Dynamic User Equilibrium in Public Transport Networks with Passenger Congestion and Hyperpaths

Valentina Trozzi[*a] Guido Gentile[b], Michael Bell[c], Ioannis Kaparias[d]

[a]*Centre for Transport Studies, Imperial College London, Skempton Building, South Kensington Campus, London SW7 2BU, UK*
[b]*Dipartimento di Ingegneria Civile, Edile e Ambientale, Universitá La Spaienza, Via Eudossiana, 18 – 00184 Rome, Italy*
[c]*Institute of Transport and Logistics Studies, University of Sydney, Newtown Campus C37 | NSW | 2006, Australia*
[d]*City Collaborative Transport Hub, City University London, Northampton Square, London EC1V 0HB, UK*

Abstract

This paper presents a Dynamic User Equilibrium for bus networks where recurrent overcrowding results in queues at stops. The route choice model embedded in the dynamic assignment explicitly considers *common lines* and strategies with alternative routes. As such, the *shortest hyperpath problem* is extended to a dynamic scenario with capacity constraints where the diversion probabilities depend on the time the stop is reached and on the expected congestion level at that time. In order to reproduce congestion for all the lines sharing a stop, the *Bottleneck Queue* model with time-varying exit capacity, introduced in Meschini et al. (2007), is extended. The above is applied to separate queues for each line in order to satisfy the First-In-First-Out principle within every *attractive set*, while allowing overtaking among passengers having different attractive sets but queuing single file.

Keywords: dynamic transit assignment; dynamic shortest hyperpaths; FIFO queues for attractive sets; Erlang pdf.

1. Introduction

Transit assignment models describe and predict the patterns of network usage by passengers and are seen as fundamental input for justifying resource allocation in transport planning. Available models differ

[*] Corresponding author. Tel.: +44-20-7594-2706.
E-mail address: v.trozzi09@imperial.ac.uk.

on the assumptions made about traveller behaviour and congestion effects. To simulate high-frequency or low-reliability services, frequency-based (FB) models are usually preferred, because it is plausible to assume that passengers do not choose a particular run and do not time their arrival at the stop to coincide with vehicle arrivals. Furthermore, these models are capable of handling easily the *common lines* dilemma [1] faced by passengers, who do not know if it is better to board the first line approaching a bus stop or keep waiting for a more convenient service. Indeed, since the seminal works of Nguyen and Pallottino [2] and Spiess and Florian [3], travel strategies (or, equivalently, hyperpaths) have been successfully exploited in FB models to represent passenger travel choice in public transport networks.

In the traditional formulation of the hyperpath concept [2-5], it is assumed that the strategy is chosen before the trip begins and, starting from the origin, it involves the iterative sequence of: walking to a public transport stop or to the destination, selecting the potentially optimal lines to board [formally defined *attractive lines*, 2] and, for each of them, the stop where to alight. If the only information available to passengers waiting at a stop is which bus arrives first and two or more attractive lines are available, the best option is to board the first approaching bus [4, 5]. As clarified by Bouzaiene-Ayari *et al.* [6] "*The outcome of such a choice is a set of simple itineraries that can diverge, only at bus stops, along the attractive lines*" and the realisation of the same travel strategy may change, from day to day, due to events at the stop, such as which attractive line becomes available first.

In the context of strategy-based assignment, passenger congestion has increasingly received attention in the literature [7-10] because recurrent queuing and overcrowding have developed into a major problem affecting high-frequency transit systems of large cities. For example, during peak-hours passengers often experience an over-saturation waiting time at stops, because they are not able to board the first vehicle of their choice set that arrives. The queue of those who remain at the stop may also increase passenger congestion for subsequent vehicle arrivals, and thus lead to great Level Of Service (LOS) variations that cannot be properly captured by static models, even when capacity constraints are considered [7-9, 11].

Despite the increasing interest in the literature, no broad agreement seems to have been achieved yet on the most appropriate method to deal with the effect of overcrowding. Consequently, this paper proposes a Dynamic User Equilibrium (DUE) model with hyperpaths on transit networks, where overcrowding may lead to the formation of First-In-First-Out (FIFO) queues of passengers at stops.

While strategy-based transit assignment with capacity constraints has been treated in a static framework [7-9, 11], only few dynamic models exist [10]. Moreover, previous work has dealt with capacity constraints in FB transit assignment with hyperpaths, either assuming that the extra waiting may affect the users' perception of service frequency, or that crowding may increase the perceived cost of travelling through the *fail-to-board probability*. The first method, formally known as *effective frequency*, has been developed only in a static framework, either considering the "practical capacity" of the transit lines [7] or strict capacity constrains [8, 9]; on the other hand the method of fail-to-board probability [10, 11] has been developed also in a dynamic framework. All the afore-mentioned methods assume that queuing takes the form of passengers mingling at the stop, without respecting any boarding priority. By contrast, it is assumed here that passenger congestion leads to the formation and dispersion of FIFO queues at transit stops.

In order to reproduce this phenomenon, it is assumed, as in Meschini *et al.* [12], that flows are macroscopic time-continuous functions, formally known as *temporal profiles* [13], and transit services are conceived as a continuous flow of supply with "instantaneous capacity" (namely, the supplied capacity is expressed in terms of passengers per hour and not passengers per vehicle), which allows representing the effect of time-discrete services as average waiting time.

Fig. 1 schematically presents the structure of the DUE model described in this work. The main inputs are: on the demand side, the time-varying Origin-Destination (OD) matrix; on the supply side, the network topology and the characteristics of the transit lines in terms of vehicle capacities, time-dependent service frequency and in-vehicle travel time.

The DUE can be generally regarded as a system of a demand model and a supply model. The former consists of:

- the *Stop Model* (SM), which yields for any given attractive line set [2] the probability of boarding each line as the first available one in the attractive set, as well as the expected (Erlang-distributed) waiting time, depending on transit lines characteristics and passenger congestion; and
- the *Route Choice Model* (RCM), which reflects the behaviour of a rational passenger travelling from an origin to a destination of the transit network for given arc performances (i.e. time-varying travel times and costs). The route choice in this paper is assumed deterministic and thus is modelled through a dynamic shortest hyperpath search.

On the other hand, the components of the supply model for dynamic assignment [14, p. 425] are:

- the *Network Flow Propagation Model* (NFPM), which aims at finding time-varying arc flows that are consistent with the arc travel times for given route choices (but not consistent with line capacities – this is the main difference between the NFPM and the Dynamic Network Loading Problem, where, instead, mutual consistence of flows and times through the APF is sought for given route choices);
- the *Arc Performance Functions* (APF), which yields for each arc the exit time at any given entry time, depending on the transit lines characteristics and the passenger flows over the network.

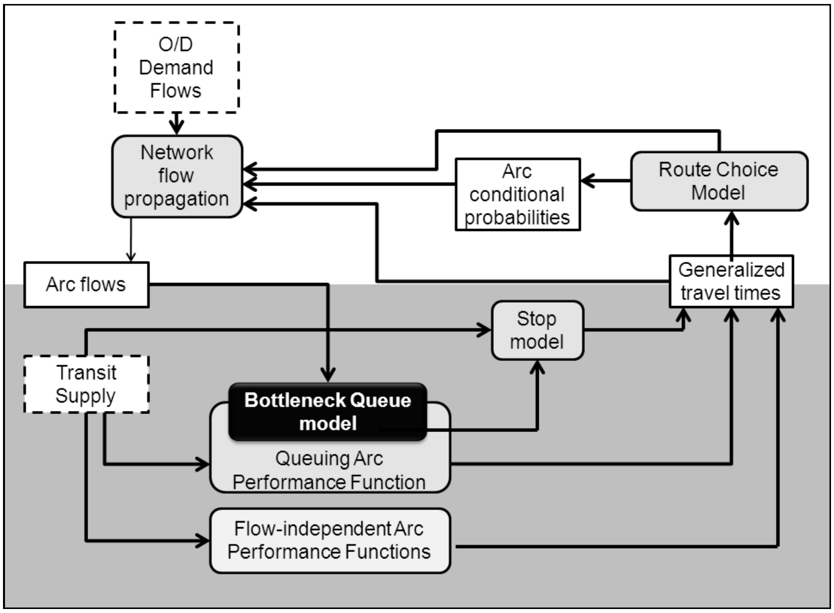

Fig. 1: Deterministic Dynamic User Equilibrium: model structure

Building on the work of [12], the present study extends the concept of hyperpaths to transit networks with passenger congestion, where the traditional assumption that passengers *always* board the first line of their choice set approaching the stop [2] does not hold because of the formation and dispersion of FIFO passenger queues. More specifically, it is assumed that, regardless of the specific line(s) that passengers might consider to board, they join a single "mixed" queue in the case of congestion. Thus, in order to reproduce overcrowding effects for all the lines sharing a stop, the *Bottleneck Queue* model with time-varying exit capacity, introduced in [12], is extended and applied to separate queues (one for each line of the stop) in order to satisfy FIFO within every *attractive set*, allowing overtaking among passengers having different attractive sets.

The rest of the paper is organised as follows. In next section, the background research is presented, while the notation and formulation is explained in Section 3. After detailing the methodology in Section 4, model implementation is presented in Section 5. Finally, Section 6 draws conclusions and identifies areas of future research.

2. Background research on transit congestion

Although recurrent passenger congestion is one of the major problems faced by public transport networks in large cities, it is only very recently that some authors have started to consider this dynamic phenomenon, and it does not seem that a full agreement has been achieved yet on the most appropriate methodology to be used. Some authors [15-18] have used the *diachronic graph* because it has the advantage of being inherently dynamic, in the sense that travel times and passenger flows on transit lines are always referred to a specific run, and thus to a certain time of the day. Moreover, this method has the additional benefit that the dynamic equilibrium reduces to a static assignment on the time-expanded network since the time dimension is explicitly reproduced by the graph topology. However, these models present a shortcoming on the demand side, because it is assumed that users would choose specific runs, rather than lines, and try to synchronise their arrival at stops with the scheduled passages of carriers. While this assumption is acceptable for services with low frequency and high regularity, it is questionable in densely connected urban networks, where passengers perceive runs of the same transit line as a unitary supply facility with high frequency or low regularity of service, since they do not consider timetables, even if available, when making their travel choices.

Alternatively, assignment models for congested transit networks can be developed using a *frequency-based* (FB) approach, which has the main advantage of requiring less detailed input data, on the supply side, and of representing more easily the *strategic* behaviour of passengers, on the demand side. In fact, FB models assume that passengers do not consider explicitly the service timetable, but make their travel choice on the basis of line average frequencies, while implicitly handling the uncertainty of the *common lines* dilemma.

Most of the research carried out in the realm of FB transit assignment with strategies (or, equivalently, hyperpaths), has dealt with overcrowding making the assumption that passengers mingle at the stop [7-11, 19, 20]. This implies that no priority is satisfied in the waiting process, and, in case of oversaturation, all passengers have the same probability to board the next arriving carrier, provided it is attractive. Under this hypothesis, a possible simple solution would consider the *effective frequency* [7], namely the line frequency perceived by waiting passengers, which decreases as the probability of not boarding its first arriving carrier increases, or the *fail to board* probability [10].

