A passenger traffic assignment model with capacity constraints for transit networks

Fabien Leurent, Ektoras Chandakas, Alexis Poulhès

▶ To cite this version:

Fabien Leurent, Ektoras Chandakas, Alexis Poulhès. A passenger traffic assignment model with capacity constraints for transit networks. 2012. hal-00687095v1

HAL Id: hal-00687095

https://enpc.hal.science/hal-00687095v1

Preprint submitted on 12 Apr 2012 (v1), last revised 12 Nov 2012 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A passenger traffic assignment model with capacity constraints for transit networks

Ektoras Chandakas[a*], Fabien Leurent[a*], Alexis Poulhés[a]

[a]*Université Paris-Est, Laboratoire Ville Mobilité Transport, Ecole des Ponts ParisTech*
6-8 avenue Blaise Pascal, 77455, Marne la Vallée, Cedex 2, France

Abstract

A passenger traffic assignment model dealing with capacity constraints for transit networks is provided. The CapTA model constitutes a framework for introducing capacity effects, in addition to those included: in-vehicle capacity in terms of seated and standing passengers, the exchange capacity of vehicles at stations and the line capacity on vehicle flows. Traffic equilibrium is intrinsically dual, or bi-layered: the upper layer pertains to passenger route choice at the network level, while the lower layer pertains to vehicle traffic along service routes and their lines of operations. The mathematical formulation is bi-layered, too, with conventional optimal strategies for passenger trip-making and innovative "line models" for the lower layer.

Keywords: Transit network; Traffic assignment; Capacity constraints;

1. Introduction

In large urban areas, the transit network is frequently submitted to heavy congestion, especially at the peak hours on working days. Under these conditions, not only may the passengers face uncomfortable conditions, due to crowding, delay and unreliability, but also the transit operation may be disrupted as a result of increased dwell time, bunching and delays, leading to the reduction of the service frequency.

Although these issues are known to the scientific community and discussed in the Transport Capacity and Quality of Service Manual (TRB, 2003), not enough attention has been given by the transportation modelers to the interplay of passenger and vehicle traffic. On the one hand, passenger traffic is addressed by models of traffic assignment to a transit network, typically in planning studies where the operating conditions are described by service routes and operation frequencies, section run time, station dwelling time and vehicle capacity (Thomas, 1991; Ortuzar and Willumsen, 2004); in some models, one or several of these features are flow-responsive. Route frequency has been associated by Lam et al (1999) to the fleet size and cycle time, considering both station and dwelling time as flow dependent. On the other hand, service operations are planned at the line on the basis of a model of vehicle running along the service trajectory, taking into account the local conditions such as line geometry and station dwelling time. The passenger traffic is considered only through its influence on the dwelling times and the vehicle load (Vuchic, 2006; Lai et al, 2011).

This paper presents an innovative traffic assignment model for a public transportation network along with a classroom application. The CapTA model (for Capacitated Transit Assignment) takes account of the configuration of transport modes, their respective nominal and effective performance, the structure of origin-

[*] Corresponding author. Tel.: +33-164-152-132; fax: +33-164-152-140
E-mail address: ektoras.chandakas@enpc.fr and fabien.leurent@enpc.fr

destination flows and the route choice behaviour. In addition to these classical factors it considers a number of traffic-related phenomena. Various phenomena have been treated, such as passengers waiting on a platform to board a vehicle, the vehicle capacity, the capacity of each route as a function of the service frequency, and the modulation of this frequency due boarding and alighting passenger flows. The paper provides a description and a mathematical definition of the network elements, such as the travel cost of the line legs. In addition, it is focused on the mathematic formulation of the traffic equilibrium at the network level.

The rest of this paper is in six parts. We begin by setting the principles of the CapTA model along with the bi-layer network representation. Then, at the service layer, the line sub-model and its treatment are described and the travel cost function is defined. At the passenger layer, the hyperpath is defined along with the route attractiveness conditions, before formulating the mathematical conditions for traffic equilibrium. An application instance is used to illustrate the various capacity effects and the conclusion focuses on on-going developments of the model.

