DYNAMIC ROUTE CHOICE IN A CONGESTED TRANSIT NETWORK WITH UNCERTAIN CARRIERS' ARRIVAL TIMES

Miss Valentina Trozzi, Centre for transport Studies – Imperial College London

Prof. Michael G.H. Bell, Centre for transport Studies – Imperial College London

Prof. Guido Gentile, Dipartimento di idraulica trasporti e Strade –Sapienza Universitá di Roma

Dr Achille Fonzone, *Centre for transport Studies – Imperial College London*

ABSTRACT

The purpose of this paper is to investigate the possibility of exploiting the hyperpath paradigm within the framework of dynamic assignment to model the route choice in a congested transit network, where passengers might encounter queues at transit stops and, thus, are not allowed to board the first approaching carrier of the attractive lines.

Keywords: dynamic route choice, hyperpath, transit network

INTRODUCTION

This paper develops a dynamic route choice model for a congested transit network where passengers behave according to a frequency-based approach, because there is a lack of information about exact vehicle arrival times at stops or because headways between carriers are so short that users perceive no advantage in timing their arrivals at stops with that of vehicles. In this case passengers can choose to board any among several competing (*common* (Chriqui and Robillard 1975)) lines in order to reach the destination, depending on which is the first available carrier.

In the static framework, where the relevant model variables, such as travel times and line frequencies, are fixed and passengers can always board the first approaching carrier, while congestion is limited to discomfort for vehicle overcrowding, the rational travel behaviour was firstly studied by Spiess (Spiess 1983; Spiess 1984), Spiess and Florian (Spiess and Florian 1989) and Nguyen and Pallottino ((Nguyen and Pallottino 1988), (Nguyen and Pallottino 1989)).

In the aforementioned works it is recognized that, under some simplifying assumptions about passengers' and carriers' arrivals (at each stop the headways of all transit lines are statistically independent with given exponential distributions, whose mean is equal to the inverse of the frequency (Nguyen and Pallottino 1988) and passengers' arrivals are uniformly distributed (Spiess and Florian 1989)), the most efficient route choice does not yield a single shortest path, but an optimal travel *strategy*, i.e. a set of routing rules that allow a passenger to reach his/her destination in the shortest possible time. The strategy is chosen before the beginning of the trip and, starting from the origin, involves the iterative sequence of: walking to a transit stop or to the destination, selecting the *attractive* lines to board and, for each of them, the stop where to alight. Moreover, based on the above definition, (Nguyen and Pallottino 1988; Nguyen and Pallottino 1989) provide a sound graph-theoretic framework for transit networks, which allows for the static representation of a strategies as a *hyperpath* that connects the origin of the trip to the destination having the diversion nodes at stops through waiting *hyperarcs*, each of which identifies a line set.

The challenge here is to investigate the possibility of developing a new model for passenger route choice within the framework of dynamic assignment by extending the hyperpath paradigm to the case where link travel times and transit frequencies vary during the day and congestion occurs at transit stops. Our model does not change basic assumption about carriers' and passengers' arrivals, nonetheless the boarding rule has to be adjusted to suit the congested scenario, where passengers cannot board the first attractive line approaching the stop, but suffer an over-saturation queuing time until the service(s) become actually available to them.

For the scope of this work, we assume the transit stop layout to be conceived so that passengers waiting for different line sets have to queue in one single row until the service becomes available to them. In particular, all passengers arriving at the stop join a *mixed queue*, where overtaking is possible only among passengers having different attractive sets, while any competition among passengers willing to board an approaching carrier is solved applying the FIFO rule (Trozzi, Haji Hosseinloo et al. 2010). The case where passengers *mingle* at the stop is not considered here, however a dynamic stop model capable of representing this situation has already been developed (Trozzi, Haji Hosseinloo et al. 2010) and can easily be embedded in the route choice model presented in this paper.

TRANSIT NETWORK AND HYPERPATHS: NOTATION AND FORMULATION

Basic Notation

As in (Meschini, Gentile et al. 2007), it is assumed here that the transit network comprises a set of lines $\Im \subseteq \aleph$ and that, together with the pedestrian network, it is represented by an *hypergraph* (Nielsen 2004). Basic definitions and general notation used to describe the hypergraph structure and the dynamic network model follow here:

HG = (N,E) = hypergraph modeling the topology of the transit-pedestrian network, where N is the set of nodes and E is the set of hyperarcs.

N = $(i \mid i = 1, 2, ..., n)$ = node set, comprising:

CN = centroid nodes, constituting the origins and destination of the journeys;
PN = pedestrian nodes;
RN = stop nodes;
LN = line nodes;
WN = waiting nodes.
Therefore $N = CN \cup PN \cup RN \cup LN \cup WN$
N_i^+ = $(j|(i,j) \in E)$; N_i^- = $(j|(j,i) \in E)$
E = $(e_1,...,e_m)$ = hyperarc set

e = $(TL(e), HD(e))$ generic hyperarc, where $TL(e) \in N$ and $HD(e) \subset N$ are respectively defined *tail* and *head* of the hyperarc. To be noticed that normal arcs are a sub-group of hyperarcs for which $HD(e) \in N$, or, equivalently, $|HD(e)| = 1$,where $|HD(e)|$ is defined as the *cardinality* of the hyperarc (Nielsen 2004). For the sake of clarity and simplicity, we will henceforth call "arcs" all the hyperarcs for which $|HD(e)| = 1$ strictly, and "hyperarcs" only those for which $|HD(e)| \geq 1$.

The hyperarc set E comprises five different subsets, namely:
PE = pedestrian arcs;
LE = line arcs;
SE = *support hyperarcs* (Trozzi, Haji Hosseinloo et al. 2010), representing all the lines serving a stop node. Namely, for the generic support hyperarc $a = (i, HD(a))$, where $i \in RN$ is a generic stop node and $HD(a) \in WN = N_i^+$ is the set of waiting nodes associated to *all* the lines serving the considered stop. (Trozzi, Haji Hosseinloo et al. 2010)
WE = *waiting hyperarcs* (Trozzi, Haji Hosseinloo et al. 2010), representing only the *attractive* lines serving the stop node. Namely $c = (i, HD(c))$ where $i \in RN$ is a generic stop node and $HD(c) \subseteq HD(a)$ represents the set of *attractive* lines serving the stop node i for passengers traveling along the hyperpath H_p at time τ ($L_{ip}^*(\tau)$).

