report calculates a series of performance measures for all routes in the system, including percentage of scheduled trips, revenue-hours, and revenue-km captured by APC; passenger boardings, passenger-km, passenger-hours, and peak load; and interesting ratios such as average passenger trip length, average passenger-km per vehicle-km, and average operating cost per passenger-km. These performance measures reflect the estimated total service in the transit property. The service provision statistics may be requested by route, in which case only statistics for each of the routes by direction and time period will be reported. If all routes are requested, the statistics will reflect all routes, but broken down by route class.

Service Performance Report (SPR)

The SPR reports (with over 30 separate reports) rank the service by four key indicators:

- passenger kilometers per revenue kilometer,
- passenger boardings,
- revenue-cost ratio, and
- on-time performance.

The ranking information is broken down by route class and time period. Routes that perform in the bottom 20% of all routes are marked.

Common Report Uses

The list of possible uses of APC data is quite long. At OC Transpo, the APC system is the primary source of all planning data, and is constantly used in the decision-making process. Some of the more common applications include:

Service Performance

The system measures the performance of all routes by route category, route and time of day in terms of average load, on-time rating and revenue-cost ratio. These performance indicators are ranked from best to worst so that high and low performing routes are identified.

Individual Route Analysis

If problems are reported about a route experiencing overloads and/or schedule adherence problems, APC data is reviewed to determine the extent and particulars of the problem. With this information, solutions to remedy the situation can be determined.

Annual Service Planning Process

OC Transpo coordinates all route changes through an annual service plan review process that is based upon an extensive public participation program. Many of the proposed changes are identified as a result of analyzing the service performance family of reports. All proposed changes are presented to the public and politicians for comment and are largely supported by APC data.

Shelter Location Analysis

Each year OC Transpo installs a number of new bus stop shelters. The long list of shelter location applications is prioritized by analyzing the APC boarding volumes at these stops.

Complaint Program

Public complaints concerning individual route performance are reviewed using APC data. APC stop-by-stop boarding/alighting data is reviewed to determine passenger usage by stop in the affected area.

Special Event Passenger Counting

The APC system is used to monitor passenger volumes of extra services operated for special events such as the three week-long Ottawa winter carnival known as Winterlude.

Ridership Control Totals for Surveys

The system provides the ridership control total figure, which is needed whenever a system-wide or area-specific on-board origin-destination survey is conducted.

The above illustrates the wide range of utilizations that OC Transpo has made of the data produced by their APC system.

INNOVATIVE PRACTICES FACILITATED BY APC DATA

Transit Priority Task Force

In 1994, a Task Force was created in the Ottawa region consisting of staff from the Environment and Transportation Department, the Planning and Development Department, the Traffic Operations/Signals Branch, and OC Transpo. Its mandate is to investigate and implement transit priority measures, and has implemented a large number of priority measures throughout the City, including bus lanes, transit ramp off/on connections from freeways, bus queue jump lanes, bus actuated signals, transit signal phases, transit signal priority, etc. The primary source of information for identifying where such measures are required is derived from the above APC system vehicle time utilization reports. This credible source of systematic data is useful in negotiations with traffic staff to determine priorities for transit priority treatment. With the assistance of this source of data, the Ottawa region has become one of the most sophisticated and extensive implementers of transit priority in North America.

Schedule Workshops for Rewriting Schedules

Different transit systems with APC systems have developed different approaches for the utilization of running time data for modifying schedules. Part of the difficulty lies in the fact that schedules are affected on the one hand by external traffic conditions and the occurrence of congestion that is sometimes predictable and sometimes random, and on the other hand by the driving behavior of bus operators. In addition, there is always a constant dichotomy between the twin objectives of reducing operating costs (achieved by tightly matching schedules to feasible running times) and ensuring customer and bus operator satisfaction through greater on-time performance. This dichotomy of objectives affects different departments in a transit system, including scheduling, service planning, and operations, each of which may perceive and react to the objectives and factors affecting their achievement differently. The debate about defining the running times for creation of the schedule is at the crux of this discussion, and often finds the different actors, management versus union, and different departments in conflict over running times and schedule design.

OC Transpo has developed an innovative format for addressing this issue through convening Schedule Workshops for specific routes. The establishment of the Scheduling Workshop process originated as a result of a comprehensive transit system and management review conducted in 1998. The objectives of the Scheduling Workshops are to: 1) determine viable running times that provide reliable service to the customers and can be maintained by operating staff under normal conditions, and 2) review and make recommendations regarding operational issues and customer needs that improve the overall level of service and schedule. Recommendations can take the form of transit priority suggestions, bus stop location changes, and/or minor re-routings.

The composition of the workshop includes bus operators who have recently worked the route under review, Transit Supervisors and Transit Schedulers, as well as union representatives, and a Facilitator. The Workshop focuses on a single route at a time, and lasts four to six weeks. The first step is to review the current schedule as well as the running times by section and time period as measured by the APC system. Second, participants review the route, time-point by time-point, and raise operational concerns or improvements for discussion. Third, fieldwork helps to review each segment's running times and operational issues. Finally, the Committee discusses and creates new running time tables, as well as recommendations for improvement, which are captured through a formal Scheduling Workshop Report. Four such workshops are typically held a year. Figure 14 shows a sample page from the Scheduling Workshop Report. Though labor-intensive, these workshops have helped to address some fundamental issues revolving around the defining of running times for schedules, in a difficult labor relations environment. These Scheduling Workshops rely significantly on the data produced by the APC system.

USING APC-GENERATED DEADHEADING RUNNING TIMES FOR INTERLINED SERVICES

Another unique aspect of OC Transpo's APC system relates to their extensive application of interlining during peak periods. As a result of the concentration of government administrative offices in Ottawa, Canada's capital, and the large penetration of OC Transpo within the government office commuting market, OC Transpo has a high peak-to-base ratio.

Over the years, it has developed an extensive network of peak hour-only regular and express services. In order to provide these services as cost-effectively as possible, it has become an aggressive implementer of interlined schedules for these services: over 50% of AM peak vehicle resources are interlined! The APC system's Deadhead Running Time Report (discussed earlier as Figure 5) has proven a valuable tool for determining deadheading times between the end and start termini of interlined services. This is a critical input to the scheduling system to ensure maximum cost-effectiveness of interlining.

CONCLUSIONS

The OC Transpo case study illustrates the extensive use of APC data made by one of the pioneers in the field. OC Transpo staff and management are intensive users of data, primarily derived from the APC system, for analysis that feeds service planning and scheduling. This analytic approach to decision making has been a significant factor in helping this transit system, serving a population of 700,000 to simultaneously achieve high market penetration (with over 118 million unlinked trips in 2001 for a fleet of 930 vehicles) and high cost-effectiveness (with a 57% operating revenue-cost ratio). The transit system will continue to explore new applications of data as it seeks to upgrade its current APC and AVL systems.

REFERENCES

1. Joel Koffman. Automatic Passenger Counting Data; Better Schedules Improve On-Time Performance. OC Transpo, August 1990.
2. OC Transpo. APC at OC Transpo: A New Dimension. OC Transpo, 1986.
3. OC Transpo. Route Scheduling Workshop, Final Report, Route 111. OC Transpo, April 2002.
4. OC Transpo. Using APC Data To Improve Schedule Performance. OC Transpo, February 1993.

FIGURE 1: APC Time Utilization Logs and Logic

Sample OC Transpo Report

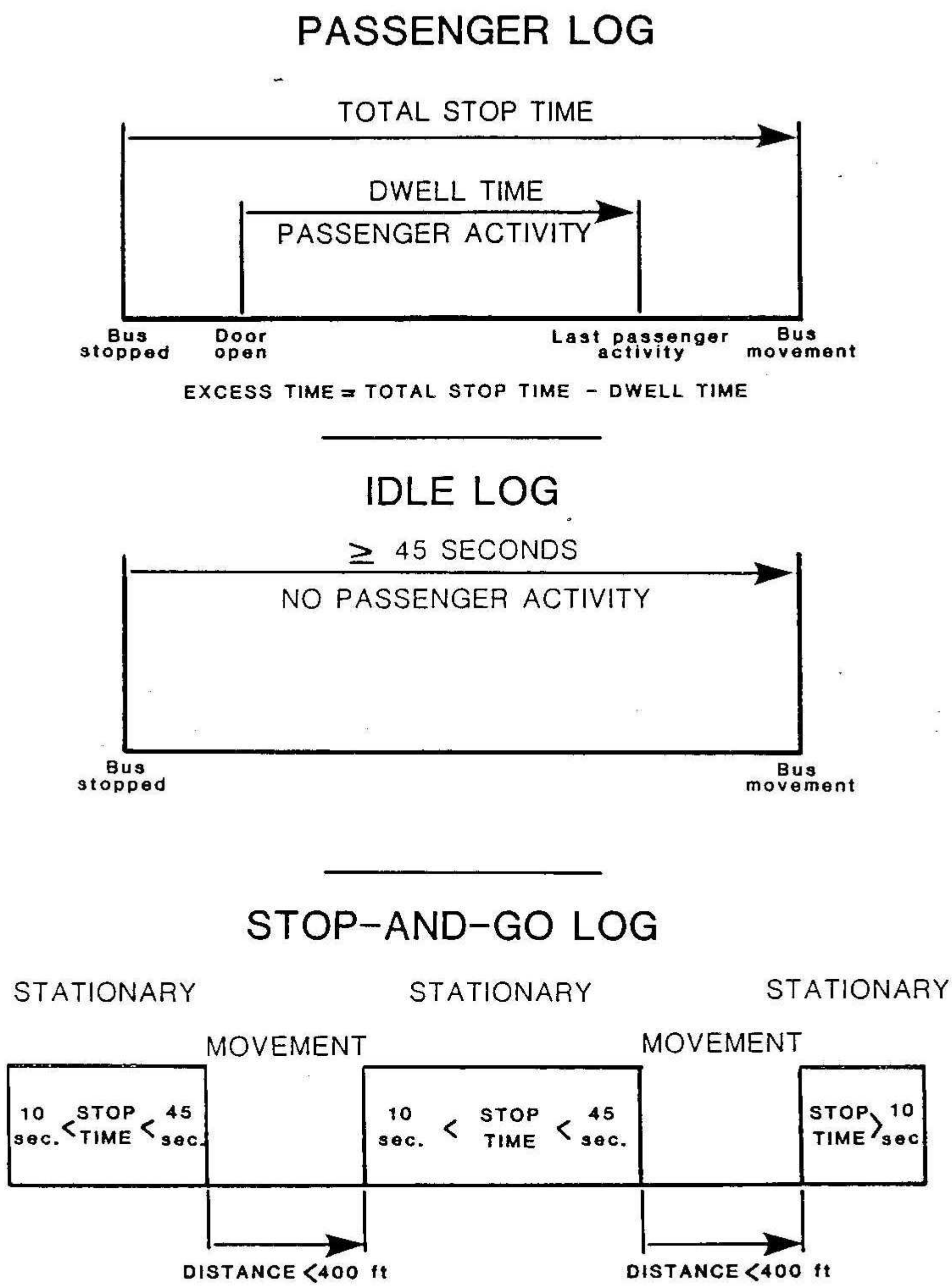


FIGURE 2: Revenue Time Utilization – Graphic Representation

Sample OC Transpo Report

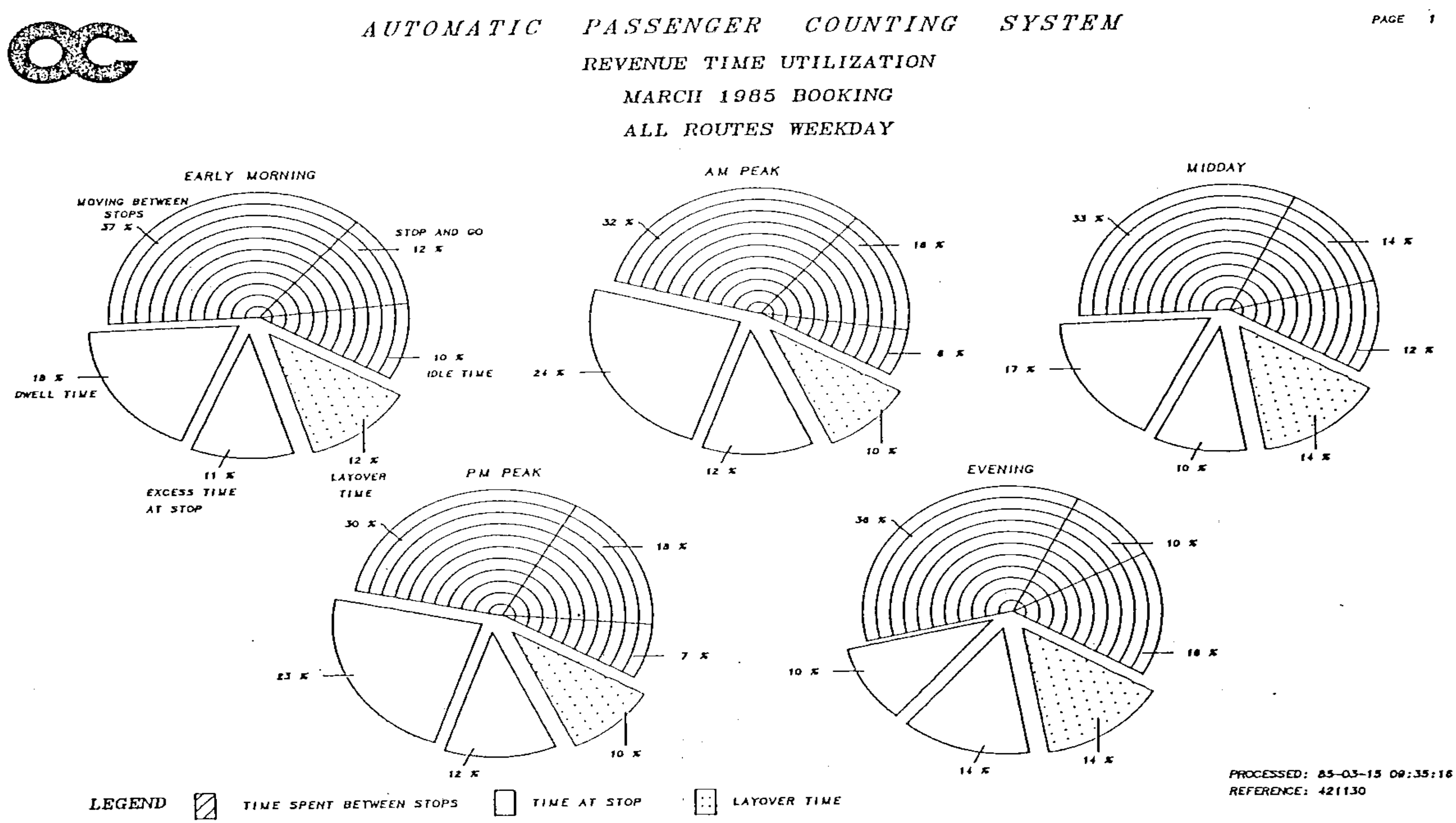


FIGURE 3: Summary of Running Times (SORT)