2. The CapTA model

2.1. Principles of the CapTA model

The assignment model takes account of various types of capacity. As public transport involves two types of traffic, passengers and service vehicles, the model distinguishes between two layers: the service layer and the passenger layer.

For the service layer, each line is treated separately and the following interactions between transit routes and passengers are considered:

- for each transit route, the passenger load per vehicle varies along the journey with passenger flow.
- for each transit route and station, the boarding and alighting passenger flows affect vehicle dwell times.
- for each line and station, the vehicle dwell time and passage constraints on the track affect the service frequencies during a given period.

The passenger layer relates to trips between origins and destinations through passenger's choice of a route on the service network. Each passenger selects the path or combination of paths that minimizes a generalized time criterion where each travel component is multiplied by discomfort coefficients, specific to the status of the individual passengers (who may be in a vehicle or not, standing or seated etc.). Thus, each passenger is modeled as a rational microeconomic agent who selects the route with the minimum cost and perceives local conditions in terms of time spent and discomfort. He is therefore sensitive to the quality of service.

The traffic equilibrium is determined by the interaction between the two layers: route choices determine the flows on the access-egress journeys; thus, the passenger layer influences the service layer. Vice-versa, the volumes on the access-egress journeys for each transit route determine the local loads on each network element, and by extension the quality of service. In this way the service layer influences the passenger layer.

2.2. A bi-layered network representation

We provide distinct network topologies for each model layer. At the passenger layer, we interest in the origin destination trips and the route choices of the passengers at the network level. The network (Fig.1b) is a directed graph $G = (N, A)$ composed of a set N of nodes n together with a set A_S of arcs a with endpoints in N. These arcs represent the state transition for a trip-maker: the set consists of a subset of private links A_M (considering among others the passenger access to a platform A_P) and the subset of the transit links composed of the line legs A_L.

At the service layer we are based on the access-egress journeys within a transit line to define the effect of the capacity constrains. Let us define the concept of a line ℓ as a subset of transit routes Z^{ℓ}, serving a set of destinations J^{ℓ}, that share the same infrastructure (track and station platforms) and cannot overtake one another. The network consists of bipartite graphs $Z^{\ell} \times J^{\ell}$ (Fig.1a) that link the routes to the stations by a set A_L of arcs. It consists for each transit service of the subsets of boarding, A_B, alighting, A_A, and sojourn, A_s, links at each station platform and the interstation arcs, A_I. The line leg arc set consists of the set $A_L = A_B \cup A_A \cup A_s \cup A_I$ with the associated nodes.

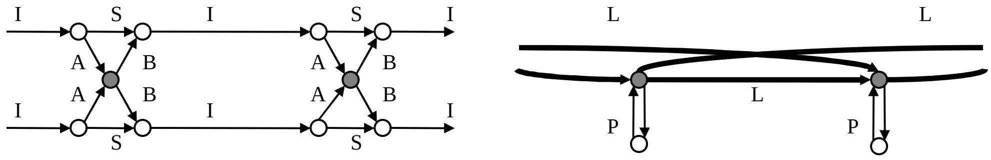

Fig. 1. A representation of the network at (a) the service layer and (b) the passenger layer

2.3. Definition of the model variables

Let us define the network flow state and the line attributes. The node set N contains origin and destination nodes, $o \in O \subset N$ and $s \in S \subset N$ respectively. An origin – destination pair $w = (o, s)$ - having an origin – destination flow of q_{os} - is associated with each destination node s through a set W^s of origin – destination (OD) pairs. If $W = \bigcup_{s \in S} W^s$ is the set of all the origin destination pairs, then the OD trip matrix is the vector of OD flow $\mathbf{q} = [q_w]_{w \in W}$. The arcs $a \in A$ are characterized by x_{as}, the transit flow per destination s. The **network flow state** is defined as a vector of arc destination flows, $\mathbf{x}_{AS} = [x_{as} : a \in A, s \in S]$.

