

Integrating Electronic Fareboxes with Other On-Board Equipment

**Final Report
Transit IDEA Project 4**

Prepared for
Transit IDEA Program
Transportation Research Board
National Research Council

Prepared by
Peter G. Furth
Northeastern University

September 22, 1995



TABLE OF CONTENTS

PROBLEM STATEMENT 1

RESEARCH APPROACH AND FINDINGS 1

 Review of Farebox Data 1

 Desirable Connections to the Farebox 2

 Review of On-Board Devices 3

 Review of Vehicle Area Networks (Van's)..... 4

 Open Integration Specifications 5

 Modeling Passenger Volume and Passenger-Miles 6

 Symmetry Assumption..... 6

 Estimating Daily Route-Level Passenger-Miles 7

 Estimating Trip-Level and Period-Level Volumes and Passenger-Miles 7

CONCLUSIONS..... 9

REFERENCES 9

APPENDIX A: Model Specification for Linking Headsign to Farebox..... 9

APPENDIX B: Model Specification for Linking Digital Speedometer to Farebox 10

APPENDIX C: Model Specification for Linking Non-Digital Speedometer to Farebox..... 11

APPENDIX D: Model Specification for Linking Door Sensor to Farebox 11

APPENDIX E: Model Specification for Linking Farebox to J1708 Vehicle Area Network 12

EXECUTIVE SUMMARY

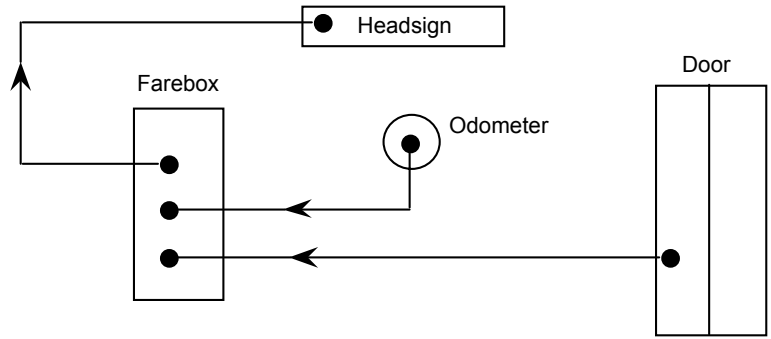
Electronic fareboxes, which can be found on the buses of almost every transit system in the U.S. and Canada, capture a large quantity of passenger use information. However, the value of this information is limited for various reasons. By integrating fareboxes with other on-board equipment, the value of the farebox data can be greatly enhanced for service monitoring, performance evaluation, planning, operations, and marketing. Historically, electronic fareboxes have been developed as stand-alone devices. In recent years, schemes for integrating fareboxes with other devices have begun to emerge. These include small scale schemes, simply connecting a farebox to another device, and broad scale schemes for connecting all electronic devices in a Vehicle Area Network (VAN). Still, the state of the practice of integrating fareboxes with other devices is far behind what is technologically possible and what is, from a management information point of view, highly desirable.

A review of farebox data systems found several opportunities for improvement through integration. One common problem is that operators sometimes fail to segment the data properly by indicating when they are changing route or direction. Integrating the farebox with the headsign would guarantee that the farebox would get any information the headsign gets about route and destination. Another problem is the need for more verification. Odometer or location stamps on farebox records would provide an excellent means of verifying route and trip segmentation. Transactional databases present the opportunity, by stamping boardings transactional records with location, to estimate passenger loads and passenger-miles. A new modeling technique was developed for this purpose based on the assumption that, over a day, the boardings profile on a route in one direction equals the alightings profile in the opposite direction.

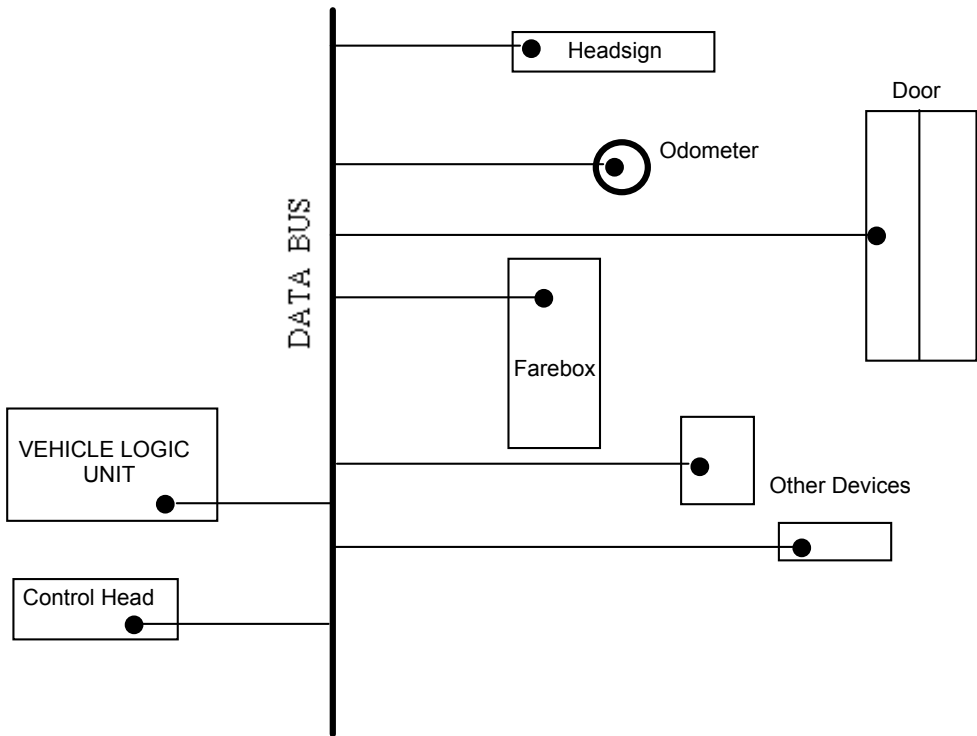
An industry review has pointed out means for integration that are technologically feasible and economical. As illustrated below, these include simple direct links connecting the farebox to the headsign, the speedometer / odometer, and the door sensor, as well as and more complex vehicle area networks connecting all of these devices to a Vehicle Logic Unit. To move the industry toward data integration, it is important that open specifications using industry standards be used. Model specifications were developed for direct links between fareboxes and three other devices, and for linking fareboxes to an SAE J1708 vehicle area network.

