23rd Australasian Transport Research Forum
Perth, Western Australia. 29 September – 1 October 1999

Roadlink: A Model for Analysing Vehicle Routes From Household Travel Surveys

Craig McPherson
Sinclair Knight Merz, Victoria

Abstract

Four-step planning models commonly represent travel demand as an aggregate set of zonal productions and attractions. While this modelling paradigm is quite satisfactory for typical strategic planning and local impact studies, there are occasions where modellers need to know not only *how many* people are travelling, but *who* and *why* they are travelling. This paper shows how household travel survey data can be integrated into a point-to-point traffic assignment technique to maintain demographic and trip purpose information throughout the modelling process. The technique is embodied in a GIS-based model, called "RoadLink". The RoadLink model is used with data from the Victorian Activity and Travel Survey (VATS) to model traffic flows in Melbourne and shown to be a reasonable predictor of general travel patterns. The model is applied in a brief case study to demonstrate the rich demographic outputs potentially available from a disaggregate assignment process.

Keywords
traffic assignment, GIS, household travel survey, VATS, disaggregate modelling, computer applications in transport

Contact Author

Craig McPherson
Sinclair Knight Merz
PO Box 2500
Malvern Victoria 3144

Phone: +61 3 9248 3331
Fax: +61 3 9248 3460
e-mail: CMcPherson@skm.com.au

Introduction

Strategic urban planners often use traffic and transit assignment models to estimate flows of vehicles and people in transport networks, and to compute associated travel times and costs. Such models provide information for other planning applications, such as benefit-cost analyses, environmental impact assessments and public transport patronage projections.

Many planning applications use aggregate measures of transport demand. Common aggregate measures include total travel from one network zone to another, accumulated vehicle flows on links and total network-wide vehicle kilometres travelled. Aggregate measures of travel demand are quite satisfactory for most strategic planning purposes and are relatively efficient for models to generate.

There are some applications, however, where aggregate measures of demand do not provide the detailed information that analysts may require. A current Australian case study provides a good example. The City Link tollway, recently opened in Melbourne, now connects three freeways that formerly terminated at the northern, western and eastern sides of the city's central activity district. City Link is privately owned and operated, with revenue raised from electronically-collected tolls. Toll collection is facilitated by vehicle-mounted transponders that identify the carrier's tollway account. Detection equipment mounted on overhead gantries detects the presence of vehicle transponders and enables the appropriate toll to be deducted from tollway users' accounts.

Critical questions in the tasks of forecasting toll revenue and marketing transponders are: "*Who* will use the tollway?", "*Why* will they use it?", "*How many* people will travel on it" and "*Where* will a transponder marketing campaign be most effective?" (*The Age*, 17 March 1998). A traditional traffic assignment could answer the "how many" and "where" questions, but would be ill-equipped to answer the "who" and "why" questions. Driver demographics could be collected by personal interviews or roadside surveys, but this would be unconnected to the outputs of assignment models. The problem is succinctly summed up by Handy (1996):

> *"Most data on household travel are collected at a disaggregate level, usually through household travel surveys, but researchers often analyze aggregate statistics, rather than individual records."*

How can the rich demographic information of a travel survey be incorporated directly into an assignment process? This paper gives an overview of the issues involved in answering this question, discusses how traditional assignment processes could be enhanced with demographic information, and presents some results of an actual working model, called *RoadLink*.

Aggregate Demand and Supply Models

The traditional four-step modelling process (see for example Ortúzar and Willumsen 1994) uses an aggregate representation of demand in each of the four steps (see Figure 1). The classical four steps are reiterated below.