Therefore $E = PE \cup LE \cup QE \cup SE \cup WE$.
Moreover, each arc $e \in LE$ is univocally associated with a line $\ell \in \Im$, and the same is true for any branch $b = (i, j)$ of the support or waiting hyperarcs.
H_p = hyperpath connecting a single origin destination pair (r, s).
P^{rs} = set of all possible hyperpaths connecting the origin destination pair (r, s): $H_p \in P^{rs}$.

Note that, in this dynamic and congested scenario, the branches of the hyperarc (Figure 1) not only represent the "*average delay due to the fact that the transit service is not continuously available over time*", but also the "*time spent by users queuing at the stop and waiting that the service become actually available to them*" (Meschini, Gentile et al. 2007). Moreover, we assume that passengers arriving at the stop join a unique, *mixed queue* regardless their particular *attractive lines' set*. In this case overtaking is possible among passengers having different attractive set; however, any competition among passengers willing to board an approaching line is solved applying the FIFO rule(Trozzi, Haji Hosseinloo et al. 2010).

Formatte
Formatte
Auto, Engl
Deleted:

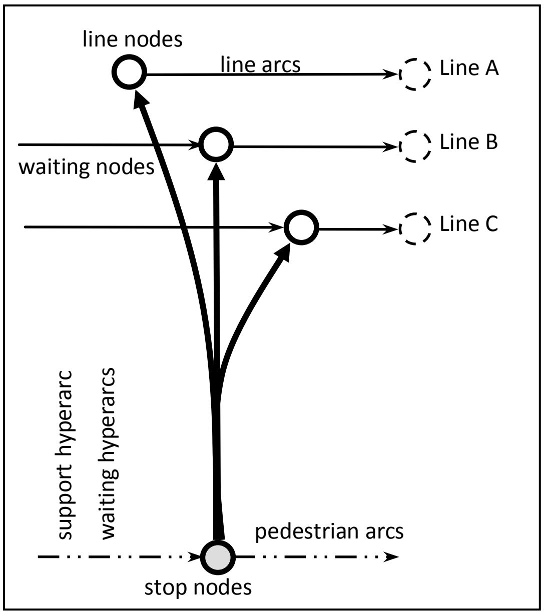


Figure 1: Representation of a stop in the hypergraph.

The above notation solely regards the topology of the network, while here after are introduced the variables describing the dynamic model. To be noticed that, because the analysis is carried out within a dynamic context, these variables are *temporal profiles* (Gentile, Meschini et al. 2004; Gentile 2006).

$x_{ij}(\tau)$ = aggregated flow entering arc $e=(i,j)\in E$ at time τ ;

$t_{ij}(\tau)$ = exit time from arc $e=(i,j)\in E$ for users entering it at time τ ;

$t_{ij}^{-1}(\tau)$ = entering time in arc $e=(i,j)\in E$ for users exiting it at time τ ;

$c_{ij}(\tau)$ = travel time of arc $e=(i,j)\in E$ for users entering it at time. Namely, if the arc is entered at time τ and the exit time is $t_{ij}(\tau)$, then $c_{ij}(\tau)=t_{ij}(\tau)-\tau$;

$\phi_{ij}(\tau)$ = mean frequency at time τ of the transit line associated with the branch $b=(i,j)$ of the hyperarc $e=(i,HD(e))$;

$AK_{ij}(\tau)$= available capacity of the line associated with the branch $b=(i,j)$ of the hyperarc $e=(i,HD(e))$ at time τ ;

$\theta_{ije}(\tau)$ = expected waiting time at time τ of the transit line associated with the branch $b=(i,j)$ of the waiting hyperarc $e=(i,HD(e))$;

$\theta_{ie}(\tau)$ = expected waiting time at the stop node i and at time τ , for the waiting hyperarc $e=(i,HD(e))$ associated to the considered stop;

$t_{ije}(\tau)$ = exit time from the branch $b=(i,j)$ of the waiting hyperarc $e=(i,HD(e))\in WE$ for users entering it at time τ ;

$t_{ije}^{-1}(\tau)$ = entering time in the branch $b=(i,j)$ of the waiting hyperarc $e=(i,HD(e))\in WE$ for users exiting it at time τ ;

$\pi_{ije}(\tau)$= diversion probability, namely the probability of using the branch $b=(i,j)$ of the hyperarc $e=(i,HD(e))$, conditional on being at node i at time τ ;

$\pi_p^{is}(\tau)$= probability of using hyperpath H_p at time τ, for users travelling from generic node i to destination s;

$g_p^{is}(\tau)$= actual cost of the hyperpath H_p for users leaving node i at time τ;

$S^{is}(\tau)$= cost of the minimal hyperpath H_p for users leaving node i at time τ.

Hyperpath formulation

When carriers' arrivals are perceived as uncertain by users and several routes are available to reach the destination from transit stops, it is assumed that passengers choose the best strategy to reach their destination rather than selecting a single path. According to Spiess (Spiess 1983; Spiess 1984), travel strategies consist in a set of rules, defined before the beginning of the journey, that are iteratively applied as the trip unfolds. Such strategies are graphically represented (Nguyen and Pallottino 1988; Nguyen and Pallottino 1989) as hyperpaths connecting the origin to the destination of the trip, where possible diversions are represented through *hyperarcs* exiting from stop nodes.

Therefore, because finding the best time-dependent strategy means, from a graphic point of view, finding the shortest (minimal) dynamic hyperpath, it is necessary here to provide a complete and formal definition of the dynamic hyperpath's structure and cost.

For the sake of clarity and simplicity, we will refer to a single rs pair, thus the notation relative to the origin and destination considered will be disregarded.

Definition.
A subgraph $H_p = (N_p, E_p, \pi_p(\tau))$ where $N_p \subset N, E_p \subset E$ and $\pi_p(\tau) = (\pi_{ije}(\tau))$ a real value vector of dimension $(|E_p| \times 1)$ is a *dynamic hyperpath* connecting origin $r \in CN_p$ and destination $s \in CN_p$, if:

1. H_p is acyclic with at least one arc;

2. node r has no predecessors and node s has no successors;

3. for every node $i \in N_p - \{r, s\}$ there is a hyperpath from r to s traversing i, and if node $i \notin R_p$ then it has at most one immediate successor;

4. the characteristic vector $\pi_p(\tau)$ satisfies the conditions:
$$\sum_j \pi_{ije}(\tau) = 1$$
$$\pi_{ije}(\tau) \geq 0 \qquad (1)$$
and the value of its component depends on the time τ they are evaluated;

5. travel times $c_{ij}(\tau)$ and traversing times $\theta_i(\tau)$ associated to arcs $(i, j) \in E_p$ and nodes $i \in N_p$ depend on the entering time τ they are evaluated.