Methods to model passenger volumes using farebox data enhanced by a location stamp were developed and successfully tested. For daily totals, estimates were found to have very small errors, and the method could be recommended for estimating passenger-miles for Section 15. At the trip level, estimates of passenger-miles and volume at the peak point were found to have average errors of 6 and 7 percent, respectively.

DEDICATED COMMUNICATION OPTIONS



NETWORK COMMUNICATION OPTION



PROBLEM STATEMENT

Electronic fareboxes, which can be found on the buses of almost every transit system in the U.S. and Canada, capture a large quantity of passenger use information. However, the value of this information is limited for various reasons. By integrating fareboxes with other on-board equipment, the value of the farebox data can be greatly enhanced for service monitoring, performance evaluation, planning, operations, and marketing. Historically, electronic fareboxes have been developed as stand-alone devices. In recent years, schemes for integrating fareboxes with other devices have begun to emerge. These include small scale schemes, simply connecting a farebox to another device, and broad scale schemes for connecting all electronic devices in a Vehicle Area Network (VAN). Still, the state of the practice of integrating fareboxes with other devices is far behind what is technologically possible and what is, from a management information point of view, highly desirable.

Transit agencies and equipment manufacturers are in need of guidance for how to integrate fareboxes with other devices. Alternative configurations corresponding to differing budgets, available equipment, and managerial objectives are needed. For each configuration, model specifications driven by the need to use open specifications and architecture and available industry standards are needed. This kind of openness will be of benefit to transit agencies, who will not end up stuck with a single supplier. In the long run, they will also be of benefit to manufacturers, who will be able to market a single system to many agencies rather than develop one agency-specific integration scheme after another.

Transit agencies, equipment manufacturers, and information system developers are also in need of a summary of ways that integrated data can be used. One particular use of farebox data that stands in need of development is in estimating passenger load and passenger-miles. Fareboxes register passengers as they board, but not as they alight. However, making use of the near symmetry that exists over a day between boardings in one direction and alightings in the opposite direction (because people tend to board on a return where they alighted on the original trip), a model of passenger activity can be developed to estimate the load and the passenger-miles from current boardings data and the historical boardings profile of a route. One aspect of this project was the development and testing of this model for such uses as FTA Section 15 passenger-miles estimation, period-level peak volume estimation in support of scheduling, and

and real time load estimation in support of operations monitoring.

RESEARCH APPROACH AND FINDINGS

REVIEW OF FAREBOX DATA

First, an informal survey was conducted to establish how farebox data is used, what problems transit agencies have encountered with farebox data, and to what extent various forms of integration would be beneficial. Farebox data systems have historically been developed by the farebox manufacturers to be self-contained, and for many smaller transit agencies they work fine. However, five reasons were identified why, particularly at larger agencies, farebox data is not as useful as it was meant to be.

The first three reasons stem primarily from operator error. First, at logon, some operators do not enter the correct codes identifying themselves, their route, and their run. This kind of problem can be overcome by training and discipline, but it requires a verification system to spot the errors, and, within the organization, a means of feedback to the operator and supervisor.

Second, operators do not always enter the correct fare type. Simplifying fare policies and farebox codes and installing card readers can go a long way to reducing this problem. Further improvement requires operator training and discipline, again requiring a means of verification and feedback. Some agencies accept the fact that there will be a small amount of undercount (of free passengers, e.g., young children, in particular) and misclassification, and sample trips using on-board observers to estimate correction factors for undercount and misclassification (*1*).

Third, operators sometimes fail to segment the data properly by entering the new route or trip code when changing route or direction. This problem is widespread. Many transit agencies cannot use farebox data to determine route level boardings or revenue, only system level totals. Route level totals are a vital input for service monitoring and performance evaluation. Having route level data also enables one to estimate passenger-miles far more efficiently because it permits stratification by route or groups of routes

routes with similar average trip lengths. Our analysis found that the availability of route level totals can reduce the Section 15 ride check sample size by a factor of two to four. Segmentation by trip is vital if the data is to be useful for scheduling and operations monitoring because it provides volumes by direction and time of day. Using historical factors to convert boardings to peak load, trip level boardings data collected automatically by the farebox system can provide estimates of peak load on every trip, or, with more statistical reliability, for route / direction / period combinations.

A fourth problem is that farebox data systems lack important verification and feedback capabilities. Transit agencies have learned that it is vital that the data be verified and that feedback be given right away to operators and their supervisors. Without this kind of feedback, it is not uncommon for passengers to be undercounted by 10 percent and for 20 percent of the data to have an invalid or incorrect route number. With strong supervision, it is not unusual for these rates to drop to 2 percent. And even with a vigorous program of supervision, there will still be some faulty data. Because fareboxes provide such a large sample, once the general quality of the data is good, the agency can afford to discard questionable passenger data, again pointing to the need for strong verification checks.

However, with the standard farebox software, only a limited degree of verification is carried out automatically, e.g., testing for invalid ID's or out of range values. Further verification requires visual inspection of daily or monthly reports for each operator or bus, and manually flagging suspect data. Larger systems find it hard to provide this level of attention. Clearly, stronger verification tests, daily verification / exception reports, and a streamlined treatment of suspect data in statistical report are needed. Valuable verification tests might include testing the driver id and run number against the dispatch and / or payroll system; comparing the data segmentation with the run schedule; comparing the time or distance elapsed between records, or the recorded location at record creation, against the run schedule; and comparing counts against route specific historical averages.