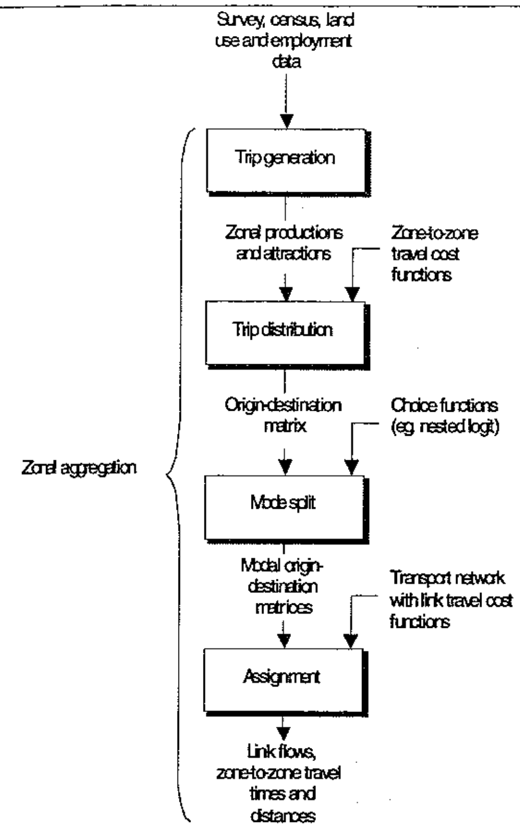

Figure 1: Aggregate four-step modelling

1. *Trip generation*: Land use zones are identified in the study area and the numbers of trips moving into and out of each zone are determined.
2. *Trip distribution*: The number of trips moving between each pair of zones is modelled.
3. *Mode split*: Trips are apportioned between the various available modes, producing an origin-destination (OD) matrix for each mode.
4. *Trip assignment*: Routes are found between each origin and destination zone, usually by minimising the user cost of each trip. The trips assigned to each link are summed to give the total flow on each of the network's links.

The inputs to this process typically include:

- travel survey and census data which describe household structure and demographics, employment location and type, and the trip-making behaviour of individuals;
- a zone system based on the underlying land use in the study area;
- mode-choice equations and parameters;
- a computer representation of the transport network;
- equations describing the cost of travel on network links as a function of the flow and the operational characteristics of the links.

From the outset of the four-step process, all the individual travel characteristics derived from the travel survey, census and land use models are absorbed into aggregate zonal trip rates. Even if discrete choice modelling techniques are used, in which individual travel decisions are modelled (see for example Ben-Akiva and Lerman 1985), the results are still typically consolidated into a zonal OD matrix prior to assignment

While the representation of demand is generally aggregate in nature, so too is the representation of supply – that is, the transport network. In the case of a road network, traffic sources and sinks are characterised by a set of **zone centroids** (there naturally being one centroid for each zone). The centroids are connected to the road system by dummy links (**centroid connectors**) Centroid connectors also represent vehicle delays in local streets, as the network itself usually comprises only major arterial road links.

With the advent of geographic information systems (GIS), there has been a trend towards more detailed representation of spatial data, including transport networks. A number of modelling software packages have now moved towards an integrated GIS and modelling paradigm, for example Caliper Corporation's TransCAD modelling package Such an approach allows the interaction of transport models with a variety of other spatial data – for example census population and employment data, satellite photographs, pedestrian and bicycle networks, regional economic projections and asset management systems.

With the popularity of GIS, ready access to detailed spatial data and exponential leaps in computer processing power, it is reasonable to assume that transport models will become more detailed in the representation of networks and of individual travel (Spear 1996). The modelling approach described in the following sections adopts this philosophy.

A disaggregate approach

In the earlier overview of conventional four-step modelling, we noted that travel demand is expressed in a zone-to-zone form early in the modelling process. The key to achieving disaggregate outputs from the assignment step is to maintain individual travel and demographic information through each of the four steps in the process In other words, the modeller only aggregates the travel data *after* the modelling process is complete. In doing so, the modeller has considerable flexibility in the way he or she groups, analyses and reports the assignment results.

How can travel demand be kept in a disaggregate form throughout the process? The approach adopted in the RoadLink model (described later) is as follows:

- express demand as a collection of point-to-point trips;
- give a each trip an expansion factor so that the volume of expanded trips is equivalent to the total trip-making population;
- determine the geographic coordinates of the origin and destination of each trip in the transport network (commonly referred to as **geocoding**); and
- assign each trip to the network, using a point-to-point assignment algorithm

Each of these steps is discussed in further detail below.

Firstly, demand is expressed as a collection of point-to-point trips, rather than zone-to-zone flows. Each trip is treated individually, with its own origin, destination, driver characteristics and trip purpose. A formal zone system is not required, since the origin and destination points of the trips will be scattered throughout the network.

The establishment of an appropriate pattern of point-to-point demand is possibly the most difficult part of the disaggregate modelling process, both in concept and practice. The method adopted in the RoadLink model is to use responses from a household travel survey directly as a point-to-point demand sample. A household travel survey is a convenient choice, since it will typically collect information on travellers' origins, destinations, trip purposes and personal characteristics. The sampling scheme and non-response rates in the survey will control the expansion factors given to each reported trip (McPherson 1999).

An appropriate survey would need to include

- a complete geographic coverage of the study area; and
- a sufficiently large sampling rate to obtain a fine resolution of origin and destination points.