In the static context, it is shown that the total travel time of the generic hyperpath H_p can be computed by explicitly taking into account all the elemental paths ℓ which form it (Nguyen and Pallottino 1988; Nguyen and Pallottino 1989). Therefore, if Q_p is the set of such paths, λ_ℓ is the probability of choosing the elemental path ℓ, and γ_ℓ is its total travel time, the travel time associated to hyperpath is H_p:

$$g_p = \sum_{\ell \in Q_p} \lambda_\ell \cdot \gamma_\ell \, . \qquad (2)$$

On the other hand, γ_ℓ can be expresses as the summation of travel and waiting times on the path's arcs and nodes:

$$\gamma_\ell = \sum_{(i,j) \in E_p} c_{ij} \cdot \delta_{ij\ell} + \sum_{i \in R_p} \theta_i \cdot \delta'_{i\ell} \, , \qquad (3)$$

where $\delta_{ij\ell}$ equals 1 if arc belongs to path ℓ and 0 otherwise, and $\delta'_{i\ell}$ equals 1 if path ℓ traverses node and 0 otherwise. Thus the following expression of the hyperpath's total travel time can be obtained:

$$g_p = \sum_{\ell \in Q_p} \lambda_\ell \cdot \left[\sum_{(i,j) \in E_p} c_{ij} \cdot \delta_{ij\ell} + \sum_{i \in R_p} \theta_i \cdot \delta'_{i\ell} \right] . \qquad (4)$$

However, in a congested network, as it is the case we are considering, travel times depend on the time the arc is entered. Consequently, it can happen the same node is traversed by different paths at different times and the travel cost associated to it has different values. Hence, the above definition of hyperpath's total travel times does not apply to the dynamic scenario.

Nevertheless, by extending to the dynamic context the recursive formula given in (Nguyen and Pallottino 1988; Nguyen and Pallottino 1989), we obtain a sequential definition of the dynamic hyperpath's travel time structure. Moreover, the equation allows for obtaining the result avoiding path enumeration, thus decreasing the computational burden.

Definition.

The total travel time of the dynamic hyperpath H_p connecting r to s and entered at time τ is sequentially defined in reverse topological (and chronological) order as:

$$g_p^{is}(\tau) = \begin{cases} 0, & \text{if } i = r \\ c_{ij}(\tau) + g_p^{js}\left(t_{ij}(\tau)\right), & \text{if } i \notin RN_p \\ \theta_{ie}(\tau) + g_p^{HD(e)s}\left(t_{iHD(e)}(\tau)\right), & \text{if } i \in RN_p \end{cases} \qquad (5)a$$

$$g_p^{is}(\tau) = \begin{cases} 0 \\ c_{ij}(\tau) + g_p^{js}\left(t_{ij}(\tau)\right) \\ \sum_{j \in L_{ip}(\tau)} \theta_{ije}(\tau) \cdot \pi_{ije}(\tau) + \sum_{j \in L_{ip}(\tau)} \pi_{ije}(\tau) \cdot \left[g_p^{js}\left(t_{ije}(\tau)\right) \right] \end{cases} , \qquad (5)b$$

where:

1. $\theta_{ie}(\tau)=\sum_{j\in L_{ip}(\tau)}\theta_{ije}(\tau)\cdot\pi_{ije}(\tau)$ is the total expected waiting (and queuing) time at stop i associated with the waiting hyperarc $e=(i,HD(e))\in WE$;

2. $g_p^{HD(e)s}(t_{iHD(e)}(\tau))$ is the remaining cost of the hyperpath, once boarded one of the attractive line at stop node i.

Note that the above formulation of hyperpaths' travel times structure and computation is independent of specific values given to the conditional probabilities $\pi_{ije}(\tau)$ and the functional form of the waiting cost $\theta_{ie}(\tau)$. These variables are specified by the stop model and depend on the particular stochastic model adopted for the passenger and the transit carrier arrivals at the stop node, and for the passengers' boarding mechanism.

THE STOP MODEL AND THE ATTRACTIVE SET

The stop model

Given a generic node $i\in RN_p$ and a time τ, the *stop model* yields the probability $\pi_{ije}{}^{s}(\tau)$ of using each branch of the associated waiting hyperarc (or, equivalently, the boarding probability for each attractive line of the transit stop). Moreover, it evaluates the expected waiting cost of any single attractive line $\theta_{ije}(\tau)$, as well as the total waiting cost $\theta_{ie}(\tau)$ associated with the stop.

In the static case, where:

- no information is provided at the stop on actual waiting times and on the available capacities of arriving carriers;
- the vehicle arrivals of different lines at the stop are statistically independent, and the same is true for the passenger arrivals with respect to vehicle arrivals;
- the headway probability distribution between two successive vehicles of the same line and hence the waiting time for a passenger randomly arriving at the stop are exponential, i.e. memoryless;

it is convenient to board the first attractive carrier that arrives the stop, instead of keep waiting (Spiess 1983; Spiess 1984).

On the other hand, if passengers have to queue until the service becomes actually available to them, the model has to be adjusted to represent the dynamic queue, which depends also on the layout of the stop and on the information eventually provided to passengers.

At this regard, we will show in the following only the results obtained for a stop served by a bunch of attractive lines, where passengers queue (therefore the FIFO rule is respected) to board the first available carrier.[1]

In this case, the passenger who arrives first at the stop is the first to board. Hence, if $x_{ij}(\tau)$ is the aggregated flow associated with the branch $b=(i,j)$ of the waiting hyperarc $e=(i,HD(e))\in WE$ at time τ, the $x_{ij}(\tau)$-th queuing passenger will have to wait for the $k_{ij}(\tau)$-th arrival, when the service will be truly available to him. An exact formula for obtaining $k_{ij}(\tau)$ should require the computation of the actual available capacity for the first $k_{ij}(\tau)$ carriers of the considered line reaching the stop (Meschini, Gentile et al. 2007). However, we assume here the temporal profile of the available capacity for the first approaching carrier is known and has the same value for all the first $k_{ij}(\tau)$ carrier arrivals. Therefore $k_{ij}(\tau)$ is equal to:

$$k_{ij}(\tau)=1+INT\left[\frac{x_{ij}(\tau)}{AK_{ij}(\tau)}\right] \qquad (6)$$

where *INT[x]* is the first integer not smaller than *x*.

