Università di Pisa
Dipartimento di Informatica

Technical Report: TR-03-14

Route choice on transit networks with on-line information at stops

Guido Gentile
Dipartimento di Idraulica, Trasporti e Strade
Università di Roma 'La Sapienza', *guido.gentile@uniroma1.it*

Sang Nguyen
Département d'Informatique et de Recherche Opérationnelle
Université de Montréal, *sang.nguyen@umontreal.ca*

Stefano Pallottino
Dipartimento di Informatica
Università di Pisa, *pallo@di.unipi.it*

November 1, 2003

ADDRESS: via F. Buonarroti 2, 56127 Pisa, Italy. TEL: +39 050 2212700 FAX: +39 050 2212726

Route choice on transit networks with on-line information at stops *

Guido Gentile
Dipartimento di Idraulica, Trasporti e Strade
Università di Roma 'La Sapienza', *guido.gentile@uniroma1.it*

Sang Nguyen
Département d'Informatique et de Recherche Opérationnelle
Université de Montréal, *sang.nguyen@umontreal.ca*

Stefano Pallottino
Dipartimento di Informatica
Università di Pisa, *pallo@di.unipi.it*

November 1, 2003

Abstract

Passengers on a transit network with common lines are often faced with the problem of choosing between either to board the arriving bus or to wait for a faster one. Many existing assignment models are based on the classical assumption that at a given stop passengers board the first arriving carrier of a certain subset of available lines. It has been shown that, in this case, an *optimal subset* of lines that minimizes the passenger travel times exists and is easily determined if the headway distributions are exponential. However, when on-line information on future arrivals of buses are posted at various stops, it is unlikely that the above classical assumption holds. Passengers may choose in this case to board a line that offers the best combination of waiting time and expected remaining time. We propose in this paper a general framework for determining the probability of boarding each available line at a stop when on-line information on bus arrivals is provided to passengers. We will also show that the classical case without on-line information may be interpreted as a particular instance of the proposed framework, and thus extend the results obtained with exponential distributions to general distributions. Numerical examples, based on various headway distributions proposed in the literature, will be used to illustrate the combined impacts of the regularity of transit lines and of the availability of information regarding bus arrivals at stops on the line loads and on the passenger travel times.

Introduction

Consider a network of buses and/or surface and underground railway lines, such that every transit line has a fixed itinerary and is described by a sequence of stops. We are interested in urban networks with *common lines*, in the sense that multiple line routes share some stops and segments of itineraries. We will focus on the case where the irregularity of transit lines is high enough so that passengers may not be able to time their arrivals at stops in order to minimize their waiting times. It is reasonable in this case to assume that line headways are statistically independent and that passengers arrive randomly at stop (see, for example, Seddon & Day 1974, Jolliffe & Hutchinson 1975, Bowman & Turnquist 1981).

Earlier works on transit systems are devoted to the analysis of service regularity and to the waiting and boarding process in presence of congestion. They are mainly aimed at developing realistic bus headway distributions and passenger waiting time distributions (e.g., Chriqui & Robillard 1975, Bowman & Turnquist 1981, Hasseltröm 1981, Larson & Odoni 1981, Gendreau 1984, Marguier & Ceder 1984).

Recent works focus more on the passenger's route choices and on the resulting assignment models. In terms of route choice, it is often assumed that each passenger has a traveling strategy that allows him

*Research grants by MIUR and INDAM-GNAMPA of Italy, and NSERC of Canada.

to reach his destination at minimum perceived cost. A simple strategy may be formally described, on the modeling network, by a *hyperpath* connecting the origin to the destination (see for instance Nguyen & Pallottino 1989). At each stop-node of the hyperpath is associated a subset of transit lines that are perceived as *attractive* by the passenger.

Existing assignment models using hyperpaths (e.g., Spiess 1984, Nguyen & Pallottino 1988, Spiess & Florian 1989, Wu, Florian & Marcotte 1994, Nguyen, Pallottino & Gendreau 1998, Bouzaïene-Ayari, Gendreau & Nguyen 2001) are all based on the following hypotheses:

- transit line headways are statistically independent and have exponential distributions,
- passengers arrive randomly at stops, and board the first arriving carrier of their set of attractive lines.

These assumptions lead to a greedy algorithm for determining a set of attractive lines that minimizes the passenger travel time. Simple methods for computing the line boarding probabilities for passengers sharing a given destination as well as complete assignment procedures have been proposed.

The new ingredient here is the introduction of *on-line information* regarding the arrival times of the line carriers at transit stops. Since the waiting time for each line can be retrieved from the displayed information, *always boarding the first arriving carrier is not the best strategy.* Note that for the sake of simplicity, we will assume that time is the only measure of disutility. However, extending the framework to handle generalized costs is a straightforward task.

1 Common lines with on-line information at stops

In this section we propose a mathematical framework for computing the probability of boarding each line at a stop when on-line information about incoming carriers are displayed. This framework can then be embedded into any existing assignment model to handle transit networks with on-line information at stops and general headways distributions.

