PASSENGER ROUTE GUIDANCE SYSTEM FOR
MULTI-MODAL TRANSIT NETWORKS

Hong K. LO, C. W. YIP

Civil Engineering Department, The Hong Kong University of Science and Technology,
Clear Water Bay, Hong Kong
clehklo@ust.hk

Brian MAK

Computer Science Department, The Hong Kong University of Science and Technology,
Clear Water Bay, Hong Kong

ABSTRACT

In many public transport oriented cities in the world, especially Asian cities, the public transport system has been developed extensively, to the extent that it has become increasingly difficult to navigate. Although inter-modal transfers are common and often necessary, a complete set of the routes across transport modes is generally not presented in a form that is accessible to travelers, as each operator would only publish its own routes. Moreover, the common nonlinear fare tables together with inter-modal fare discounts pose challenges to travelers in deciding their best routes. This study develops a multi-modal passenger route guidance system, called eFinder, to aid travelers with their combined mode-route choices. We discuss the architecture and features of this system in this study. This system forms a platform for disseminating public transit information and should complement further development and use of the public transport network by enabling travelers to make informed choices.
1 INTRODUCTION

In many transit-oriented cities around the world, the majority of the population depends on the public transportation system (PTS) for their daily travels. For example, the PTS in Hong Kong carries over 90% of the 11 million daily trips. As the PTS is being expanded and further developed, as is typically the case, the system is becoming increasingly difficult to navigate. To assist passengers in locating and using their services, transit operators often set up web sites to publish their routes. However, due to the competitive nature of privately operated services, none of the operators would publicize the routes of other operators, rendering such information to be piecemeal in nature, short of providing useful route guidance information across modes or across operators. The need of a common platform to provide routing information across modes and operators is apparent.

Even if one were equipped with perfect transit route and fare information, finding the optimal path would not be a simple task. Most transit fare structures in practice are zonal or origin-destination (OD) based, not directly proportional to the travel distance or travel time, or non-additive in nature. Moreover, the increasingly common usage of electronic smart cards for transit fare payment opens up innovative fare structures that are not possible or otherwise difficult to implement with the traditional coin-based system. Some examples include collaborative fare discount schemes for certain types of transfers between routes and modes, as are widely practiced in Hong Kong. All these add to the complexity of the fare structure in the multi-operator, multi-modal PTS.

Parallel to the emergence of a sophisticated PTS (for example, Chien and Spasovic, 2002), developing a transit route guidance system is equally important so that the PTS can be utilized to its full extent. Such a multimodal route guidance system not only allows travelers to find their ways to destinations, but more importantly, do so while well aware of the alternatives and choose accordingly, such as minimum cost, time, or number of transfers. Responding to this need, this study develops a transit route guidance system, called eFinder, to facilitate passengers’ navigating the sophisticated multi-modal PTS.

Most route guidance systems or advanced traveler information systems (ATIS) are developed for private car driving. Recently, there begin to have a few traveler information systems developed for transit users. For example, Cashin et al (2002) discussed the “Los Angeles Smart Traveler (SMART TRAVELER) Project” deployed by the California Department of Transportation (Caltrans). The system provides telephone and Internet information of traffic and transit services. There are also transit itinerary planners developed for California, as depicted in the SMART TRAVELER (http://www.dot.ca.gov/caltrans511/), such as “OTIS – trip itinerary” of San Diego (San Diego Metropolitan Transit System, 2004) and “511 Take Transit Trip Planner” of San Francisco (Metropolitan Transportation Commission and San Francisco Bay Area Transportation Partners, 2004). Travelers obtain transit services information for their trips by entering their origins, destinations, and times of departure or arrival. Other examples
of itinerary planners available on internet include TransiTrips (TransiTrips, 2004), TriMet (TriMet, 2004), and VIS’s Personnal Trip Planner (Metropolitan Transit, 2003). The emergence of these transit information services shows the increasing demand for such information. In Hong Kong, some transit operators provide route search services in their own web pages, such as “Point to Point Route Search” of Kowloon Motor Bus (http://www.kmb.hk) and “Bus Route Search” of Citybus (http://www.citybus.com.hk). However, they cover only their own routes but not the whole Hong Kong transit network.