Finally, to do anything with farebox data besides what's available in the standard system, the data must be exported to a general purpose database. While this is ostensibly a simple task, one agency found that developing a system to import, verify, edit, and arrange farebox data for useful analysis in a commercial database to be a rather involved task (2). Hopefully, agencies will share with each other applications developed for importing and processing farebox data.

DESIRABLE CONNECTIONS TO THE FAREBOX

Several transit systems have recognized that the key to getting the correct route and direction information into the farebox is the headsign (also called destination sign), since operators usually have to change the headsign with each change of direction and route. Operators are more conscientious about changing the headsign than the farebox because the headsign is seen by passengers, and the operators will get immediate feedback (complaints, comments from inspectors) if the headsign display is wrong. A few agencies have begun to integrate their headsigns and their fareboxes so that the headsign is operated from the farebox keypad. This is a very promising approach. Besides improving the quality of the farebox identifiers, it simplifies the operator's job and reduces the risk of injury to the operator who would otherwise have to stretch or stand up to use the overhead headsign keypad. However, integrating farebox and headsign will usually require adjustments to a transit's information systems, such as reconciling different route codes used by scheduling, headsigns, and publications.

Location or distance stamps on farebox records, similar to the time stamps which are already standard, is also desirable. This can be accomplished by connecting the farebox with one of the following: the speedometer (which includes the odometer), the transmission (whose output can be converted into distance traveled), or, where available, the automatic vehicle location (AVL) system. Distance or location stamps can be valuable for verifying that the data was properly segmented by simply comparing with the vehicle's or operator's run schedule. It is also possible in an advanced system to provide automatic segmentation if the AVL system knows where the bus is and its scheduled itinerary.

To stamp transactional records (stored either in a farebox with a transactional database, or in another on-vehicle computer) with location data, it will also be necessary to integrate the front door sensor. By the time passengers pay their fare, the bus is often underway. To stamp the record with the location of the last stop, it is necessary to know when the bus last opened its doors. Location or distance stamping on transactional records is the key to estimating passenger-miles and passenger loads. It is also vital that the end of each trip be properly recorded, so that the load can be "zeroed out" at that point. This requires

either an accurate AVL system with an on-board computer, or integration with the headsign and automatic verification of the segmentation location.

An on-board computer capable of managing and storing data from several devices that are integrated in a VAN is known as a vehicle logic unit (VLU). Connecting a farebox to a VLU, with the farebox providing data on each transaction, would permit an agency to create records in any format and integrate with any device it desires, and to process it independently of the farebox data system. Because a VLU's computing power dwarfs the farebox's, it makes sense for the VLU computer to serve as the host, receiving transactional data from the farebox, rather than for the VLU to send information to the farebox for storage there.

A VLU with its own primary console integrated to the farebox could also simplify and improve logs to the farebox. The VLU could tell the farebox (and the radio system and any other device) the driver ID and run number; it could store the scheduled itinerary; it could provide for automatic provision or immediate verification of some inputs (such as change in direction at the proper location). The VLU is bound to have a high capacity, easy to use method of transferring data onto and off the vehicle, and could serve as the channel for uploading data to fareboxes and other devices, such as new fare codes, tables of valid or invalid farecard serial numbers, or new headsign messages.

REVIEW OF ON-BOARD DEVICES

A review of pertinent on-board devices was conducted to establish what types of integration schemes were feasible and economical. Particular attention was paid to existing and planned examples of integration. Devices included in this review were fareboxes, radio systems, AVL systems, headsigns, annunciators, door sensors, speedometers, transmissions, PCMCIA card readers, and VLU's.

Fareboxes, while developed as self-contained systems, use digital communication links between their components, and have RS 232 and RS 485 ports that can be used for serial communication with other devices. A new farebox data system that is now being field tested stores data by transaction rather than creating trip level summary records. Farebox manufacturers take a cautious approach to integration with

other devices out of concern for the security of cash and data (which, depending on fare policy, can be almost the equivalent of cash when it is the basis for allocating revenue). They are more amenable to transmitting non-sensitive data (say, a record of a passenger boarding) to other devices than to receiving data that might have an unexpected effect on their system. They have not shown interest in upgrading their system's computers to enable them to serve as a general VLU.

Newer models of headsights are also able to connect with other devices using RS 232 or RS 485 serial ports. One transit agency, PACE, had its farebox and headsight suppliers develop a link so that, for routine route and destination changes, the headsight gets its inputs from the farebox. However, the protocol is the property of the suppliers, not of the transit agency, and is therefore proprietary. It is also not transferable to other makes of farebox or headsight. The New York City Transit Authority is procuring fareboxes that are likewise linked to the headsights using an RS 485 port. Again, a proprietary protocol is used.

Bus transmissions typically produce an output signal whose frequency is proportional to the frequency at which the transmission output shaft rotates. Speedometers receive this signal and use it to calculate speed and to drive an odometer. They have, when needed, produced an enhanced version of this transmission signal (sometimes buffered, sometimes converted to a square signal) to serve as input to AVL and cruise control. Digital speedometers capable of output to a VAN are also coming to market, although to date the demand has come from trucks rather than transit buses.