Firstly, the basics assumptions are the following:

- passengers arrive randomly at the stop,
- transit line headways are statistically independent with given distributions.

Secondly, with respect to a given destination, we assume that passengers:

- have a good estimate of the travel time associated with each line,
- can retrieve the waiting time for each line given the on-line information displayed at the stop,
- and choose to board a line that minimizes the total travel time (i.e., the sum of the waiting time and of the remaining time).

1.1 The basic framework

For each generic line $i \in L_n$, where $L_n = \{1, 2, \ldots, n\}$ is the set of lines serving a given stop, we assume that the following attributes are exogenously given:

s_i	the remaining time, i.e., the expected travel time to destination once boarded,
$f_i(w)$	the probability density function of the waiting time, where $f_i(w) = 0, \forall\ w < 0$.

Let $F_i(w)$ denote the cumulative distribution function of the waiting time and $\overline{F}_i(w)$ its complement, then it is:

$$\begin{aligned} F_i(w) &= \int_0^w f_i(v)\ dv, \\ \overline{F}_i(w) &= 1 - F_i(w) = \int_w^{+\infty} f_i(v)\ dv. \end{aligned}$$

The probability of boarding line i may be expressed as:

$$\pi_i = \int_0^{+\infty} f_i(w)\ \text{Prob}(\bigcap_{j \in L_n \setminus \{i\}} w_j \geq w + s_i - s_j)\ dw,$$

where $\text{Prob}(\bigcap_{j \in L_n \setminus \{i\}} w_j \geq w + s_i - s_j)$ is the joint probability that the total travel time using line i is better than or equal to that of any other line, when the waiting time of the next carrier of line i is equal to w. Since line headways are assumed to be independent, we may write:

$$\text{Prob}(\bigcap_{j \in L_n \setminus \{i\}} w_j \geq w + s_i - s_j) = \prod_{j \in L_n \setminus \{i\}} \text{Prob}(w_j \geq w + s_i - s_j).$$

Using:

$$\text{Prob}(w_j \geq w + s_i - s_j) = \overline{F}_j(w + s_i - s_j),$$

we can rewrite the probability of boarding line i as:

$$\pi_i = \int_0^{+\infty} \gamma_i(w)\ dw, \tag{1}$$

where $\gamma_i(w)$ denotes:

$$\gamma_i(w) = f_i(w) \prod_{j \in L_n \setminus \{i\}} \overline{F}_j(w + s_i - s_j), \tag{2}$$

and may be referred to as the probability density function of the waiting time conditional to boarding a carrier of line i. The expected waiting time conditional to boarding a carrier of line i is therefore:

$$EW_i = \int_0^{+\infty} w\gamma_i(w)\ dw.$$

Summing EW_i over all lines $i \in L_n$ gives the expected waiting time at the stop:

$$EW = \sum_{i \in L_n} \int_0^{+\infty} w\gamma_i(w)\ dw\ = \int_0^{+\infty} w \sum_{i \in L_n} \gamma_i(w)\ dw. \tag{3}$$

The sum $\sum_{i \in L_n} \gamma_i(w)$ may be considered as the probability density function of the expected waiting time at the stop, while the expected remaining time is:

$$ER = \sum_{i \in L_n} \pi_i s_i,$$

and the expected travel time is:

$$ET = EW + ER = \sum_{i \in L_n} \int_0^{+\infty} (w + s_i)\gamma_i(w)\ dw.$$

It's worth noting that a different expression of the expected travel time can be obtained. Indeed, the travel time y can be seen as the minimum of the travel times of the lines serving the stop: $y = \min\{w_i + s_i : i \in L_n\}$. Let $\phi(y)$ denote the probability density function of y and $\overline{\Phi}(y)$ the complement of its cumulative distribution function. Since the line headways are statistically independent, the probability that the travel time y is greater than or equal to a given value τ is given by the product of the probabilities that $w_i + s_i \geq \tau$, and can be expressed as:

$$\overline{\Phi}(\tau) = \text{Prob}(y \geq \tau) = \prod_{i \in L_n} \text{Prob}(w_i \geq \tau - s_i) = \prod_{i \in L_n} \overline{F}_i(\tau - s_i). \tag{4}$$

Furthermore, since $\phi(y) = 0$ for $y < 0$, then it is:

$$E[y] = \int_0^{+\infty} \tau\ \phi(\tau)\ d\tau = \int_0^{+\infty} \overline{\Phi}(\tau)\ d\tau. \tag{5}$$

Thus, combining (4) and (5) we have:

$$ET = E[y] = \int_0^{+\infty} \prod_{i \in L_n} \overline{F}_i(\tau - s_i)\ d\tau.$$

1.2 Bounded headway

If there exists a finite upper bound u_j of the probability density function $f_j(w)$, for at least one line $j \in L_n$, that is $f_j(w) = 0, \forall w \notin [0, u_j]$, then it may be useful for computational purposes to specify the upper and lower limits of the integrals of equations (1) and (3). First, observe the following:

$$\overline{F}_j(w + s_i - s_j) = \begin{cases} 1, & \text{if } w \leq s_j - s_i; \\ \int_{w+s_i-s_j}^{u_j} f_j(v)\ dv, & \text{if } s_j - s_i < w < u_j + s_j - s_i; \\ 0, & \text{if } w \geq u_j + s_j - s_i. \end{cases} \tag{6}$$

In the first case, line i dominates line j regardless of the value of the waiting time w_j (the total travel time $w + s_i$ results not greater than the remaining time s_j), while in the last case the reverse is true.