Because headways are independently and equally distributed according to an exponential distribution of known parameter $\phi_{ij}(\tau)$, then the waiting time before the $k_{ij}(\tau)$-th arrival is distributed according to a Gamma(α,β), whit parameters $\alpha=k_{ij}(\tau)$ and $\beta=1/\phi_{ij}(\tau)$.

$$f_{ij}(w,\tau)=\begin{cases}\phi_{ij}(\tau)^{k_{ij}(\tau)}\cdot e^{-\phi_{ij}(\tau)\cdot w}\cdot\dfrac{w\cdot[k_{ij}(\tau)-1]}{[k_{ij}(\tau)-1]!}, & \text{if } w\geq 0\\ 0, & \text{otherwise}\end{cases} \qquad (7)$$

where w is the stochastic variable representing the waiting time.

In this case, the internal coefficient of the corresponding hyperarc (the so-called *boarding probability*) is equal to:

$$\pi_{ije}(\tau)=\int_0^{+\infty} f_{ij}(w,\tau)\cdot\prod_{h\in HD(e)\setminus\{j\}}\overline{F}_h(w,\tau)\,dw \qquad (8)$$

where $F_h(w, \tau)$ is the complement of the cdf for the branch *b=(i, j)* of the waiting hyperarc *e =(i, HD(e))*.

While the expected waiting time associated to the considered line at the stop node *i* and the total expected waiting time for the same transit stop are respectively equal to:

[1] For those who are interested in a complete discussion of the stop model, with a comparison between the case where passengers queue or mingle at the transit stop, we refer them to [6] and [Kurauchi, F., M. G. H. Bell, et al. (2003). "Capacity Constrained Transit Assignment with Common Lines." Journal of Mathematical Modelling and Algorithms **2**(4): 309-327.

$$\theta_{ije}(\tau) = \int_0^{+\infty} w \cdot \left\{ f_{ij}(w,\tau) \cdot \prod_{h \in HD(e)\setminus\{j\}} \overline{F}_h \right\} dw \tag{9}$$

$$\theta_{ie}(\tau) = \sum_{j \in HD(e)} \theta_{ije}(\tau) \cdot \pi_{ije}(\tau) \tag{10}$$

The attractive set determination

In general, the above expressions of the diversion probabilities and expected waiting times can be applied to any subset $L \subseteq N_i^+$, however, only a specific subset $L \subseteq N_i^+$ is associated with the minimum travel time to destination.

More specifically, the set of waiting nodes comprised in the head of the waiting hyperarc is defined *attractive* and is indicated as $L_{ip}^*(\tau)$. Also, each node of the set is associated with an attractive line. Therefore, $L_{ip}^*(\tau)$ is associated with all and only the *attractive lines* of stop i at time τ, for users travelling along the hyperpath H_p.

Recalling the definition of *attractive* set given in (Nguyen and Pallottino 1988), we can write:

$$\begin{aligned} &\exists\, L_{ip}^*(\tau) \subseteq N_i^+ : \\ &\sum_{j \in L_{ip}^*(\tau)} \theta_{ije}(\tau) \cdot \pi_{ije}(\tau) + \sum_{j \in L_{ip}^*(\tau)} \pi_{ije}(\tau) \cdot \left[g_p^{js}\left(t_{ije}(\tau)\right)\right] = \\ &\min_{L_{ip}(\tau) \subseteq N_i^+} \left\{ \sum_{j \in L_{ip}(\tau)} \theta_{ije}(\tau) \cdot \pi_{ije}(\tau) + \sum_{j \in L_{ip}(\tau)} \pi_{ije}(\tau) \cdot \left[g_p^{js}\left(t_{ije}(\tau)\right)\right] \right\} \end{aligned} \tag{11}$$

Consequently, in order to determine $L_{ip}^*(\tau)$, it is in general necessary to compute $g_p^{is}\left(\tau, L_{ip}(\tau)\right)$ for all the possible subsets of N_i^+. However, it is counter-intuitive to exclude a line from $L_{ip}^*(\tau)$ if it has a shorter remaining travel time than any other attractive one at the time τ. Therefore, at least for the static case, it is possible to solve (Chriqui and Robillard 1975; Nguyen and Pallottino 1988; Spiess and Florian 1989) the combinatorial problem described above through a greedy approach. Namely, the lines are processed in ascending order of their remaining travel time and the progressive calculation of the values of $\pi_{ije}\left(\tau, L_{ip}(\tau)\right)$, $\theta_{ije}\left(\tau, L_{ip}(\tau)\right)$ and $g_p^{is}\left(\tau, L_{ip}(\tau)\right)$ is stopped as soon as the addiction of the subsequent line increases the value of $g_p^{is}\left(\tau, L_{ip}(\tau)\right)$.

By contrast, in the dynamic scenario the only exact method of finding $L_{ip}^*(\tau)$ requires the enumeration of all the possible combination of lines serving the considered stop. In fact, the first step of a greedy procedure should require the sorting of nodes $j \in N_i^+ : i \in R_p$ in an increasing order of g_p^{js}:

$$g_p^{j_1 s}(\tau) \le g_p^{j_2 s}(\tau) \le \ldots \le g_p^{j_n s}(\tau), \qquad n = \left|N_i^+\right|, \tag{12}$$

but for each line corresponding to the generic node $j \in N_i^+ : i \in R_p$ the time at which g_p^{js} is computed actually depends on which other node $j \in N_i^+ : i \in R_p$ are included in the set of attractive waiting nodes $L_{ip}^*(\tau)$ or, equivalently, in the head of the waiting hyperarc $e = (i, HD(e)) \in WE$.

In order to overcome the computational complexity rising from the exact solution of problem (11), a heuristic methods for the determination of $L_{ip}^*(\tau)$ is proposed and discussed in the following paragraph.