A transit line $\ell \in L$ is composed of a group of directed transit services $z \in Z^{\ell}$. A **transit service** has certain attributes; $f_z(\mathbf{x}_{AS})$, the frequency of the service, dependent on the network flow state, k_z the vehicle total capacity and k_z^s the seat capacity per vehicle, if a transit service is composed of homogenous vehicles. The total capacity of a transit service is flow-dependent, since $\kappa_z(\mathbf{x}_{AS}) = k_z \cdot f_z(\mathbf{x}_{AS})$ and $\kappa_z^s(\mathbf{x}_{AS}) = k_z^s \cdot f_z(\mathbf{x}_{AS})$ for total and seat capacity respectively. Therefore, the attributes $(f_{\ell}, \kappa_{\ell}, \kappa_{\ell}^s)$ of a **transit line** ℓ depend on the transit services connecting each station pair (i, j). Considering a function, $1_{\{z \in (i,j)\}} = \begin{cases} 1 & if\ z \in (i, j) \\ 0 & otherwise \end{cases}$, the line attributes at a station platform i are:

$$f_{ij}^{\ell}(\mathbf{x}_{AS}) = \sum_{z \in \ell} f_z(\mathbf{x}_{AS}) 1_{\{z \in (i,j)\}}$$

$$\kappa_i^{\ell}(\mathbf{x}_{AS}) = \sum_{z \in \ell, j > i} \kappa_i^z(\mathbf{x}_{AS}) 1_{\{z \in (i,j)\}}$$

$$\kappa_i^{\ell s}(\mathbf{x}_{AS}) = \sum_{z \in \ell, j > i} \kappa_i^{zs}(\mathbf{x}_{AS}) 1_{\{z \in (i,j)\}}$$

The **arcs** on the service network, $a \in A_S$, are defined by an average traversal cost $\hat{c}_a$ and a frequency f_a. Depending on the arc type, their average traversal cost $\hat{c}_a$ may include the average travel time in the vehicle, as well as the cost of waiting, depending to the waiting conditions. The frequency of the arcs is given through the flow dependent revised frequency $\hat{f}_a(\mathbf{x}_{AS})$ (see section 4.2).

3. The line sub-model

The line sub-model deals with a single traffic direction on a line of operation. Its inputs consist of the passenger flows for the journeys between the access and egress stations. It acts as an elaborate arc travel – cost function of the line legs by estimating the generalized time between access and egress stations of the line through a flow loading – ZIP – algorithm and a leg costing – UNZIP – algorithm.

In the following section we provide a description of the capacity effects included in the line sub-model: the platform waiting sub-model, the in-vehicle comfort sub-model and the frequency modulation sub-model. That leads to the arc-travel cost function and an assessment of the efficiency of the algorithm.

3.1. The platform waiting time sub-model

Each vehicle that stops at a given station on a given line has a residual capacity once the passengers wishing to alight have done so. There are stocks of passengers on the platform: these are differentiated according to egress station v_j, serving as the model's principal endogenous variable.

For each transit route z, the total stock of waiting passengers n_{zi} at a station i is compared to the residual capacity k'_z of each vehicle. This determines the probability of immediate boarding, π_{zi}. That probability defines the flow that can be handled by a line during the reference period and gives the waiting time by passenger. The vehicle flows between the line stations are defined as: $y_{ij}^{z} \equiv \pi_{zi} v_j$.

The transit bottleneck model, developed by Leurent (2011b) is formulated as a fixed point problem. The existence and uniqueness of a solution was demonstrated. It is solved with a Newton-Raphson algorithm.