Transit information services typically involve a route search algorithm that works with multimodal networks. This route search algorithm to a great extent defines the performance and quality of the routing information provided. Route searching in multimodal transit networks involves procedures more complicated than simply finding the shortest distance or time paths, which may involve the use of hyperpaths to handle the common line problem and waiting time of transit routes (Nguyen and Pallottino, 1988; Spiess and Florian, 1989; De Cea and Fernandez, 1993; Wu et al., 1994). In other situations, with known schedule information, the search procedure may involve schedule-based networks (Nguyen et al., 2001; Nuzzolo et al., 2001; Tong et al., 2001; Poon et al., 2004).

As for the system developed in this study, eFinder takes advantage of the State Augmented Multimodal (SAM) network structure technique (Lo et al. 2001, 2002, 2003; Lo and Yip, 2001; Yip 2001) to handle complex fare and modal transfer configurations by working with frequency-based transit services. Moreover, most existing such systems only work with routes that include only one transfer; eFinder can handle trips with unlimited number of transfers. On the other hand, it can also be set to find routes that do not exceed the number of transfers as specified by the users. Subject to the travelers’ preferences: minimum fare, minimum time, or minimum generalized cost, eFinder determines and advises the travelers with the most appropriate routes across modes and across operators. Users of eFinder input their origins, destinations, and preferences through either the Geographical Information System (GIS) interface via the Internet or through the Automatic Speech Recognition (ASR) system via both fixed or mobile telephones. The routing results are fed back to the users via these two interfaces.

When combined with the Internet, GIS multiplies its capability tremendously by allowing access to the system from any connected terminals. With wireless communication technology (such as mobile phones), access is further extended to virtually any locations. Peng and Huang (2000) categorized online public transport information system according to their content levels and functionalities, ranging from static to real time information, and from text browsing to interactive map-based and online transactions. On the other hand, speech is one of the most natural and efficient communication means. While the mobile phone is a common communication device, using mobile phones for information retrieval becomes a very attractive proposition. The system architecture of eFinder includes both GIS via internet and ASR via mobile phone as input and output interfaces.
The outline of this paper is as follows. Section 2 introduces the system architecture of eFinder. Section 3 presents the interface of the system. Section 4 describes the analytical component of eFinder system and the SAM network technique. Section 5 contains some concluding remarks and further system developments.

2 SYSTEM ARCHITECTURE

eFinder has a three-tier system architecture with trip planning and interactive GIS and ASR functions. The system architecture is summarized in Figure 1. The client tier provides interfaces for users to input their origins and destinations and for the routing results to be presented. eFinder provides the web-based GIS interface and phone-based ASR interface. For the web-based GIS interface, clients navigate the map and input their origins and destinations by entering the addresses or clicking on the map interactively. On the other hand, the ASR system allows the system to identify the clients’ locations in a few dialogues. The results of route search are also presented through the interfaces.

The middle tier connects and queries the route database according to the users’ OD inputs and composes the routing results. The middle tier is the kernel of the system. Based on origin and destination information collected through either the GIS or ASR interface, the application server determines the best transit stations or stops for boarding and alighting. It then queries the route database for the optimal routes and converts the results to a format understandable to the users. The final routing results are delivered to the users either via a GIS map or telephone messages.

The database tier mainly consists of the route database that stores the all-pairs optimal paths according to different criteria including the shortest time, cheapest, or the balanced choice. The response time of producing the optimal paths per request, therefore, does not involve calculating the paths in an online manner, but merely needs to trace back the optimal paths that are already pre-calculated and stored in a tree structure in the route database. This setup substantially reduces the response time to produce the optimal paths. The all-pairs optimal paths are pre-calculated offline by the Floyd-Warshall algorithm (Cormen et al., 1990) based on the SAM network structure, as explained in Section 4. As the transit route and fare information do not change from minute to minute, eFinder only updates the route database periodically, say on the order of days or weeks or whenever the transit system is modified.