In general, when passenger congestion occurs, the queuing protocol followed by travellers is determined by the stop layout. For example, for stations and stops with large platforms, it is correct to assume mingling passengers, so that the expected waiting time as well as the passenger distribution among the attractive lines can be calculated according to the methods presented in the afore-mentioned studies. However, in urban bus networks the stop layout is usually such that, when passenger congestion occurs, users arriving at the stop tend to respect the boarding priority of anybody who has arrived before them. Models based on the mingling queuing protocol are clearly not applicable to the latter scenario.

The problem of modelling congestion in the form of passenger FIFO queues has proven to be even more challenging than the case of mingling. To the best of the authors' knowledge, all models developed so far for FIFO queuing [6, 21-23] make use of the following *stability condition* [6]: passengers waiting at a stop would consider an attractive set that is never completely saturated, in the sense that, at least for one of the attractive lines, passengers can board the first vehicle coming. This rather strong assumption has two crucial shortcomings:

- as congestion increases, more (and hence “worse”) lines are included in the attractive set; and
- if all lines are congested, passengers would rather walk than keep waiting (even if frequencies are high, so that the extra waiting time due to congestion is, anyhow, short).

A schedule-based approach has also been used [17, 18, 24] to represent the behaviour of passengers that choose a ranked set of *runs* (that make up an ordered set of “back-up travel plans”) and know that, at each intermediate stop, they may not be able to board the best option, or that they may not be able to sit, because of capacity constraints. Thus, in this case the concept of hyperpath is applied to a different scenario: passengers know and trust the service timetable (this is one of the basic assumptions of schedule-based models) and can precisely select their best travel option; however, it is uncertain if they will be able to board/sit when congestion occurs. By contrast, the model presented here considers the case where passengers have a good estimate of congestion levels, but do not know the exact arrival time of the vehicles.

Beyond the realm of hyperpaths in transit assignment, a different approach for dealing with formation and dispersion of passengers’ FIFO queues is presented by Meschini *et al.* [12], which is based on the formulation of a FB multimodal assignment by extending an existing approach for Dynamic Traffic Assignment (DTA) [13]. This considers flows as macroscopic time-continuous functions, and hence transit services are conceived as a continuous flow of supply with “instantaneous capacity” (e.g. 1000 passengers per hour and not 100 passengers per vehicle), which allows representing the effect of time-discrete services as the average waiting time. The limitation of that study, though, is that passenger strategies are not considered in the RCM, and it is not clear how the additional waiting time due to overcrowding should be accounted for when passengers in a queue are willing to board a set of lines (formally known as *attractive set* [2]) and passengers in the same queue have different attractive sets (e.g. because they are bound for different destinations).

As in [25], in this work the additional queuing time at stops due to passenger flows exceeding line capacities is calculated by means of a bottleneck model. Also, addressing the limitation of [12], it is assumed that, in the context of commuting passengers who know by previous experience the number of carrier passages they must let go before being able to board, the route choice can still be represented as the selection of an optimal strategy. Therefore, expected waiting times and passenger distribution among attractive lines may be calculated by extending the formulas presented by Nguyen and Pallottino [2] and Gentile *et al.* [26].

For clarity and simplicity, the current study concentrates solely on the transit assignment model and only delays due to queuing are considered. The extension to a multimodal assignment, as well as to other forms of transit congestion (dwelling times depending on boarding and alighting flows, on-board discomfort depending on in-vehicle overcrowding) can be easily carried out as in [12]. The next two sections give detailed insight into the proposed methodology, including network representation, mathematical notation and problem formulation.

3. Network representation

3.1. The transit network

The topology of the transit network, including the line routes and the pedestrian network, is described through a directed *base graph* $B = (V, E)$, where $V \subset \aleph$ is the set of *vertices* ($\aleph$ is the set of positive integer numbers), and $E \subseteq V \times V$ is the set of *edges* (Fig. 2). The generic edge $e \in E$ is univocally identified by its initial vertex $TL_e \in V$, or *tail*, and its final vertex $HD_e \in V$, or *head*. The generic vertex $v \in V$ is associated with a location in space, and thus is characterised by geographic coordinates; while the generic edge $e \in E$ is characterised by λ_e and ρ_e. The topology of each line $\ell \in L$ is defined by its *route* R_ℓ. The generic section is referred to as $(R_{\ell,i\text{-}1}, R_{\ell,i})\in E$, with $i\in[2, \sigma_\ell]$, and corresponds to an edge of the base graph. For any given vertex $v\in V$ and line $\ell\in L$, the function $s(v, \ell)\in[0, \sigma_\ell]$ yields, if it exists, the index such that $R_{\ell,s(v,\ell)} = v$, and 0 otherwise.

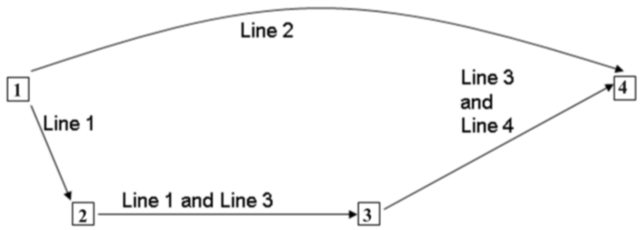

Fig. 2. Base Graph representation of a small network

Transit Network Nomenclature:

$\lambda_e\in\Re_+$: *length* of edge e ($\Re_+$ is the set of non-negative real numbers) ;

$\rho_e\in\Re_+$: *speed* on edge e; if $\rho_e = 0$ it means that a connection is not available;

$L\subset\aleph$: set of lines included in the transit network;

$\ell\in L$: generic transit line;

$R_\ell \subseteq V$: *route* of line ℓ. Ordered sequence of $\sigma_\ell\in\aleph$ (not repeated) vertices, each of which is denoted $R_{\ell,i}\in V$ with $i\in[1, \sigma_\ell]$;

$\chi_\ell\in\Re_+$: the *vehicle capacity* of its carrier;

$\varphi_\ell(\tau)\in\Re_+$: *base frequency*. Instantaneous flow of departures from the origin terminal $R_{\ell,1}$ at time $\tau\in\Re$;

$\theta_{\ell,i}(\tau)\in\Re_+$: *line time*. Time when a carrier of line $\ell\in L$, departed from $R_{\ell,1}$ at time $\tau\in\Re$, reaches the i-th vertex along its route, with $i\in[1, \sigma_\ell]$;

$\mu_{\ell,i}\in\{0,1\}$: if a *stop* is made or not at the i-th vertex along the route of line $\ell\in L$, with $i\in[1, \sigma_\ell]$;

3.2. From the base graph to the model graph

The physical topology of the transit network represented by B is not sufficiently detailed for modelling purposes. As such a *model graph* $G = (N, A)$ is introduced, where N is the set of *nodes*, $A \subseteq N\times N$ is the set of *arcs*; the generic arc $a\in A$ is univocally identified by its initial node $TL_a\in N$, or *tail*, and its final node $HD_a\in N$, or *head*, that is: $a = (TL_a , HD_a)$. The model graph is built on the grounds of the base graph, which is usually organised in a GIS database, considering the transit line routes and the transit and pedestrian speeds. Each node $i\in N$ is indeed the triplet of: a vertex $V_i\in V$, a *type* $T_i\in\{P, C, S, B, A, Q\}$, a line $L_i\in L\cup 0$. $N\subset (V\times\{P, S, B, A, Q\}\times L\cup 0)$.

Specifically, the node set and arc set are constructed as the union of the following subsets, detailed in the box below:

$N = N^P\cup N^C\cup N^S\cup N^B\cup N^A\cup N^Q$;

$A = A^P\cup A^L\cup A^D\cup A^Z\cup A^A\cup A^W\cup A^Q$.

Model graph Nomenclature:

N^P: *pedestrian nodes*, $N^P = \{(v, P, 0) : v \in V\}$;
N^C: *centroid nodes*, $N^C = \{(v, C, 0) : v \in V\}$;
N^S: *stop nodes*, $N^S = \{(R_{\ell,i}, S, 0) : \ell \in L, i \in [1, \sigma_\ell-1], \mu_{\ell,i} > 0\}$;
N^B: *boarding nodes*, $N^B = \{(R_{\ell,i}, B, \ell) : \ell \in L, i \in [1, \sigma_\ell-1]\}$;
N^A: *alighting nodes*, $N^A = \{(R_{\ell,i}, A, \ell) : \ell \in L, i \in [2, \sigma_\ell]\}$;
N^Q: *queuing nodes*, $N^Q = \{(R_{\ell,i}, Q, \ell) : \ell \in L, i \in [1, \sigma_\ell-1], \mu_{\ell,i} > 0\}$;
A^P: *pedestrian arcs*. Represent walking time: $A^P = \{(i, j) : i \in N^P, j \in N^P, e = (V_i, V_j) \in E, \rho_e > 0\}$;
A^L: *line arcs*. Represent in-vehicle travel time:
$A^L = \{(i, j) \in A: i \in N^B, j \in N^A, V_i \equiv R_{\ell,k}, V_j \equiv R_{\ell,k+1}, \ell \in L, i \in [1, \sigma_\ell-1])\}$;
A^D: *dwelling arcs*. Represent the time spent by the bus at a stop while passengers alight/board
$A^D = \{(i, j) \in A: i \in N^A, j \in N^B, V_i \equiv R_{\ell,k}, V_j \equiv R_{\ell,k+1}, \ell \in L, i \in [2, \sigma_\ell-1])\}$;
A^Z: *dummy arcs*. Introduced for algorithmic purposes to identify more easily arcs representing the waiting process. They connect the line stops and the pedestrian network: $A^Z = \{(i, j) \in A: i \in N^P, j \in N^S, V_i \equiv V_j\}$;
A^A: *alighting arcs*. Represent the time passengers need to disembark:
$A^A = \{(i, j) \in A: i \in N^A, j \in N^P, V_i \equiv V_j\}$;
A^Q: *queuing arcs*. Representing the *over saturation* waiting time, i.e. the "time spent by users queuing at the stop and waiting that the next service becomes actually available to them" [12]: $A^Q = \{(i, j) \in A: i \in N^S, j \in N^Q, V_i \equiv V_j\}$.
A^W: *waiting arcs*. Represent the *under saturation* waiting time, i.e. the average delay due to the fact that the transit service is not continuously available over time: $A^W = \{(i, j) \in A: i \in N^Q, j \in N^B, V_i \equiv V_j\}$;
FS_i : *forward star* of node $i \in N$. Set of arcs sharing the same tail i: $FS_i = \{a \in A: TL_a = i\}$;
BS_j : *backward star* of node $j \in N$. Set of arcs sharing the same head j: $BS_j = \{a \in A: HD_a = j\}$.

3.3. Hypergraph and hyperpaths

The *Network Flow Propagation Model*, which yields the inflow temporal profile of each arc, exploits the above graphic representation of the transit network as a directed graph G. On the other hand, the *Stop Model*, which yields the arc conditional probabilities at each node for every destination, and the *Route Choice Model* relate to a different graphic structure, that is a directed *hypergraph* $H = (N, F)$, built on the graph G. Let F be the set of *forward hyperarcs* [27], henceforth simply referred to as *hyperarcs*, included in hypergraph H. We assume: $F = A \setminus \{A^W \cup A^Q\} \cup A^H$, such that:

- A^H: is the set of *waiting hyperarcs* [28]. Represent the total expected waiting time (i.e. under saturation plus over saturation delay) for the considered set of *attractive lines* serving a stop: $A^H \subseteq \{(i, j) \in F: i \in N^S, J \subseteq N^B, j \in J, V_i \equiv V_j\}$.
- HFS_i : the *hyper-forward star* of node $i \in N$, i.e. the set of hyperarcs sharing the same stop tail i $HFS_i = \{a \in A^H: TL_a = i\}$.