If external trips (that is, trips that originate or terminate outside the study area) are not collected by the survey, then a pattern of external trips will need to be derived by other means.

Once a demand profile has been created, each trip's endpoints are geocoded to appropriate points in the network. In a conventional computer representation, a network consists of a set of links connected at node points. The nodes will commonly occur at intersections, but may also be present at mid-block points to maintain the geographic accuracy of a curving road. The geocoding process determines the node that best matches a stated trip endpoint (usually – but not always – the closest node to the endpoint).

The final step in the disaggregate modelling process is to assign each of the individual trips onto the network. This is done in a point-to-point fashion, as individual paths are generated for each origin-destination pair. The following section contrasts the disaggregate assignment process with some of the more commonly used assignment techniques.

Disaggregate assignment and path-finding algorithms

Conventional equilibrium assignment methods are generally formulated as optimisation programs in which drivers choose routes in such a way that "... no driver can reduce his [or her] journey time by choosing a new route." (Wardrop 1952). Extending this in more general terms, drivers seek to minimise some perceived travel cost (of which time is a component). Readers interested in assignment algorithms are invited to consult one of the many texts on the subject (for example, Sheffi 1985, Bell and Iida 1997, Ran and Boyce 1996).

A key component of any assignment procedure is the technique used to find optimum paths in the network. The optimum path problem is well-documented in the operations research literature. In a zone-based trip assignment, the path-finding algorithm needs to find optimum paths from each zone to every other zone – in other words a many-to-many mapping of paths

Disaggregate assignment, on the other hand, requires a one-to-one mapping from single origins to single destinations. Each trip record in the demand data set is assigned in a point-to-point manner by determining a single path from origin to destination. The classical many-to-many and one-to-many algorithms, such as those of Floyd (1962) and Dijkstra (1959), are often not the most efficient choice for one-to-one path-finding. The method used in the RoadLink model is an efficient heuristic path-finding technique based on the A* algorithm (Hart, Nilsson and Raphael 1968). The A* (pronounced "A-star") algorithm limits the amount of searching for alternative paths in the network by "intelligently" guiding the search from the origin towards the destination (see Figure 2). A comprehensive discussion of the heuristic functions that guide the path-finding is presented in Pearl (1984).

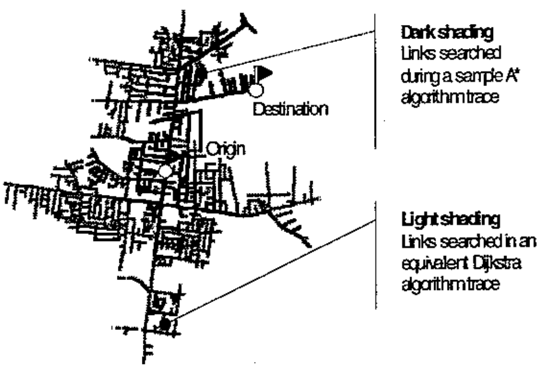

Figure 2: A comparison of conventional (Dijkstra) and A* algorithm efficiency

As a further refinement of the path-finding process, the RoadLink model uses route information provided by the survey respondents to enhance the realism of the paths. For example, if a survey respondent said that she travelled from central Melbourne to the inner eastern suburb of Hawthorn using Hoddle Street, Barkers Road and Glenferrie Road, then the RoadLink model would ensure the path included these named streets. A complete description of the algorithm and associated difficulties (such as misspelled street names) is beyond the scope of this paper, but is discussed in detail in McPherson (1999).

The path-finding algorithm is applied to each origin-destination pair, resulting in a pattern of paths across the network. The path pattern will overlap both spatially and temporally: spatially because many paths will share the same links, and temporally because many journeys will share the same time periods during the day. In this respect, RoadLink is a dynamic model because it assigns portions of each trip to different time periods, depending on the trip's departure time and duration.

In a conventional assignment, the flow pattern is expressed as a vector (or table) of link flows for each time period. Most popular assignment methods then use an iterative process of updating the travel costs on each link and generating new paths based on the new costs. The new link costs are typically governed by a speed-flow relationship which relates the capacity, free-flow speed, flow and travel time on the link. As flow increases, travel times also increase, making the link less attractive for travel. In the RoadLink model, Akçelik's time-dependent travel time relation was used, because it has a more rigorous theoretical basis than some of the other commonly used speed-flow relationships (Akçelik 1991). The Akçelik functions were calibrated on the basis of free-flow speeds and capacities used by the Victorian state road authority in their strategic highway model of Melbourne.