Annunciator systems announce the route and destination as well as next stop. While next stop announcements can be operator actuated, systems are being marketed that rely on automatic vehicle location as determined either using the annunciator's own location capability or by an independent AVL system. They are meant to be integrated with headsights and interior signs, either directly using standard serial ports (RS 232 or RS 485) or by connection to a VAN, and can likewise be connected to other devices. Because of the large amounts of data required for audio messages, one manufacturer uses Personal Computer Memory Card Industry Association (PCMCIA) cards for entering the day's itinerary and messages. PCMCIA cards can store large amounts of data, and, with added microelectronics, find application as fax / modems, paging devices, and global satellite positioning receivers. Because of the

computing power needed, an advanced annunciator's computer can serve as the VLU, using a second PCMCIA card to store and off-load data about the day's operations.

Radio based AVL systems have historically involved a good deal of integration with other devices, including the speedometer, the radio, and various mechanical systems for reporting mechanical alarms (e.g., overheating). Historically, they have not had provision for on-vehicle data storage, instead transmitting data in real time via the radio. The transmission capacity is severely limited by the number of radio channels an agency is allowed to use. To use the radio as a means of transmitting farebox data other than alarms or exceptions (e.g., a security alarm or a "cashbox full" message) would cripple the AVL system by forcing a drastic reduction in the polling rate. Historical AVL systems that have simply transmitted various inputs to a central computer that use these inputs to calculate location are not suitable for providing location input to either a farebox or an annunciator. However, an AVL system with its own on-vehicle storage and computing facility is well suited to integration with a farebox. The likely means for integration is through a VAN, since the AVL has to be connected with many devices, with the AVL computer serving as VLU. Data transfer onto and off vehicle could be accomplished through various means, including RF or infrared link, floppy disk, and PCMCIA card.

Electronic door systems include sensors connected to microswitches that open and close circuits to indicate (a) whether or not the door is fully closed, and / or (b) whether or not the door is fully open. These signals are usually used in connection with mechanical systems, e.g. to keep brakes engaged until doors are fully closed, or to ensure that a door is fully open before a lift begins operation. They have also served as input to automatic passenger counting (APC) systems, and could serve as input to a farebox or a VLU based system.

REVIEW OF VEHICLE AREA NETWORKS (VAN'S)

The Society of Automotive Engineers (SAE) has a set of standards, originally developed for trucks, for VAN's. Known as the J1708 "family" of standards, they include the standards J1708 (basic protocol) (3), J1587 (message content and format)(4), and J1455 (operating environment)(5). J1587 was revised in

1994 to accommodate for transit data, resulting in the first ever industry standard for a VAN for transit. Bill Kronenburger of Houston Metro was instrumental in this development, and has authored several documents describing the VAN concept (6, 7). The backbone of the J1708 VAN is a data bus (a simple twisted pair of copper wires) to which each device is linked to receive and transmit (or to receive only) via a modified RS 485 port. While there is provision for a (virtual) direct link to be temporarily established between a pair of the devices, for the most part messages are simply broadcast onto the data bus, where they become available to every other device. Devices must be programmed to send messages in the standard format at the appropriate time, and to recognize and receive messages that apply to them.

In terms of network topology, the VLU is just another node on the network. However, it will usually play a special role, although this role is not specified by any standard. The VLU will typically be programmed to manage data communications (e.g., initialize other devices at logon, request diagnostic and other data, provide software and data uploads) and to store data accumulated from the network over the day. In the configuration envisioned by Houston Metro, the VLU will have excess computing power and an appropriate operating system (e.g., UNIX) so that it can also serve as a platform for the software of other systems. In this way, another system (e.g., AVL or an annunciator system) would not have to provide its own physical computer, but would instead install its software on the VLU platform, where it would run in parallel with the VLU. When systems share the VLU host computer, communication with the VLU need not be through the VAN but through the host computer's operating system.

OPEN INTEGRATION SPECIFICATIONS

While proprietary schemes for integrating devices have been developed, they do not meet the challenge of the transit industry. Proprietary integration schemes often require license fees, tie the agency to keeping the same supplier of the integrated devices, and cannot be shared with other agencies who might wish to issue an open specification to invite competition. Furthermore, open specifications, in the long run, are in the manufacturers' best interests, because they vastly increase the likelihood that a system for integration developed for one agency can be reused at other agencies. Industry experience has shown again and again

again and again that system development (except for very small systems) for a single customer costs more than the customer wants to pay and the supplier wants to invest.

There are two keys to open specifications. The first is the use of available industry standards. The emergence of the J1708 standards is a great step forward for the industry. Agencies desiring to link devices through a VAN can simply reference the J1708 standards, and manufacturers can develop the necessary software and communication ports. J1455 is particularly valuable as a means of specifying the operation environment a device must be able to operate under, since the applications for which it was developed (trucking) is a lot closer to the transit environment than the military applications envisioned by many electronics standards. However, it is not sufficient to specify that "a device shall be J1708 compatible" since there is some vagueness in the standards as to what exactly is expected of a device in terms of messages it should be able to transmit and receive, and how it is to respond to various messages. These details should be included in a specification.

The second key to an open specification is that the transit agency should make the specification. If it is modified through interaction with suppliers, which may certainly be appropriate, the agency should insist on ownership of the final specification, precluding proprietary protocols. Transit agencies, which are not in competition with each other, freely share specifications with each other so that even if specifications evolve they remain open to the industry as a whole.

Model specifications for a variety of integration schemes were developed. Agencies that are not in a position to procure a VLU will want to use direct links between a few devices. Model specifications are given for connecting the farebox with a headsign (Appendix A), a digital speedometer (Appendix B), a non-digital speedometer (Appendix C), and a door sensor (Appendix D). Agencies that are procuring an AVL system or an advanced annunciator system are in a position to use their computer as a VLU, and will want to communicate via a VAN. Model specifications for connecting a farebox to the J1708 VAN are given in (Appendix E). It covers two alternative levels of integration. In Level A the farebox only transmits data to the network. In Level B, the farebox also receives and responds to messages from the network, enabling the VLU to exercise some control over the farebox. A further level, in which the farebox keypad is never used (the VLU console is used to operate the farebox) is not covered.