The transportation information system (TIS) contains the up-to-date transit system information including transit schedule, station location, transit fare, travel time, etc. The TIS is typically maintained by the transport department or to be obtained individually from each of the transit operators and composed together. Other than calculating the optimal paths, the analytical model transforms the transit route, fare, and transfer
information from the TIS to the SAM network structure for the optimal path calculation.

3 USER INTERFACES

3.1 Web-Based Geographical Information System (GIS) Interface

Passengers input their origins, destinations and preferences to the system through the Geographical Information System (GIS) map interface via the Internet or through the Automatic Speech Recognition (ASR) system via the telephone. Moreover, the GIS facilitates the manipulation, analysis, and presentation of the spatial information. The prototype system of eFinder employs Intergraph GeoMedia® Web Enterprise as the web-based GIS engine for the presentation of map and route results, and for the input of locations through the map interface.

The Graphical User Interface (GUI) of eFinder is shown in Figure 2. The main frame shows an interactive map. Users navigate the map by clicking on it and using the directional arrows. They can also zoom in the map at various levels of details and select their locations of interest by clicking on the map. The frame on the left hand side provides users another interface to input their origins and destinations hierarchically, by area, district, and finally address or landmark. Users then select their preferences: fastest, cheapest, or eFinder (i.e., minimum generalized cost considering the combination of fare, travel time, transfer penalties, etc.) routes.

The optimal routes and corresponding stations are shown through the same map interface in the main frame. By gradually zooming in, one gets to the details of showing the exact stations for boarding, alighting, or transfer (see Figure 3). The bottom frame shows the detailed information of the resulting routes including the suggested modes, routes, travel time and fare of each segment, and the transfer details such as modal change locations, frequency of mode being transferred to, and any applicable interchange fare discounts, etc.

3.2 Automatic Speech Recognition (ASR) System Interface

eFinder includes the server-based automatic speech recognition (ASR) technology to allow users query the system with their voice. Users simply tell the system any landmarks that are closest to their locations and destinations, then the system will process the query and answer with the routing results including the modes, routes, fare, the station locations, transfers if any, etc.

Voice query is particularly useful when a computer is not readily accessible especially while en-route. On the other hand, telephones or mobile phones are ubiquitous. Moreover, it is more natural to interact with eFinder by voice. Speech recognition will be accomplished on the server side, thus relieving a user from expensive computers, smart
phones, or personal digital assistants (PDA). Server-based ASR has the further advantage of decoupling the implementation of technology advances from the functionalities of clients’ machine.

To enable voice queries, our prototype is equipped with the Intel's computer telephony (CT) card, Dialogic D/41 JCT-LS. The Dialogic CT card is hooked up to the telephone fixed line on one end, and connects to our ASR system on the other end. When a user calls eFinder, the CT card answers the call without human intervention, and then passes the voice query to our ASR system. If the ASR decides that the query is valid, it will pass it to eFinder. The search result from eFinder is converted back to speech using a simple concatenative speech synthesizer (Dutoit 1997) and is played back to the caller via the CT card.

The ASR system in eFinder currently recognizes areas and landmarks spoken in Cantonese (a dialect of Chinese spoken by Chinese in southern China, including Hong Kong). Each location consists of 2 to 12 Chinese characters. According to the Chinese phonology, each character consists of only one syllable, and each Chinese syllable is divided into two sound segments commonly known as the initial segment and the final segment. The initial segment consists of an optional consonant and the final segment consists of a vowel followed by one to two optional consonants. Altogether, there are 19 distinct initials and 53 distinct finals in Cantonese. These 72 (= 19+53) Cantonese sounds are modeled using the continuous-density hidden Markov model (CDHMM) (Rabiner, 1989). Each initial CDHMM has 3 states, and each final model has 5 states. Observations at each CDHMM state are modeled by a 20-mixture Gaussian distribution. In the final system, there are four more CDHMMs to capture short pause, silence, garbage noise, and non-Cantonese speech sounds. The CDHMMs are trained from a large telephone speech corpus collected by the Chinese University of Hong Kong (Ching, 1997). The corpus consists of more than 10,000 Cantonese words spoken by more than 300 speakers of both genders. All speeches are band-filtered with a lower cutoff of 300Hz and a higher cutoff of 3400Hz which represent the pass-band of common telephone transmission channel. After signal processing, 39 features are extracted from every 10ms of speech. An in-house test was performed to gauge the performance of our ASR system. The recognition accuracy is over 90%. Notice that the ASR system is speaker-independent. Thus, one may further improve the recognition performance for a user by speaker adaptation techniques (Gauvain and Lee, 1994; Leggetter and Woodland, 1995; Kuhn et al., 2000).