3.2. The in-vehicle comfort sub-model

Each transit user considers two distinct categories of in-vehicle comfort; sitting – with a unit cost $\underline{c}$ - and standing – with a cost $\overline{c}$ -, while $\underline{c} \leq \overline{c}$. The rider seeks to minimize travel cost, by trying to sit, since it is a less inconvenient situation. The in-vehicle comfort sub-model transposes the Seat Capacity Model developed in Leurent (2010) to the line approach of the CapTA model.

We consider two behavioural rules to all passengers. First, the standing riders have the same motivation to sit, whatever their egress station. The passengers of the same class (boarding or on-board) have equal probability to sit. Secondly, the on-board passengers at a station have a priority to occupy an available seat over the boarding passengers. As a result, the in-vehicle perceived cost is a random variable related to the sitting probabilities at each station – subject to the in-vehicle flow vector $y_z = [y_{ij}^{z} : z \in l,\ (i, j) \in A_l,\ i < j]$. The average cost per service leg is calculated through a recursive leg-based algorithm.

3.3. The frequency modulation sub-model

Each vehicle occupies an operation block for a certain time plus the safety headways. The time set aside limits the route frequencies on the line. The constraint is greatest at the stations, where vehicles stop. Their occupancy time depends on the dwell time and the operational time. If the doors stay open for as long as it takes for all the passengers wishing to do so to board or alight, the dwell time will depend on the volume of boarding and alighting passengers.

The frequency modulation model, given in Leurent et al (2011) computes the station platform occupation time H', related to the dwell time per vehicle and the service frequency of upstream arrivals, f_z^{+}. If that time is longer than the reference period, H, the service frequency is reduced, by a modulation factor,

$\eta \equiv \min\{1; H/H'\}$. That factor is applied to the frequency of all transit routes using the station infrastructure, whether it stops or not. The modulated frequencies f'_{zi} are propagated downstream along each transit route.

3.4. The arc travel-cost function

The arc travel cost is calculated according to the line leg flows and depends on the network flow state x_{AS} according to the cost-flow function: $\mathbf{x}_{AS} \mapsto c_A(\mathbf{x}_{AS}) = [c_a(\mathbf{x}_{AS}) : a \in A]$

That cost reflects the impact of various capacity effects, both at a node and on a transit line, at the line leg arc. The former is induced by the local arc flow $x_a = \sum_{s \in S} x_{as}$, by restricting the function, $c_a(\mathbf{x}_{AS}) = c_a(x_a)$. The latter concerns the line legs, $a \in A_l$, by restricting to the flow of the legs, $\mathbf{x}_\ell = [x_a : a \in A_L]$. Thus, $c_a(\mathbf{x}_{AS}) = c_a(\mathbf{x}_\ell)$.

The cost flow function on the CapTA model consists of executing for each transit line two algorithms in sequence. First, the ZIP – flow loading – algorithm starts at the origin and moves forward, dealing with every station in the topological order of the line. At each station it assembles the vehicle flows and applies the capacity constraints. The bottleneck sub-model yields the probability to board $\mathbf{p}_z = \mathbf{P}(\mathbf{x}_\ell)$ and consequently the vector of in-vehicle flows $\mathbf{y}_z = [y_{ij}^z : z \in \ell,\ (i, j) \in A_L,\ i < j]$ for the services of the transit line. The function can be made continuous by enforcing on each route, for each station a strictly positive through arbitrarily small minimal residual capacity. The in-vehicle comfort calculates the probabilities of a passenger to occupy a seat at the station, differentiating the on-board, $\mathbf{p}_z^0 = \mathbf{P}^0(\mathbf{y}_z)$, and the boarding, $\mathbf{p}_z^+ = \mathbf{P}^+(\mathbf{y}_z)$, passengers, according to the in-vehicle flow vector $\mathbf{y}_z$. The continuous approximation also applies for the in-vehicle comfort function. Finally, the frequency modulation adjusts the frequency downstream according to the in-vehicle flows $\mathbf{f}_l = \mathbf{f}([\mathbf{y}_z]_{z \in \ell})$.