Sample OC Transpo Report

PROCESSED: 2001-08-27 10:01:19

AUTOMATIC PASSENGER COUNTING SYSTEM
SUMMARY OF RUNNING TIMES
WEEKDAY SERVICE

BOOKING: JAN00
ROUTE: 1 SOUTH KEYS-DOWNTOWN/CENTREVILLE — PERIOD: 2000-01-04 TO 2000-04-22
DIRECTION: 1 NBND Pattern 1 — DAY OF THE WEEK: WEEKDAY

	GREN STAT to S.K. STAT			S.K. STAT to BANK JOHN			BANK JOHN to B.B. TERM			B.B. TERM to BANK GROV			BANK GROV to LANS PARK		
TIME PERIOD	Sch.	Mean	Obs	Sch.	Mean	Obs	Sch.	Mean	Obs	Sch.	Mean	Obs	Sch.	Mean	Obs
04:01-05:00	1.0	1.4	2	3.0	4.1	1	7.0	9.5	1	3.0	3.5	1	1.0	1.0	1
05:01-06:00	1.0	1.0	4	3.0	3.9	5	7.0	6.9	5	3.0	2.8	5	1.0	1.2	5

FIGURE 4: Distribution of Running Times (DORT)

Sample OC Transpo Report

PROCESSED: 2002-07-30 15:34:14

AUTOMATIC PASSENGER COUNTING SYSTEM
DISTRIBUTION OF RUNNING TIMES BY SECTION
WEEKDAY SERVICE

BOOKING: SEP01
PERIOD: 2001-09-02 TO 2001-12-15
DAY OF WEEK: MON TUE WED THU FRI

ROUTE(S): 1
DIRECTION: N/A
Pattern(s): ALL
STOPS: CD935 TO RA910

							RUNNING TIME (MINUTES)																		
TIME PERIOD	SCHED TIME	MEAN	MED.	STDDEV	AVG SPEED	NO. OBS	12	14	16	18	20	22	24	26	28	30	32	34	36	00	00	00	00	00	0
00:01-01:00	15.0	14.8	14.5	2.1	22.2	3	1	1		1															
05:01-06:00	14.0	13.5	13.5	0.8	24.0	3	1	2																	
06:01-07:00	20.3	18.0	18.2	1.6	18.1	11			4	3	4														
07:01-08:00	21.2	19.6	19.4	1.4	16.6	13				5	5	3													
08:01-09:00	23.0	24.8	25.3	1.7	13.1	12						1	5	4	2										
09:01-10:00	23.0	20.9	20.6	2.1	15.6	10				1	6	1	2												
10:01-11:00	23.0	21.6	22.0	1.6	15.0	6					3	2	1												
11:01-12:00	23.0	21.6	21.6	1.7	15.1	10				1	3	3	3												
12:01-13:00	23.1	23.1	23.3	2.8	14.2	11				1	1	3	3	2	1										
13:01-14:00	24.0	21.4	21.0	2.2	15.3	12				2	4	3	3												
14:01-15:00	24.0	24.9	24.7	2.3	13.1	13						3	6	2	1	1									
15:01-16:00	24.0	23.5	23.5	3.0	14.0	20				1	2	6	5	4	1	1									
16:01-17:00	24.0	25.7	25.5	2.2	12.7	17						1	5	6	4	1									
17:01-18:00	24.0	27.1	27.9	3.9	12.2	16					1	2	2	3	3	2	2	1							
18:01-19:00	22.3	28.3	33.4	8.1	12.2	4						2						1	1						
19:01-20:00	21.0	20.6	19.9	2.0	15.8	3					2	1													
20:01-21:00	20.5	19.1	19.4	0.5	16.9	2				1	1														
21:01-22:00	19.5	19.0	18.8	1.0	17.0	4				3	1														
22:01-23:00	16.0	18.4	18.6	0.2	17.6	2				2															
23:01-24:00	15.0	15.2	15.7	1.9	21.5	3		1	2																

FIGURE 5: Deadheading Running Times (DHRT)

Sample OC Transpo Report

PROCESSED: 2001-08-01 10:20:15

AUTOMATIC PASSENGER COUNTING SYSTEM
DEADHEAD RUNNING TIMES
WEEKDAY SERVICE

BOOKING: JAN00
PERIOD: 2000-01-04 TO 2000-04-22
DAY TYPE: WEEKDAY
TIME: ALL DAY

ROUTE: 1 SOUTH KEYS-DOWNTOWN/CENTREVILLE

FROM NODE		TO NODE		SCHEDULED TIME	MEAN TRAVEL TIME	DIFFERENCE	STDDEV	NO OBS
4501 HERON	RIVERSIDE	6300 BILLINGS	BR.TERMINA	12	9	-3	0.0	1
5575 NICHOLAS	TIME POINT	4815 INNES	ORLEANS	29	29	0	3.7	3
5575 NICHOLAS	TIME POINT	9001 ST.LAUREN	T NORTH	13	12	-1	0.0	1
5575 NICHOLAS	TIME POINT	9005 MERIVALE		26	11	-15	0.0	1
7080 HURDMAN	STATION	6300 BILLINGS	BR.TERMINA	11	13	2	0.0	1
7330 GREENBORO	STATION	9005 MERIVALE		15	31	16	38.2	15
9005 MERIVALE		5575 NICHOLAS	TIME POINT	26	31	5	4.4	3
9005 MERIVALE		6300 BILLINGS	BR.TERMINA	13	15	2	2.8	3
9005 MERIVALE		7330 GREENBORO	STATION	16	25	9	9.9	10

FIGURE 6: Revenue Time Utilization (RMOC)

Sample OC Transpo Report

ROUTE	RUN	FROM_STOP	TO_STOP	DATE	ACT_DEP_TIME	SCHED_DEP	MOVING_TIME	STOPGO_TIME	IDLE_TIME	DWELL_TIME	EXCESS_TIME	TOTAL_TIME	SCHEDULED_TIME	DISTANCE	LOAD
1	1-1A	CD155	CA080	3/6/00	6:05:15	6:05:00	0:03:03	0:01:15	0:00:00	0:00:01	0:00:10	0:03:03	0:05:00	1.4	0
1	1-1A	CD155	CA080	3/7/00	6:06:45	6:05:00	0:04:09	0:00:00	0:00:00	0:00:02	0:00:34	0:04:09	0:05:00	1.4	0
1	1-2A	CD155	CA080	4/10/00	6:21:45	6:20:00	0:04:23	0:00:00	0:00:00	0:00:07	0:00:16	0:04:23	0:06:00	1.4	1
1	1-3A	CD155	CA080	4/11/00	6:36:30	6:35:00	0:03:11	0:00:00	0:00:00	0:00:44	0:00:05	0:03:11	0:06:00	1.4	5
1	1-3A	CD155	CA080	4/13/00	6:36:45	6:35:00	0:01:25	0:00:00	0:00:00	0:00:03	0:00:31	0:01:25	0:06:00	1.4	1
1	1-4A	CD155	CA080	3/14/00	6:49:45	6:50:00	0:03:26	0:00:00	0:00:00	0:00:21	0:00:33	0:03:26	0:06:00	1.4	2
1	1-4A	CD155	CA080	4/14/00	6:52:00	6:50:00	0:02:31	0:00:00	0:01:00	0:00:11	0:00:10	0:02:31	0:06:00	1.4	4
1	1-5A	CD155	CA080	1/5/00	7:05:45	7:05:00	0:01:55	0:00:45	0:01:39	0:00:11	0:00:30	0:03:29	0:06:00	1.4	3
1	1-5A	CD155	CA080	3/14/00	7:04:15	7:05:00	0:02:55	0:00:45	0:00:50	0:00:05	0:00:12	0:03:40	0:06:00	1.4	1
1	1-7A	CD155	CA080	1/18/00	7:36:00	7:35:00	0:03:32	0:00:00	0:00:49	0:00:21	0:00:09	0:03:32	0:06:00	1.4	5
1	1-7A	CD155	CA080	4/13/00	7:35:30	7:35:00	0:02:19	0:01:30	0:00:56	0:00:09	0:00:07	0:02:19	0:06:00	1.4	3
1	1-7A	CD155	CA080	4/14/00	7:35:30	7:35:00	0:03:44	0:00:00	0:00:00	0:00:13	0:00:11	0:03:44	0:06:00	1.4	2
1	1-7A	CD155	CA080	4/19/00	7:35:15	7:35:00	0:03:38	0:01:30	0:00:00	0:00:11	0:00:09	0:03:38	0:06:00	1.4	3
1	1-1A	CD155	CA080	3/6/00	7:53:16	7:50:00	0:03:09	0:00:00	0:00:00	0:00:17	0:01:03	0:03:09	0:06:00	1.4	4
1	1-1A	CD155	CA080	3/7/00	7:47:45	7:50:00	0:04:07	0:00:00	0:01:08	0:00:37	0:00:35	0:04:52	0:06:00	1.4	11
1	1-2A	CD155	CA080	4/10/00	8:05:30	8:05:00	0:03:19	0:00:00	0:00:00	0:00:26	0:00:16	0:03:19	0:06:00	1.4	7
1	1-3A	CD155	CA080	4/11/00	8:24:14	8:20:00	0:00:44	0:05:00	0:01:17	0:00:03	0:00:12	0:04:29	0:06:00	1.4	2
1	1-3A	CD155	CA080	4/13/00	8:22:59	8:20:00	0:03:05	0:00:00	0:00:52	0:00:22	0:01:12	0:03:57	0:06:00	1.4	2
1	1-4A	CD155	CA080	3/14/00	8:39:00	8:35:00	0:00:53	0:01:30	0:01:02	0:00:25	0:00:26	0:02:43	0:06:00	1.4	6
1	1-4A	CD155	CA080	4/14/00	8:34:15	8:35:00	0:02:44	0:01:15	0:00:54	0:00:27	0:00:32	0:03:59	0:06:00	1.4	10
1	1-5A	CD155	CA080	1/5/00	8:49:00	8:50:00	0:04:15	0:00:00	0:01:48	0:01:19	0:00:53	0:06:03	0:06:00	1.4	3
1	1-5A	CD155	CA080	3/14/00	8:52:01	8:50:00	0:01:19	0:00:00	0:00:47	0:00:00	0:00:00	0:01:19	0:06:00	1.4	10
1	1-7A	CD155	CA080	1/18/00	9:21:14	9:20:00	0:02:19	0:01:00	0:00:00	0:00:41	0:01:01	0:03:52	0:06:00	1.4	8
1	1-7A	CD155	CA080	4/13/00	9:24:45	9:20:00	0:01:58	0:00:00	0:00:00	0:00:03	0:00:10	0:01:58	0:06:00	1.4	11

FIGURE 7: Running Time Utilization by Section (RTUBSEC)

Sample OC Transpo Report

```
PROCESSED: 2001-08-22 10:00:43
                                                     AUTOMATIC PASSENGER COUNTING SYSTEM
                                                        REVENUE TIME UTILIZATION
                                                               WEEKDAY SERVICE
                                                                                           BOOKING: JAN00
FROM STOP: CA080 BANK QUEEN NS                                                              PERIOD: 2000-01-04 TO 2000-04-22
TO STOP:   RF330 SOUTHGATE 3308 FS                                                        DAY TYPE: WEEKDAY
ROUTE:     1 SOUTH KEYS-DOWNTOWN/CENTREVILLE                                                  TIME: 06:30-09:30
```

	Average Time per Trip Minutes	Std Dev	% of Time
Moving between stops	16.18	1.77	55.87
Stop and go time	2.79	2.02	9.63
Idle time	1.34	0.81	4.62
Dwell time	3.07	1.23	10.61
Excess time	5.58	0.78	19.28
TOTAL	28.96		100%

Layover time	0.00
Average sched time per trip	29.55
Total distance (KM)	9.22
Average moving speed (KM/HR)	34.20
Average total speed (KM/HR)	19.10
Total trips captured:	20

FIGURE 8: Route Direction Summary (RDSUM)