On each assignment iteration, a new flow vector is generated for the computed link travel costs. The final flow vector is then taken as a linear combination of the flow vectors from the iterations. The main deterministic assignment methods differ in how these linear combinations are derived.

In a disaggregate approach, it is important not to separate the flow pattern into a vector of individual link flows. Doing so would entail discarding the connectivity of the complete path and the demographic information related to the trip. Instead, the disaggregate process stores complete paths on each assignment iteration so that full path and traveller information is maintained. An iterative process may still be used –path-finding, updating travel costs and applying a linear combination of flows – but the linear combinations apply to a vector of paths, rather than a vector of separate flows.

Storing complete paths, rather than link flows, comes at a considerable cost. The amount of computer storage required to maintain the paths is many times greater than that of a conventional assignment. The problem is exacerbated if a detailed network containing many links is used. In the RoadLink model, which embodies the principles of disaggregate assignment, the number of iterations able to be completed was severely limited. Three iterations of a 50,000 path data set with a 230,000 link network required approximately 300 megabytes of storage space. Despite this limitation, it is anticipated that future computer processing power and memory technology will comfortably deal with path storage requirements.

The following sections discuss how these principles were put into practice – firstly the source of travel survey information, then the model itself and finally a case study of the model's application.

The Victorian Activity and Travel Survey

The travel survey used in this study is the Victorian Activity and Travel Survey (VATS). Conceived in 1993 by the Transport Research Centre, then at the University of Melbourne and later at RMIT University, VATS is an ongoing household travel survey that collects detailed travel information within the Melbourne Statistical Division. The survey is one of the largest of its type in the world, with approximately 5000 households responding each year since the project's inception at the end of 1993. For an overview of the survey's operation and derivation of expansion factors see Richardson, Ampt and Meyburg (1995) or McPherson (1999).

Each responding household reports each trip made by occupants using a 24 hour travel diary. All travel by all modes is considered in the survey: car, public transport and bicycle trips – even walking the dog! Each trip in the VATS sample is weighted by an expansion factor to reflect total population movements, according to Australian Census figures. In each year of data collected, approximately 20,000 car driver trips are reported

Each VATS car driver trip has a variety of data associated with it. The variables of particular interest in the disaggregate assignment process are listed in Table 1 below.

Table 1: VATS data used in the disaggregate assignment process

Variable	Description
Trip identifier	Unique identifier that allows the trip to be linked to individual and household information collected in the survey
Origin and destination	Coordinates of the trip's origin and destination
Streets used	Up to six names of streets used between the origin and destination
Weight	The trip's expansion factor
Departure time	Time of day, day of week and date of trip commencement

The trip identifier can be used to link the trip with other information collected in the survey, such as driver age, vehicle type and trip purpose. This raises the important issue of maintaining the privacy of individuals, particularly where individual trips are being modelled. In VATS, home addresses are not stored with the data, and a random component is introduced to the geographic coordinates of home locations so that the individual residences cannot be deduced from trip origins and destinations.

The trip origin and destination coordinates are used to geocode the trip endpoints to a detailed GIS representation of the Melbourne road network. The path-finding algorithm uses the nominated street names to fit the path to actual streets the driver used, enhancing the realism of the final path. The weight is used to compute the traffic flow contribution of the trip and the departure time allows the model to take into account congestion effects at different times of the day.

The RoadLink user interface

The RoadLink software package is a modelling tool that implements on-screen mapping, a road network inventory, path-finding algorithms and disaggregate assignment. The software was designed with a graphical GIS-style interface in line with recent trends in transport planning software and computer operating systems.

The interface provides users with the following features:

- an interactive network map display, including panning, zooming and selection of nodes and links with a mouse or other pointing device;
- network editing facilities, including link creation, deletion, splitting, merging and moving;
- link and intersection attribute editing features, enabling users to view and set parameters such as link capacities and intersection controls;
- artificial intelligence techniques for estimating road and intersection properties where these are unknown;
- animated interactive path-finding, allowing the user to experiment with different path-finding algorithms and parameters;
- disaggregate assignment processing; and
- a powerful query interface that enables users to construct complex select-link analyses and thematic maps

The software was written in the C++ language for the Microsoft Windows operating system. The following figures demonstrate some of the RoadLink interface's features