Because a waiting hyperarc $h \in A^H$ is univocally identified by a singleton *tail* $TL_h \in N^S$ and by a set *head* $HD_h \subseteq N^B$, the hyperarc may be conceived as made up of "branches", each of which represents the total waiting and queuing time before boarding one specific line of the attractive set (more rigorously, a branch of the hyperarc represents the total waiting time conditional to boarding that specific line as the first available at the stop among all the lines in the considered attractive set). In the remainder of the paper, the following notation will be used to identify the branch of a hyperarc: $a \subseteq h \cdot \in A^H$.

The alternatives considered in the RCM are hyperpaths, and are defined as follows: a *hyperpath k* connecting origin $o \in N$ to destination $d \in N$ is a sub-hypergraph $H_{k,o,d} = (N_k, F_k)$ of H, where $N_k \subset N$, $F_k \subset F$, such that:

- $H_{k,o,d}$ is acyclic;
- o has no predecessors and one successor hyperarc;
- d has no successors and at least one predecessor hyperarc;

Each other node has one successor hyperarc and at least one predecessor hyperarc.

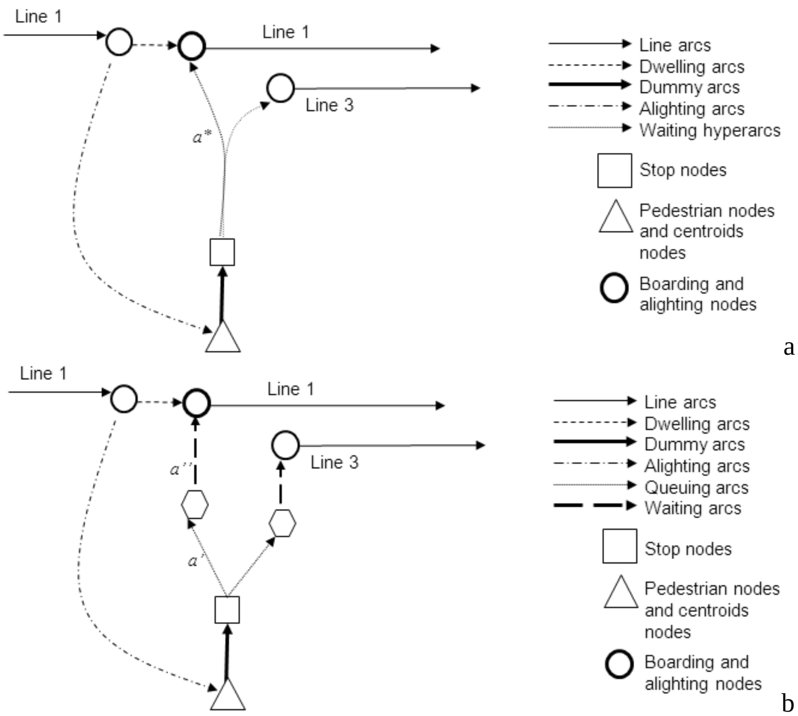

Fig. 3. Model graph (a) and hypergraph (b) representation of the same network of Stop 2 of Fig. 2

It should be noted here that, in contrast to what is assumed in [12, 25], the supply and demand models are represented by means of different graphic structures. This theoretical complication has important advantages from the algorithmic side, as it allows great flexibility in representing different phenomena that are relevant only in the demand or supply model.

For example, it is assumed that when travelers make their travel choices (demand model), they do not distinguish between the under-saturation waiting and the over-saturation queuing time. Moreover, they tend to group together partially overlapping lines that share the same stop and consider them as a unique travel alternative, with its own total waiting (plus queuing) time and travel time upon boarding. The hyperarcs of the hypergraph H are crucial to represent this combined waiting plus queuing time for a *set* of attractive lines considered together as a unique travel alternative.

On the other hand, when passenger flows are propagated across the network by the NFPM (supply model) it is very important to separate the waiting from the queuing time in order to ensure consistency between the flow that leaves a queuing arc and the capacity available on the vehicle; consequently, in the model graph *G* it is chosen to distinguish waiting and queuing arcs. Moreover, "combining" the total waiting times of several lines through a hyperarc is not effective in the NFPM, because it is necessary to build a model with separate queues. For example, if overcrowding occurs at a bus stop, the layout is usually such that all passengers wait in a single "mixed" queue, regardless the specific line(s) they may consider to board. Assume the first passenger in the queue does not board the first bus arriving at the stop because it is not attractive for him/her. Then, if there is available capacity on board, the first passenger can be overtaken by the second, third, etc. in the queue, if the service is in their attractive set. This overtaking among passengers with different attractive sets of lines, would violate the FIFO discipline of the "mixed" queue, but can be disregarded by developing a model with separate queues.

If the same graphic structure (hypergraph) is to be used to model the supply and demand when hyperpaths are considered in the RCM for dynamic assignment, then the under-saturation waiting and over-saturation queuing time have to be represented in the inverse order, as in [25]. Although this allows for representing separate queues even when passengers queue in a single file, this modelling assumption is questionable from a phenomenal point of view, since exactly the opposite occurs in reality, and does not guarantee the same flexibility of two separate models (supply and demand) that are referred to different graphic structures (model graph and hypergraph).

In order to clarify the difference between *G* and *H*, Fig. 3 shows the different representation of Stop 2 in Fig. 2 in the model graph and hypergraph. It is important to notice here that each branch of the hyperarc ($a^* \subseteq h \cdot \in A^H$) in the hypergraph *H* is associated with a line and corresponds to a queuing arc ($a' \in A^Q$) and the following waiting arc ($a'' \in A^W$) in the model graph *G*, such that $V_{TL_{a^*}} = V_{TL_{a'}}$ and $V_{HD_{a^*}} = V_{HD_{a''}}$.

4. Methodology

4.1. Problem definition

The extension of the first principle of Wardrop to a dynamic scenario allows for formulating the strategy-based dynamic transit assignment model studied here as a *User Equilibrium* (UE) that represents configurations in which no user can improve his/her travel cost at the time he/she is travelling by unilaterally changing hyperpath. The DUE can be specified as a fixed-point problem [14, p. 305, 464-467] by combining the supply and demand models or, equivalently, Network Loading Map ([13, 29], which combines the RCM and NFPM) and APF, as done here.

Therefore the following sub-components need to be specified and will be examined in the remainder of this section:

- *Arc Performance Functions*, that yield, for given passenger flows and transit supply, the temporal profiles of exit time for each arc;
- *Stop Model*, that yields for each destination, the temporal profiles of diversion probabilities $p_{a|h}(\tau)$, when passenger flows and transit supply are known;
- *Route Choice Model*, which reflects the behaviour of a rational passenger travelling from an origin to a destination for given profiles of: generalised costs, travel times and diversion probabilities. For each destination, it yields the *(hyper)arc conditional probability*, i.e. the probability that the considered (hyper)arc is included in the minimal hyperpath;
- *Network Flow Propagation Model*, which calculates temporal profiles of link flows that are consistent with line capacities for given *(hyper)arc conditional probabilities*.

Dynamic User Equilibrium Nomenclature:

$\varphi_a(\tau)$: *instantaneous frequency* (instantaneous flow of carriers) of the line corresponding to a, if a is entered at time τ. $\varphi_a(\tau) \neq \infty \ \forall \ a \in A^W \in A$, or $a \subseteq h \in A^H \in F$;

$t_a(\tau)$: *exit time* from arc a for users entering it at time τ;

$t_a^{-1}(\tau)$: *entry time* to the arc a for users exiting it at time τ;

$\kappa_a(\tau)$: *congestion parameter*, expressed as the number of vehicle arrivals that passengers, reaching the stop corresponding to the vertex V_{TLa} at time τ, have to wait before boarding the line associated with a. $\kappa_a(\tau)>0 \ \forall \ a \in A^Q \in A$, or $a \subseteq h \in A^H \in F$.

$w_h(\tau)$: *expected waiting time* for passengers reaching the stop of vertex V_{TLh} at time τ and considering the attractive line set represented by hyperarc $h \in A^H$;

$c_a(\tau)$: *travel cost* on arc a for users entering it at time τ:$a \in A^P \cup A^L \cup A^D \cup A^Z \cup A^A \in F$;

$c_h(\tau)$: *travel cost of the waiting hyperarc* for passengers reaching the stop of vertex V_{TLa} at time τ and considering the attractive line set represented by hyperarc $h \in A^H \in F$

$\eta_a(\kappa_a(\tau))$: *value of time* on a for users entering it at time τ, $a \in A^P \cup A^L \cup A^D \cup A^Z \cup A^A \in F$ or it may be a branch of an hyperarc $h \in A^H$. η_a is a function of the congestion at time τ;

$t_{a|h}(\tau)$:*conditional exit time*, for passengers reaching the stop of vertex V_{TLa} at time τ and considering the attractive line set represented by waiting hyperarc $h \in A^H$, relative to the event of boarding the line L_{HDa} associated with the branch $a \in h$;

$p_{a|h}(\tau)$:*conditional probability*, for passengers reaching the stop of vertex V_{TLa} at time τ and considering the attractive line set represented by waiting hyperarc $h \in A^H$, relative to the event of boarding the line L_{HDa} associated with the branch $a \in h$;

$g_{i,d}(\tau)$: *total travel cost* from node i to destination $d \in N^C$ at time τ;

$r_{a,d}(\tau)$: *diversion probability* of using hyperarc $a \in F$ for passengers entering it at time τ and directed to destination $d \in N^P$, among the hyperarcs of its tail forward star;

$q^{dem}_{TL_a,d}(\tau)$: *instantaneous demand flow* from TL_a to $d \in N^C$ at time τ. It is $\neq 0$ only if $TL_a \in N^C$;

$q^{in}_a(\tau)$: *instantaneous in-flow* of passengers entering arc $a \in A$ at time τ;

$q^{in}_{a,d}(\tau)$: *instantaneous in-flow* of passengers entering arc $a \in A$ at time τ and directed to destination $d \in N^C$;

$q^{IN}_a(\tau)$:*cumulative in-flow* of arc $a \in A$ at time τ, resulting from the network loading;

$q^{out}_a(\tau)$: *instantaneous out-flow* of passengers leaving arc $a \in A$ at time τ;

$q^{out}_{a,d}(\tau)$: *instantaneous out-flow* of passengers leaving arc $a \in A$ at time τ and directed to destination $d \in N^C$;

$q^{OUT}_a(\tau)$: *cumulative out-flow* of arc $a \in A$ at time τ, resulting from the network loading;

$e_a(\tau)$: *instantaneous exit capacity* of arc $a \in A$ at time τ;

$e^{CUM}_a(\tau)$: *cumulative exit capacity* of arc $a \in A$ at time τ;

$\mathbf{q^{dem}}$: *vector of demand flows*;

q: *vector of instantaneous arc* (*in/out*) *flows*;
g: *vector of travel cost to destinations*;
r: *vector of conditional probabilities.*

4.2. Arc Performance Functions

4.2.1. Flow independent performances

The APF of each arc $a \in A$ determines the temporal profile of the exit time for given entry time τ depending on the inflows of the arc itself and of its adjacent arcs at previous instants resulting from the network loading. Thus, in general the travel costs are not separable in either time or space.