Sample OC Transpo Report

PROCESSED: 2001-08-15 10:31:45

AUTOMATIC PASSENGER COUNTING SYSTEM
ROUTE DIRECTION SUMMARY
WEEKDAY SERVICE

BOOKING: JAN00
PERIOD: 2000-01-04 TO 2000-04-22
ROUTE: 1 SOUTH KEYS-DOWNTOWN/CENTREVILLE
TIME: ALL DAY
DIRECTION: 0 SBND
DAY OF THE WEEK: WEEKDAY
PATTERN: 3

RUN	TERMINAL NODES START	TERMINAL NODES END	SCHEDULED DEPAR	SCHEDULED ARRIV	SCHEDULED R/T	ACTUAL DEPAR	ACTUAL ARRIV	ACTUAL R/T	ADHERENCE DEPAR	ADHERENCE ARRIV	ADHERENCE R/T	TOTAL ONS	R/C RATIO	MAX LOAD	MAX LOAD POINT	DATE OF OBSERVATION
1001A	NICH T.P.	GREN STAT	06:05	06:43	38	06:05	06:36	31	0	- 7	- 7	16	24.7	13	CF520 BANK GROVE FS	2000-03-06 MON
						06:07	06:36	29	2	- 7	- 9	16	24.7	12	CF700 BANK GLEBE NS	2000-03-07 TUE
						*AVERAGE:		30	1	- 7	- 8	16	24.7	13 (2)		
1002A	NICH T.P.	GREN STAT	06:20	07:03	43	06:22	06:58	36	2	- 5	- 7	22	19.1	15	CH090 BANK MCLEOD NS	2000-04-10 MON
						*AVERAGE:		36	2	- 5	- 7	22	19.1	15 (1)		
1003A	NICH T.P.	GREN STAT	06:35	07:18	43	06:37	07:14	37	2	- 4	- 6	41	35.6	15	CH090 BANK MCLEOD NS	2000-04-11 TUE
						06:37	07:14	37	2	- 4	- 6	16	13.9	10	CF520 BANK GROVE FS	2000-04-13 THU
						*AVERAGE:		37	2	- 4	- 6	29	24.8	13 (2)		
1004A	NICH T.P.	GREN STAT	06:50	07:33	43	06:50	07:30	40	0	- 3	- 3	41	35.6	19	CF620 BANK HOLMWOOD FS	2000-03-14 TUE
						06:52	07:27	35	2	- 6	- 8	60	52.1	28	CH090 BANK MCLEOD NS	2000-04-14 FRI
						*AVERAGE:		38	1	- 5	- 6	51	43.9	24 (2)		
1005A	NICH T.P.	GREN STAT	07:05	07:48	43	07:06	07:43	37	1	- 5	- 6	42	36.5	27	CF500 BANK CHESLEY FS	2000-01-05 WED
						07:04	07:45	41	- 1	- 3	- 2	37	32.1	17	RA940 BILLINGS BRIDG 4B	2000-03-14 TUE
						*AVERAGE:		39	0	- 4	- 4	40	34.3	22 (2)		

FIGURE 9: Passenger Activity and Load by Stop (PALBSTP)

Sample OC Transpo Report

```
PROCESSED: 2001-07-18 15:34:09
                                                  AUTOMATIC PASSENGER COUNTING SYSTEM
                                                  PASSENGER ACTIVITY AND LOAD BY STOP
                                                            WEEKDAY SERVICE
                                                                                          BOOKING: JAN00
ROUTE:         1 SOUTH KEYS-DOWNTOWN/CENTREVILLE                                           PERIOD: 2000-01-04 TO 2000-04-22
DIRECTION:     0 SBND                                                                        TIME: ALL DAY
BSI STOP :     From  CD155
               To    CA080
RUN:          ALL                                                                 DAY OF THE WEEK:  MON TUE WED THU FRI
PATTERN: ALL
```

BSI STOP	LOCATION	ACTIVITY	MINIMUM	MAXIMUM	MEDIAN	AVERAGE	STD DEV	OBSERVED	STOPS	% SAMPLED	TOTAL ACTIVITY
CD155	NICHOLAS RIDEAU FS	ONS	0	15	1	1.7	3.0	53	28	87	102
		OFFS	0	0	0	0.0	0.0	53	28	87	0
		LOAD	0	15	1	1.7	3.0	53	28	87	
CD935	RIDEAU 4A	ONS	0	27	9	9.6	5.9	55	54	86	616
		OFFS	0	6	0	0.5	1.3	55	54	86	31
		LOAD	0	30	11	10.8	6.1	55	54	86	
CB030	WELLINGTON #3 CHATEAU	ONS	0	1	1	0.1	0.2	55	3	86	3
		OFFS	0	1	0	0.0	0.1	55	3	86	1
		LOAD	0	30	11	10.8	6.1	55	3	86	
CB050	WELLINGTON METCALFE#1 NS	ONS	0	16	1	0.8	2.2	55	27	86	51
		OFFS	0	4	0	0.2	0.6	55	27	86	10
		LOAD	0	39	9	10.3	7.4	55	27	86	
CB080	WELLINGTON 1ST OCONNOR NS	ONS	0	1	1	0.2	0.4	55	17	86	10
		OFFS	0	1	0	0.1	0.3	55	17	86	8
		LOAD	0	39	10	10.7	7.3	55	17	86	
CA080	BANK QUEEN NS	ONS	0	10	1	1.0	2.1	55	21	86	64
		OFFS	0	8	0	0.4	1.2	55	21	86	24
		LOAD	0	42	11	11.7	7.9	55	21	86	

FIGURE 10: Point Check Report (PTCHECK)

Sample OC Transpo Report

PROCESSED: 2001-08-14 15:47:17

AUTOMATIC PASSENGER COUNTING SYSTEM
POINT CHECK
WEEKDAY SERVICE

BOOKING: JAN00
STOP: CA100 BANK SLATER NS
PERIOD: 2000-01-04 TO 2000-04-22
ROUTE: 1 SOUTH KEYS-DOWNTOWN/CENTREVILLE
TIME: ALL DAY
DIRECTION: 1 NBND
DAY OF THE WEEK: MON TUE WED THU FRI

RUN	PAT	TIME EXPECTED	TIME ACTUAL	DIFFERENCE	ONS	OFFS	LOAD AT DEPARTURE	DATE
1001A	3	22:32	22:34	2	0	0	2	2000-03-07 TUE
			AVERAGE	3	0	1	4 (2)	
1002A	3	23:00	N/A					
1020P	3	23:30	23:30	0	0	4	16	2000-02-21 MON
1001A	3	24:00	24:03	3	0	0	2	2000-03-06 MON
1002A	3	24:30	N/A					
86004A	1	05:32	05:32	0	1	3	1	2000-03-18 SAT
			05:30	-2	2	5	5	2000-03-25 SAT
			05:32	0	0	3	0	2000-04-01 SAT
			AVERAGE	-1	1	4	2 (3)	
7001A	1	06:02	06:02	0	0	0	0	2000-02-05 SAT
118003A	4	06:22	06:20	-2	0	1	2	2000-04-08 SAT
118001A	4	06:52	06:50	-2	0	0	1	2000-02-19 SAT
118004A	4	07:22	07:22	0	0	3	8	2000-02-26 SAT
			07:21	-1	0	1	4	2000-03-11 SAT
			07:22	0	1	4	5	2000-03-25 SAT
			AVERAGE	0	0	3	6 (3)	

FIGURE 11: Passenger Volume at Screenlines (PVASCR)

Sample OC Transpo Report

```
PROCESSED: 2001-02-06 15:27:55
                                                   AUTOMATIC PASSENGER COUNTING SYSTEM
                                                PASSENGER VOLUMES AT SCREENLINES BY ROUTE
                                                      ALL ROUTES WEEKDAY BY TIME PERIOD
BOOKING: JAN00                                                                                         DAY OF THE WEEK:  MON TUE WED THU FRI
SCREENLINE: SL10  BILLINGS

                                                                 I N B O U N D
```

HOUR ROUTE	1ST STOP	2ND STOP	1ST STOP DESCRIPTION	03:01-06:30	06:31-09:30	09:31-15:30	15:31-18:30	18:31-03:00	24 TOTAL
1	RA020	RA910	BANK RIVERSIDE FS	30	245	606	369	205	1455
5	RA020	RA910	BANK RIVERSIDE FS		126	133	129	61	449
			TOTAL	30	371	739	498	266	1904

NOTE: a blank field indicates that no service is scheduled in the time period

FIGURE 12: Passenger Activity by Area (PABARE)

Sample OC Transpo Report

PROCESSED: 2001-09-18 15:09:22

AUTOMATIC PASSENGER COUNTING SYSTEM
PASSENGER ACTIVITY BY AREA BY TIME PERIOD
SUNDAY SERVICE

BOOKING: JAN00
AREA: GA03 From Ottawa PERIOD: 2000-01-04 TO 2001-10-01
DAY TYPE: SUNDAY

		03:01-06:30		06:31-09:30		09:31-15:30		15:31-18:30		18:31-03:00		TOTAL	
ROUTE	STATISTICS	ONS	OFFS	ONS	OFFS	ONS	OFFS	ONS	OFFS	ONS	OFFS	ONS	OFFS
8	MINIMUM	0	0	0	0	0	0	0	0	0	0	0	0
	MAXIMUM	0	0	1	1	3	3	2	5	2	3	3	5
	MEDIAN	0	0	0	0	1	1	1	1	0	1	0	1
	AVERAGE	0.0	0.0	0.14	0.14	1.10	0.72	0.69	1.00	0.33	0.73	0.65	0.70
	STD DEV	0.0	0.0	0.38	0.38	1.04	0.88	0.62	1.42	0.52	0.68	0.83	0.89
	OBSERVED	0	0	7	7	21	21	9	9	19	19	56	56
	% SAMPLED	0	0	88	88	88	88	75	75	90	90	86	86
	TOTAL ACT.	0	0	1	1	26	17	8	12	7	15	42	45
	SCHEDULED	0	0	8	8	24	24	12	12	21	21	65	65
TOTAL	MINIMUM	0	0	0	0	0	0	0	0	0	0	0	0
	MAXIMUM	0	0	1	1	3	3	2	5	2	3	3	5
	MEDIAN	0	0	0	0	1	1	1	1	0	1	0	1
	AVERAGE	0.0	0.0	0.14	0.14	1.10	0.72	0.69	1.00	0.33	0.73	0.65	0.70
	STD DEV	0.0	0.0	0.38	0.38	1.04	0.88	0.62	1.42	0.52	0.68	0.83	0.89
	OBSERVED	0	0	7	7	21	21	9	9	19	19	56	56
	% SAMPLED	0	0	88	88	88	88	75	75	90	90	86	86
	TOTAL ACT.	0	0	1	1	26	17	8	12	7	15	42	45
	SCHEDULED	0	0	8	8	24	24	12	12	21	21	65	65

FIGURE 13: Overall Service Provision (OSP)

Sample OC Transpo Report

PROCESSED: 2001-08-14 13:08:36

AUTOMATIC PASSENGER COUNTING SYSTEM
OVERALL SERVICE PROVISION
ALL ROUTE SUNDAY

BOOKING: JAN00

21:31-03:00

SERVICE CHARACTERISTICS	TRANSITWAY	BASE	CROSSTOWN	REGULAR LOCAL	TOTAL	EXPERSS	INTERLINE LTD STOP	TOTAL	GRAND TOTAL
# RUNS SCHEDULED	16	49	19	23	41	0	6	6	102
# SCH ONE-WAY TRIPS	58	114	34	58	92	0	12	12	276
% ONE-WAY TRIPS CAPTURED	66	48	56	45	49	0	92	92	54
# ADDITIONAL TRIPS CAPTURED	16	32	11	10	21	0	4	4	73
# SCH REVENUE KMS	1912	2194	686	658	1344	0	76	76	5526
% SCH REVENUE KMS CAPURED	61	45	50	38	44	0	48	48	50
# SCH INTERTRIP DEADHEAD KMS	1	43	7	77	84	0	24	24	151
# SCH GARAGE DEADHEAD KMS	262	530	116	239	355	0	61	61	1209
# SCH VEHICLE KMS	2175	2766	809	975	1783	0	162	162	6886
# SCH REVENUE HRS	46.8	81.9	28.4	23.3	51.7	0.0	2.7	2.7	183.1
% SCH REVENUE HRS CAPURED	61	44	51	45	48	0	73	73	50
# SCH INTERTRIP DEADHEAD HRS	1.0	3.4	1.6	6.1	7.7	0.0	1.6	1.6	13.7
# SCH GARAGE DEADHEAD HRS	5.9	14.4	4.3	5.7	9.9	0.0	1.5	1.5	31.6
# SCH LAYOVER HRS	10.5	22.4	5.1	5.0	10.1	0.0	1.5	1.5	44.5
# SCH VEHICLE HRS	64.1	122.0	39.4	40.1	79.4	0.0	7.3	7.3	273.0
AVE REVENUE SCHEDULED SPEED	40.9	26.8	24.1	28.3	26.0	0.0	27.7	27.7	30.2
AVE OVERALL SCHEDULED SPEED	33.9	22.7	20.5	24.3	22.4	0.0	22.0	22.0	25.2
TOTAL # UNLINKED PASSENGERS	2527	3125	828	384	1298	0	42	42	7107
TOTAL # PASSENGER KMS	22071	15304	4048	1312	5532	0	222	222	44571
TOTAL # PASSENGER HRS	550	608	166	44	215	0	7	7	1417
AVE PASSENGER TRIP LENGTH	8.7	4.9	4.9	3.4	4.3	0.0	5.3	5.3	6.3
AVE PASSENGER TRIP TIME	13:04	11:40	12:02	06:53	09:57	00:00	09:39	09:39	11:58
AVE MAXIMUM PASSENGER LOAD	14	7	7	2	4	0	3	3	7
AVE PASSENGER KM PER REV KM	11.5	7.0	5.9	2.0	4.1	0.0	2.9	2.9	8.1
AVE PASSENGER HR PER REV HR	11.8	7.4	5.8	1.9	4.2	0.0	2.5	2.5	7.7
AVE PASSENGER KM PER VEH KM	10.1	5.5	5.0	1.3	3.1	0.0	1.4	1.4	6.5
AVE PASSENGER HR PER VEH HR	8.6	5.0	4.2	1.1	2.7	0.0	0.9	0.9	5.2
TOTAL OPERATING EXPENSE	4223	7394	2566	2100	4667	0	247	247	16530
AVE COST PER PASSENGER KM	0.19	0.48	0.63	1.60	0.84	0.00	1.11	1.11	0.37
AVE COST PER PASSENGER HR	7.67	12.17	15.45	47.64	21.67	0.00	36.68	36.68	11.67
AVE COST PER UNLINKED PASS	1.67	2.37	3.10	5.46	3.59	0.00	5.90	5.90	2.33
# TRIPS EXCEEDING SEATED LD	0	0	0	0	0	0	0	0	0