The model specifications for direct links have been developed following two principles. First, be as consistent as possible with industry practice. Second, where new protocols are needed, use J1708 protocols since manufacturers will be developing J1708 communication facility for other transit and trucking procurements. The model specification for connection to the J1708 VAN is based in part on a specification developed by Houston Metro.

The model specifications are not complete in the sense of including everything needed for a bid document. For example, they do not address important contract issues such as acceptance testing. They also do not specify the physical connection hardware. There are several commercially available RS 485 ports, for example. Houston Metro has been sharing its specification for connection hardware among transit agencies, and it may become an SAE standard. The purpose of these model specifications is to lay out a communication protocol that will not only meet transit agency needs, but will also guide system development by manufacturers of fareboxes, headsign, speedometers, and door sensors. If manufacturers respond to these specifications (either as they now appear or embedded in transit agencies' future bid documents) and develop the communications facility contained therein, they will soon begin marketing those capabilities to recoup their investment, and integration will spread quickly through the industry.

Finally, it should be noted that SAE standards are subject to revision using established procedures. Agencies and manufacturers who feel that it would be in the best interest of the industry can request revisions to the J1708 standards. Based on this study it would appear that two minor revisions seem appropriate with respect to farebox PID's contained in J1587. One suggested change is for the transmission update period for PID 502 (new words are shown in italics): Transmitted at the start, end, in service, and out of service event, *and on change*. For PID 503, a suggested change is that characters a and b be simply agreed upon by the operating agency and the fare collection equipment manufacturer. This will provide more flexibility for differing fare policies. With two characters, boarding passengers can be assigned to one of more than combinations of 65,000 fare category, passenger type, and transaction type. In the current specification, a maximum of 12 transaction type, 16 passenger categories, 12 fare detail types, and 12 upgrade detail types are supported. This appears overly restrictive.

MODELING PASSENGER VOLUME AND PASSENGER-MILES

Electronic fareboxes do register nearly all boardings by fare category, but unlike automatic passenger counters, they do not count alightings. One commonly used means of estimating passenger-miles is to multiply total boardings, as counted by the farebox system, by average passenger trip length (8). This trip length factor is estimated from a sample of ride checks (on / off counts) as the ratio of the total passenger-miles in the sample to total boardings in the sample. This is a substantially more effective means of estimating passenger-miles for Section 15 than either of the FTA published methods, especially if boardings totals are known by route and factors are estimated for each route, or groups of similar routes. The same methodology can be applied to estimating peak loads from trip level boardings counts, using route / direction / period specific ratio factors developed from a sample of ride checks on each route.

The drawback to these estimation procedures is that, while they make efficient use of the the essentially limitless sample of boardings counts provided by the farebox, they still rely on expensive ride check samples. A method that was developed that relies solely on farebox data, provided that (1) the data was stored by transaction, and (2) each transactional record contains an odometer stamp.

Symmetry Assumption

In response to the unavailability of alightings data, a daily symmetry assumption is proposed: *The boarding pattern for a route in one direction is equivalent to the alighting pattern in the opposite direction over the course of an entire day.* At the home end, the stop at which people board is the same stop at which they alight on the ride home. Likewise, at the destination end, people board for their return trip where they alighted earlier in the day. To test this assumption, a full day's set of on-off counts was obtained for six Los Angeles Metro bus routes: 10/11, 60, 105, 236/240, 251/252, and 418.

Complicating the comparison of the boarding and alighting patterns is the lack of perfect one-to-one correspondence between stops in opposing directions. A route may have a greater number of stops in one direction or the bus may not stop at the same intersection in both directions. A more serious complication is that the alignment of many bus routes, especially among the test routes, combines a trunk with a number of route deviations (called branches by Los Angeles Metro). For example, around 3 pm a few trips serve a number of stops on a route deviation that passes by a school. Because these deviations are small, many passengers would just as readily use a trunk stop or a deviation stop, depending on where the bus goes at the time of day they are traveling.

To account for these complications, a single “model route” was developed for each direction of the Los Angeles routes. The model route follows the trunk, allocating deviation stops to the closest trunk stop. The number of stops on the model route is the same in both directions and all the stops have a one-to-one correspondence. Where necessary, inbound and outbound stops were matched through a proportional allocation. For example, if an inbound stop falls half-way between two outbound stops, 50 percent of the passenger activity at the inbound stop is allocated to each outbound stop.

A statistical test of the validity of the symmetry assumption was made using the Kolmogorov-Smirnov goodness-of-fit test. Observed alighting profiles were compared with the hypothesized alighting profile, which is the opposite direction's boarding profile, scaled to match the total volume. The Kolmogorov-Smirnov test is more appropriate than the Chi-square goodness-of-fit test since the Chi-square test does not recognize how close stops are to each other, but simply regards each stop as a separate category. A rider that is misplaced by one stop is punished in the same manner as one that is misplaced by many stops. In contrast, with the Kolmogorov-Smirnov test a misallocation of alightings is serious only if it results in a large difference in volume somewhere along the route. Its test statistic is the greatest arithmetic difference between the observed and hypothesized volume profile along the route. (Because volume equals cumulative boardings, which are known, minus cumulative alightings, the difference in volumes is the same as the difference in cumulative alightings, which would be the standard test statistic.) Of the twelve profiles tested (one for each route and direction), only two were rejected at

the 95% significance level. These results, especially considering the adjustments made due to route deviations, seem to validate the symmetry foundation needed for estimating volumes and passenger-miles.