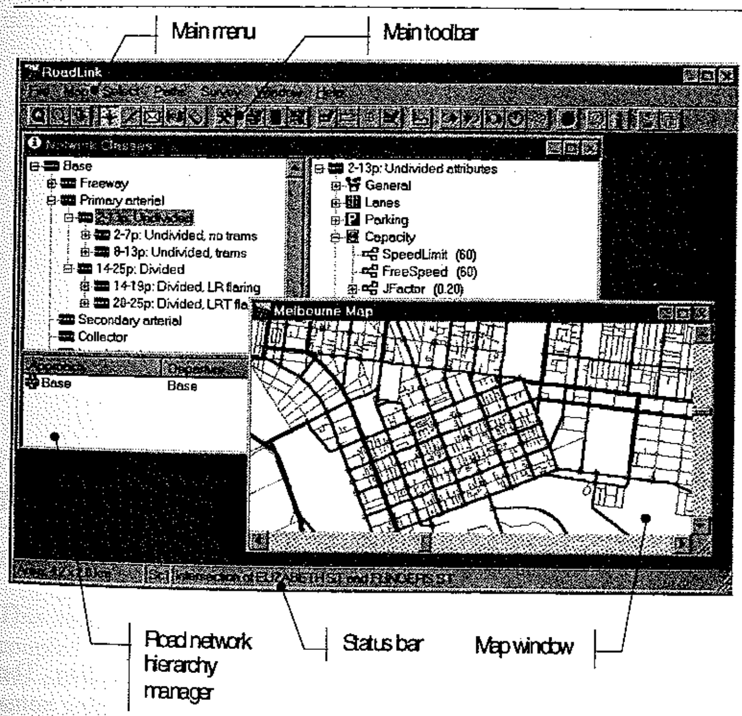

Figure 3: The main RoadLink window

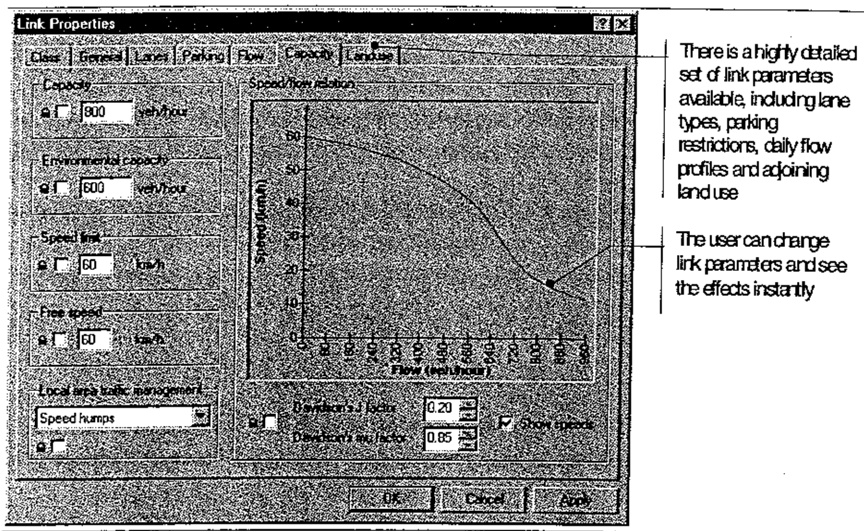

Figure 4: Input of link properties

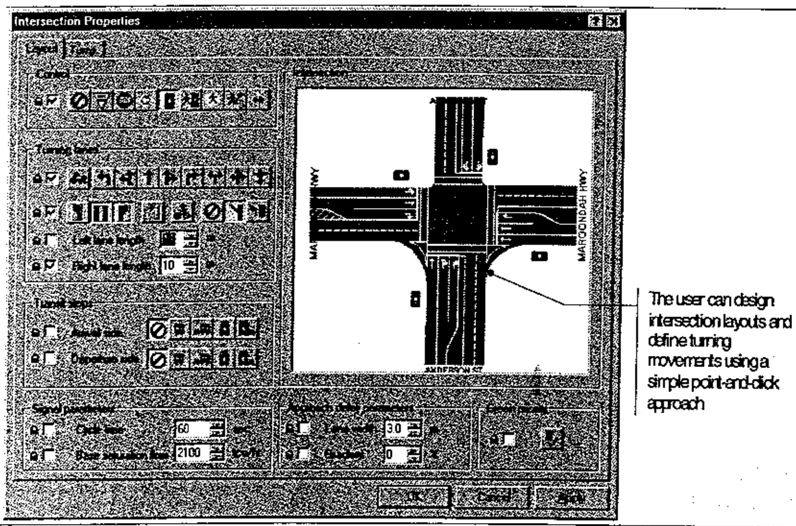

Figure 5: Input of intersection properties

Experimental Results

RoadLink was used to assign car driver trips from two years of VATS data. The assignment was performed on a highly detailed Melbourne network containing some 230,000 links and 80,000 intersections. Due to the heavy storage requirements, only three assignment iterations were possible. In addition, the assigned VATS data did not contain any external trips (with origins or destinations outside Melbourne), nor did they contain a full representation of commercial vehicles. The resulting modelled flows therefore were expected to underestimate actual flows by approximately 10-20%.