FIGURE 14: Route Scheduling Workshop Report

Sample OC Transpo Report

ROUTE 111 - DAILY
RUNNING TIME BY SECTION
EASTBOUND DIRECTION

Section: Lincoln Fields to Baseline Station	**Eastbound**			
	Scheduled	Data Avg.	Recommend	Comments
Early Morning < 06:00	4	4	4	
A.M. Peak 06:00 to 09:00	4	4-5	4	
Midday 09:00 to 15:00	5	4	4	
P.M. Peak 15:00 to 18:00	5	4-5	5	
Early Evening 18:00 to *22:00*	5	4-5	5	
Late Evening > *22:00*	4-5	4	4	

Section: Baseline Station to Meadowlands/Merivale	**Eastbound**			
	Scheduled	Data Avg.	Recommend	Comments
Early Morning < 06:00	5	6	5	
A.M. Peak 06:00 to 09:00	11-13	9-11	11	
Midday 09:00 to *15:30*	12-14	11	11-12	9.00 - 12.00 = 11 12.00 - 15.30 = 12
P.M. Peak *15:30* to *18:30*	13	10-12	13	
Early Evening *18:30* to 21:00	12	10	12	
Late Evening > 21:00	11	9	10-11	21.00 - 22.00 = 11 >22.00 = 10

APPENDIX G: Hermes-Eindhoven's Experience with Automatic Data Collection, Operations Control, Management Information, and Passenger Information Systems[*]

Eindhoven is the largest city in the province North-Brabant in the Netherlands with about 200,000 inhabitants. The population of the metropolitan area and the region focused on Eindhoven are 300,000 and 700,000, respectively. Two private transit companies, subsidized and regulated by local and regional governments, offer bus service in Eindhoven. The BBA operates regional lines to the west, and Hermes operates city bus lines and regional lines to the east (see Figure 1).

Hermes runs several types of bus lines in the southwest part of the Netherlands. They include the *'city bus'* (a high frequency connection during day hours between city neighborhoods and the center), the *'region bus'* (a more or less direct connection between rural towns and larger cities), the *'neighborhood bus'* (a demand responsive service following a more or less fixed route connecting small rural communities and providing transfers opportunities to city and regional lines), the *'night bus,'* and the *'Interliner'* (a high quality, fast, long distance route with a limited number of stops connecting important railroad stations that lack a direct rail connection).

FIGURE 1: Regional and City Bus Lines in Eindhoven

---

[*] This report was prepared by Theo H.J. Muller of the Traffic and Transportation Research Laboratory, Faculty of Civil Engineering and Geo-sciences, Delft University of Technology, Delft, the Netherlands.

RELATION OF INFORMATION INFRASTRUCTURE TO LOCAL TRANSIT SERVICE IMPROVEMENT INITIATIVES

Automatic data collection and analysis in Eindhoven has been driven by the data needs of several initiatives to improve transit service. The chief initiatives have been a compact central bus station with dynamic bay assignment, a focus on operational control, and a program of transit priority.

To meet these needs for real-time location information, all buses in Eindhoven are equipped with on-board computer systems provided by Peek Traffic. They have two-way communication with a roadside automatic vehicle identification (AVI) system using two-way Philips VECOM electro-magnetic buried loops. Through the AVI system, the on-board processors are loaded each day on pull-out with updated line descriptions (with stop and loop distances) and timetables (scheduled times at the stops). Bus location is tracked by the on-board computers using AVI loops as signposts (each loop emits a unique code detected by passing buses), by dead reckoning, and by matching to stop and loop locations. Door sensors tell the on-board computer when the doors open and close.

Also through the roadside AVI system, buses send their location and schedule deviation information directly to local traffic controllers to request signal priority, and to the dynamic compact bus station processor to request a loading bay. The functionality of the on board systems has been extended with a data collection function to register a great number of defined events with time and odometer stamps for off-line analysis.

Some buses also have an ***audio-visual passenger information*** system that displays and announces next stops and transfer points using the on-board computer's bus location algorithm.

Dynamic Compact Bus Station

Eindhoven's combined bus and railway station is a focal point for both regional and city lines. This station is a very busy transfer point, connecting two major railway lines, nineteen regional bus routes, and twenty city lines. To improve passenger comfort, safety, and security, and to minimize the amount of space needed on the edge of downtown, a compact bus station with dynamic bay assignment has been developed. Its concept is to separate and optimize the processes of alighting, layover, passenger transfer and waiting, and boarding, as illustrated in Figure 2.

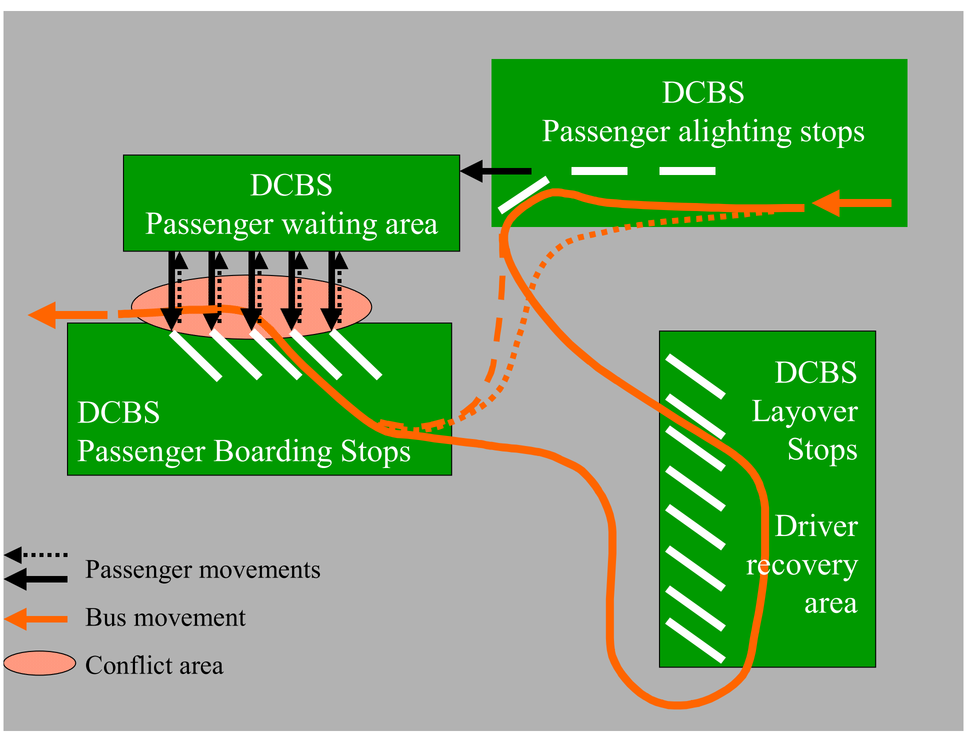


FIGURE 2: Passenger and Vehicle Movements at a Dynamic Compact Bus Station

The *alighting process* takes place near the passenger waiting areas for buses and trains. Because the alighting process should take as little time as possible, the bus looks for a free alighting stop, unloads, and then proceeds to the layover area.

For the *layover process*, buses park near an operators' lounge, which can be located at some distance from the passenger waiting area. The actual layover time depends on the schedule, the punctuality deviation at arrival, and the running time needed to and from the layover area. Late vehicles skip the layover area (see dashed line in Figure 4); and through-going late buses will minimize their time at the station by skipping both the alighting and layover areas, proceeding directly to the boarding area (dotted lines for bus and passenger movements). For the station to handle as many buses as possible during busy periods, long layovers can be shifted entirely outside the station area.

The *boarding process* takes place at a set of bays near the passenger waiting area. Assignment of routes to bays is not static, but is dynamic, based on real-time location and schedule adherence information transmitted by buses in advance of their arrival at the station. The processor at the station control center assigns to each bus a loading bay using an algorithm that constantly optimizes bus bay allocation, taking into account bus length (standard vs. articulated), bay length (some bays can serve multiple buses), preferred bay (e.g. closest to the passenger waiting area, or at a bay normally used by a certain bus company or by certain lines), and, for through-going lines that use the same bay for alighting and boarding, time till departure. About 5 minutes before departure, the selected boarding bay is communicated to the bus operator on the in-vehicle display, and to passengers in the waiting area on information panels much as used in airports. At the scheduled departure time, an in-ground green light illuminates to facilitate an on-time dispatch.

The compact *passenger waiting and transfer area* shared by all bus routes increases passengers' comfort – they wait in a sheltered area instead of scattered at berths around the station. It increases passenger security, especially during times of low demand. The well-designed paths within the station for bus and passenger traffic improve traffic safety (fewer conflicting bus and pedestrians streams, and low speed in the conflict area). The information system shortens passengers' needed orientation time and walking distance.

Figure 3 shows the dynamic compact bus station's control tower and boarding area.

FIGURE 3: Dynamic Compact Bus Station (DCBS), Eindhoven

The "Speed, Regularity, and Punctuality" Program

The Eindhoven Regional Government (SRE) has developed a coherent "Regional Traffic and Transportation Plan" to reduce the growth of congestion and to promote public transport and bike use (http://www.sre.nl/engels/eng.htm). Its aim is to guarantee the accessibility of the region without harming the environment by guiding the growth of car traffic and promoting transit and bike use. One of their programs is the DRS-system, a program of investments aimed at improving transit's operational speed (**D**oorstroming), regularity (**R**egelmaat), and punctuality (**S**tiptheid). Many bus lanes have been implemented on congested route sections, and almost all intersections along bus routes have traffic signal controllers and detectors that can provide absolute priority (almost no delay at intersections). The development of intelligent systems in support of conditional priority is also part of this program.

Further projects currently being investigated by the SRE are automatic guided bus (double articulated) and passenger information displays at important stops.

A Focus on Operational Control

With the help of the Delft University of Technology's Traffic and Transportation Research Laboratory and the cooperation of the operating companies, the local authorities have developed a strong emphasis on operational control as a means of improving transit service quality. An example of this focus is a map-based display, updated nightly and available on the web by password to government and operating company officials, showing average speed on each route segment and on-time performance at each stop (see Figure 4). Colors varying from bright green to bright red indicate performance. There is also a similar display that indicates malfunctioning detectors.