For added insights, assume that the lines in L_n are indexed in non-decreasing order of the remaining time, that is: $s_1 \leq s_2 \leq \ldots \leq s_n$, and denote $L_i = \{j \in L_n : j \leq i\}$ and $\overline{us} = \min\{u_j + s_j : j \in L_n\}$. This leads to the following expression for the line probability:

$$\pi_i = \begin{cases} \int_0^{\overline{us}-s_i} \gamma_i(w)\ dw, & \text{if } s_i < \overline{us}; \\ 0, & \text{otherwise.} \end{cases}$$

Then, the subset of the lines with positive probability is L_t, where $t = \max\{i : i \in L_n, s_i < \overline{us}\}$.

2 Common lines without on-line information at stop

Consider equations (1)-(2) and assume that all remaining times are equal: $s_i = s_j, \forall\ i, j \in L_n$, with $i \neq j$. We obtain:

$$\pi_i = \int_0^{+\infty} f_i(w) \prod_{j \in L_n \setminus \{i\}} \overline{F}_j(w)\ dw. \tag{7}$$

In this case, the passenger is indifferent to which line to board, and thus π_i denotes now the probability that a carrier of line i is the first one reaching the current stop. If L_n is indeed the attractive line set, the above equation reduces to the probability that a passenger boards line i under the classical assumption, which holds when no additional information is available. The classical common lines' model can be then seen as a particular instance of the general model with on-line information. In addition, this holds for general headway distributions, while existing results are only valid for the exponential distributions.

It's worth noting that the expected waiting time can also be expressed as:

$$EW = \int_0^{+\infty} \prod_{i \in L_n} \overline{F}_i(w)\ dw. \tag{8}$$

The above equation, firstly derived in Gendreau (1984), can be shown to be equivalent to equation (3) with $s_i = s_j$.

When there is some bounded headway, the upper limit of integration $+\infty$ in (7) and (8) can be replaced by $\overline{u} = \min\{u_j : j \in L_n\}$, because at least one carrier must arrive within time $\overline{u}$.

3 The Erlang headway distribution case

Depending on the probability density functions of the waiting times, expressions (1) and (3) may or may not have a closed form, and hence may or may not be easy to compute. For practical purposes, we will now focus on a particular family of headway distributions that can be applied to a large number of cases, namely the Gamma functions with integer parameter, which are also known as Erlang distributions.

First, note that for passengers who arrive randomly at a stop, the waiting time probability density function is related to the headway distribution through the formula (Larson & Odoni 1981):

$$f_i(w) = \frac{\int_w^{+\infty} g_i(h)\ dh}{E[h_i]} = \lambda_i(1 - G_i(w)), \tag{9}$$

where h_i is the headway of line i, $g_i(h)$ its probability density function, $G_i(h)$ its cumulative distribution function, and $1/E[h_i] = \lambda_i$ is the inverse of the mean headway, referred to as the frequency.

In the following, we will assume that the headway h_i of each line i is a random variable with an Erlang distribution having integer shape parameter m_i and expected value $1/\lambda_i$; that is:

$$g_i(h) = \begin{cases} \frac{\mathrm{e}^{-m_i\lambda_i h}\ (m_i\lambda_i)^{m_i}\ h^{m_i-1}}{(m_i-1)!}, & \text{if } h \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

The above distribution describes the sum of m_i independent Poissonian processes, so that, in this sense, the memory of the overall process lasts for m_i elementary events. Clearly the shortest is the memory of the process the highest is its irregularity. In order to represent this way the arrival of a given carrier that is correlated only with the arrivals of the m_i previous carriers of the same line, several approximations and assumptions are involved (Larson & Odoni 1981), which we adopt in the following without further discussion. Note that for $m_i = 1$, the Erlang distribution reduces to the exponential distribution, while for $m_i \to +\infty$, it yields the deterministic headway. Both these extremal distributions are widely used in transit models, while the Erlang distribution with general shape parameter has been extensively investigated in Gendreau (1984) and in Marguier & Ceder (1984). Introducing the Erlang headway into formula (9) yields the following probability density function of the waiting time:

$$f_i(w) = \begin{cases} \lambda_i\ \mathrm{e}^{-m_i\lambda_i w} \sum\limits_{k=0}^{m_i-1} \frac{(m_i\lambda_i w)^k}{k!}, & \text{if } w \geq 0; \\ 0, & \text{otherwise,} \end{cases}$$

and the following complement of the cumulative distribution function:

$$\overline{F}_i(w) = \begin{cases} \mathrm{e}^{-m_i\lambda_i w} \sum\limits_{k=0}^{m_i-1} (1 - \frac{k}{m_i}) \frac{(m_i\lambda_i w)^k}{k!}, & \text{if } w > 0; \\ 1, & \text{otherwise.} \end{cases}$$