4 ANALYTICAL MODEL

4.1 SAM Network Structure

State augmented multimodal (SAM) network is a network model for passenger flows in a multi-modal network. Its network structure combines the networks of different modes and intrinsically includes the multi-modal network characteristics such as sectional fares, interchange discounts and probable transfers. eFinder employs the SAM network
structure for the calculation of optimal paths. Each node in the SAM network includes four variables to denote the location and transfer information in order to capture nonlinear cost functions and probable transfer rules. Links in the SAM network are divided into two kinds, “directed in-vehicle links” and “transfer links”. The details are summarized as follows.

4.1.1  SAM Nodes

Contrasting from the conventional and simple way presenting nodes, four state variables are used to describe each SAM node, including:

Location (i): this state variable corresponds to the identity (ID) of each bus stop or transit station. It is used to represent the location of the SAM node.

Transfer state (s): this state variable is related to the adopted transfer rules, to be explained later.

Number of prior transfers (n): it indicates the number of prior modal transfers and is used to find routes subject to any restrictions on the number of transfers allowed. If there is no limit imposed on the number of transfers allowed, this variable can be ignored.

Alight or aboard indicator (l): it specifies the status of a trip segment. It is equal to 1 (0) at the start (end) of an in-vehicle link; or, correspondingly, at the end (start) of a transfer link. This state variable is used for modeling modal transfers.

Denote $N$ as the entire set of SAM node. Each node is represented as $[i, s, n, l]$, containing the four state variables defined above. The notations of origin and destination nodes, where passengers enter and leave the network respectively, require special attention. We denote an origin node as $[i, 0, 0, 0]$ and a destination as $[i, 0, 0, 1]$ where $i$ in both cases designates their physical locations. The middle two 0’s in these two cases do not carry any physical meaning other than specifying them as an origin or a destination. The last state variable denotes the status of starting or ending a transfer link.

4.1.2  Transfer Rules

The set of links connecting the SAM nodes is divided into two subsets, “directed in-vehicle links” and “transfer links”. In a multi-modal network, travelers may transfer a number of times before reaching their final destinations. These multi-modal paths consist of sequences of transfer links and directed in-vehicle links. In constructing the multi-modal network, one must be careful in establishing the connectivity. To better reflect actual transfer behavior, the SAM network structure limits the modal transfers to a set of probable ones by embedding the possible transfer state transitions.

For exposition purposes, we define a set of probable transfer states as shown in Figure 5. Each node in the figure represents a transport mode, whereas each arrow indicates a probable transfer. Transfer state 0 is reserved for the start (or origin) and end (or destination) of a trip. Referring to
Figure 5, a traveler can start with auto (transfer state 1, mode 1), subway (transfer state 2, mode 2), or bus (transfer state 3, mode 3). Say, once subway is chosen for the first segment, the subsequent probable transfer state is limited to bus. Undoubtedly, a traveler who chooses to take the subway on the first segment may arrive at his destination without any transfers, completing the transfer state sequence of 0-2-0. Associated with each transfer state is a transport mode, for example, mode 1 (auto) is associated with transfer state 1 whereas mode 3 (bus) is associated with transfer state 4.

What kinds of transfers are probable obviously a function of the locality and its practices or habits. There is no restriction on how these rules are defined. For the case of Hong Kong, eFinder allows all transfers except for transfers within the same route.