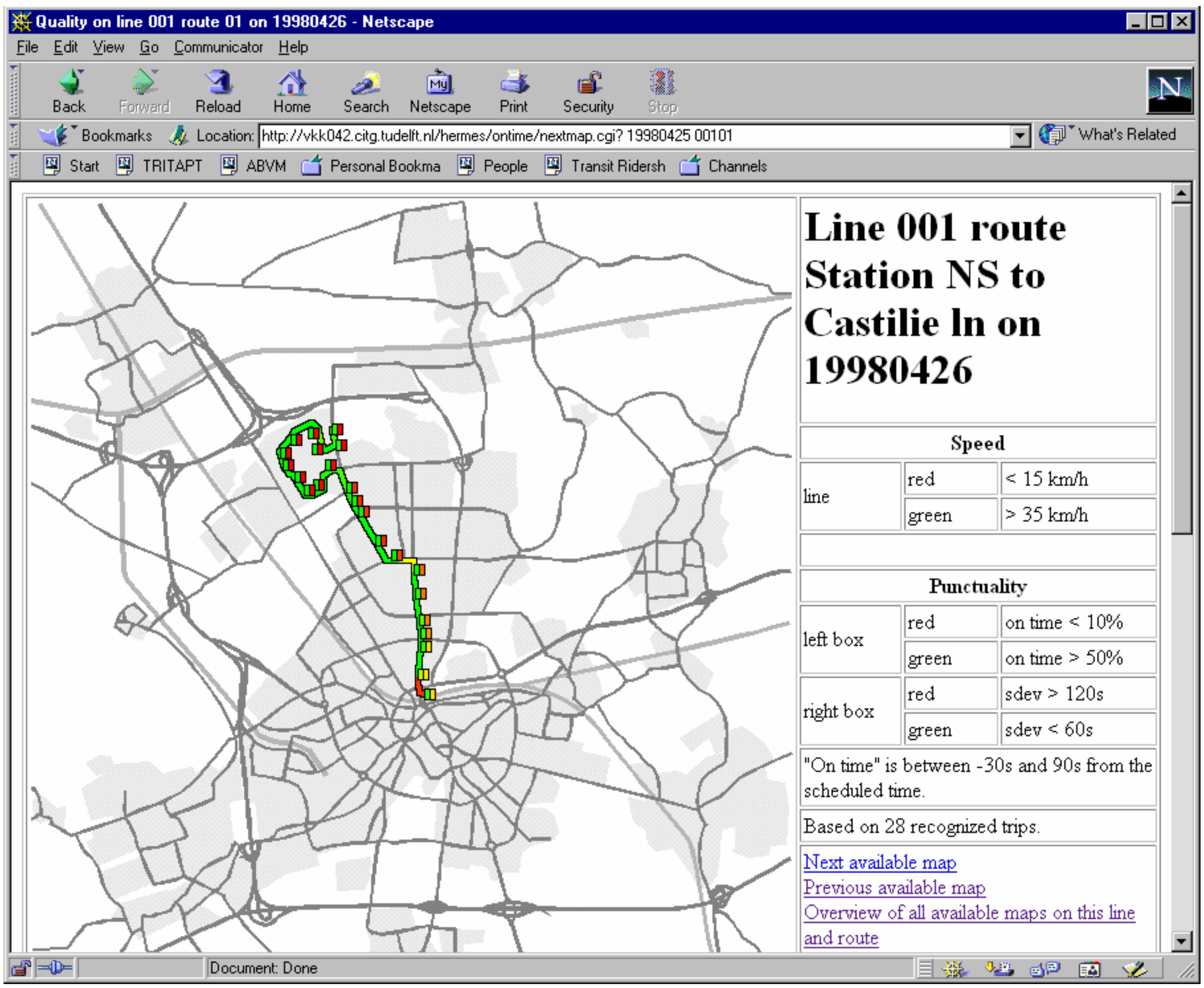


FIGURE 4: Daily Service Quality Report Distributed by Internet

The focus on operational control is manifested in a preference for conditional (schedule-based) priority at traffic signals over absolute priority, and in scheduling practices that emphasize and promote operational control.

Conditional Priority at Traffic Signals

Absolute priority increases operational speed. However, pre-emption for early vehicles does not improve punctuality and regularity. Besides, frequent pre-emption can result in disturbed traffic flows (long queues and much car delay). The regional government and the Hermes Company have agreed to use conditional priority instead, requesting priority only if a vehicle is behind schedule. Individual buses request priority directly to local intersection controllers (decentralized). The on-board processor determines the bus location and the schedule deviation. When passing a vehicle identification loop, the on-board processor transmits its bus and route number and a priority code (early, on time, late). The on-time definition is 0 s early to 30 s late by default, although managers can vary the late threshold by intersection (e.g., to limit priority requests at critical intersections). At each intersection controller, managers can set which buses get priority (all buses? on-time and late buses? late buses only?) and the signal control tactics allowed (truncate or skip conflicting streams, insert an extra bus stream green, extend green) when serving the bus and when recovering from an interruption.

The system of conditional priority requires a finely tuned schedule making process. If the scheduled time at an intersection is too tight, the bus will always be late and therefore always request priority; if too loose, the bus will never be late and therefore never request priority. Either way, the push-pull benefit of conditional priority is lost.

Scheduling for Operational Control

While conditional priority has so far been implemented on only one route, operational control is still exercised on every route by having operators respect scheduled departure times at timepoints, not departing early. Consistent with European practice, each bus stop has a scheduled departure time and therefore serves as a timepoint.

As a real-time aid, the Eindhoven on-board setup includes an operator's display that continuously shows the schedule deviation in increments of 0.1 minute.

However, more important than real-time aids is a realistic, attainable scheduled running time based on an analysis of large amounts of operational data. An attainable scheduled running time and a correct distribution of running time over route segments enable drivers to run the route punctually with only minor adjustments to speed or dwell time, especially when supported by conditional priority at intersection. The scheduled departure time at each stop should guarantee that, most of the time, a bus that is just a few seconds early and therefore held (and, on routes with conditional priority, denied priority) will still be able to reach the terminal stop on time. To promote operator compliance – because operators care about arriving on time and enjoying their layover break, both the scheduled running time for the whole route and the scheduled completion time from each stop to the end of the line should have a high degree of attainability, e.g., 80%. As traffic conditions change during the day, the scheduled running time and its distribution over the route segments should fit the expected traffic conditions. Scheduled running times should be static only for periods of the day in which actual running times are homogeneous.

Because running times vary enormously across the day as well as day to day, an extensive amount of data is needed to establish statistically reliable scheduled running times for the route and each segment. Eindhoven gathers operational data from 100% of its buses using its on-board computers and AVI system, and then uses the TriTAPT analysis program to analyze that data to support schedule planning. TriTAPT was developed over the last 20 years by the Delft University of Technology's Traffic and Transportation Engineering Laboratory, which has provided technical support to the City of Eindhoven and to Hermes in its implementation. In addition to scheduled time analyses, TriTAPT also provides management information about operational characteristics (run times, delay times, stopping times), service quality (punctuality, regularity, speed), and service quantity (unlinked trips, passenger kilometers). TriTAPT passenger-count related features are still only planned for implementation in Eindhoven pending the acquisition of automatic passenger counters.

Expected Passenger Benefits of Operational Control

Operational control, including a finely tuned schedule based on a high attainability condition, conditional priority, and operator holding at frequent timepoints, results in a reliable transit operation with small punctuality deviations. As passengers recognize the smaller variation in bus arrival times, they will confidently arrive closer to the scheduled departure time, decreasing passenger waiting time. As in-vehicle time becomes more reliable, passengers will reduce the slack time built into their planned trip, especially for those trying to make a transfer or arrive at an appointed time. In many cases, the decrease in passenger waiting and slack time will exceed the small increase in scheduled running time needed to provide operational control, making transit more attractive and competitive (smaller and more reliable overall travel time). At the same time, the more reliable running time reduces the necessary layover length to ensure an on-time departure for the next trip, limiting or eliminating the cost to the operating agency of using a high attainability condition in establishing scheduled running times.

INFORMATION FLOWS

Figure 5 shows the chief flows of data related to operation performance in the Hermes bus company.

The Hermes bus company uses the Hastus package for scheduling routes, vehicles, and personnel. Schedule planners load the Hastus database with information about available vehicles and personnel, routes, stop locations, departure times, and running times between timepoints. An automatic conversion program uses the Hastus line description and timetable together with a dataset of AVI detector positions and a dataset with calendar information about holidays, special events and so on (which schedule is valid on each day) to create the framework of TriTAPT "one-day-one-route" datasets. These "empty" datasets (containing only schedule information at this point) will later be filled with data recorded in the on-board computers and analyzed by TriTAPT.

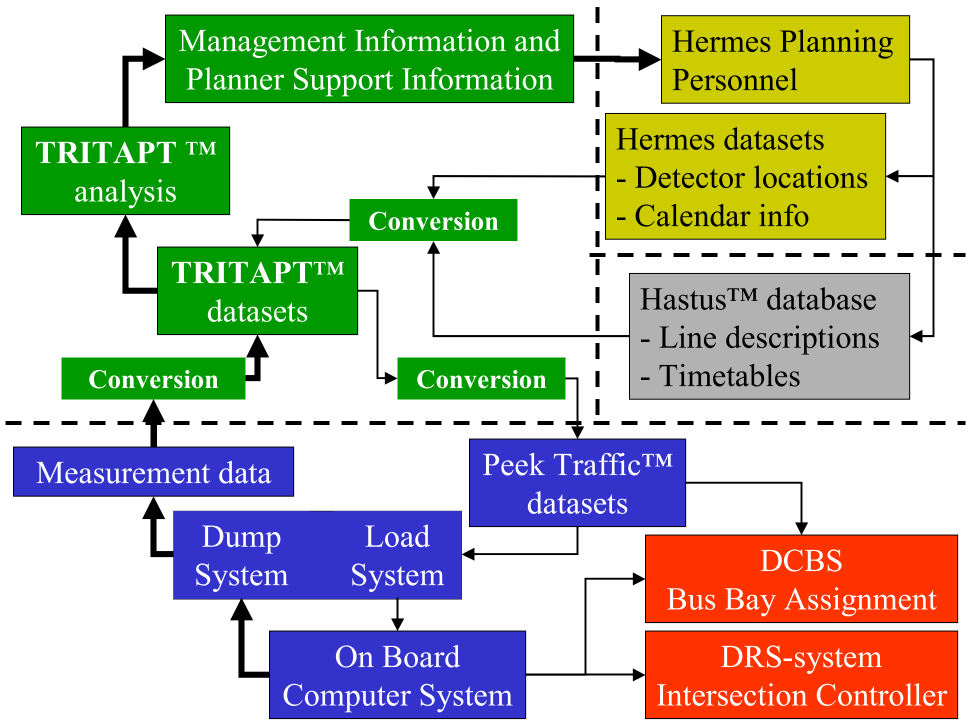


FIGURE 5: Data Flow in the Hermes system

A second conversion program converts the empty TriTAPT dataset into an empty dataset in the format required by the Peek Traffic on-board computers. Through the AVI system, a bus communicates its route / run, and the appropriate empty dataset is loaded at pullout.

As a bus operates, its on-board computer makes time- and location-stamped records each time the bus opens and closes its doors, passes an AVI loop, and passes a crawl speed threshold. It also sends real-time information to intersection controllers and to the dynamic compact bus station processor. At the end of the day, full datasets are uploaded from the on-board computers on pull-in through the AVI system.

A third conversion program converts the measurement data into TriTAPT databases ready for analysis. One TriTAPT dataset is created for each line-route-day ("route" in this context implies both direction and line variation, if any). TriTAPT datasets summarize data to the level of "stop module," which is a bus stop and the segment following it until the next bus stop. Summary measures include arrival and departure times, delay time (time spent below crawl speed and away from a stop), and holding or control time (time spent at a stop in excess of normal dwell time, for early buses only). The data structure will also accommodate passenger counts, if available, and other user-defined stop- or segment-level variables. Table 1 shows the data structure.

ANALYSES IN SUPPORT OF SCHEDULING

This case study will focus on TriTAPT analyses used to support schedule planning. Examples of other TriTAPT analyses used by Hermes, including trajectory plots and graphs showing delay distribution, can be found in Muller and Furth (2001).

One complication of using observed running time to make a new schedule is the fact that the performance observed is affected by the schedule then in place, because operators hold when they are early. To undo that bias, the processing of event records identifies holding time (also called control time) as time spent at a stop in excess of normal dwell time when the bus is ahead of schedule. Most analyses related to proposing a new schedule use "net" running time, defined as running time excluding time spent holding.

Figure 6 shows gross (includes holding time) and net running times for each scheduled trip on a route during a user-selected set of hours (in this case, 7:00-12:00) and over a user-selected set of days (in this case, 20 weekdays). Average values are presented by bars (gray is net running time and black is holding time), and individual observations are shown as green (net running time) and red (gross running time) marks.

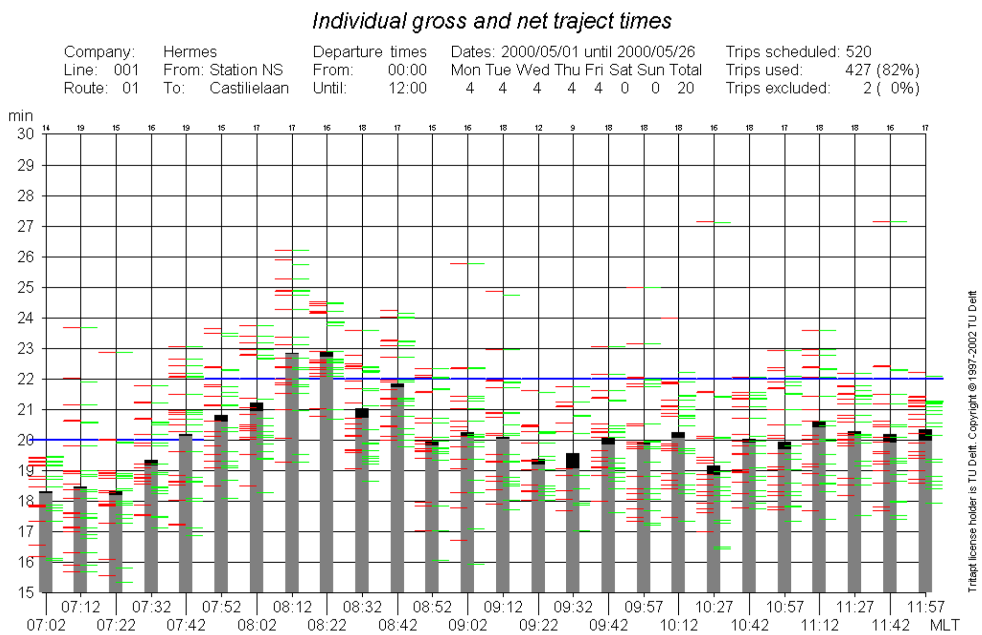


FIGURE 6: Measured Gross and Net (Excluding Holding) Running Times by Trip

Figure 7 displays how the TriTAPT program recommends scheduled running times, simultaneously finding homogeneous periods of the day. In Figure 7, mean and maximum running times for each scheduled trip are indicated by the gray bar and red arrow. The two jagged lines indicate the user-selected lower and upper percentile bounds for running time, in this case 75- and 95-percentile. The route's original schedule had the same 22 minute running time from 7:45 am until 12 noon (blue line). One can see that during the peak period this schedule was not attainable for most trips, while off-peak, the schedule was more than attainable. The analysis, based on net observed running times, searches for whole-minute running times that lie between the given percentile bounds (horizontal blue bars). At the same time it identifies periods that can be scheduled with the same running time.