Heuristic method for the attractive set determination

Because the lines cannot be exactly ordered in increasing order of remaining total travel time when the set of attractive lines is not known yet, it is suggested the value is computed for every line as if it were the only attractive one, namely:

$$\tilde{g}_p^{j_1 s}\left(t_{ij_1|j_1}(\tau)\right) \le \tilde{g}_p^{j_2 s}\left(t_{ij_2|j_2}(\tau)\right) \le \ldots \le \tilde{g}_p^{j_n s}\left(t_{ij_n|j_n}(\tau)\right), \qquad n = \left|N_i^+\right|, \tag{13}$$

where $t_{ij_k|j_k}(\tau)$ is the exit time from the waiting hyperarc formed by the only branch $b = (i, j_k)$, for users entering it at time τ, and the $\tilde{g}$ symbol reminds that the remaining total travel time computed here is approximated by considering only one *potential attractive* line at a time.

Furthermore, in the dynamic scenario congestion does not allow to consider waiting time at transit stops as exponentially distributed, but in the stop model it has been shown such times are distributed according to an Erlang pdf. Albeit the basic assumption allowing for adopting the greedy approach is not respected, the heuristic method proposed in the following performs the determination of the attractive set by adding one line at a time.

Heuristic method for determining the attractive set.
For computational reasons, the greedy approach is exploited for the update of $L_{ip}^*(\tau)$, even if it doesn't assure the minimum value $g_p^{is}\left(\tau, L_{ip}(\tau)\right) = S_p^{is}(\tau)$ is selected.

(1) Initialization

Sort the nodes $j \in N_i^+$ in increasing order of $\tilde{g}$:

$$\tilde{g}_p^{j_1 s}\left(t_{ij_1|j_1}(\tau)\right) \le \tilde{g}_p^{j_2 s}\left(t_{ij_2|j_2}(\tau)\right) \le \ldots \le \tilde{g}_p^{j_n s}\left(t_{ij_n|j_n}(\tau)\right), \qquad n = \left|N_i^+\right|.$$

Set $L_{ip}^*(\tau) := \{j_1\}$

$S_p^{is}(\tau) := \theta_{i|j_1}(\tau) + \tilde{g}^{j_1 s}\left(t_{ij_1|j_1}(\tau)\right)$

k := 2

(2) Updating $L^*_{ip}(\tau)$

while ($k \le n$)and $\tilde{g}^{j_k s}\left(t_{ij_1|j_1}(\tau)\right) < S^{is}_p(\tau)$ **do**:

$HD(e) := L^*_{ip}(\tau) \cup \{j_k\}$, where $e = (i, HD(e)) \in WE$ is the waiting hyperarc;

$$S^{is}(\tau) := \theta_{ie}(\tau) + \sum_{j_k \in HD(e)} \left\{ \pi_{ij_k e} \cdot \left[c_{j_k \sigma(j_k)}\left(t_{ij_k e}(\tau)\right) + S^{\sigma(j_k)s}\left(t_{j_k \sigma(j_k)}\left(t_{ij_k e}(\tau)\right)\right)\right]\right\}$$

$L^*_{ip}(\tau) := L^*_{ip}(\tau) \cup \{j_k\}$

$k := k + 1$

ROUTE CHOICE MODEL

If all transit passengers can be assumed to be rational decision makers, they will travel only along *shortest hyperpaths* connecting r to s at the departing at time τ. This assumption is formally stated by *dynamic* Wardrop's first principle, as formulated in (Gentile, Meschini et al. 2004) and (Gentile 2006), that we extend here to the dynamic hyperpath choice:

- if the hyperpath $p \in P^{rs}$ is used at time τ, namely $\pi_p(\tau)$>0, then $g^{rs}_p(\tau) \le S^{rs}(\tau)$
- on the other hand, if (sub)hyperpath $p \in P^{rs}$ is not used at time τ, then its actual cost is bigger than the minimum one

This condition can be formally expressed by:

$$S^{rs}(\tau) = \min\left\{g^{rs}_p(\tau) : p \in P^{rs}\right\} \quad (14)$$

$$\pi_p(\tau) \cdot \left(g^{rs}_p(\tau) - S^{rs}(\tau)\right) = 0 \quad (15)$$

Subject to:

$$\sum_{p \in P^{rs}} \pi_p(\tau) = 1$$
$$\pi_p(\tau) \ge 0 \quad (16)$$

To be noticed here that, if there is more than one shortest hyperpath connecting the origin-destination pair, the characteristic vector $\pi_p(\tau)$ solving the system (14)$\div$(16) is not unique.

In a static context, the shortest hyperpath search can be easily performed backward from the destination because the *concatenation property* holds true, namely any sub-hyperpath of a minimal hyperpath is itself minimal. Therefore, the deterministic route choice is solved through a recursive local search by applying *generalized Bellman's equations* (Nguyen and Pallottino 1988; Nguyen and Pallottino 1989). However, it is known from literature records that, the concatenation property can be exploited in the *dynamic shortest paths* selection only when the FIFO rule is respected, i.e. for a generic path ℓ connecting r to s it is always true that:

$\forall \tau' > \tau \Rightarrow T_\ell(\tau') > T_\ell(\tau)$

and the same applies to any generic arc of the considered path.

Thus, for the scope of our work, we have to prove that the FIFO rule allows to apply the *dynamic generalized Bellman's equation* for the shortest dynamic hyperpath search.

Preposition.

If $p \in P^{rs}$ is a shortest hyperpath from r to s for users departing at time τ, then the sub-hyperpath p_1 from r to any intermediate node i is itself the shortest for users travelling from r to i and departing at time τ, and the sub-hyperpath p_2 from any intermediate node i to s is itself the shortest for users travelling from i to s and departing at time $T_{p1}(\tau)$.

Proof.

If by contradiction there is a hyperpath $q \in P^{ri}$ that is faster than p_1 for users departing at time τ, then they can travel along q and arrive at the intermediate node i at time $T_q(\tau) < T_{p1}(\tau)$. However, for the FIFO rule $T_{p2}\left(T_q(\tau)\right) < T_{p2}\left(T_{p1}(\tau)\right) = T_p(\tau)$. Therefore, q is not the fastest path to arrive in i when departing from r at time τ, which contradicts the first hypothesis.

Proposition.