It is not difficult to show that, for the present case, π_i, EW and ET may be written as a sum of integrals where each term has the following form:

$$\int \beta\ \mathrm{e}^{-\alpha x}\ x^r\ dx = -\beta\ \mathrm{e}^{-\alpha x} \sum_{k=0}^{r} \frac{r!}{k!}\ \frac{x^k}{\alpha^{r-k+1}} + c,$$

where α and β are real parameters, while r is an integer parameter and c is an additive constant. Therefore, for the Erlang distributions, the computations required by the proposed model can be performed in closed form. In Gendreau (1984) an efficient algorithm is proposed to compute the expected waiting time in the standard case without on line information. This scheme can be extended to the computation of the line probabilities and of the expected travel time in the case with on-line information.

In the next two sections, we will take a closer look to the two cases where all lines have an exponential or a deterministic headway distribution, respectively, as these pervade existing transit models.

3.1 The exponential case

Consider first the exponential headways. In this case, we have for each line i an exponential waiting time distribution with probability density function:

$$f_i(w) = \begin{cases} \lambda_i\ \mathrm{e}^{-\lambda_i w}, & \text{if } w \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

We then have the following complement of the cumulative distribution function:

$$\overline{F}_j(w + s_i - s_j) = \begin{cases} \mathrm{e}^{-\lambda_j(w+s_i-s_j)}, & \text{if } w > s_j - s_i; \\ 1, & \text{otherwise.} \end{cases}$$

To obtain the results for the case without on-line information it suffices to set the remaining times of all the different lines to a fixed value in the above expression, or equivalently to eliminate these times all

together. This leads respectively to the well known expressions for the line probabilities and the expected waiting time:

$$\pi_i = \int_0^{+\infty} f_i(w) \prod_{j \in L_n \setminus \{i\}} \overline{F}_j(w)\ dw = \int_0^{+\infty} \lambda_i \mathrm{e}^{-\lambda_i w} \prod_{j \in L_n \setminus \{i\}} \mathrm{e}^{-\lambda_j w}\ dw = \lambda_i \int_0^{+\infty} \mathrm{e}^{-(\sum_{j \in L_n} \lambda_j) w} dw = \frac{\lambda_i}{\Lambda_n},$$

$$EW = \int_0^{+\infty} \prod_{i \in L_n} \overline{F}_i(w)\ dw = \int_0^{+\infty} \prod_{i \in L_n} \mathrm{e}^{-\lambda_i w}\ dw = \int_0^{+\infty} \mathrm{e}^{-(\sum_{j \in L_n} \lambda_j) w} dw = \frac{1}{\Lambda_n},$$

where $\Lambda_n = \sum_{j=1}^{n} \lambda_j$ denotes the combined frequency of all lines.

3.2 The deterministic case

With deterministic headways, we have a uniformly distributed waiting time for each line i with probability density function:

$$f_i(w) = \begin{cases} \lambda_i, & \text{if } 0 \le w \le 1/\lambda_i; \\ 0, & \text{otherwise,} \end{cases}$$

where $1/\lambda_i = h_i = u_i$. Consequently, equation (6) becomes:

$$\overline{F}_j(w + s_i - s_j) = \begin{cases} 1, & \text{if } w \le s_j - s_i; \\ 1 - \lambda_j(w + s_i - s_j), & \text{if } s_j - s_i < w < u_j + s_j - s_i; \\ 0, & \text{if } w \ge u_j + s_j - s_i. \end{cases}$$

Again, we may easily derive the following results for the case without on-line information:

$$\pi_i = \int_0^{\overline{u}} f_i(w) \prod_{j \in L_n \setminus \{i\}} \overline{F}_j(w)\ dw = \lambda_i \int_0^{\overline{u}} \prod_{j \in L_n \setminus \{i\}} (1 - \lambda_j w)\ dw,$$

$$EW = \int_0^{\overline{u}} \prod_{i \in L_n} \overline{F}_i(w)\ dw = \int_0^{\overline{u}} \prod_{j \in L_n} (1 - \lambda_j w)\ dw.$$

4 The attractive line set

The framework for the computation of the line probabilities and the expected waiting time can be applied to any given subset $L \subseteq L_n$ of all the lines serving a given stop. Let ET^L denote the expected travel time from the considered stop to the destination corresponding to L. In this section we will address the problem of determining the *attractive line set*, which is a subset $L^* \subseteq L_n$ that yields the least expected travel time, once known the remaining times of the lines belonging to L_n.

We first analyze the case with on-line information. By definition in this case the expected travel time is $ET^L = E[\min\{w_j + s_j : j \in L\}]$. Since, for each $L \subseteq L_n$, $\min\{w_j + s_j : j \in L_n\} \le \min\{w_j + s_j : j \in L\}$, we have $ET^{L_n} \le ET^L$. This shows that the expected travel time corresponding to L_n is not greater than that corresponding to any of its proper subsets. Then, the attractive line set L^* can be obtained from L_n by eliminating all the lines i having probability $\pi_i = 0$. It follows that, in the case of bounded headways $L^* = L_t$ (see subsection 1.2), while in the case of unbounded headways $L^* = L_n$.