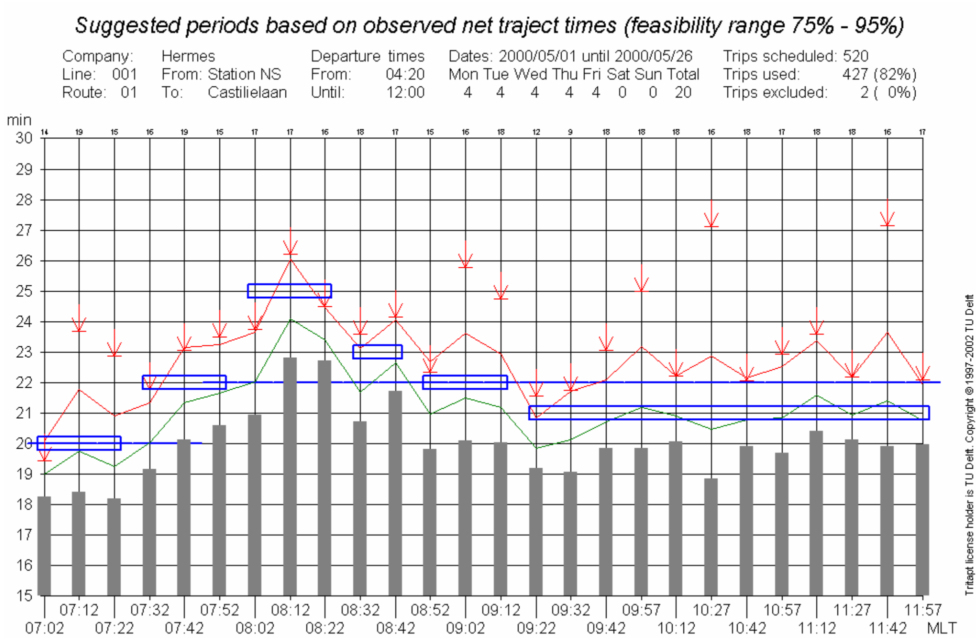


FIGURE 7: Proposed Homogeneous Periods and Scheduled Running Times

Next, for each homogeneous period of the day, segment running times should be determined. Applying the same attainability condition for the remainder of route from each timepoint determines the scheduled running time from each timepoint to the end of the line; then, by subtraction, it yields running time by segment. Figure 8 shows the distribution of a 21 minute schedule route time, based on net running observed on all trips between 9:20 and 12 noon over a 4-week period. The TriTAPT display makes the result easier to visualize by showing the bus stops (on the horizontal axis) spaced in proportion to their distance along the route.

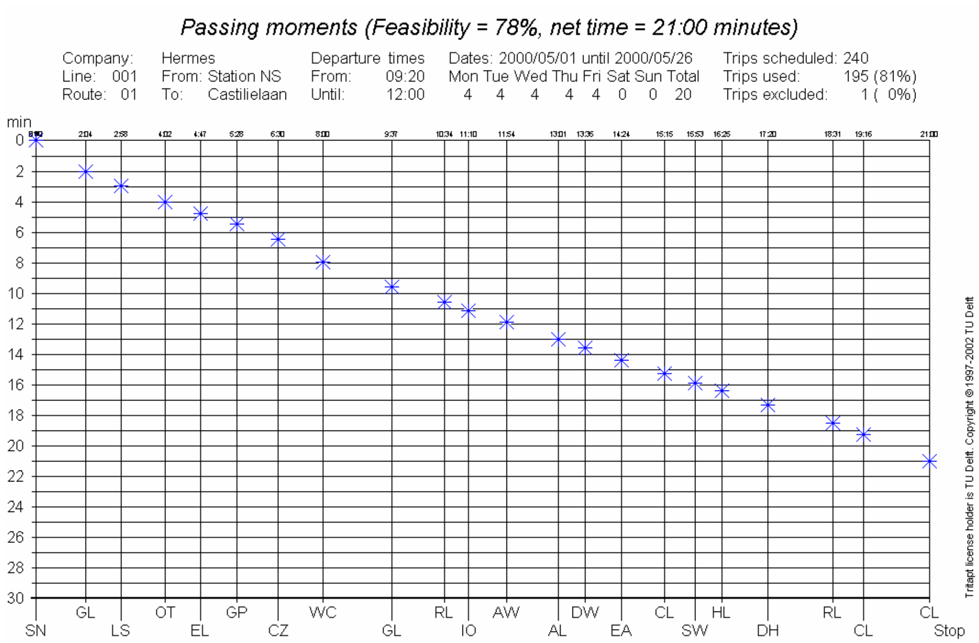


FIGURE 8: Proposed Stop-by-Stop Distribution of Scheduled Running Time

For the example of Figure 8, holding at timepoints until the scheduled departure time ("passing moment") gives an operator a 78% chance of arriving on time at the terminal stop. The higher the attainability, the more operators will cooperate (for the same example, a 22 minute route time offers 90% attainability). However, a very high attainability will slow down operation more than even operators will accept. A compromise could be the 85% value.

The analysis of proposed running time distribution over segments can be visualized in the display shown in Figure 9, showing proposed departure times (passing moments) in relation to the distribution of observed net running time, with every observed trip's running time shifted to an on-time start. This example shows proposed running times with a 90-percentile attainability. One can visualize that, with a good schedule, holding, and an on-time start, early trips (those above the scheduled time) will hold to follow the schedule, resulting in most trips being right on time, with only a small percentage arriving late.

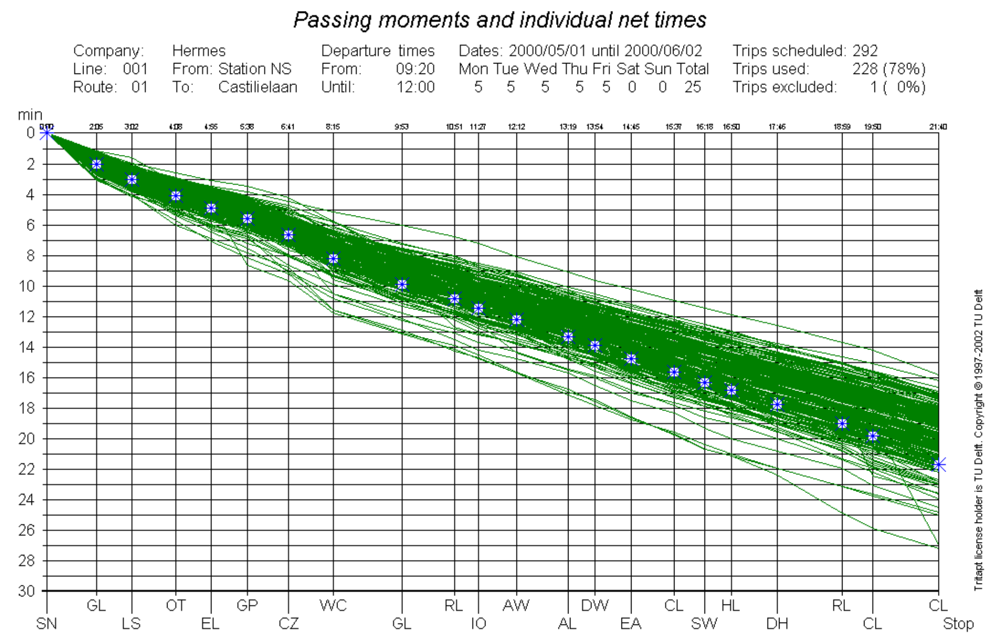


FIGURE 9: Scheduled Departure Times in Relation with Uncontrolled Trajectories

REFERENCES

1. Muller, Th.H.J. and P.G. Furth. "Trip Time Analyzers: Key to Transit Service Quality." *Transportation Research Record 1760*, pp. 10-19, 2001.

TABLE 1: Data Structure for a TriTAPT Basic Dataset

Data format

TriTAPT 10	Number of items in basic dataset
[0] magic 3751950737	1 Keyword
[1] version 2	2 Version items
[2] company Hermes	3 Name of the transit company
[3] line 001	4 Line number
[4] route 01	5 Route number
[5] date 2000/01/21	6 Date
[6] timetable 2	7 Number of timetable items
[7] modules 22	8 Number of stop modules
[8] beacons 26	9 Number of beacons
[9] trips 73	10 Number of trips
[1] version 2	Version items
[1 0] major 1	1 Major version number
[1 1] minor 0	2 Minor version number
[6] timetable 2	Timetable items
[6 0] name Summer service	1 Timetable name
[6 1] valid 3	2 Number of validity items
[6 1] valid 3	
[6 1 0] from ????/??/??	1 Start date
[6 1 1] until ????/??/?	2 End date
[6 1 2] default_days_of_week 7	3 Number of default days
[6 1 2] default_days_of_week 7	
[6 1 2 0] Monday 0	
[6 1 2 1] Tuesday 0	
[6 1 2 2] Wednesday 0	
[6 1 2 3] Thursday 0	
[6 1 2 4] Friday 0	
[6 1 2 5] Saturday 0	
[6 1 2 6] Sunday 0	
[7] modules 22	Number of stops
[7 0] module 4	1 number of items to describe a stop
* * * * *	
[7 21] module 4	
[7 0] module 4	Stop items
[7 0 0] stop_id 64000010	1 Stop identification
[7 0 1] distance 0	2 Distance from start of route
[7 0 2] name Station NS	3 Name of stop
[7 0 3] short_name SN	4 Abbreviation
[8] beacons 26	Number of beacons
[8 0] beacon 2	1 Number of beacon items
* * * * *	
[8 25] beacon 2	

[8 *] beacon 2	Number of beacon items
[8 * 0] code 30015	1 Beacon code
[8 * 1] distance 170	2 Distance from start of route
[9] trips 73	Number of trips
[9 0] trip 13	1 Number of trip items
* * * * *	
[9 72] trip13	
[9 *] trip 13	Number of trip items
[9 * 0] trip_id 1	1 Trip number
[9 * 1] shift 27	2 Shift number
[9 * 2] observed_shift -1	3 Shift observed
[9 * 3] driver_id-1	4 Driver identification
[9 * 4] vehicle_id 3108	5 Vehicle identification
[9 * 5] flags 4	6 Number of measurement flags
[9 * 6] first_stop 0	7 First stop to be analyzed
[9 * 7] last_stop 21	8 Last stop to be analyzed
[9 * 8] vehicle_seats -1	9 Number of seats
[9 * 9] vehicle_stands -1	10 Number of standing places
[9 * 10] vehicle_max_load -1	11 Absolute vehicle capacity
[9 * 11] stop_events 22	12 Number of stop-modules to measure
[9 * 12] beacon_events 26	13 Number of beacons to measure
[9 * 5] flags 4	Number of trip measurement flags
[9 * 5 0] trip_is_recognized 1	1 Trip is recognized Y/N
[9 * 5 1] vehicle_counted_pass 0	2 Passengers are counted
[9 * 5 2] dont_use_times 0	3 Trip excluded for time calculations
[9 * 5 3] dont_use_passengers 0	4 Trip excluded for passr calculations
[9 * 11] stop_events 22	Number of stop module
[9 * 11 0] stop_event 11	1 Number of items per stop module
* * * * *	
[9 * 11 21] stop_event 11	
[9 * 11 *] stop_event 11	Number of measure items
[9 * 11 * 0] MAT 07:02:00	1 Scheduled arrival time at stop
[9 * 11 * 1] MLT 07:02:00	2 Scheduled departure time at stop
[9 * 11 * 2] MAA 07:01:11	3 Measured arrival time at stop
[9 * 11 * 3] MLA 07:02:42	4 Measured departure time at stop
[9 * 11 * 4] TSA 45	5 Measured stopping time
[9 * 11 * 5] TDSA 0	6 Measured delay time at stop
[9 * 11 * 6] TCSA 4	7 Measured control time at stop
[9 * 11 * 7] TDMA 0	8 Measured delay time between stops
[9 * 11 * 8] TCMA -1	9 Measured control time between stops
[9 * 11 * 9] NPAA 0	10 Measured number of alightings
[9 * 11 * 10] NPBA 0	11 Measured number of boardings
[9 * 12] beacon_events 26	Number of beacons
[9 *12 0] beacon_event 1	Number of beacon items
* * * * *	
9 * 12 25] beacon_event 1	
[9 * 12 *] beacon_event 1	
[9 * 12 * 0] MPMA 07:03:04	1 Measured passing time

APPENDIX H: The Transit Management Information System of HTM, The Hague *

The Hague is the residence of the Dutch Government and of the International Court of Justice. There are 450,000 inhabitants in the municipality, and about 1 million in the metropolitan area. Its location is shown in Figure 1.

FIGURE 1: Location of The Hague

The two major transit companies running services in The Hague are the "Haagsche Tramweg Maatschappij" (HTM) and "Connexxion." HTM operates mainly bus and tramlines in the city and Connexxion operates the regional bus lines.

All HTM buses and trams have been equipped with on-board processors for four reasons:

- To request priority at intersections. Control at almost every signalized intersection is vehicle actuated, with bus detection causing preemption to prioritize public transport.
- All bus and tramline operators in the Netherlands use a common (zone based) ticketing system called "Strippenkaart." Those cards can be bought at many places like post offices, tobacco shops, and information centers. To allocate ticket revenues fairly between transit companies, the transit companies have to measure unlinked trips and passenger kilometers. To meet this need, about 25% of the HTM vehicles have APC's connected to the onboard processors.