Estimating Daily Route-Level Passenger-Miles

Using the symmetry assumption, total passenger-miles was estimated for each of the the six Los Angeles bus routes. Because of route deviations, the distance between two stops may be differ from one trip to another. We again used a linear model route, assigning to each stop a single consistent distance from the head of the route. To determine the distances, each deviation was isolated from the trunk at its junction points, and the distances between junction points was measured along the deviation (d_v) and along the trunk (d_t). From the route schedule, the frequency with which the bus travels between those junction points along the deviation, q_v , and along the trunk, q_t , are determined. For the model route, the distance between junction points is the average of the deviation and trunk distances, weighted by their corresponding frequencies, or $(q_v d_v + q_t d_t) / (q_v + q_t)$. Distances between intervening points were scaled accordingly.

Passenger-miles estimates were calculated for each model route by summing over all the segments the estimated volume along that segment multiplied by segment length. As described earlier, estimated volume on a segment equals cumulative boardings up to that segment minus cumulative estimated alightings, and alightings at each stop are estimated to equal actual boardings at that stop in the opposite direction, factored so that total daily estimated alightings in each direction equals total daily boardings.

When passenger-miles estimates were compared to the actual passenger-miles calculated directly from the data (with actual mileage on route deviations), the relative errors in the estimates for the six routes were found to be quite small: -1.3%, 0.4%, 1.3%, 1.9%, 2.6%, and 3.5% (where a positive value indicates an overestimate), for an average absolute relative error of 1.8%. Because such a method, using farebox data, is not subject to sampling error, these estimates are well within the FTA Section 15 standard of 10% precision at the 95% confidence level.

While there appears to be nothing inherent in the model that would tend to cause overestimation, a test was made to determine whether, in view of the preponderance of overestimates, the model is biased, or whether the six observed errors could be explained by random error alone. A null hypothesis that there is no systematic error in the model was accepted at the 95% confidence level.

Estimating Trip-Level And Period-Level Volumes And Passenger-Miles

Symmetry in travel cannot be assumed for a period smaller than a day. For example, the number of boardings in a direction in the morning is not mirrored by the number of alightings in the opposite direction during the same morning period. The methodology for estimating daily passenger-miles cannot be directly applied to estimating trip-level or period-level volumes and passenger-miles.

Passenger travel patterns are best described by an origin-destination (O-D) matrix. Each cell in the matrix indicates the number of trips between a pair of stops. Each row total is the number of boardings at a particular stop, and each column total the number of alightings at a stop. Our approach was to estimate the trip-level O-D matrix, and reduce it to boardings and alightings totals, from which are determined the volume profile and passenger-miles.

The doubly constrained gravity model was chosen to estimate the number of trips between stops i and j , t_{ij} . The model is mathematically represented as $t_{ij} = s_{ij}A_iB_j$ where s_{ij} is a seed describing the propensity to make a trip as a function of its distance d_{ij} . A_i is a row-specific factor representing the popularity of each origin stop, and B_j is a column-specific factor representing the popularity of each destination stop.

Previous modeling efforts have used as seeds small O-D surveys or mathematical fits to observed trip length data. We chose a propensity function that would not require an O-D sample on each route, one that includes the standard exponential decline of propensity with distance, yet increases with distance for the first half mile or so to account for competition with walking. An exponential decay model was assumed for both walk and for total non-automobile travel, albeit with different decay rates. The bus propensity is taken to be the difference between total travel propensity and walk propensity, resulting in the propensity function

$$s_{ij} = e^{-\beta d_{ij}} (1 - e^{-\alpha d_{ij}})$$

To utilize this model, the parameters α and β needed to be estimated. Alpha was estimated to be 0.5 using a maximum likelihood methodology from O-D data collected on several Boston bus routes. The maximum likelihood methodology gave better results for this propensity function than for propensity functions used in other modeling efforts, including the exponential and gamma functions. As explained in Navick and Furth (9), β is not needed to generate trip level O-D matrices and can be set equal to zero.

To estimate trip-level O-D matrices for given direction, first a triangular seed O-D matrix is generated for each trip of the day. The seed matrix is simply the matrix of distance based propensities for O-D pairs over which travel is possible in the chosen direction (it contains zeroes on and below the diagonal), and is the same for every trip. These seed matrices are then stacked one below the next. Target row totals are boardings at each stop on each trip, as would be registered by a farebox with a location stamp. Target column totals (these totals include cells of all the trips) are the daily total alightings determined using the symmetry assumption: total daily boardings in the opposite direction at each stop, scaled so that total estimated alightings match total boardings. The matrix is then balanced using an iterative procedure known as the biproportional method. Rows are first scaled as needed to match the target row totals, then columns are scaled to match their target totals. This will unbalance the rows, so the procedure is repeated until both the row and column totals match their target values. Convergence is guaranteed.

To test how well trip level volumes are modeled, we did two comparisons. First, we compared actual to estimated passenger-miles for all the trips in the daily schedule for the six Los Angeles routes. The average absolute relative passenger-mile error for all trips was 5.9%, with overestimates more or less balancing underestimates. The second comparison was of load at each trip's maximum load point (as determined from the true, not estimated, volumes). On average, the relative absolute error was 7.3%. This is more accurate than many estimates now being used for scheduling in transit agencies across the country. For example, we know of a few statistical studies that found point checks to have average measurement errors around 13%; on top of that, one should add sampling error. Ride checks have little

measurement error, but are subject to large sampling errors since only a tiny fraction of a years' trips are typically sampled. An estimation procedure based on enhanced farebox using this model would have no sampling error. It would require a large one-time effort to implement the model, and a small ongoing effort to verify its estimates, and would provide a wealth of operations data.