Note that the result just obtained does not hold for the case without on-line information. In this case, the determination of the attractive line set L^* requires the calculation of ET^L for all possible subsets of L_n. However, from a behavioral standpoint it is somewhat counterintuitive to consider an attractive set where any member of this set may have a larger remaining time than that of an excluded line. Therefore, in the sequel we will assume that passengers consider only the sets L_j, with $j = 1, \ldots, n$, as feasible line sets.

On this basis, a simple algorithm to determine the attractive line set is to calculate in turn ET^{L_i}, for $i = 1, \ldots, n$, and then choose: $L^* = \operatorname{argmin}\{ET^{L_i} : i = 1, \ldots, n\}$. It is well known (Spiess 1984, Nguyen & Pallottino 1988) that, in the particular case of exponential headways, a greedy approach can be adopted, where the progressive calculation of the values ET^{L_i} can be stopped as soon as adding the next line produces an increase of the expected travel time.

5 On the use of the proposed model in transit assignment

So far, we have assumed that the remaining times s_j, $j \in L_n$, are given constants and have proposed a method for determining the attractive line set, as well as the corresponding line probabilities and expected travel time, for a single transit stop. Actually, a transit trip may involve transfers, so that, in this case, the remaining times are themselves expected travel times resulting from downstream line choices.

In order to obtain an optimal strategy in the framework of transit assignment without recurring to an explicit enumeration of hyperpaths, the following *optimality principle* must hold: "all sub-strategies of an optimal strategy are optimal themselves". To prove that the optimality principle holds, we will investigate the sensitivity of the model with respect to changes of the remaining times.

In the case with on-line information, by definition, we have:

$$ET = E[\min\{w_j + s_j : j \in L_n\}] = \int_0^{+\infty} \dots \int_0^{+\infty} \min\{w_j + s_j : j \in L_n\} \prod_{j \in L_n} f_j(w_j)\, dw_1 \dots\, dw_n.$$

The function ET is continuous in the space $\Re^n$ of the remaining time variables $s_j, j \in L_n$. Meanwhile, note that the function $\min\{w_j + s_j : j \in L_n\}$ is not everywhere differentiable. Indeed, for any given w_i, we have:

$$\frac{\partial \min\{w_j + s_j : j \in L_n\}}{\partial s_i} = \begin{cases} 1, & \text{if } w_j \in (w_i + s_i - s_j, +\infty) \quad \forall\, j \in L_n \setminus \{i\}; \\ 0, & \text{if } \exists\, j \in L_n \setminus \{i\} :\ w_j \in [0, w_i + s_i - s_j). \end{cases}$$

Where i is not the only line yielding the minimum total travel time, so that there exists another line $j \in L_n \setminus \{i\}$ such that $w_j = w_i + s_i - s_j$, the derivative is not well defined. However, at these points the left partial derivative is 1 and the right partial derivative is 0. By splitting each integral defined on $[0, +\infty)$ into the sum of two integrals defined on $[0, w_i + s_i - s_j]$ and on $[w_i + s_i - s_j, +\infty)$, all the points such that $w_j = w_i + s_i - s_j$, with $j \in L_n \setminus \{i\}$, are on the boundary of the intervals of integration. These points form a set of null measure. Thus, by eliminating the zero valued integrals we obtain:

$$\frac{\partial ET}{\partial s_i} = \int_0^{+\infty} f_i(w) \left(\prod_{j \in L_n \setminus \{i\}} \int_{w + s_i - s_j}^{+\infty} f_j(w_j)\, dw_j \right) dw.$$

The nested form reduces to a product of integrals since the probability density functions are independent from each other. Taking into account equations (1) and (2), we obtain:

$$\frac{\partial ET}{\partial s_i} = \pi_i. \tag{10}$$

Equation (10) holds true also in the case without on-line information, where we have:

$$ET = EW + \sum_{i \in L_n} \pi_i s_i.$$

In fact, based on (7) and (8), EW and π_i, $i = 1, \dots, n$, are independent of the remaining times.

Based on (10), since line probabilities are non-negative, the optimality principle holds for the proposed model. Indeed, the expected travel time is minimized when all the remaining times are minimized.

6 Numerical examples

6.1 Common lines at a single stop

In order to evaluate the effect of on-line information at a stop on the line loads, we first apply the proposed framework to a very simple case with three lines having the same headway $E[h] = 20$ min and the following remaining times: $s_1 = 8$ min, $s_2 = 14$ min, $s_3 = 16$ min. In the case without on-line information, regardless the level of irregularity, all the three lines result to be attractive, and because they have the same mean headway they will have the same probability. On the contrary, in the case with on-line information it may happen that, based on the displayed waiting times, it is preferable to wait for

a faster line instead of boarding an incoming slower one. This produces higher probabilities for faster lines. In the above example we have in fact that the probability of the fastest line varies from 52.9% for the exponential case to 66.5% for the deterministic case; conversely, the probability of the slowest line varies from 20.1% to 12.5%. Thus, in the deterministic case providing the on-line information doubles the number of passengers choosing the fastest line and almost 2/3 leave the slowest line.