In this particular formulation, except for the queuing arcs, all other arcs have a flow-independent exit time provided by the following equations, based on the exogenous inputs:

$$t_a(\tau) = \frac{\lambda_{ij}}{\rho_{ij}} + \tau \qquad \forall\ a \in A^P, \tag{1}$$

$$t_a(\tau) = \theta_{\ell,s}\left(\theta^{-1}_{\ell,s-1}(\tau)\right) \qquad \forall\ a \in A,^L \tag{2}$$

$$\varphi_a(\tau) = \frac{\varphi_\ell\left(\theta^{-1}_{\ell,s}(\tau)\right)}{1 - \dfrac{\partial \theta^{-1}_{\ell,s}(\tau)}{\partial \tau}} \qquad \forall\ a \in A^W, \tag{3a}$$

$$t_a(\tau) = \delta_a + \frac{1}{\varphi_a(\tau)} + \tau \qquad \forall\ a \in A^W, \tag{3b}$$

$$t_a(\tau) = \delta_a + \tau \qquad \forall\ a \in A^Z \cup A^D \cup A^A, \tag{4}$$

where: $ij = (V_{TL_a}, V_{HD_a})$; $\ell = L_{HDa} = L_{TLa}$ is the line corresponding to HD_a and TL_a; $s = s(L_{HDa}, V_{HDa})$; s -1 = $s(L_{TLa}, V_{TLa})$; δ_a is a constant representing alighting, boarding dwelling time and, for algorithmic purposes, also the travel time on dummy arcs.

4.2.2. Bottleneck queuing model with variable exit capacity

The queuing process can be seen as a “gate system”. As soon as passengers reach the stop they join a queue that can be thought of as being behind a gate; whoever is at the front of the queue passes through the gate (so the queuing time due congestion is over) and starts waiting for the next arrival (this is the under-saturation delay due to the discontinuity of the service).

Equation (3b) only considers the latter contribution, while the contribution due to overcrowding is considered in the performance function of queuing arcs, which in turn depends on the current length of the queue at the stop by means of a *Bottleneck Queuing Model* with time-varying exit capacity (equations (5) - (10)).

Although these equations may appear somehow complicated, their algorithmic implementation [13] is fairly straightforward, and only a conceptual explanation is given here.

With reference to Fig. 3b, consider the waiting arc *a''*. Its exit capacity at τ may be calculated as in equation (5), where $b \in A^D$ is the dwelling arc that enters the same boarding node of the waiting arc *a''*, and the out-flow is defined as in equation (6) by applying the FIFO and flow conservation rules [30].

$$e_{a''}(\tau) = \chi_\ell \cdot \varphi_{a''}(\tau) - q_b^{out}(\tau) \tag{5}$$

$$q_b^{out}(\tau) = q_b^{in}\left(t_b^{-1}(\tau)\right) \cdot \frac{\partial t_b^{-1}(\tau)}{\partial \tau} \tag{6}$$

$$e_{a'}\left(t_{a''}^{-1}(\tau)\right) = \frac{e_{a''}(\tau)}{\dfrac{\partial t_{a''}^{-1}(\tau)}{\partial \tau}} \tag{7}$$

The exit capacity of $a'' \in A^W$ coincides with the instantaneous capacity available on board (5) at the same time τ. On the other hand, for those who leave $a' \in A^Q$ at $t_{a''}^{-1}(\tau)$ and leave $a'' \in A^W$ at τ, the exit capacity $e_{a'}\left(t_{a''}^{-1}(\tau)\right)$ does not coincide with the instantaneous capacity available on board at the same time $e_{a''}\left(t_{a''}^{-1}(\tau)\right)$, but the exit capacity $e_{a''}(\tau)$ needs to be propagated backwards in time (7).

The temporal profiles of the exit capacity and inflows of the queuing arc $a' \in A^Q$ are then used to obtain the cumulative values of exit capacity and inflows (8).

$$q_{a'}^{IN}(\tau) = \int_0^\tau q_{a'}^{in}(\vartheta) d\vartheta, \quad e_{a'}^{CUM}(\tau) = \int_0^\tau e_{a'}(\vartheta) d\vartheta \tag{8}$$

$$q_{a'}^{OUT}(\tau) = \min\left\{q_{a'}^{IN}(\vartheta) + e_{a'}^{CUM}(\tau) - e_{a'}^{CUM}(\vartheta) : \vartheta \le \tau\right\} \tag{9}$$

$$q_{a'}^{IN}(\tau) - q_{a'}^{OUT}(\tau) = \int_\tau^{t_{a'}(\tau)} e_{a'}(\vartheta) d\vartheta \tag{10}$$

At this point, the cumulative flow leaving arc $a' \in A^Q$ is calculated in the spirit of the bottleneck model (9). If it is assumed that the queue at time τ began at a previous instant $\vartheta \le \tau$, then $q_{a'}^{OUT}(\vartheta) = q_{a'}^{IN}(\vartheta)$, and from ϑ to τ the cumulative flow of passengers that leave the arc $a' \in A^Q$ is $e_{a'}^{CUM}(\tau)$ - $e_{a'}^{CUM}(\vartheta)$, then equation (9) yields the cumulative number of passengers that have left the queue at time τ as the minimum among each cumulative outflow that would occur if the queue began at a previous instant $\vartheta \le \tau$.

On the other hand, the number of passengers queuing on arc $a' \in A^Q$ at time τ, which is $q_{a'}^{IN}(\tau)$ - $q_{a'}^{OUT}(\tau)$, based on equation (10) can be expressed also as the integral of the exit capacity from τ to the exit time $t_{a'}(\tau)$. The latter is consistent with the temporal profile of variable exit capacity of the queuing arc and is the output of the above bottleneck model.

During this period of time $[\tau, t_{a'}(\tau)]$ some buses of the line associated with $a' \in A^Q$ approach the stop, but queuing passengers cannot board them because of capacity constraints. Therefore, the number of bus passages that the passengers must let go before boarding is given by equation (11), where $\lfloor x \rfloor$ indicates the floor function of x.

$$\kappa_{a'}(\tau) = 1 + \left\lfloor \int_\tau^{t_{a'}(\tau)} \varphi_{a'}(\vartheta) d\vartheta \right\rfloor \tag{11}$$

This result will be, then, used as an input of the SM. Considering Fig. 3, if a^* is a branch of $h \in A^H$ such that $V_{TL_{a*}} = V_{TL_{a'}}$ and $V_{HD_{a*}} = V_{HD_{a''}}$, then $\kappa_{a*}(\tau) = \kappa_{a'}(\tau)$.

4.3. Stop Model

Once congestion is known for each line, the *Stop Model* yields: the diversion probabilities $p_{a|h,}(\tau)$; the total (waiting plus queuing) time spent at the stop $w_{h,}(\tau)$ for users that have selected the hyperarc h at time τ; and the exit time from $a \subseteq h$ $t_{a|h,}(\tau)$, conditional to the event of boarding the line corresponding to a if users have selected he attractive set corresponding to h.

When making their choice, passengers do not distinguish between under-saturation and over-saturation waiting time. Moreover, they tend to group together partially overlapping lines that share the same stop and consider them as a unique travel alternative, with its own total waiting (plus queuing) time and travel time upon boarding. w_h is this combined waiting (plus queuing) time for the *set* of attractive lines represented by the hyperarc h.

The model is formally described by equations (12) – (14) [31]:

$$p_{a|h}(\tau) = \begin{cases} \int_0^{+\infty} \gamma_a(\tau)\, dw, & a \subseteq h \\ 0, & a \not\subset h \end{cases}, \tag{12}$$

$$t_{a|h}(\tau) = \frac{1}{p_{a|h}(\tau)} \int_0^{+\infty} w \cdot \gamma_a(\tau)\ dw\,, \tag{13}$$

$$w_h(\tau) = \sum_{a \subset h} p_{a|h}(\tau) \cdot t_{a|h}(\tau)\,, \tag{14}$$

where $\gamma_a(\tau)$ is the probability to board the line associated with a at time τ and is defined as:

$$\gamma_a(\tau) = f_a(w,\tau) \cdot \prod_{b \subset h} \bar{F}_b(w,\tau),\ a \subseteq h\ , \tag{15}$$

and $f_a(w,\tau)$ and $\bar{F}_a(w,\tau)$ are respectively the Probability Density Function (PDF) and the survival function at time τ of the total time w before boarding the line associated with branch $a \subseteq h \in A^H$.

Given the conventional assumptions about passenger and carrier arrivals [2, 3], the total waiting time w before boarding the κ^{-th} carrier of the line associated with a may be described as a time dependent stochastic variable with Erlang distribution of parameters $\kappa_a(\tau)$ and $\varphi_a(\tau)$. Equation (16) shows the Erlang PDF of w, where, in order to improve readability, $\kappa_a(\tau)$ and $\varphi_a(\tau)$ are expressed as κ_a and φ_a.

$$f_a(w,\tau) = \begin{cases} \dfrac{\varphi_a{}^{\kappa_a} \cdot \exp(-\varphi_a \cdot w) \cdot w^{\kappa_a - 1}}{(\kappa_a - 1)!}, \text{if } w \geq 0 \\ 0, \text{otherwise} \end{cases} \tag{16}$$

It should be noted here that the results of the *Stop Model* are not destination specific and are used as an input by the *Route Choice Model*.

4.4. Route Choice Model

The *Route Choice Model* calculates (hyper)arc *conditional probabilities* on the grounds of: travel costs, travel times, *diversion probabilities* and *total waiting costs*. As for the *Stop Model*, the total waiting time w_h is associated with the hyperarc h.

Travel costs are assumed to be dependent on travel times through the value of time (VOT) function η_a, as in equations (17) – (19). The VOT is the sum of a constant parameter α_a and an increasing function of the congestion $\beta_a(\kappa_a(\tau))$, which is non-null only if: $a \subseteq h \in A^H$.

$$\eta_a(\kappa_a(\tau)) = \alpha_a + \beta_a(\kappa_a(\tau)) \tag{17}$$

$$c_h(\tau) = \sum_{a \subset h} p_{a|h}(\tau) \cdot \left(t_{a|h}(\tau) - \tau\right) \cdot \left(\alpha_a + \beta_a\left(\kappa_a(\tau)\right)\right) \tag{18}$$

$$c_a(\tau) = \left(t_a(\tau) - \tau\right) \cdot \alpha_a \quad \forall\ a \in A^P \cup A^L \cup A^D \cup A^Z \cup A^A \tag{19}$$

In order to avoid explicit hyperpath enumeration, the formulation of the RCM can be seen as a recursive local choice that starts from the destination $d \in N^C$ and proceeds backwards to the origin(s). At each intermediate node $i \in N$, passengers select the arc / hyperarc of its (hyper)forward star that minimises the total travel cost to d, as in equation (20).

$$g_{i,d}(\tau) = \min \begin{cases} 0, \text{if } i = d \\ c_a(\tau) + g_{HD_a,d}\left(t_a(\tau)\right) : a \in FS_i, \text{if } i \notin N^S \\ c_h(\tau) + \sum_{a \subseteq h} p_{a|h}(\tau) \cdot \left(g_{HD_a,d}\left(t_{a|h}(\tau)\right)\right) : h \in HFS_i, \text{if } i \in N^S \end{cases} \tag{20}$$

As said, passengers do not distinguish between the under-saturation delay and the over-saturation queuing time, but consider the total waiting time at the stop. Nonetheless, one would expect that the additional queuing time due to denied boarding will be perceived more negatively than the normal waiting time. Consequently, when calculating the cost of waiting c_h, the contribution of "total waiting time" brought by each line is weighted by the value of time, which is an increasing function of the congestion parameter κ_a.