The foregoing procedure estimated trip-level passenger loads after the day is done. To estimate loads in real time – say, in order to report via a radio system when load has exceeded a certain threshold so that operations adjustments can be made – another, non-iterative procedure would be used. Using an aggregation of daily estimated O-D matrices, a family of alightings distributions can be generated for each boarding stop, with each distribution indicating the fraction alighting at that stop. A family of alightings distributions can be generated for each period, or even each trip. A simpler variation of this procedure is to generate a survival profile indicating the fraction of passengers on board on each segment (regardless of where they boarded) that do not alight at the following stop. A single survival profile for the whole day could be generated from daily estimated volume profiles; alternatively, period level or even trip level survival profiles could be generated from estimated trip level volume profiles. Either way, estimated passenger volume can tracked along the route and corrective action taken as appropriate.

CONCLUSIONS

Many ways have been outlined by which integration with other on board devices could enhance the quality and usefulness of farebox data. An industry review has pointed out means for integration that are technologically feasible and economical, including simple direct links and more complex vehicle area networks. To move the industry toward data integration, it is important that open specifications using industry standards be used. Model specifications were developed for direct links between fareboxes and three other devices, and for linking fareboxes to a J1708 vehicle area network.

Methods to model passenger volumes using farebox data enhanced by a location stamp were developed and successfully tested. For daily totals, estimates were found to have very small errors, and the method could be recommended for estimating passenger-miles for Section 15. At the trip level,

estimates of passenger-miles and volume at the peak point were found to have average errors of 6 and 7 percent respectively. These methods, which are capable of providing highly detailed estimates of passenger activity free of sampling error, are an example of what can be done with integrated farebox data.

REFERENCES

1. M. K. Martin. *Using Farebox Data to Accurately Account for SCCTA Bus Ridership*. Report to the Santa Clara County Transportation Authority, 1993.
2. T. C. Dawson and D. Ryan. *Integrating Electronic Fareboxes at the CTA*. APTA Operations and Service Planning Symposium, 1993.
3. Society of Automotive Engineers. *Serial Data Communications Between Microcomputer Systems in Heavy-Duty Vehicle Applications*. Standard SAE J1708, 1993.
4. Society of Automotive Engineers. *Joint SAE/TMC Recommended Practice for Electronic Data Interchange Between Microcomputer Systems in Heavy-Duty Vehicle Applications*. Standard SAE J1587, 1994.
5. Society of Automotive Engineers. *Joint SAE/TMC Recommended Environmental Practices for Electronic Equipment Design (Heavy/Duty Trucks)*. Standard SAE J1455.
6. W. Kronenburger. *An Overview of the IVHS / FHWA / SAE Standard for a Vehicle Area Network*. FTA Vehicle Area Network Subcommittee draft release, 1994.
7. W. Kronenburger. Miscellaneous materials related to vehicle area networks, 1995.
8. P.G. Furth and B. McCollom. Using Conversion Factors to Lower Transit Data Collection Costs. In *Transportation Research Record 1144*, pp. 1-6, 1987.
9. D.S. Navick and P.G. Furth. Distance Based Model for Estimating a Bus Route O-D Matrix. In *Transportation Research Record 1433*, pp. 16-23, 1994.

APPENDIX A: MODEL SPECIFICATION FOR LINKING HEADSIGN TO FAREBOX

1. Purpose: To enable route and destination information to be sent from the farebox to the headsign, making the farebox keypad the place by which the operator changes the headsign message for route / destination changes. The headsign shall still be operable using its own keypad in case the farebox doesn't work, or to display non-routine messages.

2. The farebox shall transmit messages, and the headsign receive messages, using an EIA (Electronics Industries Association) RS 485 port.

3. The message format shall conform to SAE J1708 Section 5.

4. Only one type of message shall be sent. It shall contain route and destination information. Its format shall conform to SAE J1587, parameter 501 (described in Appendix A.501). Specifically, the format, character by character, shall be

MID PID1 PID2 n1 a1 b1 b2 b3 ... PID2 n2 a2 c1 c2 c3 ... Checksum

where

MID = 196 (module ID for farebox)

PID1 = 255 (parameter ID for page 2 extension)

PID2 = 245 (page 2 ID for signage message)

n1 = number of characters including only a1 and b1, b2, ...

a1 = R (indicating that what follows is a route code)

b1, b2, ... = ASCII route code where b1 is the most significant character

n2 = number of characters including only a2 and c1, c2, ...

a2 = D (indicating that what follows is a destination code)

c1, c2, ... = ASCII destination code where c1 is the most significant character

Checksum = the two's complement of all the characters in the message, excluding the Checksum itself

MID, PID1, PID2, n1, n2, and checksum are unsigned short integers. The remaining characters are ASCII alphanumeric.

5. The message shall be transmitted every 10 sec. The headsign shall update its display upon receipt of the message.

6. The agency shall specify a default route and destination code to be used when the farebox is powered up (typically the "Out of Service" code). The agency's farebox system manager shall have the facility to set and change these codes.

7. The farebox shall determine the route and destination code to transmit to the headsign based on operator inputs to the farebox in one of the following ways:

a. The route and / or destination code shall be entered directly to the farebox, either by direct keying or by means of an operator's selection from a menu of codes.

b. A code not directly entered into the farebox shall be determined from farebox inputs by means of a lookup table stored in the farebox system memory. The agency's farebox system manager shall have the facility to set and change the lookup table.

c. For agencies that use only a route code or only a destination code on their headsigns, the unused code shall either stay at its default value or be set equal to the other code.

APPENDIX B: MODEL SPECIFICATION FOR LINKING DIGITAL SPEEDOMETER TO FAREBOX

1. Purpose: To enable distance traveled (i.e., odometer) information to be sent from the speedometer to the farebox so that the farebox can stamp the records it creates with an odometer reading.

2. The speedometer shall transmit messages, and the farebox receive messages, using an EIA (Electronics Industries Association) RS 485 port.