4.1.3 Transfer Links

Every pair of alighting and boarding nodes of two different modes that are within walking distance of each other are connected by transfer links. Consider two stations, \( i_1 \) and \( i_2 \), of two different routes, \( b_1 \) and \( b_2 \), respectively. If the distance between two stations is within walking distance, passengers are possible to transfer from one route to the other. Two columns of nodes are constructed, as shown in Figure 6. The left hand nodes represent those alighting route \( b_1 \) at \( i_1 \) as indicated by the alighting/boarding variable \( l = 0 \) and the right hand ones those boarding route \( b_2 \) at \( i_2 \) with \( l = 1 \). Graphically, transfer links are shown as arrows between the left and right hand nodes in Figure 6. Associated with each transfer link is its head node and tail node. Mathematically, from each tail (or alighting) node \([i,s,n,0]\), a set of transfer links is constructed to connect to this set of head (or boarding) nodes \([i,g,n+1,1]\) where (i) \( g \in \tau(s) \), and (ii) \( n+1 \leq R \). Only those boarding nodes with probable transfer states from \( s \) are connected. Moreover, the number of transfers is increased from \( n \) at the alighting node to \( n+1 \) at the boarding nodes, which must be less than or equal to the maximum number of allowable transfers (\( R \)). In the case if the constraint of maximum number of transfers is not imposed, one can simply ignore criterion (ii) or put \( R \) to be a very large number.
Two special nodes, origin and destination, are represented as \([i, 0, 0, 0]\) and \([i, 0, 0, 1]\) in Figure 6. Their associated departing or arriving links are shown dotted in Figure 6; whereas en route transfers are shown as solid arrows. To capture the interchange fare discount, the transfer link connecting the alighting node of the first mode and boarding node of the second mode have a negative fare value corresponding to the transfer discount. For example, suppose passengers can enjoy a fare discount of $1 for transferring from subway (state 2) to bus (state 4), the transfer links connecting state 2 to state 4 will have a -$1 fare.

4.1.4 Direct In-vehicle Links

Transit fares are usually not directly proportional to the travel distance or time. Composing the origin-destination based fare on a link by link approach is in general not possible. That is, one cannot simply add up the individual link fares to obtain the origin-destination fare. In addition, travelers may have nonlinear valuation of travel time – small amounts of time have lower value whereas large amounts of time are very valuable. Lo and Chen (2000) referred to these situations as non-additive route costs. To allow for flexibility in capturing these nonlinear or non-additive route cost structures, we construct a direct in-vehicle link between each pair of connected nodes with the same state \(s\) and prior transfer number \(n\). For each boarding node \([i, s, n, 1]\) on the sub-network of transport mode \(b\), where \(b = \eta(s)\), a direct in-vehicle link is constructed to connect each of the reachable alighting nodes with the same transition state \(s\): that is, \([j, s, n, 0]\) where \(i, j \in U_b, i \neq j\). The variables \(i\) and \(j\) denotes its start and end locations; whereas \(s\) represents that it is a direct link with transfer state \(s\). The second and third state variables \(s\) and \(n\) of SAM nodes remain unchanged as the travel is on the same modal sub-network involving no change of state or transfer. The fourth state variable changes from \(l = 1\) at the start of the direct in-vehicle link to \(l = 0\) at the end of it.

4.1.5 Link Costs

eFinder calculates the fastest, cheapest or eFinder (minimum generalized cost) routes according the preferences of passengers, as enabled by the SAM Network. Each direct in-vehicle link and transfer link are attributed by travel time, waiting time, and fare. For a direct in-vehicle link, its travel time captures the in-vehicle travel time on the link segment whereas its fare is set to be equal to the fare from its tail node to its head node. In this way, the point-to-point fare structure is fully represented.

The travel time of a transfer link is estimated by the walking distance between the two connected stations. Its waiting time is set to be a fraction of the headway of the boarding mode. Its fare is equal to zero if there is no interchange discount; equal to the negative value of the interchange discount if any.
4.2 Shortest Path Algorithm

To prevent excessive response time for calculating the optimal paths online, all-pairs shortest paths from stations to stations are pre-calculated offline and stored in the route database server. The route database is updated periodically whenever the transit system is modified. The shortest paths are stored with a tree structure through the notation of predecessors. To search for an optimal path, the application server queries all stations within walking distance from the destination, then traces back from the one with lowest general cost. Since the calculation of shortest paths is preprocessed, each route query involves only tracing back the predecessors sequentially. The process time is typically with 1-3 seconds. This approach is thus suitable for online application in large-scale networks.