Figure 6 is a plot of actual and modelled 24 hour traffic flows across screenlines that intercepted all the major north-south, east-west, inner, outer and Mornington Peninsula flows. The plot suggests that the VATS sample is a strong predictor of basic traffic movements across the city, but actual flows are underestimated by approximately 20%. The underestimation of flow is probably due to the missing external and commercial vehicle trips as mentioned earlier, but could also be influenced by the weights used to expand the survey sample.

For the purposes of rough modelling of link flows, the results were scaled by a factor of approximately $\frac{1}{0.8}$ to remove the bias. A plot of actual and scaled daily link flows is shown in Figure 7, showing that the bias has indeed been compensated for. This plot shows a reasonably strong correlation between actual and modelled flows, but with a level of scatter that would indicate some further model calibration is required. The results, however, are quite encouraging and indicate that a household travel survey certainly has potential to form a direct input to the disaggregate assignment process. A more extensive validation of the RoadLink model, including travel time comparisons, is given in McPherson (1999).

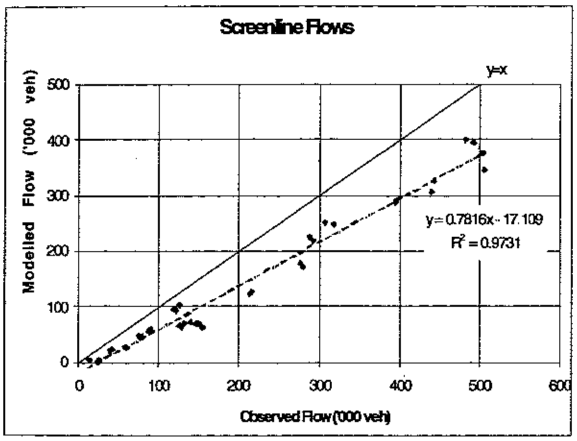

Figure 6: Modelled and observed screenline flows (24 hour)

Case study: Burwood Highway

This section presents some typical outputs available from a disaggregate assignment coupled with VATS data. The chosen case study is the Burwood Highway corridor in the eastern suburbs of Melbourne. Burwood Highway is one of Melbourne's major radial arterials that facilitates the east-west movement of traffic in the eastern suburbs. At its outer eastern end, it serves the outer suburbs of Belgrave, Upwey and Ferntree Gully in the foothills of Mount Dandenong. Traffic volumes are only moderate in this section, being around 20,000 vehicles per day (Vic Roads 1996). It continues through a number of large residential suburbs, linking finally to Toorak Road and the South Eastern Freeway, which carry large volumes of traffic to the centre of the city.

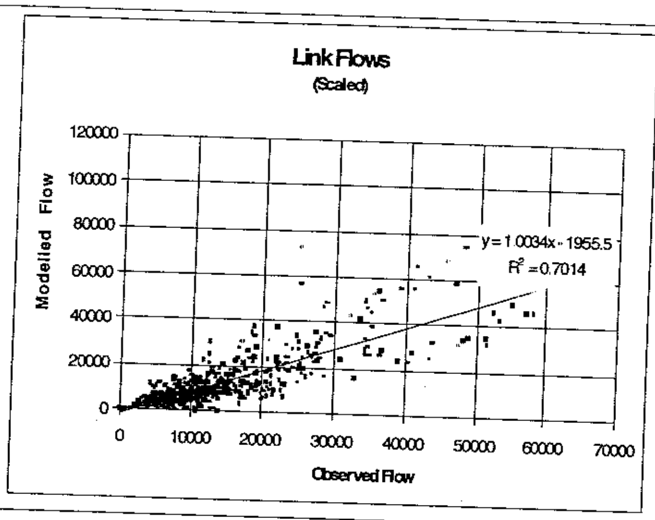

Figure 7: Modelled and observed individual link flows (24 hour – scaled)

The purpose of the case study is not to draw any specific conclusions about travel patterns on Burwood Highway, but rather to give examples of how demographic information can be linked with path data to enhance the information available from a household travel survey.