3. The message format shall conform to SAE J1708 Section 5.

4. Only one type of message shall be sent. It shall contain odometer information. Its format shall

shall conform to SAE J1587, parameter 245 (described in Appendix A.245). Specifically, the format, character by character, shall be

MID PID n a a a a Checksum

where

MID = 145 (module ID for road speed indicator)

PID = 245 (parameter ID for total vehicle distance)

n = 4 = number of characters following excluding Checksum

a a a a = total vehicle distance, with each bit representing 0.16 km (0.1 mi)

Checksum = the two's complement of all the characters in the message, excluding the checksum itself

MID, PID, n, and Checksum are unsigned short integers. a a a a is an unsigned long integer.

5. The message shall be transmitted every 10 sec.

APPENDIX C: MODEL SPECIFICATION FOR LINKING NON-DIGITAL SPEEDOMETER TO FAREBOX

1. Purpose: To enable distance traveled (i.e., odometer) information to be sent from the speedometer to the farebox so that the farebox can stamp the records it creates with an odometer reading.

2. The speedometer shall continuously transmit a TTL compatible clock signal whose frequency is proportional to the ground speed.

3. The farebox shall receive the speedometer signal into a 32 bit counter which will count the number of cycles in the speedometer signal since the farebox was turned on. The farebox computer shall either interrogate the counter when an odometer reading is needed, or program the counter to transmit the count at least once every 10 sec.

4. The farebox shall calculate distance traveled since it was turned on using the count and conversion factors stored in farebox memory related to tire diameter and rear axle ratio. The farebox system shall provide for an easy way to enter and adjust those conversion factors, such as dip switches corresponding to standard tire sizes.

APPENDIX D: MODEL SPECIFICATION FOR LINKING DOOR SENSOR TO FAREBOX

1. Purpose: To enable the farebox to know when to mark the odometer reading in a farebox system that keeps transactional records. The odometer reading at the time the door was last opened shall be the odometer reading stamped on a transactional record, so that that transactions that occur after the bus has left the stop will be marked with the stop location rather than the location at the moment of the transaction.

2. Two circuits shall connect the door sensor to the farebox. In the first circuit, positive voltage ($5\text{ V}\pm$) shall indicate that the door is fully closed; ground shall indicate that the door is not fully closed, or that there is an error condition. In the second circuit, positive voltage ($5\text{ V}\pm$) shall indicate that the door is fully open; ground shall indicate that the door is not fully open, or that there is an error condition.

In most door systems, the middle door (if there is one) opens and closes with the rear door, making a link from the middle door unnecessary. In many transit systems, the rear door is opened only when the front door is opened; where this is the case, a link to the front door only is needed. Otherwise a link will only be needed from both the front and rear doors.

3. The farebox computer shall maintain the door status of each door in an 8-bit register based on the voltage in the two circuits. These definitions follow in part those of SAE J1587, parameter 379 (described in Appendix A.379). Bits 8-3 will all be set to 1. Bits 2-1 will be as follows:

- 00 Door fully closed (and not fully open)
- 01 Door fully open (and not fully closed)
- 10 Error condition (circuits indicate that door is fully open and fully closed)

11 Door neither fully open nor fully closed

4. The farebox computer shall maintain the status of the full set of doors in an 8-bit register based on the door-specific registers. Bits 8-3 will all be set to 1. Bits 2-1 will be as follows:

00 All doors fully closed

01 At least one door fully open

10 Error condition and no doors fully open

11 No doors fully open and at least one door not fully closed

5. Within the farebox computer, the register containing the status of the full set of doors shall be accessible to logic for determining when doors open, so that instructions to record odometer reading when the doors open can be issued.

APPENDIX E: MODEL SPECIFICATION FOR LINKING FAREBOX TO J1708 VEHICLE AREA NETWORK

Depending on the desired level of interaction of the farebox with the rest of the vehicle's data systems, the farebox may be required to support more or fewer of the possible facilities of a J1708 network. Two levels of interaction are likely to be desired by a transit agency: At level A, the farebox only transmits data. At Level B, the farebox also receives and responds to messages broadcast on the VAN by the vehicle logic unit (VLU).

1. The farebox shall be able to connect to the vehicle's data bus in conformity with SAE J1708 and its related standards. For Level A: The farebox shall only transmit data onto the data bus. For Level B: The farebox shall both transmit to and receive data from the data bus.

2. The farebox shall support the following MID's:

196 (fare collection unit)

188 (vehicle logic unit) (Level B only)

3. The farebox shall transmit the following PID's to the network:

502 service detail

503 point of sale detail

378 farebox alarms

259 restart response (Level B only)

243 component identification (report make, model , serial number) (Level B only)

194 diagnostic code (transmit when farebox detects an error condition) (Level B only)

196 response to diagnostic data / count clear request (Level B only)

Note. The transmission update period for parameter 502 shall be different from the J1587, as follows (the change is shown in italics): Transmitted at the start, end, in service, and out of service event, *and on change.*

Note. Some change in parameter 503 may also be appropriate. See text for details.

4. For Level B only, the farebox shall receive and respond to the following PID's when contained in messages from the VLU (MID 188):

257 cold restart request

258 warm restart request

000 general request

128 request to specific module

195 request for diagnostic data or to clear a diagnostic code count

251 clock time (used to reset farebox clock)

252 date (used to reset farebox clock)

507 operator ID (used as input to farebox system for stamping farebox records)

502 service detail (input needed at farebox logon)

Note: If service detail data (contained in PID 502) is entered via the farebox keypad, the farebox should transmit PID 502; but if the service detail data is entered otherwise and is transmitted via the data bus to the farebox, the farebox should receive and respond to, but not transmit, PID 502.

5. PID's 502 and 503 contain some characters that are to be agency defined, or jointly agreed upon by

xxx

by the agency and the farebox manufacturer. They depend upon fare categories and other aspects of the fare policy. They should either be included in the original specification, or it should be indicated that those characters will be agreed upon sometime during the course of the project.