This small example shows that the on-line information may have a significant impact on the use of the transit lines, which must be taken into consideration when simulating and dimensioning the transit network for planning purposes.

We now analyze the impact of the on-line information on travel and waiting times and the sensitivity of the two models with respect to the service irregularity. Consider the example given in Table 1, where for simplicity all lines are assumed to have the same Erlang shape parameter m. The results are reported in Tables 2-3.

line i	mean headway $E[h_i]$	remaining time s_i
1	20 min	30 min
2	15 min	40 min
3	10 min	45 min

Table 1: Line attributes

m	π_1	π_2	π_3	EW	ER	ET
1	0.587	0.257	0.156	6.81	34.92	41.73
2	0.653	0.233	0.114	6.52	34.22	40.74
3	0.687	0.218	0.095	6.55	33.61	40.16
4	0.707	0.209	0.084	6.60	33.35	39.95
10	0.755	0.185	0.060	6.85	32.75	39.57
$+\infty$	0.805	0.160	0.035	7.27	32.12	39.39

Table 2: Line probabilities and travel costs with on-line information

m	π_1	π_2	π_3	EW	ER	ET
1	0.429	0.571	0	8.42	35.45	43.87
2	0.419	0.581	0	6.98	35.73	42.71
3	0.413	0.587	0	6.48	35.79	42.28
4	0.409	0.591	0	6.25	35.84	42.09
10	1	0	0	11.02	30.00	41.02
$+\infty$	1	0	0	10.00	30.00	40.00

Table 3: Line probabilities and travel costs without on-line information

For a same service regularity, the expected travel time when on-line information is available is systematically lower than that of the case without. The relative difference of the expected travel time between the two cases may decrease or increase when the regularity increases, and in general is not monotonic, as shown by the above example where we obtain the improvements $\Delta ET = (ET_{\text{without info}} - ET_{\text{with info}})/ET_{\text{without info}}$ reported in Table 4.

With on-line information, we note that the faster line 1, which is also the less frequent, has a high share (always greater than 50%) which increases monotonically with the regularity of the service. For the standard case when on-line information is not available, the faster line has a lower share (always smaller than 50%) when the service is irregular, while it is the only one to be utilized when the service is regular.

As expected, in both cases the more the service is regular the lower is the expected travel time. In addition, this improvement is monotonic. It also appears that the impact of service irregularity on the expected travel time is considerably lower when on-line information is available at the stop. Indeed, the gap in terms of expected travel time between the deterministic and the exponential headways decreases from 9.67% to 5.94%.

m	ΔET
1	4.88%
2	4.61%
3	5.01%
4	5.08%
10	3.53%
$+\infty$	1.53%

Table 4: Improvements due to the on-line information

In the case with on-line information, the expected waiting time increases with the regularity of the service as passengers accept to wait longer in order to board a line that takes less time to reach the destination. Note that in the case without on-line information, the number of attractive lines decreases when the regularity increases.

6.2 The attractive line set

As explained in section 4, in the case with on-line information the attractive line set may be retrieved by performing the computation with respect to the set of all lines available at the stop and by simply eliminating the lines having a null probability to be chosen.

Different is the case without on-line information. In this case, the determination of the attractive line set can be solved heuristically with the *sequential* approach. That is, all the lines are added in increasing order of their remaining times and the attractive line set is the set of the sequence with the least expected travel time. It's worth noting that the greedy approach (Chriqui & Robillard 1975), based on stopping the processes as soon as adding the next line produces an increase of the expected travel time, can be applied only to the exponential headways (as shown in Marguier 1981).

For instance, consider the example given in Table 5. For exponential headways, as shown in Table 6, we have: $ET^{1} = 57.00$, $ET^{1,2} = 49.10$ and $ET^{1,2,3} = 40.47$. While for deterministic headways, as shown in Table 7, we have: $ET^{1} = 42.00$, $ET^{1,2} = 42.30$ and $ET^{1,2,3} = 41.14$. In the latter case, the greedy approach produces the line set $\{1\}$ while the attractive line set is $\{1, 2, 3\}$.

line i	mean headway $E[h_i]$	remaining time s_i
1	30 min	27 min
2	50 min	38 min
3	5 min	40 min

Table 5: Line attributes

line set	info	π_1	π_2	π_3	EW	ER	ET
1	N	1	0	0	30.00	27.00	57.00
2	N	0	1	0	50.00	38.00	88.00
3	N	0	0	1	5.00	40.00	45.00
1 2	N	0.618	0.372	0	18.25	30.85	49.10
1 3	N	0.136	0	0.830	3.82	36.89	40.71
2 3	N	0	0.081	0.875	3.93	38.07	42.01
1 2 3	N	0.125	0.068	0.771	3.68	36.79	40.47
1 2 3	Y	0.430	0.071	0.480	5.60	33.49	39.09