This dynamic version of the concatenation property referring to the departure times, allows us to prove that the following dynamic generalized Bellman's equation, where the temporal point of view is the exit time from the node, is equivalent to the problem expressed by equation (14) .

$$S^{is}(\tau) = \begin{cases} 0 & \text{, if } i = s \quad \text{(17)a} \\ \min\limits_{j \in N^+_i} \left\{ t_{ij}(\tau) + S^{js}\left(t_{ij}(\tau)\right) \right\} & \text{, if } i \notin RN_p \quad \text{(17)b} \\ \min\limits_{L^*_{ip}(\tau) \subseteq N^+_i} \left\{ \sum\limits_{j \in L^*_{ip}(\tau)} \theta_{ije}(\tau) \cdot \pi_{ije}(\tau) + \sum\limits_{j \in L^*_{ip}(\tau)} \pi_{ije}(\tau) \cdot \left[S^{js}\left(t_{ije}(\tau)\right) \right] \right\} & \text{, if } i \in RN_p \quad \text{(17)c} \end{cases}$$

Proof.

Equation (17)b is valid for $i \notin RN_p$ and can be proved ad in (Gentile, Meschini et al. 2004).

Consequently, the only part of the dynamic generalized Bellman equation that needs to be proved is equation (17)c.

In general, all the hyperpaths belonging to the set P^{is} can always be partitioned on the basis of their first (hyper)arc. In our case, because node $i \in RN_p$, the hyperpath is partitioned on the basis of its first waiting hyperarc, allowing for re-writing equation as(14):

$$S^{is}(\tau)=\min\left\{g_p^{is}(\tau): HD(e)\subseteq HD(a), p\in P_e^{is}\right\}, \tag{18}$$

where:

$e=(i,HD(e))\in WE, HD(e)\subseteq N_i^+$

$a=(i,HD(a))\in SE, HD(a)=N_i^+$

P_e^{is} is a subset of hyperpaths belonging to P^{is} that begin with the waiting hyperarc e.

As the hyperpaths in P_e^{is} coincide with hyperpaths in P^{js} but for the waiting hyperarc e, and given the definition of dynamic hyperpath's cost (5), it is possible to write:

$$S^{is}(\tau)=\min_{p\in P^{HD9e)s}}\left\{\min_{L_{ip}^*(\tau)}\left\{\sum_{j\in L_{ip}^*(\tau)}\theta_{ije}(\tau)\cdot\pi_{ije}(\tau)+g_p^{HD(e)s}\left(t_{iHD(e)}(\tau)\right)\right\}\right\} \tag{19}$$

But, recalling equation (5)b, and because the associativity holds for minimization, the right part of the equation becomes:

$$\min_{L_{ip}^*(\tau)}\left\{\min_{p\in P^{HD9e)s}}\left\{\sum_{j\in L_{ip}^*(\tau)}\theta_{ije}(\tau)\cdot\pi_{ije}(\tau)+\sum_{j\in L_{ip}^*(\tau)}\pi_{ije}(\tau)\cdot\left[g_p^{HD(e)s}\left(t_{iHD(e)}(\tau)\right)\right]\right\}\right\} \tag{20}$$

Moreover, because $\sum_{j\in L_{ip}^*(\tau)}\pi_{ije}(\tau)\cdot\theta_{ije}(\tau)$ is common to all the terms of the inner minimization, on the basis of equation (14), we obtain:

$$S^{is}(\tau)=\min_{L_{ip}^*(\tau)\subseteq N_i^+}\left\{\sum_{j\in L_{ip}^*(\tau)}\theta_{ije}(\tau)\cdot\pi_{ije}(\tau)+\sum_{j\in L_{ip}^*(\tau)}\pi_{ije}(\tau)\cdot\left[S^{js}\left(t_{ije}(\tau)\right)\right]\right\} \tag{21}$$

FORMULATION OF THE DYNAMIC SHORTEST PATH ALGORITHM

Given the proofs above, we can now develop the following recursive procedure to find the shortest dynamic hyperpath, for every departure time/arrival.

The analysis period $[0,\Theta]$ is here divided into M time intervals $\left(\tau^0,\tau^1,...,\tau^m,...,\tau^M\right)$, with τ^0 = 0 and τ^M = 0. The period between two successive intervals is called *temporal layer* (Chabini 1998; Gentile, Meschini et al. 2004). Moreover, the generic time profile, such as the line frequency or the available capacity, is given here through a piece-wise linear function defined by the values taken at such instants, for example we have:

$$\phi_{ij}(\tau)=\phi_{ij}^{\,m},\quad \tau\in\left(\tau^{m-1},\tau^m\right],\quad m=1,...,M\,. \tag{22}$$

On the other hand, we assume that outside the analysis period, namely $(\Theta,0)$, the network behaves as if it were is static and all the travel variables stays equal

In order to find the shortest dynamic hyperpath all the nodes are visited, in any topological order, starting from the last temporal layer to the value assumed for $\tau = \tau^0$.

When the computation is performed, the values of $\pi_{ije}(\tau), \theta_{ije}(\tau), \theta_{ie}(\tau), t_{ije}(\tau), c_{ij}(\tau), t_{ij}(\tau)$ and $S^{is}(\tau)$ are known for any time τ and for any node i, allowing for finding the shortest dynamic hyperpath for every departure time (Chabini 1998).

The following example explains how the procedure works.

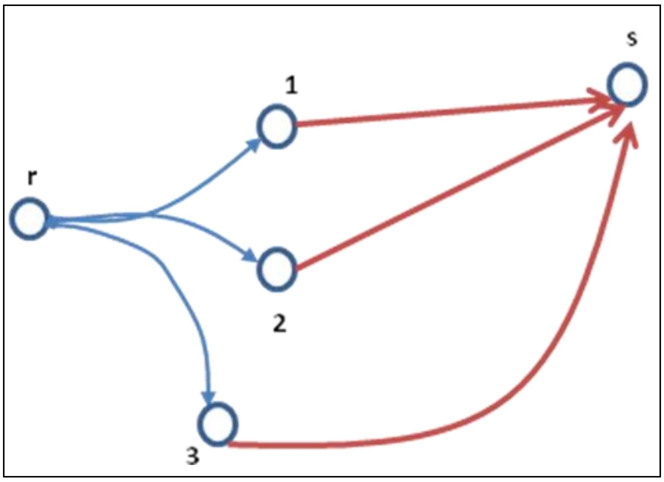


Figure 2: *Example network*.

Example.

For the scope of the example, we consider the simple network depicted in Figure 2 and the analysis period [08:00, 08:20], where each layer lasts only one minute.
We assume, the temporal profiles of frequencies and travel times are known and the number of the first available carrier, for any line, has already been computed according to equation (6). The tables **Error! Reference source not found.** show the data given:

Deleted:

Table Ia: Temporal profiles for the travel variables needed.