Then, the UNZIP – leg costing – algorithm moves backwards, treating each station at a reverse topological order. Initially, it evaluates the in-vehicle comfort by leg recursively as described in Leurent (2010), yielding a continuous approximation $G_{ij}^{z\varepsilon}(\mathbf{y}_z)$. Combining the other effects, the cost of the line leg $a \approx (i, j)$ along a line ℓ, will be then characterized by the function $c_a = C_{i,j}^{\ell}([\mathbf{P}^{\varepsilon}(\mathbf{x}_\ell)]_{z \in \ell}, [\mathbf{f}(\mathbf{x}_\ell)]_{z \in \ell}, [G_{ij}^{z\varepsilon}(\mathbf{y}_z)]_{z \in \ell})$, or:

$$c_{ij}^{\ell\varepsilon} = [\alpha + \textstyle\sum_{z \in (i,j)} G_{ij}^{z\varepsilon} f'_{zi} \pi_{zi}^{\varepsilon}] / \textstyle\sum_{z \in (i,j)} f'_{zi} \pi_{zi}^{\varepsilon} \quad if\ q_{ij}^{+} \geq 0 \tag{1}$$

Since $a \in A_L$ is a line leg arc, it is noted that $c_a^{\varepsilon} = c_a^{\varepsilon}(\mathbf{x}_{AS})$.

The evaluation of the arc cost through the ZIP and UNZIP algorithms is efficient with minimal complexity. If a line consists of S_ℓ stations and Z_ℓ transit services, the complexity of the ZIP algorithm is $O(S_\ell \times Z_\ell)$ and of the UNZIP algorithm $O(S_\ell{}^2)$. The complexity of the line sub-model is not greater than $O(S_\ell{}^2)$.

4. The passenger route choice at the network level

4.1. The attributes of the arcs and the hyperpath description

The arcs of the service network, $G = (N, A_S)$, have an average traversal cost $\hat{c}_a$ with an arc split defined by an flow dependent revised frequency $\hat{f}_a(x_{AS})$ (see section 4.2). The average arc traversal cost of the set of line legs $a \in A_L$ includes the weighted average in-vehicle journey cost and the cost of traversing a passenger stock for a saturated flow. Additional waiting due to the discontinuity of the transit services is considered among the platform access arcs $a \in A_p$.

The route choice to a destination is described by a hyperpath. A hyperpath $h = (\hbar, \hat{h})$ with a destination node s is defined through a pair, the arc set $\hbar$ and the routing field $\hat{h}$. The routing field is a mapping of the set A_S onto $[0,1]$, where for an arc a of the hyperpath it will stand $\hat{h}_a > 0$. The routing proportions of the outgoing arcs $a \in A_m^+$ at a node m are $\hat{h}_a = 1$ for the arcs $a \in A_S \setminus A_p$. The routing proportions of the platform access arcs $a \in A_p$ depend on the revised frequency of the line legs served at the platform.

The travel cost from a node n to a destination s along a hyperpath $h \in H_{ns}$, for a network flow state x_{AS}, is the average travel cost along its paths. Let $w_m^h(x_{AS}) \equiv \alpha_m / \sum_{a \in A_m^+} 1_{\{a \in \hbar\}} \hat{f}_a(x_{AS})$ the waiting delay at a node m for outgoing platform arcs, $a \in A_p$, or $w_m^k \equiv 0$ otherwise. If $R_{ns}(h)$ is the set of the elementary paths r along h from n to s, $\hat{h}_r = \prod_{a \in r} \hat{h}_a$ the path flow proportion and $c_a(x_{AS})$ the travel cost of an arc a related to network flow state x_{AS}, it will stand:

$$C_{ns}(h, x_{AS}) = \sum_{r \in R_{ns}(h)} \hat{h}_r \cdot [\sum_{m \in r} w_m^h(x_{AS}) + \sum_{a \in r} c_a(x_{AS})] \qquad (2)$$

The waiting factor α_m depends on the arrival distribution of the vehicles and the passengers at node m.