The deterministic RCM can also be formulated as the following complementarity problem, formally known as *Wardrop inequalities*, referred to the decision of passengers leaving node $i \in N$ at time τ and directed toward destination $d \in N^C$:

$$\left.\begin{aligned} & r_{a,d}(\tau) \cdot \left(c_a(\tau) + g_{HD_a,d}\left(t_a(\tau)\right) - g_{TL_a,d}(\tau)\right) = 0, \forall a \in FS_i \\ & r_{a,d}(\tau) \geq 0, \forall a \in FS_i \\ & \sum_{a \in FS_i} r_{a,d}(\tau) = 1 \end{aligned}\right\} \text{, if } i \notin N^S \tag{21.1}$$

$$r_{h,d}(\tau)\cdot\left(c_h(\tau)+\sum_{a\in h}\left(p_{a|h}(\tau)\cdot g_{HD_a,d}\left(t_{a|h}(\tau)\right)\right)-g_{TL_a,d}(\tau)\right)=0,\forall h\in HFS_i$$

$$r_{h,d}(\tau)\geq 0,\forall h\in HFS_i \qquad \text{, if } i\in N^S \tag{21.2}$$

$$\sum_{h\in HFS_i} r_{h,d}(\tau)=1$$

$$r_{a',d}(\tau)=\sum_{h\in HFS_{TLa}}\left(r_{h,d}(\tau)\cdot p_{a*|h}(\tau)\right),\forall a'\in FS_i\cap A^Q \text{, if } i\in N^S \tag{21.3}$$

Equation (21.3) transforms the output of the RCM (hyperarc conditional probabilities, referred to elements of the hypergraph *H*) in the input of the NFPM (arc conditional probabilities, referred to elements of the model graph *G*). Thus, with reference to Fig. 3, the conditional probability of $h\in A^H \in F$ and the diversion probability of its branch a^* are transformed into the conditional probability of the arc $a'\in A^Q \in A$.

4.5. Network Flow propagation Model

This model gives the temporal profile of inflow and outflow on the grounds of the time-dependent travel demand and the results of the *Route Choice Model*.

The flow is propagated forward across the network, starting from the origin node(s). Once the intermediate node *i* is reached, the flow propagates along its forward star, according to equation (22).

The inflow $q^{in}_{a,d}(\tau)$ on arc $a\in A$ at time τ directed to destination $d\in N^C$ is given by the arc conditional probability $r_{a,d}(\tau)$ multiplied by the flow on node *i* (i.e. TL_a) at time τ. The latter is given, in turn, by the sum of: the flows $q^{out}_{b,d}(\tau)$ that leave each arc $b\in BS_i$ at time τ; and the demand flow $q^{dem}_{i,d}(\tau)$ at the same time. $q^{out}_{b,d}(\tau)$ is calculated as in equation (6) by applying the FIFO and flow conservation rules [30].

The arc flows directed towards different destinations are, then, combined together as in equation (23).

$$q^{in}_{a,d}(\tau)=r_{a,d}(\tau)\cdot\left(q^{dem}_{i,d}(\tau)+\sum_{b\in BSi}q^{out}_{b,d}(\tau)\right),\ a\in FS_i \tag{22}$$

$$q^{in}_{a}(\tau)=\sum_{d\in N^C}q^{in}_{a,d}(\tau) \tag{23}$$

4.6. Formulation of the Dynamic User Equilibrium as a Fixed-Point Problem

In order to formulate the DUE as a Fixed-Point Problem (FPP) with implicit path enumeration, the RCM (20) and the NFPM (22) are expressed in compact form respectively by equations (24) and (25) (RCM), and (26) and (27) (NFPM).

$$\mathbf{g} = \mathrm{g}(\mathbf{c}, \mathbf{t}, \mathbf{p}) \tag{24}$$

$$\mathbf{r} \in \mathrm{r}(\mathbf{g}, \mathbf{c}, \mathbf{t}, \mathbf{p}) \tag{25}$$

$$\mathbf{q} = \mathrm{q}(\mathbf{r}, \mathbf{t}, \mathbf{q}^{\mathbf{dem}}) \tag{26}$$

$$\mathbf{e} = \mathrm{e}(\mathbf{q}, \mathbf{t}) \tag{27}$$

It should be noted here that, as in [12], because the RCM considered is deterministic, when more than one arc in the forward star of a node minimises the total travel time from the node to destination, than the set of arc conditional probabilities solving equations (21) is not unique. Consequently, in equation (25), the symbol "=" is substituted by "∈". On the other hand, differently from [12], the compact formulation of the RCM clearly shows that the vector of conditional probabilities (**r**) also depends on the vector of diversion probabilities (**p**).

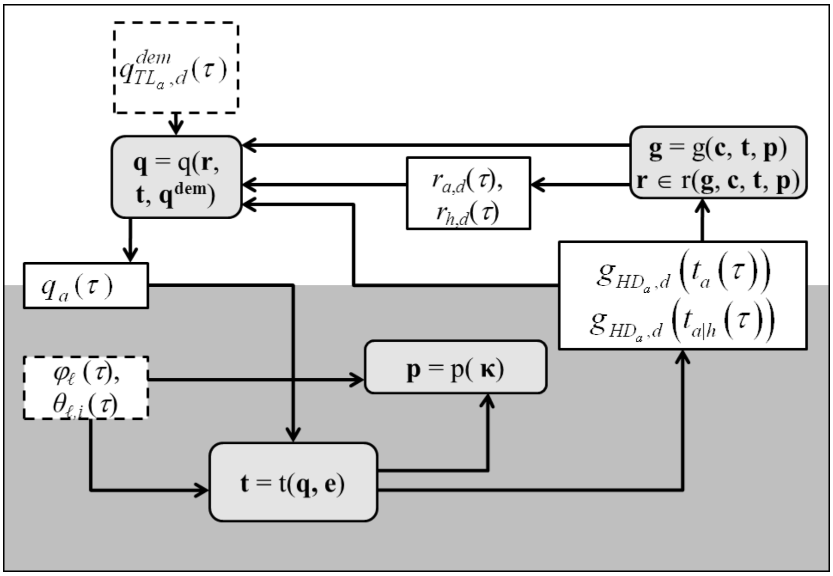

Fig. 4. Variables and models of the fixed-point problem formulation for the strategy-based Dynamic User Equilibrium of Fig.1.

Combining the RCM (24) – (26) and the NFPM (26), the following Network Loading Map ([13, 29]) is obtained

$$[\mathbf{q}, \mathbf{e}] \in \xi(\mathbf{t}, \mathbf{c}, \mathbf{p}, \mathbf{q}^{\mathbf{dem}}) \tag{28}$$

On the other hand, the APFs also imply equations (29) and (30), while the diversion probabilities calculated by the SM may be expressed in compact form as in (31). Thus, if the generalised arc cost functions (18), (19) are consolidated as in equation (32), then exit times, arc costs and diversion probabilities, which represent the supply, are expressed as in (33)

$$\mathbf{t} = \mathrm{t}(\mathbf{q}, \mathbf{e}) \tag{29}$$

$$\boldsymbol{\kappa} = \kappa(\mathbf{t}) \quad (30)$$

$$\mathbf{p} = \mathrm{p}(\boldsymbol{\kappa}) \quad (31)$$

$$\mathbf{c} = \mathrm{c}(\mathbf{p}, \boldsymbol{\kappa}, \mathbf{t}) \quad (32)$$

$$[\mathbf{t}, \mathbf{c}, \mathbf{p}] = \upsilon(\mathbf{q}, \mathbf{e}) \quad (33)$$

In conclusion, the DUE is obtained as a FPP combining equations (28) and (33):

$$[\mathbf{q}, \mathbf{e}] \in \varsigma\,(\mathbf{q}, \mathbf{e}, \mathbf{q}^{\mathbf{dem}}) \quad (34)$$

Fig. 4 summarises the FPP and shows the inputs and outputs of the sub-models included in the DUE represented in Fig. 1.

Following [14], it is possible to prove the existence of the network equilibrium because all the maps and functions, defined over the nonempty, compact, and convex set of arc flows, are (upper semi) continuous. Conversely, it is not possible to prove mathematically the uniqueness of the equilibrium because the problem does not have separable APFs (as in [32]), but the generalised travel cost for queuing arcs depends on the link flow on the queuing arc considered, as well as on adjacent dwelling arcs (equations (5) – (7)).

Finally, it is important to note that the formulation of a deterministic UE, as in this case, implies the assumption that users have a full and correct perception of generalised travel costs and choose travel alternatives with minimum cost. Thus, at the equilibrium, the same *od* pair may be connected by several minimal hyperpaths or the total generalised travel cost may be minimised through a split of the demand flows among different strategies, as shown in [8, 33].

5. Model implementation and numerical examples

In order to implement the strategy-based dynamic assignment procedure detailed in Section 4, a solution procedure is developed, which extends to the dynamic setting the original formulation given by Nguyen and Pallottino [2] and Spiess and Florian [3] in their seminal works on static strategy-based transit assignment.

Obviously, in the case of interest the algorithm structure is more complicated because the presence of congestion requires an additional feedback loop expressing the dependency of generalised travel times on flows. This loop, formalised with the FPP illustrated in Section 4.6, can be solved, as usual, by means of the MSA. Thus, the algorithm includes 3 parts:

- Part 0: Equilibrium arc flows initialisation to zero.
- Part 1: Hyperpath search.
  For every possible destination, shortest all-to-one hyperpaths (or hypertrees) are found scanning the network in reverse topological order;
- Part 2: Assign demand according to shortest hyperpaths.
  For every possible origin, the travel demand is loaded scanning the network in topological order. Loading flows are obtained;
- Part 3. MSA

If the difference between equilibrium and loading flows is greater than a fixed quantity, apply the MSA.

Because the proposed model requires to consider explicitly the time dimension, a solution algorithm can be devised extending the Decreasing Order of Time (DOT) method, proposed by Chabini [34] for the Dynamic Shortest Path Problem (DSPP) solely, to the problem of strategy-based dynamic assignment.

Therefore, although the proposed model has a continuous time representation, a discrete-time representation for its numerical solution has been adopted. The main idea is to divide the analysis period $AP = [0, \Theta]$ into T time intervals, such that $AP = \{\xi^0, \xi^1, ..., \xi^\tau, ..., \xi^{T-1}\}$, with $\xi^0 = 0$ and $\xi^{T-1} = \Theta$, and to replicate the network along the time dimension, forming a time-expanded hypergraph HG_T containing *vertices* in the form (i, ξ), and *edges* in the form $((i, \xi), (j, t_{ij}(\xi)))$.

If time intervals are short enough to ensure that the exit time of a generic edge $t_{ij}(\xi^\tau)$ is not earlier than the next interval $\xi^{\tau+1}$, for $\tau \leq T$-2, it is ensured that the network is cycle-free and the *vertex* chronological order is equivalent to the topological one.