---

* This report was prepared by Theo H.J. Muller, of the Traffic and Transportation Research Laboratory, Faculty of Civil Engineering and Geo-sciences, Delft University of Technology, Delft, the Netherlands.

- The ongoing privatization of transit operation in the Netherlands is organized by concession. Transit agencies tender for services requested by the regional transit authorities. More and more, they have to prove the quantity (vehicle trips run) and quality (punctuality and regularity) of the executed services. Automated data collection is the preferred method of service monitoring.
- The information can be used for planning purposes and process management (quality management).

REAL TIME PASSENGER INFORMATION

One of the ITS applications is to inform passengers in real time about the expected arrivals of trams and busses at stops. Together with ACIS Nederland, HTM announced the ability to check with an SMS messaging service the prediction of the arrival of the next tram or bus at the specific stop requested by the traveler. The information is also available on the HTM website at http://mobiel.htm.nl/acislive. The system has been operational since January 25, 2002.

FIGURE 2: Real-time Arrival Predictions Via Internet.

MANAGEMENT INFORMATION

Recently the HTM has installed the TriTAPT package for operations data analysis. The HTM's main objective is to make transit more efficient and effective. To avoid vehicle capacity failures, route frequencies are based not only on average vehicle occupancy (load), but also its variation. If the variation can be reduce by punctuality and regularity control, the headways can increase without decreasing the service quality.

Figure 3 shows the frequency offered and the corresponding static occupancy per trip (number of passengers in the vehicle at the busiest section on the route). Each bar indicates the average static occupancy of a trip of the day measured by APC's. The green ticks show the individual observations (on average, there were 8 observations per scheduled trip in the 12-day period selected). The distance between two bars is the schedule headway. The variation in vehicle occupancy per trip (variation in individual values indicated by the tics of a certain trip of the day) and between trips (varying height of successive bars) is obvious. Given the capacity of the vehicle and the policy headway the measured occupancy helps to plan adjusted headways.

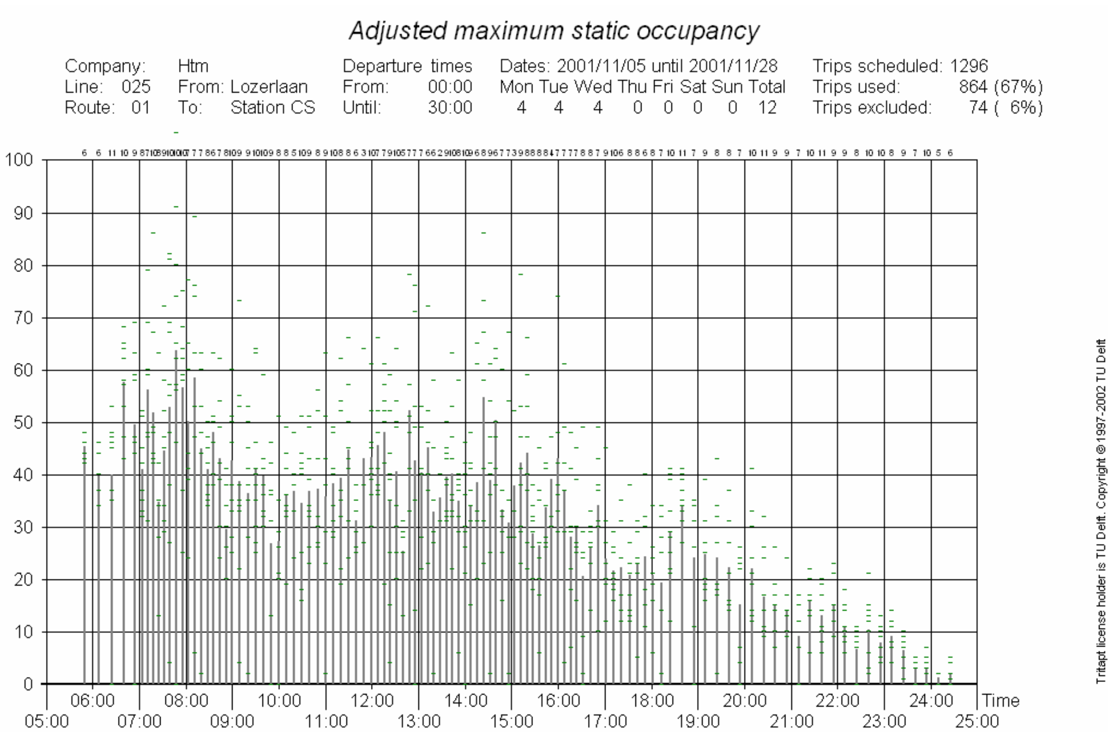


FIGURE 3: Static Occupancies (Mean and Individual Observations) by Scheduled Trip

The main causes of the occupancy variation are the random passenger arrivals and, even more so, the irregularity of operation (bunching). The variation in demand is only local at a stop. An extremely busy stop can be followed by a quiet stop. A long or short headway, however, influences the probability of more or less passengers at all following stops.

The irregularity I is expressed as the absolute percentage of headway deviation from the timetable headway (h_T):

$$I = |h - h_T| / h_T$$

The irregularity can be shown by a graph (see Figure 4).

Information about the average occupancy per trip of the day and its variation can also result in tuned headways per trip. The TRITAPT program will be extended to use the occupancy information to suggest to planners the optimal headway.

The combination of management information of productivity and operational quality is the HTM's starting point to improve schedules and services.

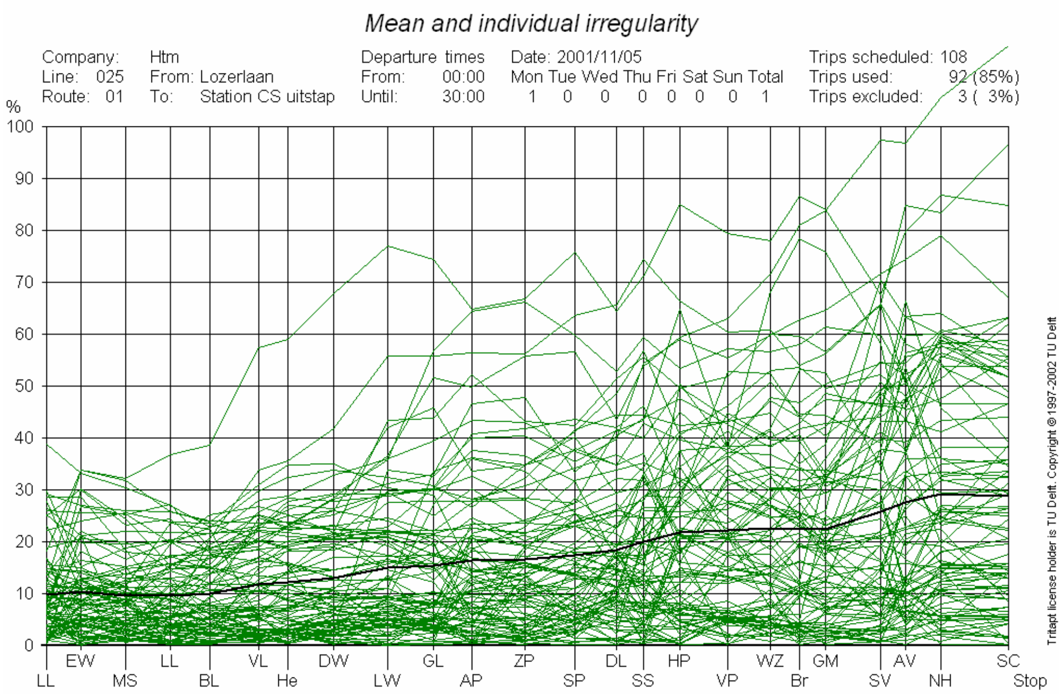


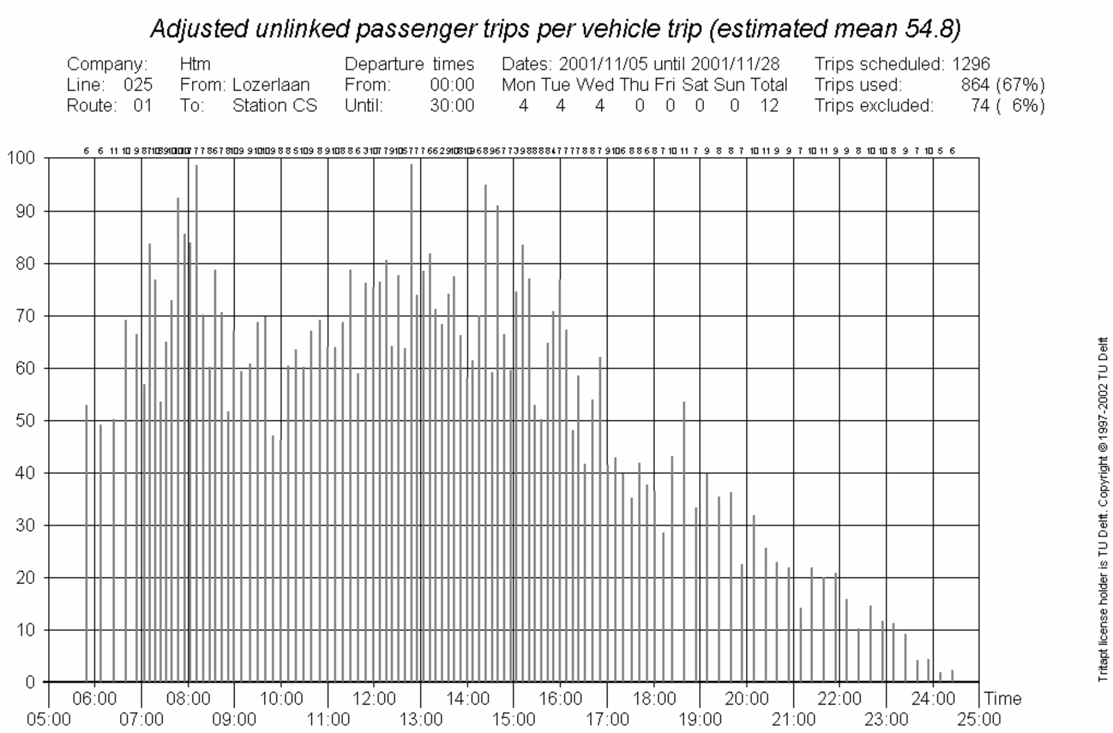
FIGURE 4: Irregularity in Service Execution

A regular or punctual (includes regular) service should decrease the fluctuations in the static occupancy (busiest section on the route). As the frequency is based on the probability of occupancy failures (overcrowding), a reduction in the variance can result in increased headways without a lower service quality.

The TriTAPT package also has to provide planners with route time information to determine and to tune schedules during various periods of the day and to distribute the route time over the stops. This has to result in attainable schedules to encourage schedule adherence (by holding) while limiting the risk of late arrivals at the terminal stop (see also the Eindhoven case study).

The management information of HTM is also focused on unlinked trips (see Figure 5) and passenger kilometers. As many of the vehicles are equipped with automatic passenger counters (switch steps), the segment occupancy and segment length are multiplied to find the production in passenger kilometers. Figure 5 shows the unlinked passenger trips by scheduled vehicle trip (108 scheduled vehicle trips per day). The average number of unlinked passenger trips is 55. Figure 6 shows the passenger kilometers per scheduled trip and the average per trip (169 passenger kilometers). The trips run with the policy headway of 15 minutes (after 6 pm) have a lower production. The average passenger trip length is 3 km (169/55). This information can be used to distribute Strippenkaart revenues, as well as negotiate local subsidies for run trips at the policy headway.

The tendency to privatize agencies requires a system to prove that scheduled trips are really run according to the quality of the service requirements. This means that trips are measured in detail. Unmeasured trips can be explained by actual fall out or by malfunctioning of the on-board equipment.



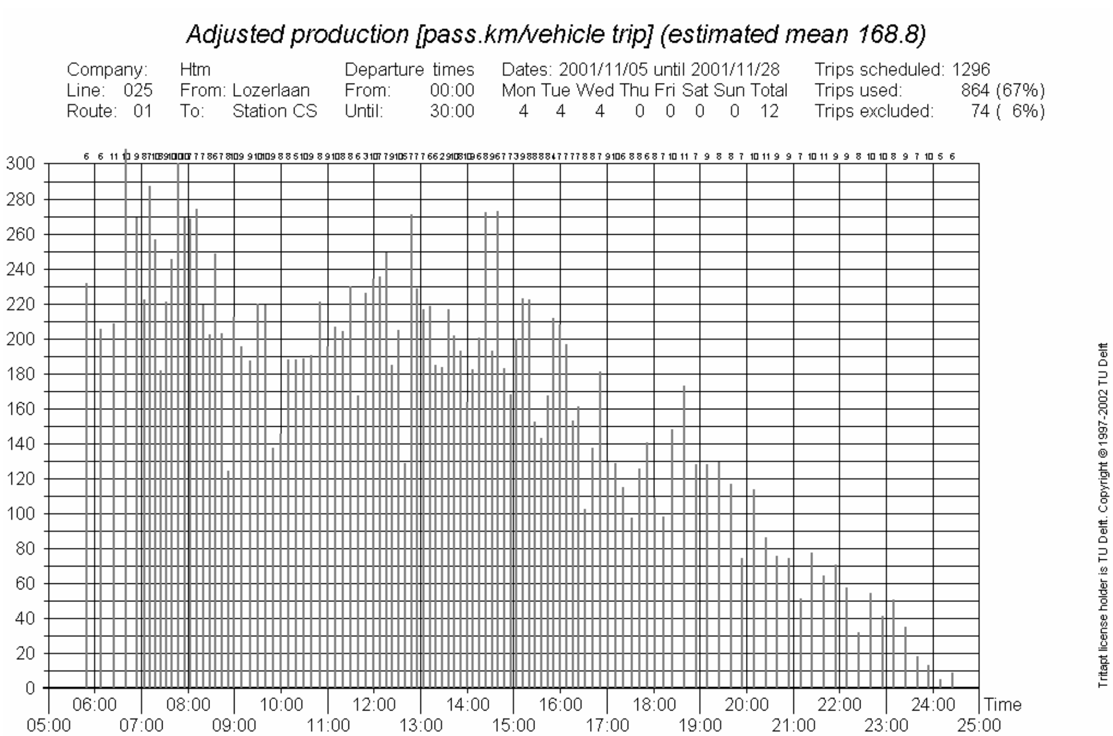


Figure 5. Passenger kilometers per trip

APPENDIX I: Metro Transit's Integrated AVL-APC System*

From 1999 to early 2002, Metro Transit (Minneapolis) benefited from a pilot AVL system installed on 250 buses as part of a larger ITS initiative called the Orion project. This project gave Metro Transit valuable experience in understanding AVL system design, and it provided them with a good archive of operations data for which they developed some creative analysis programs.