Table 6: The exponential case

It's worth noting that, by rearranging the order of the lines according to the expected travel time of the line considered separately ($\{3, 1, 2\}$ in Table 6 and $\{1, 3, 2\}$ in Table 7), the greedy approach is successful. Moreover the sets of the sequence thus obtained are optimal compared to any other set with

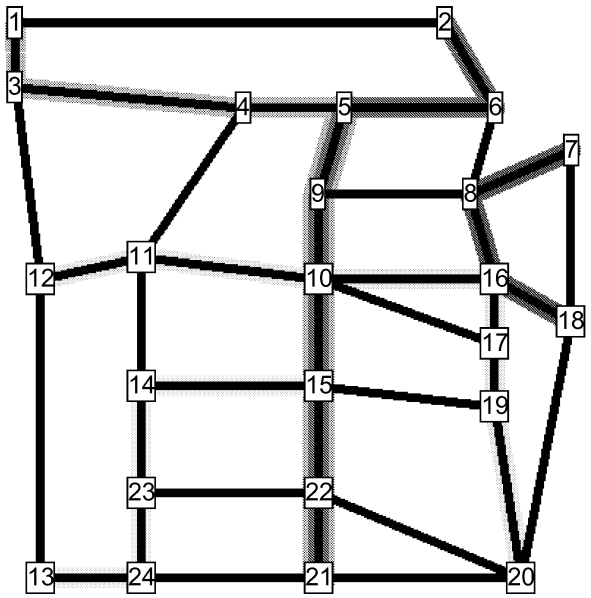
line set	info	π_1	π_2	π_3	EW	ER	ET
1	N	1	0	0	15.00	27.00	42.00
2	N	0	1	0	25.00	38.00	63.00
3	N	0	0	1	2.50	40.00	42.50
1 2	N	0.700	0.300	0	12.00	30.30	42.30
1 3	N	0.083	0	0.917	2.36	38.92	41.28
2 3	N	0	0.050	0.950	2.42	39.90	42.32
1 2 3	N	0.081	0.047	0.872	2.29	38.86	41.14
1 2 3	Y	0.510	0.049	0.441	5.08	33.29	38.37

Table 7: The deterministic case

the same number of lines. This is not proved to be a general result, however we were unable to find a counterexample.

When the on-line information is provided, the attractive line set is $\{1, 2, 3\}$ both in the case of exponential headways and in that of deterministic headways. Comparing the results relative to the attractive line set obtained for the corresponding cases with and without on-line information, we observe a remarkable difference in the line loads. In particular for deterministic headways, in the case without on-line information almost all users (87.2%) choose line 3 and just a few (8.1%) choose line 1, while in the case with on-line information line 1 is choosen by more than half users (51.0%) and the share of users choosing line 3 decreases to 44.1%.

6.3 Assignment on a small network



line i	mean headway $E[h_i]$	speed V_i	itinerary of line i
A	10 min	15 km/h	1 3 4 5 9 10 15 22 21
B	20 min	40 km/h	2 6 5 9 10 15 22 21
C	10 min	15 km/h	12 11 10 16 17 19 20
D	10 min	15 km/h	13 24 23 14 15
E	10 min	15 km/h	7 8 16 18

Figure 1: Transit network nodes and arcs

The transit network of Sioux Falls has been considered for a numerical application of the proposed framework. On a base network of 76 directed arcs and 24 nodes (all of which are centroids), the itineraries

of 5 bus lines are defined, covering the entire network (each node is served by at least one line). In particular, the original network presented in LeBlanc (1988) has been slightly modified (see Figure 1) by moving the south terminal of line B from node 5 to node 21 so that line A and line B overlap along the North-South direction between nodes 5 and 21. In addition, line B is designed as a low frequency express line, while all other lines have a slow speed and a frequency that is twice higher. This yields a supply configuration more suitable for the investigation of the impact on the passenger flow patterns when on-line information about the carriers' arriving times is available. The analysis is performed considering three Erlang headway distributions with different shape parameter: $m = 1$ (exponential headway), $m = 4$, and $m \to +\infty$ (deterministic headway).

	info	Y	Y	Y	N	N	N
	m	1	4	$+\infty$	1	4	$+\infty$
	total time	4419 h	3946 h	3670 h	4441 h	3980 h	3703 h
stops	passengers	share B	share B	share B	share B	share B	share B
5-9	838	0.345	0.360	0.387	0.321	0.331	0.365
9-10	1110	0.375	0.388	0.406	0.325	0.326	0.351
10-15	1648	0.443	0.466	0.472	0.322	0.297	0.348
15-22	795	0.491	0.537	0.572	0.330	0.314	0.445
22-21	275	0.516	0.593	0.615	0.331	0.345	0.673
21-22	275	0.531	0.582	0.607	0.342	0.327	0.673
22-15	795	0.508	0.548	0.557	0.342	0.316	0.425
15-10	1648	0.475	0.496	0.492	0.350	0.318	0.355
10-9	1110	0.496	0.502	0.505	0.424	0.438	0.430
9-5	838	0.505	0.515	0.525	0.452	0.492	0.480

Table 8: Total times and shares on line B

The results reported in Table 8 refer to an *all or nothing* network loading. The total time decreases substantially with the service regularity, but the improvement due to the presence of on-line information is not significant (less than 1%). The share of the fast line B increases with the service regularity and with the presence of on-line information. This result is more pronounced at the southern stops where the common lines A and B are in direct competition for a greater number of O/D pairs.