One simple output readily available from a disaggregate assignment is a 'select-link' plot of vehicle paths. A link (or series of links) is selected, and all traffic using the link is traced to its origins and destinations. Figure 8 shows a select-link plot with the routes of all respondents who use any part of Burwood Highway. This plot demonstrates significant flows in the immediate vicinity of the highway, with flows gradually dispersing in more distant areas. At the western end of the highway (Toorak Road) a large proportion of traffic diverts to the South Eastern Freeway en route to the city. An interesting feature of this plot not normally available from traffic assignments is the infiltration of flow into local streets (shown by the fine offshoots from the flow on the main arterials).

Disaggregate assignment is well-suited to select-link analyses, since all vehicle paths are stored as part of the assignment process. The modeller simply extracts those paths from the path database that meet the required select-link criteria. In the above example, we extracted *all* weekday trips using Burwood Highway – but we could, for example, have just as easily selected morning peak commuter trips, or non-business trips, or single-occupant vehicle trips or trips that had travel times less than ten minutes.

Cross-sectional traffic profiles are other primary outputs from the disaggregate assignment process. In a cross-sectional profile, the modeller chooses links of interest, then examines the composition of drivers, trip purposes, travel times and vehicle types using the links. Figure 9 shows a cross-sectional profile of trip purposes between Blackburn and Springvale Roads, a section of Burwood Highway with fairly high traffic volumes.

Figure 8: 24 hour weekday flow pattern of Burwood Highway trips

Cross-sectional profiles can be compared at different locations along a traffic corridor to give a longitudinal profile of the corridor's use. Figure 10, for example, shows the number of vehicles on chosen sections of Burwood Highway that are travelling on an education-related purpose. This figure shows a high concentration of education trips near Station Street where Deakin University's Burwood campus is located.

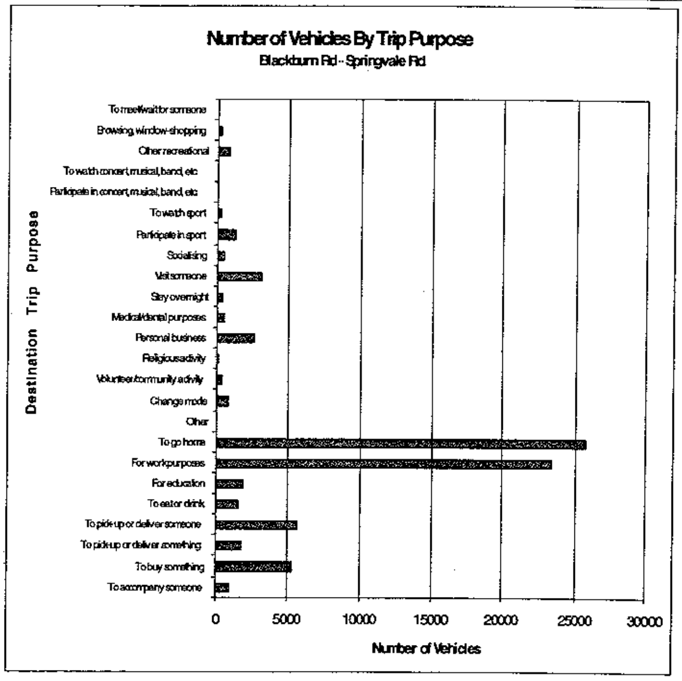

Figure 9: Trip purpose profile

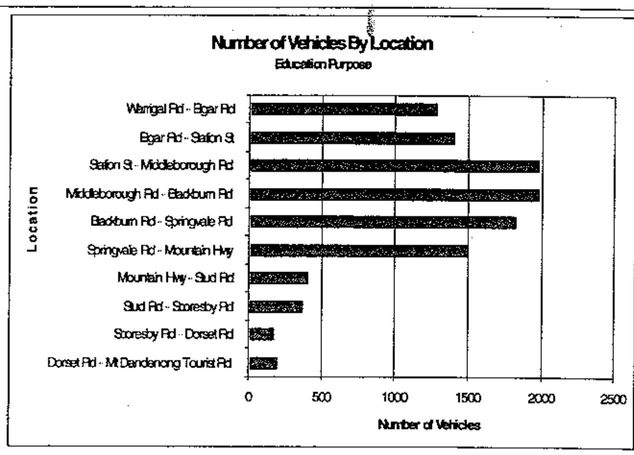

Figure 10: Longitudinal profile of education trips along Burwood Highway

There are naturally many other ways the demographic and vehicle data from a household travel survey could be used in similar route analyses. For instance, different path-finding criteria could be used for different types of driver or trip purpose. (This technique was actually used in a recent study of driver route-selection behaviour in response to the CityLink tollway in Melbourne). Disaggregation allows researchers and modellers to experiment with the interactions between driver characteristics and route selection.