One thing to note is that although some transfer links contain the interchange discounts expressed as negative values, the direct links preceding those transfer links must have a higher positive fare value. Thus the SAM network structure prevents the occurrence of negative cost cycles, causing the problem of infinite path loops. Finally, the Floyd-Warshall all-pairs shortest path algorithm is used to find the all-pairs shortest paths. The algorithm has a run time of order $O(|N|^3)$ where $|N|$ is the number of SAM nodes.

5 CONCLUDING REMARKS

In this study, we contend that parallel to the emergence of a sophisticated public transit system, developing a transit route guidance system is equally important so that the system can be fully utilized. Indeed virtually all transit operators in Hong Kong have developed web sites and hotline enquiry services to publicize their services. However, publishing the routes and schedules is one thing, finding optimal routes across operators and across modes is quite another. The system needs to be able to work with nonlinear fare structures, sometimes complex inter-modal transfer fare discounts; pays attention to transfers rules and routing preferences of travelers; and maintains the routing databases for all operators and modes. The system must also deploy convenient ways of accessing the system, both for offline pre-trip planning and more preferably, while travelers are en route. The prototype system eFinder successfully demonstrates these functionalities. In particular, taking advantage of the SAM network structure, eFinder works with the complex fare structure and travelers’ preferences as in Hong Kong. Moreover, eFinder is enabled by two interfaces, both web-based GIS map interface and mobile-phone based automatic speech recognition.

Most of the components for this system are available from off-the-shelf technology. This study develops a system architecture that integrates these components together. Indeed, as a new prototype, there are several directions wherein further research and development are beneficial. Firstly, the common line issue (Spiess & Florian, 1989) in transit routing needs to be duly incorporated. Common lines are bus routes that serve the same corridor.
so that passengers can take any one of them for trips along the corridor. Passengers may take first arriving bus route. On the other hand, they may like to transfer at a stop with more common lines so as to increase their choices and shorten their waiting time.

Secondly, the automatic speech recognition interface is valuable as passengers can call the system while en-route. The more convenient way to communicate with the system is to talk to the system (via ASR) for inputs, while the routing results are fed back as texts or graphics so that passengers can keep a copy of the result for reference later on. The short message service (SMS) or multimedia message service (MMS) can be useful in this regard. Linking up eFinder with these messaging methods is an important development direction.

Thirdly, providing real time information based on current congestion and network conditions is another important future development. This will put a tight requirement on finding a good route in a short time, such as in seconds. This may require trading the solution quality with the solution time. That is, the system quickly finds a good, rather than an optimal, solution. Which part of the calculation can be performed offline and stored, and which part must be performed in real-time need to be carefully allocated.

Finally, if one considers the system not only as a way to offer transit route information to passengers, but a way to collect passenger preferences and origin-destination data, then it can be used to build up a passenger demand database to calibrate transit assignment models, such as the SAM model (Lo et al. 2001, 2002, 2003, 2004; Lo and Yip, 2001; Yip 2001). Such models, once calibrated, can be used to study pricing and operation strategies for individual companies or the entire multi-modal network as a whole.

In Summary, the system offers an interactive way for connecting transit operators and passengers. It can be used both for disseminating transit service information to passengers as well as for collecting modal preferences and OD information from passengers. The system offers lots of room for further refinement and development.

ACKNOWLEDGEMENT

This study is sponsored by the Competitive Earmarked Research Grants, HKUST6083/00E and HKUST6161/02E, of the Hong Kong Research Grant Council and the Sino Software Research Institute Applied Research Award, SSR101/02.EG06.

REFERENCES


Figure 1. System architecture of eFinder
Figure 2  The web page display of eFinder
Figure 3  The eFinder display with zoom-in features
Figure 4  The four state variables associated with a node

$i, s, n, l$

- location
- number of prior transfer
- transfer state
- alight or aboard
Figure 5. An example of probable transfer states
Figure 6. Augmented nodes and transfer links in the SAM network at a location $i$