A recent decision to upgrade the radio system to the 800 MH band became the opportunity to acquire a new fleet-wide AVL system. It features a "smart bus" design with a J1708/1587 vehicle area network, customer information signs displaying the estimated arrivals of the next buses, and computer aided dispatching. It will also include APCs in 100 coaches (about 12% of Metro's peak fleet requirement). In late 2002, the new system is partway through implementation.

Metro Transit's practice with AVL and APC includes these five highlights:

- *Dual data capture, with primary reliance on radio transmission.* All data captured on the bus, including location data and passenger counts, is transmitted by radio to the central computer and stored. Routine polling for location occurs at intervals of 40 to 60 s. Event messages, including timepoint departure and APC records when a door closes, are transmitted independent of the polling cycle. In event of radio failure, messages are recorded in the on-board computer, uploaded at night, and inserted into the database of poll messages.
- *Timepoints captured on all buses; stops captured on APC buses only.* The AVL system tracks bus location in real-time, notes when a timepoint is crossed, and transmits a "timepoint departure" message by radio to the central computer. Door sensors enable the bus to know the stop it's serving. That information is used to inform the APC processor, but it does not trigger a stop record except as part of an APC record.
- *Single database for AVL and APC.* AVL and APC data are all part of the same data stream, and will be stored in different tables in a single database. Reports will account for the difference in sample size between timepoint data, captured on all buses, and stop data, captured only on the APC-instrumented buses.
- *Automatic operator sign-on and sign-on verification.* The on-board computer receives operator sign-on information (operator ID and run ID) from the garage dispatch center via the data radio. In the new system, operators will be able to verify their sign-on information on the mobile data terminal.
- *Creative analysis and reporting.* Metro Transit's Service Analysis department developed several reports using the Orion AVL data, including an analysis of running time that extended to proposing new segment running times. In planning for their new AVL-APC system, they reviewed reports offered by the vendor and reports used at Tri-Met and developed specifications for new analyses and reports whose development will be the shared responsibility of Metro Transit and the vendor.

CONFIGURATION AND DATA CAPTURE

Orion Project AVL (1999-2002)

The Orion project AVL system operated from 1999 to early 2002 on 250 buses. It was intended as a pilot project, as part of a larger ITS effort called the Orion project, and was a testbed for 3M's venture into the AVL market. Location was determined using GPS. Buses got sign-on and schedule information automatically from the dispatch center at the garage over the wireless local area network used there for data transfer. Buses tracked their progress against the schedule. The mobile data terminal indicated to the operator by a GPS clock display and colored lights whether the bus was early, on time, or late.

Buses reported their location via radio polling every 40 s. This information was used in the control center for real time monitoring. Buses also recorded events in the on-board computer, including door opening and closing at timepoints. Event data was uploaded nightly at the garage by wireless local area network. Metro Transit staff used the archived event data to analyze operational performance.

* This case study was prepared by Peter G. Furth of Northeastern University based on information received from Erin Mitchell, Janet Hopper, and Gary Nyberg of Metro Transit and Dave McFarland of Siemens.

New AVL-APC System (in delivery 2002)

The new system, supplied by Siemens, follows the "smart bus" concept with a J1708/1587 vehicle area network. Connected to the network are an on-board computer ("in-vehicle logic unit") that tracks vehicle location, a mobile data terminal with keypad and display, a GPS receiver, door sensors, and a radio. In addition, 100 buses, about 12% of the peak fleet requirement, will have APC's connected to the network with door and passenger sensors. They will be able to add stop announcement functionality in the future as a separate project.

Operator sign-in and schedule information is transmitted automatically to buses each morning over the data radio. Operators verify their sign-in information, correcting errors and reporting corrections by radio to the control center. This facility should help reduce the incidence of buses having incorrect operator and run ID's, making tracking easier.

Like the Orion system, the Siemens system features dual data recording. The primary means of data recording, serving both real-time and off-line applications, is a radio wide area network. Routine poll messages, including time, location, and identifiers, are sent at 60 s intervals; that is expected to improve to 40 s. Event messages are sent independently of the polling cycle. All events are stamped with current GPS coordinates and time and are stored to a database. A timepoint event includes timepoint ID, departure time, and schedule deviation. On buses with APC's, stop events include stop ID, ons, offs, and door open and close time. On buses without an APC, no record is made of stop arrival or departure time (except at stops linked to a timepoint), even though the on-board computer knows the stop.

The secondary means of data recording is the on-board computer. In case of radio failure, messages that could not be transmitted by radio are recorded in the on-board computer's memory. The on-board computer is connected to the central computer by a wireless local area network allowing high speed data transfer each night at the garage. The wireless LAN is used to download schedule and to upload any records written during the day.

Data Recording by Radio vs. On-Board Storage

The system that Metro Transit has purchased is capable of recording data in two ways: transmitting over the air for storage in the central computer, and recording in the on-board computer. At Metro Transit, as at several other agencies served by the same vendor, event records are all sent over the air, with on-board storage used only for backup during radio failure. The vendor claims it will be able to provide 40 s polling with 225 peak buses per channel, even with 12% of the buses transmitting APC records at each stop and all buses transmitting timepoint records. The advantage of this configuration is that it keeps on-board computer memory requirements to a minimum, lowering the cost of equipping a bus.

However, recording via the radio system has its limitations. Some members of the Metro Transit staff are interested in getting speed data, such as maximum speed between stops, which will not be provided in the new system. Siemens customers at other transit agencies have asked for still more: speed records every 2 seconds or so to permit more accurate investigations into accidents. This further level of detail cannot be accommodated using radio traffic; it requires on-board event recording and storage. By increasing the memory in the on-board computers, Metro Transit's new system could be configured for routine on-board event recording. (Their on-board computers now have 64K of memory, most of which is needed for schedule and smartcard data; higher capacity memory modules, up to 512K, could be installed if requested.) The tradeoff is clear: the lower cost of limited on-board data storage brings with it the limitations of radio-based data recording.

AVL/APC vs. Farebox Data Stream

In current practice, operators key into the farebox to indicate wheelchair lift use. Data on lift use is currently analyzed from the farebox data stream. In the new AVL system, a wheelchair cycle automatically triggers an event in the AVL/APC data stream. (Unfortunately, the record doesn't indicate whether the wheelchair is boarding or alighting.) The advantage of recording lift use in the AVL/APC data archive is that the AVL/APC database will have an easier user interface and will be the primary source for analysis of both demand and running time data.

ARCHIVED DATA STORAGE AND ANALYSIS

Experience with Orion AVL Data

From the archive of event records uploaded from Orion AVL buses, timepoint departure records were extracted and entered into an Excel file. An Excel spreadsheet was set up with a header row for each route variation containing timepoint names; records (rows) were then inserted under the appropriate header for each observed trip, indicating the time at which the observed trip passed each timepoint, as well as trip identifiers. The data are sorted so that all the observations of a given scheduled trip are together.

Running time analysis programs were created as Excel macros. One analysis shows, for each scheduled trip, the difference between schedule and average observed running time on each timepoint segment. Another analysis proposes new segment running times. The algorithm used is that the proposed segment running time equals the average observed running time (rounded to whole minutes), subject to the following limits: the scheduled time may not be increased by more than 2 minutes, or decreased by more than 1 minute, or fall below 1 minute.

Plans with New AVL-APC Data

All radio-transmitted data is stored in a database for off-line analysis of operational performance. Events recorded in the on-board computers during times of radio failure are uploaded each night and inserted into the database.

Data on bus location at the moment of polling will be not be used in off-line analysis (except for playback). Running time and schedule adherence analysis will be based on timepoint departure records. Stop-level records from APC-equipped buses will be used for analyzing passenger demand, but not for analyzing running time or schedule adherence, except for determining whether a particular observation was so far off-schedule as to make it unreliable.

In acquiring its new smart bus system, Metro Transit called for reports on schedule adherence and on running time by timepoint segment and route. However, the RFP did not specify the content or format of the reports, allowing bidders maximum freedom to use reports they had already created. The standard set of reports that come with the selected vendor's system was too basic to meet Metro Transit's needs. Metro Transit and the vendor are therefore both working to develop a limited set of more advanced analysis programs. The agreed on division of responsibility is that the vendor will develop stored procedures, and Metro Transit will develop programs that take the tables created by the stored procedures and format them as desired. The logic of this arrangement is that the vendor doesn't want to customize report formats for each customer; however, they see the value of developing generic procedures that could be used at other transit agencies. Metro Transit is still considering what software to use in this final stage of analysis; one possibility is Excel macros, which they used with Orion AVL data, but a more likely software will be Crystal Report Writer, the agency standard.

These reports will run from a Metro Transit database. Since operations data rolls off the Siemens historical database every 90 days, Metro Transit will copy the necessary data over to a Metro Transit database in Sybase daily. There are several benefits to this:

- Having the AVL and APC data on a database separate from the real-time operations ensures that the integrity of the transit control center's computer system is not compromised due to data processing.
- Speed of processing should be faster as data exists on summary tables.
- Information Services and Service Development can develop a schedule for how long to keep schedule and ridership data live before archiving.
- Potential ease of integration with other agency databases with ridership and schedule adherence data, such as farebox data and point checks (conducted by district supervisors and data collectors to record loads and times at timepoints).

To help them formulate the analyses that they would like to see with their new AVL-APC system, Metro staff learned from their experience with the Orion AVL data, and consulted with Tri-Met, whom they recognized as an industry leader in the analysis of archived AVL-APC data. They prepared specifications and prototypes for over 20 reports. Those dealing with operational data include the following.

- Average schedule deviation by timepoint-route-direction.
- Operational performance by scheduled trip, including

- percentage of timepoint observations early / on-time / late
- comparison of scheduled and average observed running time
- comparison of scheduled, average observed, and minimum (by labor contract) layover
- average speed
- average number of stops
- average dwell time
- average lift uses

- Departure deviations for a single route, including:
  - histogram of departure deviations, aggregated over the entire day and all timepoints
  - table showing percentage of early / on-time / late departures by period-direction, aggregating over all timepoints
- On-time performance (percentage of early / on-time / late departures, aggregating over all timepoints and trips) by route-direction.

Because the database holds both AVL and APC data, statistics based on timepoint data – captured on all buses – can be combined with statistics based on stop-level data – captured only on buses with APC's – such as dwell time and passenger volume. Of course, timepoint-based statistics will have a far greater sample size than stop-based statistics, something the analysis software will account for.

Metro Transit plans to make the result table sortable on every field. Sorting on a measure of performance is helpful in identifying problem areas. For example, by sorting on average departure deviation, the timepoint-route-direction combinations for which buses are most late and most early will appear at the extremes of the list. Alternatively, by sorting on route, one can focus on each route's performance.

Reports dealing with APC data include the following.

- Stop summaries (a table showing, for each scheduled trip in a route-direction, ons, offs, load, and arrival time by stop, averaged over all observations of that scheduled trip in the specified date range).
- Segment utilization reports that summarize passenger counts to the segment level and find performance measures by segment such as the boardings to revenue-hours ratio and the passenger-miles to revenue-miles ratio (which is average load).
- Trip summaries. For each scheduled trip in a route-direction, one report shows average ons and offs and calculates boardings per revenue-hour and per revenue-mile. Another shows the point of maximum average load and the load and load factor (load divided by seats) at that point. A graph shows average peak load by scheduled trip. An interesting summary measure for the day is number of minutes above the loading standard. For through-routes, maximum load is identified on each side of downtown.
- Load exceptions. A sortable list of all scheduled trips whose average load at the peak point exceeds the load standard.

Also requested are reports of sample coverage, indicative of concerns about rotating the subset of APC instrumented buses. One such report is a list of blocks not yet sampled by an APC-equipped bus. Another will show, for each block-route combination, the percentage of trips covered at least once and the maximum number of times any single trip in that block-route was observed (a proxy for the number of times that block-route was sampled).

LOCATION MATCHING ISSUES

Analysts using archived Orion AVL data have learned to suspect the accuracy of running time data for the first and last segment of a trip, because the unpredictable trajectories buses can follow at the end of the line make it hard to determine the moment and location at which a trip ends and the next one starts. They expect to face this issue with the new AVL system as well, and look for improved algorithms to resolve these thorny end-of-line issues.