	08:00	**08:01**	**08:02**	**08:03**	**08:04**	**08:05**	**08:06**	**08:07**	**08:08**	**08:09**
$c_{1s}(\tau)$	5	5	5	5	5	6	6	6	6	6
$c_{2s}(\tau)$	3	3	3	3	3	4	4	4	4	4
$c_{3s}(\tau)$	7	7	7	7	7	7	7	7	7	7
$k_{r1}(\tau)$	2	2	2	2	2	1	1	1	1	1
$k_{r2}(\tau)$	3	3	3	3	3	2	2	2	2	2
$k_{r3}(\tau)$	1	1	1	1	1	1	1	1	1	1
$\varphi_{r1}(\tau)$	0.5	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
$\varphi_{r2}(\tau)$	0.33	0.33	0.33	0.33	0.33	0.25	0.25	0.25	0.25	0.25
$\varphi_{r3}(\tau)$	0.08	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.07	0.07

Table Ib: Temporal profiles for the travel variables needed.

	08:10	**08:11**	**08:12**	**08:13**	**08:14**	**08:15**	**08:16**	**08:17**	**08:18**	**08:19**
$c_{1s}(\tau)$	7	7	7	7	7	7	7	7	7	7
$c_{2s}(\tau)$	4	4	4	4	4	5	5	5	5	5

$c_{3s}(\tau)$	8	8	8	8	8	9	9	9	9	9
$k_{r1}(\tau)$	1	1	1	1	1	1	1	1	1	1
$k_{r2}(\tau)$	2	2	2	2	2	2	2	2	2	2
$k_{r3}(\tau)$	1	1	1	1	1	1	1	1	1	1
$\varphi_{r1}(\tau)$	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
$\varphi_{r2}(\tau)$	0.33	0.33	0.33	0.33	0.33	0.20	0.20	0.20	0.20	0.20
$\varphi_{r3}(\tau)$	0.08	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.07	0.07

Iteration 1: time layer 08:19.

(1): Initialization

Set: $S^{ss}(08{:}19) = 0$

$S^{is}(08{:}19) = \infty \ \forall i \neq s$

(2): Update labels for non-stop nodes, in any topologic order.

If: $i \notin RN$ and $S^{is}(\tau) > c_{ij}(\tau) + S^{js}(t_{ij}(\tau))$

Then: $S^{is}(\tau) := c_{ij}(\tau) + S^{js}(t_{ij}(\tau))$

$S^{1s}(08{:}19) = 7$
$S^{2s}(08{:}19) = 5$
$S^{2s}(08{:}19) = 9$

(3):- Update labels for stop nodes, in any topologic order

(3.1): Heuristic greedy initialization

Compute: $\theta_{rj_k|j_k}(\tau)$,

$t_{rj_k|j_k}(\tau)$

$\tilde{g}^{j_k s}(t_{ij_k|j_k}(\tau))$

$\theta_{r1|1}(08.19) = 3$, $t_{r1|1}(08.19) = 08.22$, $\tilde{g}^{1s}(08.22) = 7$

$\theta_{r2|2}(08.19) = 10$, $t_{r2|2}(08.19) = 08.29$, $\tilde{g}^{2s}(08.29) = 5$

$\theta_{r3|3}(08.19) = 15$, $t_{r3|3}(08.19) = 08.34$, $\tilde{g}^{3s}(08.34) = 9$

Rank: j_k in increasing order of $\tilde{g}^{j_k s}(t_{ij_k|j_k}(\tau))$

j_1 = node 2
j_2 = node 1
j_3 = node 3

Set: $L^*_{rp}(08.19) := j_1$ = node 2

$S^{rs}(08.19) := \theta_{rj1|j1}(08.19) + \tilde{g}^{j1s}(08.29) = 15$

$k := 2$

(3.2): Heuristic Greedy Update of $L^*_{rp}(08.19)$

$k \leq 3$

$\tilde{g}^{1s}(08.22) < S^{rs}(08.19)$

Then do:

Set: $HD(e) := L^*_{rp}(08.19) \cup \{j_2\}$

Compute: $\pi_{rj_k}(\tau)$

$\theta_{rj_k}(\tau)$

$\theta_{re}(\tau)$

$S^{j_k s}(\tau)$

$S^{rs}(\tau)$

$\pi_{rj_1}(08.19) = 0.140625$

$\pi_{rj_2}(08.19) = 0.859375$

$\theta_{rj_1}(08.19) = 0.527344$

$\theta_{rj_2}(08.19) = 2.05078125$

$\theta_{re}(08.19) = 1.836547887$

$S^{j_1 s}(08.19) = 5$

$S^{j_2 s}(08.21) = 7$

$S^{rs}(08.19) = 8.555297887$

Set: $L^*_{rp}(08.19) := L^*_{rp}(08.19) \cup \{j_2\}$

$k := k + 1$

end.

$k \leq 3$

$\tilde{g}^{3s}(08.34) > S^{rs}(08.19)$

After repeating the procedure above, for all the temporal layers in reverse chronological order, we obtain the result

Table IIa: Results obtained for all the time layers of analysis

	08:00	**08:01**	**08:02**	**08:03**	**08:04**	**08:05**	**08:06**	**08:07**	**08:08**	**08:09**
$\tilde{g}^{1s}(\tau)$	6	6	6	6	6	6	6	6	7	7
$\tilde{g}^{2s}(\tau)$	4	4	4	4	4	4	4	4	4	5
$\tilde{g}^{3s}(\tau)$	8	8	9	9	9	7	7	7	9	9
$\pi_{rj1}(\tau)$	0.18	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.11
$\pi_{rj2}(\tau)$	0.82	0.82	0.82	0.82	0.82	0.89	0.89	0.89	0.89	0.89
$\theta_{rj1}(\tau)$	0.78	0.78	0.78	0.78	0.78	0.3	0.3	0.3	0.3	0.3
$\theta_{rj2}(\tau)$	2.73	2.73	2.73	2.73	2.73	1.48	1.48	1.48	1.48	1.48
$\theta_{re}(\tau)$	2.38	2.38	2.38	2.38	2.38	1.35	1.35	1.35	1.35	1.35
$L^*_{rp}(\tau)$	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2
$S^{rs}(\tau)$	7.02	7.02	7.3	7.3	7.3	7.13	7.13	7.13	7.13	8.02