Thus, in Part 1 of the algorithm HG_T is scanned starting from the last temporal layer to the value assumed for $\xi = \xi^0$ and, within the generic layer, no topological order is respected. When a generic vertex (i, ξ^τ) is visited, its forward star is scanned in order to set the minimal travel cost to the destination and the successive edge by means of equation (20). On the other hand, in Part 2 and 3 HG_T is scanned in chronological order without respecting any topological order within the same time layer. Firstly (Part 2), when a generic vertex (i, ξ^τ) is visited, the flow is loaded along its forward star by means of equation (22), then (Part 3) if the convergence criterion is not met, the MSA is applied. Finally the APF and congestion parameters are updated for every edge according to equations (1) – (11).

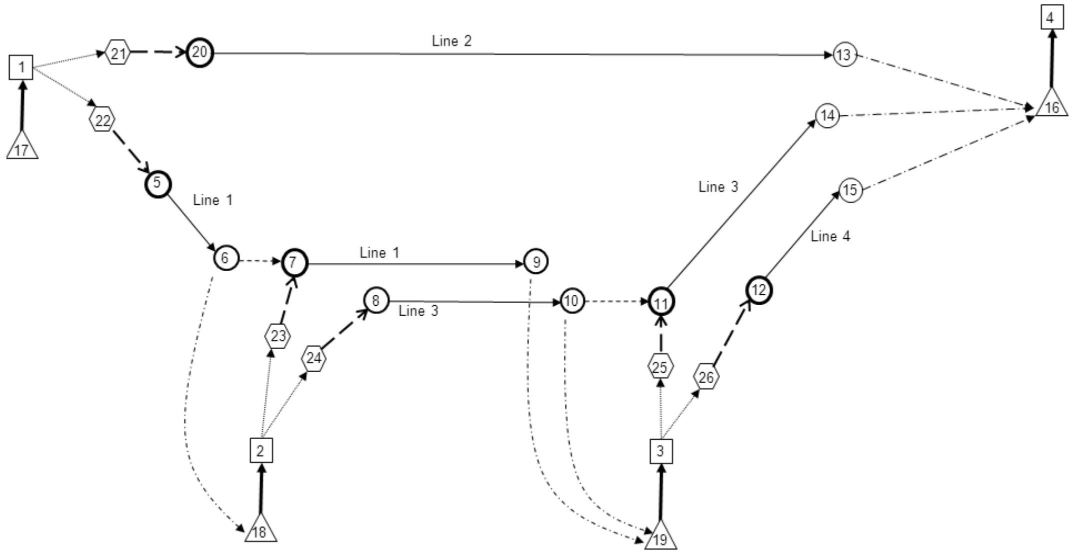

Fig. 5. Model graph representation of the example network.

The proposed methodology and solution algorithm have been applied to solve dynamic assignment problems for the small example network depicted in Fig. 2, whose model graph is shown in Fig. 5. Nodes 1, 2, 3, 4 $\in N^S$ and represent, respectively, Stop 1, Stop 2, Stop 3 and Stop 4 of Fig. 2, while nodes 17, 18, 19, 16 $\in N^C$ and represent centroids connected, respectively, to Stops 1, 2, 3, and 4. Also the two different route sections connecting Stop 2 and Stop 3 are represented by distinct line arcs (7, 9) and (8, 10), and similarly the two route sections connecting Stop 3 and Stop 4 are represented by distinct line arcs (11, 14) and (12, 15).

In order to highlight the different hyperpath selection when passenger queues arise, travel times and frequencies are assumed to stay constant during the analysis period [07:30-09:00], and are displayed in Table 1, together with the vehicle capacity. The analysis period is divided in one-minute intervals, and the travel time on alighting, dwelling and dummy arcs is assumed to be of one minute. Finally, for every arc/branch of hyperarc, η_a is assumed to be constant and equal to 1.

Table 1. Frequencies, in-vehicle travel times and vehicle capacity of the lines associated with line arcs of the network.

Line	Route section	Frequency (vehicles/min)	In-vehicle travel time (min)	Vehicle capacity (passengers)
2	Stop 1 – Stop 4	1/6	25	50
1	Stop 1 – Stop 2	1/6	7	50
1	Stop 2 – Stop 3	1/6	6	50
3	Stop 2 – Stop 3	1/15	4	50
3	Stop 3 – Stop 4	1/15	4	50
3	Stop 3 – Stop 4	1/3	10	25

Table 2. Time-dependent OD matrix during the analysis period for the first example considered

Origin Centroid	Destination Centroid	Travel demand (passengers/min)
17	16	5
18	16	7
19	16	7
16	16	0

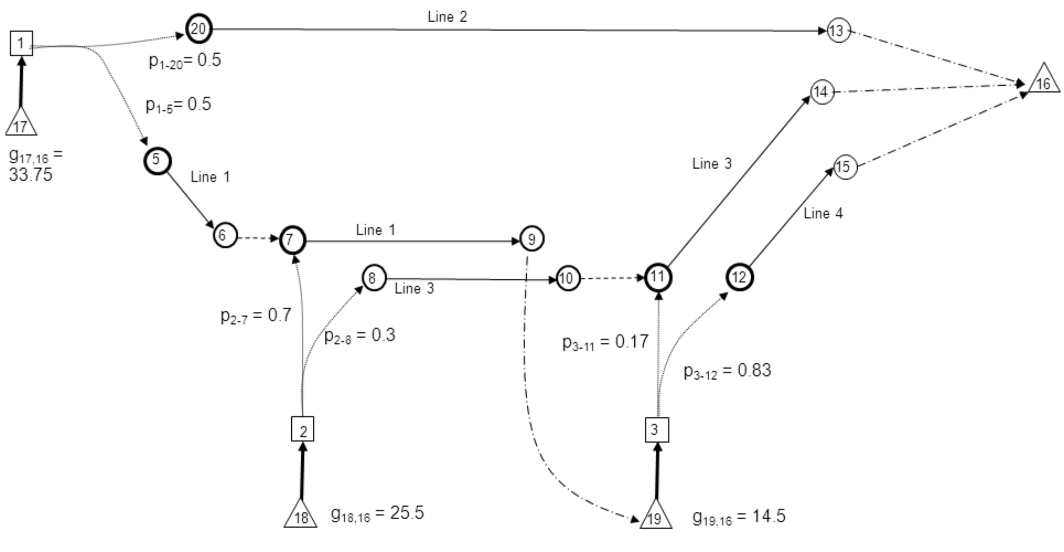

Fig. 6. All-to-one shortest hyperpaths in uncongested conditions. The figure also depicts diversion probabilities and travel times to destination when the network is not congested. For clarity the diversion probability is here indicated as p_{i-j}, where i is the stop node where passengers are waiting and j is the boarding node corresponding to the attractive line considered.

In the first example studied, the only destination considered is node 16 and it is assumed that the OD matrix is in the form given by Table 2. In this setting, at the beginning of the analysis period [07:30 – 07:55] no congestion occurs in the network and the model yields the same results that could be obtained by applying static models as in [2, 3]. Fig. 6 shows the hypergraph representation of the all-to-one shortest hyperpaths in uncongested conditions.

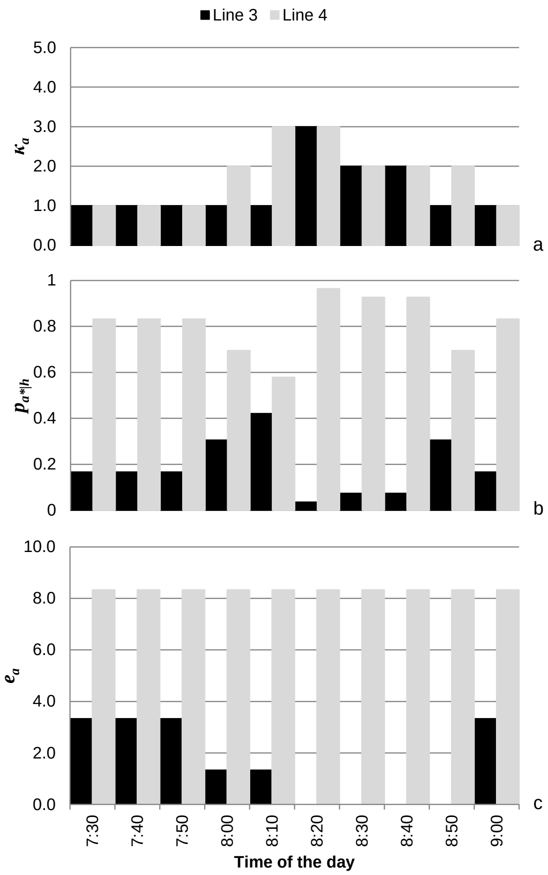

Fig. 7. Variation of congestion parameter κ_a (a), boarding probability $p_{a^*|h}$ (b) and exit capacity e_a (c) at Stop 3 during the analysis period [07:30 – 09:00]. a represents respectively queuing arc (3, 25), corresponding to Line 3, and queuing arc (3, 26), corresponding to Line 4. On the other hand, a^* represents a branch of the waiting hyperarc in the hypergraph H, and corresponds, respectively to Line 3 and Line 4.

When congestion occurs, static models (with and without capacity constraints) are not able to reproduce the *dynamic* phenomenon of the formation and dispersion of FIFO queues, nor its effect on route choice. Namely, with reference to the model graph of Fig. 5, a passenger who is at origin node 17 at 07:30 and boards Line 1 reaches node 7 at 07:46 and is joined by other passengers who, from stop node 2, board the same line. These travellers have to disembark through arc (9, 19) and reach stop 3 at 07:54. Therefore, from this moment onwards the total flow from Stop 3 to the destination exceeds the total available capacity and, when the next vehicle arrives, Line 3 and Line 4 become heavily congested (Fig. 7a and Fig. 7c). The decreased available capacity of Line 3, combined with a lower frequency of the service, results in a fall of its diversion probability, while the diversion probability of Line 4 increases (Fig. 7b).

It is important to note here that the value of the diversion probability solely depends on the frequency of the line and on its congestion level at the considered stop. On the other hand, the inclusion of a line in the attractive set depends on its total travel time upon boarding.

The analysis of congestion patterns at Stop 3 suggests that the model is able to simulate “forward effects”, i.e. effects produced by what happened upstream in the network at an earlier time of the day (passengers boarding Line 1 at 07:30), on what happens downstream at a later time (queue of passengers, wishing to board Line 3, that occurs at stop 3 at 07:55).

Additionally, the model also simulates “backward effects”, namely effects produced by what is expected to happen downstream in the network at a later time, on what happens upstream at an earlier time. The analysis of Line 1 helps to clarify this concept.

Line 1 never becomes congested at Stop 1 or 2 (Fig. 8a and Fig 9a). However, because at 08:12 a long queue forms at Stop 3 for Line 3, then, since 07:53, the travel time upon boarding Line 1 from Stop 1 increases to 35 minutes (Fig. 10) and therefore it is excluded from the attractive set of Stop 1 (Fig. 8b). Line 1 is included again from 08:25 onwards, namely when the travel time upon boarding decreases again because, by the time Stop 3 will be reached (08:44), congestion on Line 3 dissipates.