Other numerical results, which are not reported here for brevity, show that the impact of service regularity and the presence of on-line information is less relevant for short distance O/D pairs, as the relative weight of waiting time with respect to remaining time decreases. For this reason, on real networks the impact of providing on-line information on flow patterns may be light, yet definitely relevant.

Conclusion

We have developed in this paper a general framework for investigating passenger's route choice in transit networks with and without on-line information at stops on incoming line carriers' arrival times.

When common lines are available at a transit stop, passengers do not necessarily board the first arriving carrier if such information is provided to them. Assuming that the ultimate passenger's objective is to minimize his total travel time (waiting time plus remaining time), a model is proposed for determining the probability of boarding at a given stop each among a given set of lines and the expected waiting time when headways have any independent distribution.

In addition, the classical common lines' model, based on the assumption that passengers always board the first arriving carrier of their attractive line set, may be interpreted as a particular instance of the proposed framework. In fact, when the remaining times are ignored, the formulas developed here express the probability that the first arriving carrier belongs to a given line and the corresponding waiting time. Therefore, the present framework also extends the existing models to cover general headway distributions.

Numerical examples, concerning a single stop as well as a small network, are presented to validate the framework. Using the Erlang distribution with various shape parameters, and in particular with the limit parameters corresponding to deterministic and exponential headways, the case with on-line information is contrasted with the case without it. Drastic differences are observed in terms both of travel costs and of the proportion of passengers boarding slow and fast lines. Moreover, the impact of service regularity

is remarkable in both cases. These results imply that the existing models are inadequate to deal with the presence of on-line information at stops and to represent correctly the effect of service regularity, and thus may lead to significantly distorted traffic patterns for planning purposes.

References

Bouzaïene-Ayari, B., M. Gendreau & S. Nguyen 2001, 'Modelling bus stops in transit networks: A survey and new formulations', *Transportation Science* **35**, 304–321.

Bowman, L.A. & M.A. Turnquist 1981, 'Service frequency, schedule reliability and passenger wait times at transit stops', *Transportation Research A* **15**(6), 465–471.

Chriqui, C. & P. Robillard 1975, 'Common bus line', *Transportation Science* **9**(2), 115–121.

Gendreau, M. 1984, Une étude approfondie d'un modèle d'équilibre pour l'affectation des passagers dans les réseaux de transport en commun, PhD thesis, Département d'informatique et de recherche opérationnelle, Université de Montréal, Canada.

Hasseltröm, D. 1981, Public Transportation Planning – A Mathematical Programming Approach, PhD thesis, Department of Business Administration, University of Gothenburg, Sweden.

Jolliffe, J.K. & T.P. Hutchinson 1975, 'A behavioral explanation of the association between bus and passenger arrivals at a bus stop', *Transportation Science* **9**(3), 248–282.

Larson, R.C. & A.R. Odoni 1981, *Urban Operations Research*, Prentice-Hall, Englewoods Cliffs.

LeBlanc, L.J. 1988, 'Transit system network design', *Transportation Research B* **22**, 383–390.

Marguier, P.H.J. 1981, Optimal strategies in waiting for common bus lines, Master's thesis, Department of Civil Engineering, M.I.T., Cambridge, USA.

Marguier, P.H.J. & A. Ceder 1984, 'Passenger waiting strategies for overlapping bus routes', *Transportation Science* **18**(3), 207–230.

Nguyen, S. & S. Pallottino 1988, 'Equilibrium traffic assignment for large scale transit networks', *European Journal of Operational Research* **37**(2), 176–186.

Nguyen, S. & S. Pallottino 1989, Hyperpaths and shortest hyperpaths, *in* B. Simeone, ed., 'Combinatorial Optimization', Vol. 1403 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, pp. 258–271.

Nguyen, S., S. Pallottino & M. Gendreau 1998, 'Implicit enumeration of hyperpaths in logit models for transit networks', *Transportation Science* **32**(1), 54–64.

Seddon, P.A. & M.P. Day 1974, 'Bus passenger waiting times in Greater Manchester', *Traffic Engineering and Control* **15**(9), 442–445.

Spiess, H. 1984, Contributions à la théorie et aux outils de planification des réseaux de transport urbain, PhD thesis, Département d'informatique et de recherche opérationnelle, Université de Montréal, Canada.

Spiess, H. & M. Florian 1989, 'Optimal strategies: A new assignment model for transit networks', *Transportation Research B* **23**(2), 83–102.

Wu, J.H., M. Florian & P. Marcotte 1994, 'Transit equilibrium assignment: a model and solution algorithms', *Transportation Science* **28**(3), 193–203.