Conclusion

In this paper, we have briefly examined the aggregate or zonal paradigm commonly used in four-step transport models. Zone-to-zone traffic flows are a convenient way of expressing travel demand and assigning this demand onto a transport network. While this modelling paradigm is quite adequate for typical strategic planning and local impact studies, there are occasions where modellers need to know *who* is travelling and *why* they are travelling, rather than simply *how many* people are travelling.

The subsequent sections of the paper discussed the use of a disaggregate trip-by-trip assignment method that enables modellers to determine demographic and trip purpose information about vehicle flows in a transport network. A working model, RoadLink, was developed which demonstrated that household travel survey data can be used with a disaggregate assignment technique to produce useful flow and demographic profiles of transport facility use in a large urban area.

Disaggregate assignment and the RoadLink model could be applied in a variety of further applications, including:

- greenhouse gas traffic emissions studies (for example, in the determination of vehicle types and engine sizes using each road and intersection in a network);
- marketing and outdoor advertising;
- analysis of catchment areas for selected traffic corridors and intersections;
- evaluating the impact of changes to the road network;
- understanding how a road (or other transport facility) is used by the driving population.

The next few years may see further trends towards disaggregate transport modelling. In a forward-looking paper summarising research directions for the United States Travel Model Improvement Program, Spear (1996) notes:

> *"... detailed network information is needed to obtain more precise measures of transportation level of service parameters.... This network information can be derived and utilized through GIS technology which will allow models to be calibrated using point-to-point trip records."*

Perhaps disaggregate models, such as RoadLink, will assist us in moving in this direction.

Acknowledgments

The author would like to thank Professor Tony Richardson for supervising this research and providing many avenues for lateral thought. Thanks also goes to the Transport Research Centre at RMIT University, Sinclair Knight Merz Pty Ltd, and Dr Russell Thompson at the University of Melbourne for support and assistance in this project.

References

Akçelik, R (1991) Travel time functions for transport planning purposes: Davidson's function, its time-dependent form and an alternative travel time function *Australian Road Research 21* (3), 49-59

Bell, M G H and Iida, Y (1997) *Transportation Network Analysis* Chichester: John Wiley and Sons

Ben-Akiva, M E and Lerman, S R (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand* Cambridge Massachusetts: MIT Press

Dijkstra, E W (1959) A note on two problems in connexion with graphs *Numerische Mathematik 1*, 269-271

Floyd, R W (1962) Algorithm 97: shortest path *Communications of the ACM 5*, 345

Handy, S (1996) Methodologies for exploring the link between urban form and travel behavior *Transportation Research 1D* (2), 151-165

Hart, P E Nilsson, N J and Raphael, B (1968) A formal basis for the heuristic determination of minimum cost paths *IEEE Transactions on System Science and Cybernetics SSC-4*, 100-107

McPherson, C D (1999) *Disaggregate Assignment of Household Travel Survey Data* unpublished PhD thesis: University of Melbourne

Ortúzar, J D and Willumsen L G (1994) *Modelling Transport* (2nd ed) New York: John Wiley & Sons

Pearl, J (1984) *Heuristics: intelligent search strategies for computer problem solving* Menlo Park California: Addison Wesley

Ran, B and Boyce, D E (1996) *Modeling Dynamic Transportation Networks: An Intelligent Transportation System Oriented Approach* Berlin: Springer-Verlag

Richardson, A J Ampt, E S and Meyburg, A H (1995) *Survey Methods for Transport Planning* Melbourne: Eucalyptus Press

Sheffi, Y (1985) *Urban Transportation Networks* Englewood Cliffs: Prentice-Hall

Spear, B D (1996) New approaches to transportation forecasting models. A synthesis of four research proposals *Transportation 23*, 215-240

The Age (1998) Getting City Link to pay its way *The Melbourne Age newspaper*, 17 March 1998

Vic Roads (1996), *Burwood Highway, Springvale Road to Belgrave, Principal Traffic Routes Investigation Final Report* Kew Victoria: Vic Roads

Wardrop, J G (1952) Some theoretical aspects of road traffic research *Proceedings of The Institution of Civil Engineers Part II* (1), 325-378