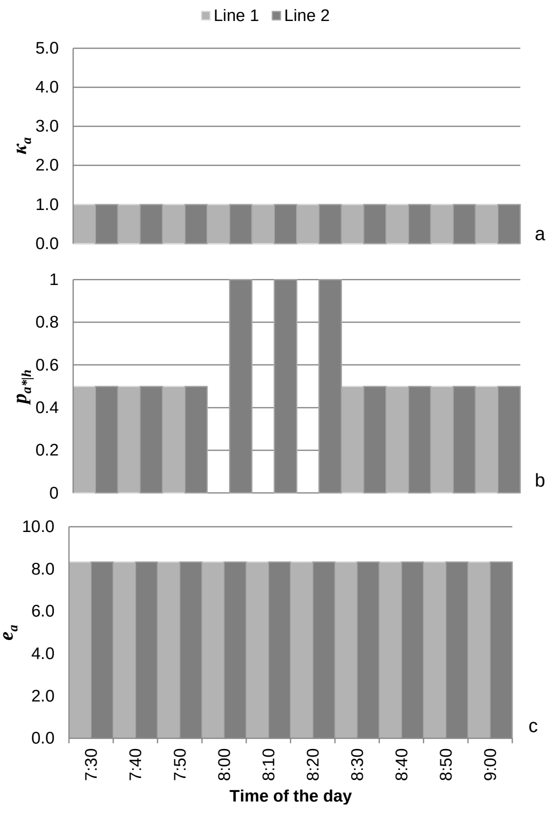

Fig. 8. Variation of congestion parameter κ_a (a), boarding probability $p_{a^*|h}$ (b) and exit capacity e_a (c) at Stop 1 during the analysis period [07:30 – 09:00]. a represents respectively queuing arc (1, 5), corresponding to Line 1, and queuing arc (1, 20), corresponding to Line 2. On the other hand, a^* represents a branch of the waiting hyperarc in the hypergraph H, and corresponds, respectively to Line 1 and Line 2.

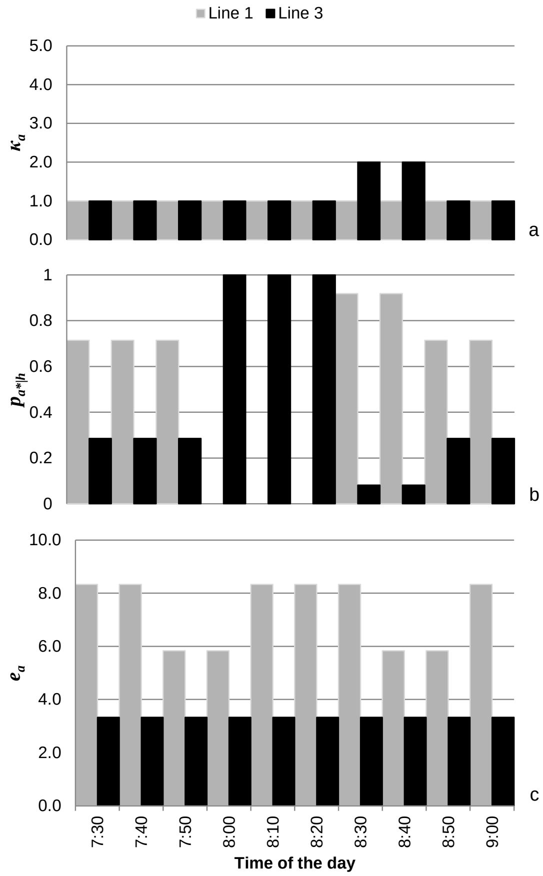

Fig. 9. Variation of congestion parameter κ_a (a), boarding probability $p_{a^*|h}$ (b) and exit capacity e_a (c) at Stop 2 during the analysis period [07:30 – 09:00]. a represents respectively queuing arc (2, 23), corresponding to Line 1, and queuing arc (2, 24), corresponding to Line 3. On the other hand, a^* represents a branch of the waiting hyperarc in the hypergraph H, and corresponds, respectively to Line 1 and Line 3.

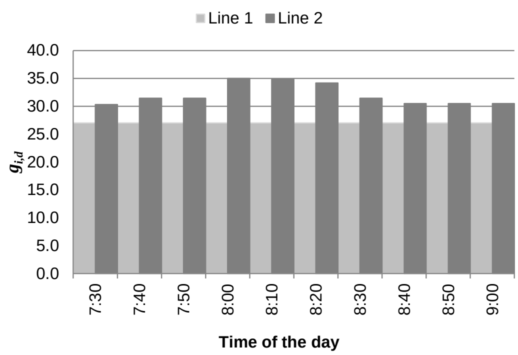

Fig. 10. Travel time to destination ($g_{i,d}$) upon boarding Line 1 or Line 2 from Stop 1 during the analysis period**.**

Similarly, at Stop 2 (Fig. 9b) starting from 08:00 passengers would board only Line 3. Should they board Line 1, they would reach Stop 3 at 08:09, when a queue for boarding Line 4 arises and, consequently, the travel time upon boarding Line 1 increases to 23.4 minutes. Afterwards (at 08:33, Fig. 10a), because Line 3 becomes congested at Stop 2, Line 1 is reintroduced in the attractive set of Stop 2 (Fig. 9b).

Fig 8c and Fig 9c complete the example and respectively depict the available capacity of Line 1 and Line 2 at Stop 1 and Line 1 and Line 3 at Stop 2.

Table 3. Time-dependent OD matrix during the analysis period for the second example considered. The travel demand is expressed in passengers per minute

	Destination Centroid		
Origin Centroid	16	17	18
17	5	5	1
18	4	0	3
19	6	0	0
16	5	5	1

In the second example studied, the OD travel demand matrix is given by Table 3. In this case the travel demand is not only directed to node 16 (centroid connected to Stop 4), but also to nodes 18 and 19 of Fig. 5 and is assumed to be constant during the analysis period [07:30 – 09:00]. The solution algorithm (Fig. 11) converges to ε = 0,001 in 30 iterations and to ε = 0,0001 in 67 iterations.

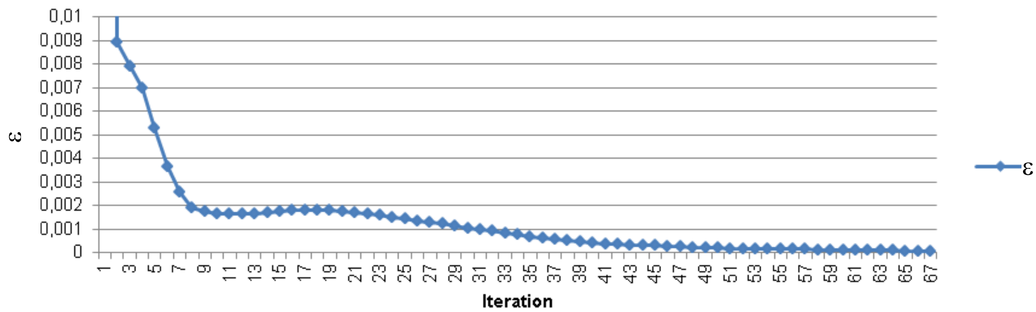

Fig. 11. Algorithm convergence.

The results show that the only queue in the network occurs at Stop 3, where between 08:25 and 08:55 passengers have to wait the second passage of Line 4 if they want to board this service. Therefore the diversion probabilities and, thus, the inflow on waiting arcs, at Stop 3 is greatly affected by congestion, as depicted in Fig. 12, and the inflow on arc (3, 13) increases (08:20 – 08:50) when passengers know that, by the time the next carrier of Line 4 arrives, it will be full.

Notwithstanding the queue at Stop 3, the increase in total travel time from node 20 and node 7 to node 16 is not remarkable and, as opposed to the first example, Line 1 is always kept in the attractive sets of Stop 1 and Stop 2, whichever the final destination node is. As a consequence (Fig. 13 and Fig. 14), the inflow on queuing arcs (1, 21) and (1, 22) at Stop 1 and (2, 23) and (2, 24) at Stop 2 stays constant during all the analysis period.

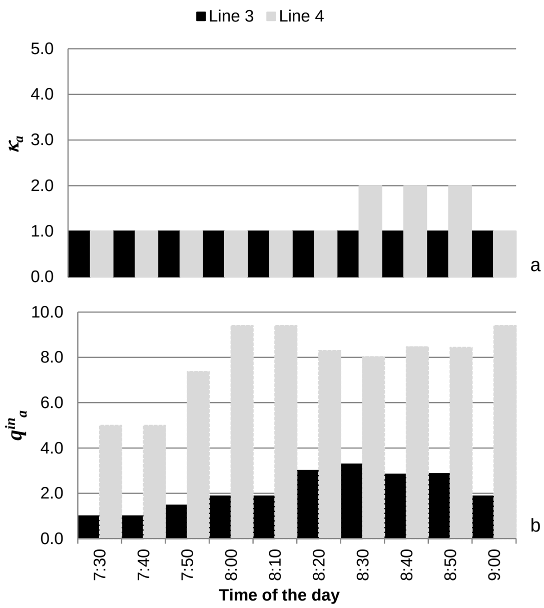

Fig. 12. Variation of congestion parameter κ_a (a) and inflow q^{in}_a (b) at Stop 3 during the analysis period [07:30 – 09:00]. *a* represents respectively queuing arc (3, 25), corresponding to Line 3, and (3, 26), corresponding to Line 4.

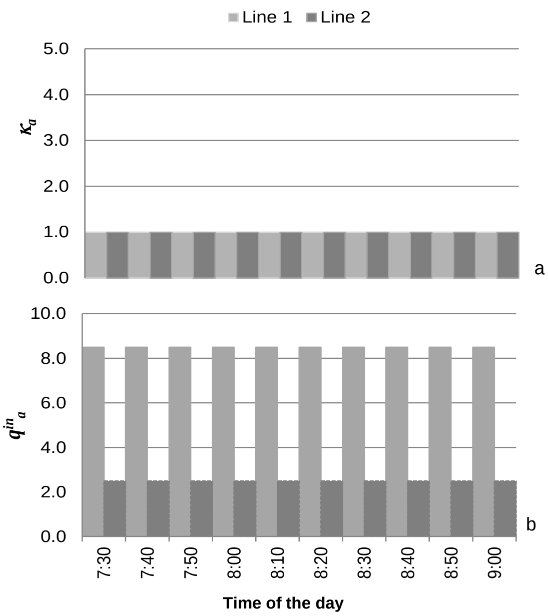

Fig. 13. Variation of congestion parameter κ_a (a) and inflow q^{in}_a (b) at Stop 1 during the analysis period [07:30 – 09:00]. *a* represents respectively queuing arcs (1, 22), corresponding to Line 1, and (1, 21), corresponding to Line 2.