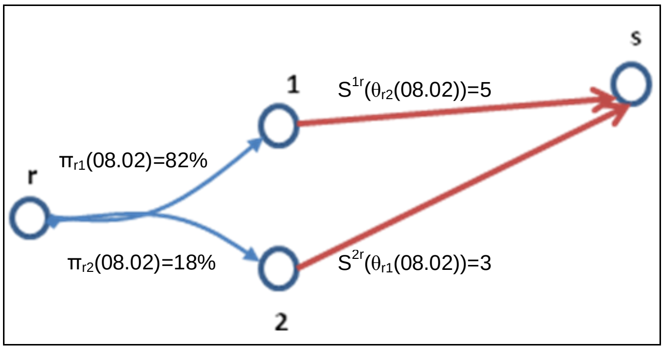
Table IIb: Results obtained for all the time layers of analysis

	08:10	**08:11**	**08:12**	**08:13**	**08:14**	**08:15**	**08:16**	**08:17**	**08:18**	**08:19**
$\tilde{g}^{1s}(\tau)$	7	7	7	7	7	7	7	7	7	7

$\tilde{g}^{2s}(\tau)$	5	5	5	5	5	5	5	5	5	5
$\tilde{g}^{3s}(\tau)$	9	9	9	9	9	9	9	9	9	9
$\pi_{rj1}(\tau)$	0.25	0.25	0.25	0.25	0.25	0.14	0.14	0.14	0.14	0.14
$\pi_{rj2}(\tau)$	0.75	0.75	0.75	0.75	0.75	0.86	0.86	0.86	0.86	0.86
$\theta_{rj1}(\tau)$	0.75	0.75	0.75	0.75	0.75	0.53	0.53	0.53	0.53	0.53
$\theta_{rj2}(\tau)$	1.5	1.5	1.5	1.5	1.5	2.05	2.05	2.05	2.05	2.05
$\theta_{re}(\tau)$	1.31	1.31	1.31	1.31	1.31	1.84	1.84	1.84	1.84	1.84
$L^{*}_{rp}(\tau)$	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2	j_1, j_2
$S^{rs}(\tau)$	7.56	7.56	7.56	7.56	7.56	8.56	8.56	8.56	8.56	8.56

Given the results listed in the tables above, it is now possible to compute the dynamic shortest hyperpath for *any* departure time. Indeed, for any τ, we can retrieve: the lines included in the attractive set of any stop node, the internal coefficient of the waiting hyperarc, the remaining travel time towards the destination

For instance, for the departure time τ =08.02



CONCLUSIONS

The main achievement of this paper has been the development of a new route choice model in a congested transit network, where passengers might encounter queues at transit stops and, thus, are not allowed to board the first approaching carrier of the attractive lines.

Moreover, if no information about exact vehicle arrival times at stops is assumed, or if headways between carriers are so short that users perceive no advantage in timing their arrivals at stops with that of vehicles, then users can choose to board any among several competing (*common* (Chriqui and Robillard 1975)) lines in order to reach the destination, depending on which is the first available carrier.

In this framework, it has been proved for the static case that the best travel behaviour (Spiess 1983; Spiess 1984) does not simply require the selection of the shortest path leading to the destination, but the selection of the best *travel strategy*.

Therefore, in this work we have tried to seek implications of extending the travel strategy approach to the dynamic context, verifying also how the definition of the hyperpath and its cost structure change when congestion occurs.

It has been shown that the only way to define hyperpath's cost is through a recursive formula (dynamic generalized Bellman equation) that works backwards in reverse topological order. Such formula can be applied also in this case because, as we have demonstrated, the *concatenation property* holds for the shortest dynamic hyperpath as for the static one.

On the other hand, the time-dependency of travel variables does not allow relying on the greedy approach to select attractive lines, if not as a heuristic method.

Future developments of our research, therefore, include: a comparison, in terms of algorithm complexity e precision, between the heuristic greedy and the exact approach for determining attractive lines at transit stops; an investigation of how should the route choice change if the networks include also (reliable) services for which headways are deterministically known; a development of a dynamic transit assignment.

REFERENCES

Chabini, I. (1998). "Discrete dynamic shortest path problems in transportation applications: complexity and algorithms with optimal running time." Transportation Research Record: Journal of the Transportation Research Board **1645**: 170-175.

Chriqui, C. and P. Robillard (1975). "Common bus lines." Transportation Science **9**(2): 115-121.

Gentile, G. (2006). Dynamic Traffic Assignment: a continuous macroscopic model. CORS / Optimization Days Monréal, Canada.

Gentile, G., L. Meschini, et al. (2004). Fast heuristics for continuous dynamic shortest paths and all-or-nothing assignment. AIRO 2004, Lecce, Italy.

Kurauchi, F., M. G. H. Bell, et al. (2003). "Capacity Constrained Transit Assignment with Common Lines." Journal of Mathematical Modelling and Algorithms **2**(4): 309-327.

Meschini, L., G. Gentile, et al. (2007). A frequency based transit model for dynamic traffic assignment to multimodal networks. Proceedings of 17th International Symposiumon Transportation and Traffic Theory London, Elsevier, London, UK.

Nguyen, S. and S. Pallottino (1988). "Equilibrium traffic assignment for large scale transit networks." European Journal of Operational Research **37**(2): 176-186.

Nguyen, S. and S. Pallottino (1989). Hyperpaths and shortest hyperpaths. Combinatorial Optimization. Berlin / Heidelberg, Springer **Volume 1403/1989:** 258-271.

Nielsen, L. R. (2004). Route Choice in Stochastic Time-Dependent Networks. Department of Operation Research. Aarhus, University of Aarhus**:** 168.

Spiess, H. (1983). On optimal route choice strategies in transit networks. U. d. M. Centre de recherche sur les transports. **Publication 286**.

Spiess, H. (1984). Contributions á la téorie at aux outils de planification des résaux de transport urbain. Department d'Informatique et de recherche opérationelle, Université de Monteréal. **Ph.D. Thesis**.

Spiess, H. and M. Florian (1989). "Optimal strategies: A new assignment model for transit networks." Transportation Research Part B: Methodological **23**(b): 83-102.

Trozzi, V., S. Haji Hosseinloo, et al. (2010). Dynamic hyperpaths: the stop model. International Conference on Models and Technology for Intelligent Transportation Systems. G. Fusco. Rome, Italy, Aracne.