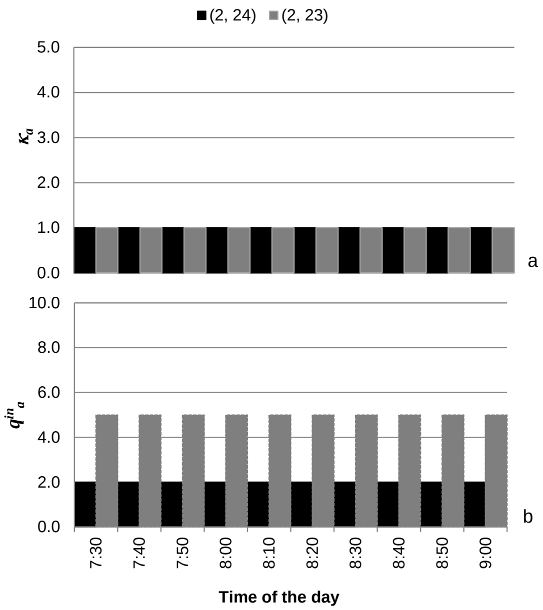

Fig. 14. Variation of congestion parameter κ_a (a) and inflow $q^{in}{}_a$ (b) at Stop 2 during the analysis period [07:30 – 09:00]. *a* represents respectively queuing arcs (2, 23), corresponding to Line 1, and (2, 24), corresponding to Line 3.

6. Conclusion

A vast literature, produced since the late 1980s, has demonstrated that in FB networks with partially overlapping services, the route choice can be best modelled as a shortest hyperpath problem rather that single path. However, a broad agreement on how to most appropriately represent dynamic congestion phenomena and their impacts on travel choices has not been attained so far. Therefore, the major contribution of this work has been the consideration of the effects of the formation and dispersion of FIFO queues at transit stops in networks, where users have no explicit information of the service schedule, should any schedule exist, and minimise their travel time to destination by selecting the dynamic shortest hyperpath.

A FB approach has been chosen to build the model, as it handles more suitably the inherent uncertainty faced by passengers who travel in densely connected transit networks, in which several alternatives may be available from the same stop/station. While FB models have been exploited so far to reproduce, mainly, the effect of mingling queuing on travel strategies, the concept of travel strategies has been extended here in order to address the case where passenger congestion results in FIFO queues at transit stops. Consequently, a Bottleneck queuing model with time-varying exit capacity is adopted to evaluate the queuing time at transit stops. On the other hand the *Route Choice* needs to be modelled in the form of all-to-one dynamic hyperpath search.

To this aim four main theoretical implications have been addressed:

- From a graph theory point of view, the hypergraph H, which is built on the model graph G, is introduced. $H = (N, F)$, where the set of *forward hyperarcs* is formally defined as $F = A \setminus \{A^W \cup A^Q\} \cup A^H$. This is a substantial innovation with respect to [25], where a first hypothesis of dynamic assignment with hyperpaths is developed. In this previous work, the use of the same graphic structure for the demand and supply models leads to a representation of the waiting / queuing process which is questionable from the phenomena point of view and does not allow the great flexibility that can be attained detaching completely the supply and demand models as done here.

- The *Stop Model* and *Route Choice Model* refer to the hypergraph *H*, whose *waiting hyperarcs* graphically represent the process of waiting for the set of attractive lines. When making their travel choices, passengers do not distinguish between the under-saturation delay, due to the inherent transit service discontinuity, and over-saturation queuing time. Thus the total waiting cost w_h for a waiting hyperarc *h* includes both components and depends on the lines that are included in the attractive set. In order to account for the increased perceived cost of waiting when boarding is denied because of capacity contraints, it is assumed the VOT for hyperarcs it an increasing function of congestion by means of the parameter κ_a.
- The *Network Flow Propagation Model* and *Arc Performance Functions* refer to the model graph *G*. In order to ensure that inflows/outflows are consistent with arc exit times, in the NFPM and APF, the waiting arcs $a \in A^W$ only represent the under-saturation delay, while the cost of queuing, as it results from the *Bottleneck Queuing Model*, is represented by queuing arcs $a \in A^Q$. The queuing time calculated in the APF is used to calculate congestion parameters $\kappa_a(\tau)$, which are used, in turn, in the SM to derive specific *diversion probabilities* and *waiting times*.
- The stop layout considered is such that passengers wait in a unique “mixed” FIFO queue, regardless of the specific line(s) they may consider to board. Thus, while passengers having the same attractive set respect a FIFO rule, there can be overtaking between those who consider different sets of attractive lines.

The resulting DUE is capable of reproducing the results from the static model by Spiess and Florian [3] in the uncongested scenario with time-independent (constant) travel variables, but clearly demonstrates the changes in the hyperpath, total travel cost and flow pattern when these become dynamic.

Future work will concentrate on extending the approach to other transport modes, where the usual assumption of Poisson-distributed carrier arrivals is not applicable. Moreover, the new model will also be tested on large transit networks, in order to verify the value of heuristics solution methods (for example when selecting the set of attractive lines at each intermediate stop) and algorithm speed-up techniques.

7. References

1. Chriqui, C. and P. Robillard, *Common bus lines.* Transportation Science, 1975. **9**(2): p. 115-121.
2. Nguyen, S. and S. Pallottino, *Equilibrium traffic assignment for large scale transit networks.* European Journal of Operational Research, 1988. **37**(2): p. 176-186.
3. Spiess, H. and M. Florian, *Optimal strategies: A new assignment model for transit networks.* Transportation Research Part B: Methodological, 1989. **23**(b): p. 83-102.
4. Spiess, H., *On optimal route choice strategies in transit networks*, U.d.M. Centre de recherche sur les transports, Editor 1983.
5. Spiess, H., *Contributions á la téorie at aux outils de planification des résaux de transport urbain*, in *Department d'Informatique et de recherche opérationelle*1984, Université de Monteréal.
6. Bouzaïene-Ayari, B., M. Gendreau, and S. Nguyen, *Modeling Bus Stops in Transit Networks:A Survey and New Formulations.* Transportation Science, 2001. **35**: p. 304-321.
7. De Cea, J. and E. Fernandez, *Transit Assignment for Congested Public Transport Systems: An Equilibrium Model.* Transportation Science, 1993. **27**(2): p. 133-147.
8. Cominetti, R. and J. Correa, *Common-Lines and Passenger Assignment in Congested Transit Networks.* Transportation Science, 2001. **35**: p. 250 - 267.
9. Cepeda, M., R. Cominetti, and M. Florian, *A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria.* Transportation research Part Part B: Methodological, 2006. **40**(6): p. 437-459.

10. Schmöcker, J.-D., M.G.H. Bell, and F. Kurauchi, *A quasi-dynamic capacity constrained frequency-based transit assignment model.* Transportation Research Part B: Methodological, 2008. **42**: p. 925-945.
11. Kurauchi, F., M. Bell, and J.-D. Schmöcker, *Capacity Constrained Transit Assignment with Common Lines.* Journal of Mathematical Modelling and Algorithms, 2003. **2**(4): p. 309-327.
12. Meschini, L., G. Gentile, and N. Papola, *A frequency based transit model for dynamic traffic assignment to multimodal networks*, in *17th International Symposiumon Transportation and Traffic Theory* R. Allsop, M.G.H. Bell, and B.G. Heydecker, Editors. 2007, Elsevier: London. p. 407-436.
13. Bellei, G., N. Papola, and G. Gentile, *A within-day dynamic traffic assignment model for urban road networks.* Transportation Research Part B: Methodological, 2005. **39**: p. 1-29.
14. Cascetta, E., *Transportation Systems Analysis - Models and Applications*. Second ed. Springer Optimization and Its Applications, ed. D.-Z. Du2009, New York: Springer US. 742.
15. Nuzzolo, A., F. Russo, and U. Crisalli, *Transit Network Modelling: the schedule-based dynamic approach*2003: Franco Angeli - Collana Trasporti.
16. Papola, N., Filippi, F., Gentile, G. Meschini, L., *Schedule-based transit assignment: a new dynamic equilibrium model with vehicle capacity constraints*, in *Schedule-Based Modeling of Transportation Networks* N.H.M. Wilson and A. Nuzzolo, Editors. 2009, Springer US. p. 145-170.
17. Sumalee, A., Z. Tan, and W.H.K. Lam, *Dynamic stochastic transit assignment with explicit seat allocation model.* Transportation Research Part B: Methodological, 2009. **43**(8–9): p. 895-912.
18. Hamdouch, Y., Ho, H. W., Sumalee, A., Wang, G., *Schedule-based transit assignment model with vehicle capacity and seat availability.* Transportation Research Part B: Methodological, 2011. **45**(10): p. 1805-1830.
19. Marcotte, P. and S. Nguyen, *Hyperpath formulations of traffic assignment problems*, in *Equilibrium and Advanced Transportation Modelling*1998, Kluwer Assinipi Park, Norwell, MA. p. 175-199.
20. Leurent, F., E. Chandakas, and A. Poulhès, *User and service equilibrium in a structural model of traffic assignment to a transit network.* Procedia Social and Behavioral Sciences, 2011. **20**: p. 495–505.
21. Gendreau, M., *Étude approfondie d'un modéle d'équilibre pour l'affectation des passangers dans les réseaux de transport en commun*, in *Department d'Informatioque et de Recherche Operationnelle*1984, Université de Montréal: Montreal.
22. Bouzaïene-Ayari, B., *Modélisation des arrêts multiples d'autobus pour les réseaux de transport en commun.*, in *Département de Génie Industriel*1988, École Polytechnique de Montréal: Montréal.
23. Leurent, F. and V. Benezech, *The Passenger Stock and Attractivity Threshold model for traffic assignment on a transit network with capacity constraint*, in *Transportation Research Board, 90th Annual Meeting*2011: Washington, D.C.
24. Hamdouch, Y. and S. Lawphongpanich, *Schedule-based transit assignment model with travel strategies and capacity constraints.* Transportation Research Part B: Methodological, 2008. **42**: p. 663-684.
25. Trozzi, V., G. Gentile, M.G.H. Bell and I. Kaparias. *Dynamic User Equilibrium Hyperpaths in Bus Networks with Passengers Queues*. in *LATSIS 1st European Symposium on Quantitative Methods in Transportation Systems* . 2012. Lausanne, Switzerland.
26. Gentile, G., S. Nguyen, and S. Pallottino, *Route Choice on Transit Networks with Online Information at Stops.* Transportation Science, 2005. **39**(3): p. 289-297
27. Gallo, G., Longo, P., Pallottino, S., Nguyen, S., *Directed Hypergraphs and Applications.* Discrete Applied Mathematics, 1993. **42**(2-3): p. 177-201.

28. Gentile, G., N. Papola, and L. Persia, *Advanced pricing and rationing policies for large scale multimodal networks.* Transportation Research Part A: Policy and Practice, 2005. **39**(7-9): p. 612.
29. Gentile, G., L. Meschini, and N. Papola, *Spillback congestion in dynamic traffic assignment: a macroscopic flow model with time-varying bottlenecks.* Transportation Research Part B: Methodological, 2007. **41**: p. 1114-1138.
30. Cascetta, E., *Transportation systems engineering: theory and methods*2001.
31. Trozzi, V., Kaparias, I., Bell, M.G.H., Gentile, G., *A dynamic route choice model for public transport networks with boarding queues.* Transportation Planning and technology, 2013. **36**(1): p. 44-61.
32. Cantarella, G.E., G. Gentile, and P. Veloná. *Uniqueness of Stochastic User Equilibrium*. in *5th IMA Conference on Mathematics in Transportation*. 2010. London, UK.
33. Schmöcker, J.-D., *Dynamic Capacity Constrained Transit Assignment*, in *Department for Civil and Environmental Engineering*2006, Imperial College London.
34. Chabini, I., *Discrete dynamic shortest path problems in transportation applications: complexity and algorithms with optimal running time.* Transportation Research Record: Journal of the Transportation Research Board, 1998. **1645**: p. 170-175.