4.2. A bi-layer route choice and the bundle attractiveness

The passenger route choice presented previously corresponds to a bi-layer route choice. At the lower layer, a passenger at a platform will choose the line leg with the shortest path. The arc travel cost, c_{ij}^{ℓ}, is composed of a waiting time – related to the discontinuity of the service and the presence of a saturated flow – and the weighted average of the perceived in-vehicle travel time of the transit services composing the line leg. Therefore, if we consider u_{js} the perceived travel cost from node j to the destination s, the cost from node i to the destination, will be: $c_{i(s)} = c_{ij}^{\ell} + u_{js}$. At the platform level, if the demand exceeds capacity, the passengers wait in a stock that does not dissipate after the departure of a vehicle. The presence, of a non dissipating stock absorbs the discontinuous characteristic of the transit services. A transition phase exists between the two traffic states.

At the upper level of the path choice, a passenger faces a choice among alternative lines to reach the destination s. A line combination may occur in order to minimize the travel time to the destination. Let us provide the appropriate definitions.

Definition 1 - Discontinuity attenuator: *If a' is a line leg arc, $a' \in A_L$, and a is the platform arc leading to that, let us define the discontinuity attenuator as $\beta_a \equiv \phi(w_{a'} - 1/f'_{a'})$ where ϕ is a continuous function decreasing from $\phi(0) = 1$ to $\phi(x) = 0$ for all values $x \geq \varepsilon$, a small argument. The discontinuity attenuator corresponds to the transition between a discontinuous uncongested service and a continuous congested one.*

For the line leg a transition phase is distinguished between uncongested and saturated flow states. Even though it stems from the line leg, it is applied to the platform access arc.

Definition 2 - Revised Frequency: *A transition phase exists between the uncongested discontinuous and the congested continuous service and the line combination at the station level may occur in a revised way. Each arc a is then associated to a revised frequency $\hat{f}_a \equiv f_a / \beta_a$ for $\beta_a > 0$; otherwise $\hat{f}_a \equiv \infty$.*

According to that definition, an arc $a \approx (i,j)$, delivers a minimum time of $c_{ij}'^{\ell} = u_a + w_a - (\alpha / f_a) \cdot \beta_a$ and a mean time of $u_a + w_a$, or $c_{ij}'^{\ell} + \alpha / \hat{f}_a$. In addition, the combined frequency of an attractive bundle and the routing proportions of a path depend on the revised frequency.

Definition 3 - Bundle Attractiveness: *Assuming that the lines are not saturated, a line* ℓ *is attractive in respect to a bundle* B *if, denoting* $c_{ij}'^{\ell}$ *the minimum time of a line leg, it holds:* $c_{ij}'^{\ell} + u_{js} \le c_{i(s)}^{B} + \alpha_m / \hat{f}_B$

Where $c_{i(s)}^{B}$ denotes the minimum time from node i to the destination s by the bundle B.

5. The traffic equilibrium at the network level

After the necessary mathematical formulations for the network elements and the arc travel cost function, we define the traffic equilibrium. We first establish the feasible flow states and their hyperpath representation. Then, we define the traffic equilibrium in the form of a Nonlineal Complemementarity Problem (NCP). The existence of an equilibrium state can be demonstrated through a Variational Inequality Problem (VIP).

5.1. Feasible flow states and hyperpath representation

Definition 4 – Feasible network flow state: *A network flow state* $\mathbf{X}_{AS} = [x_{as}]_{a \in A, s \in S}$ *is feasible if it is non-negative and the conservation of flow by destination* $s \in S$ *is satisfied at every node.*

The second condition corresponds to the equation, $\sum_{a \in A_m^+} x_{as} = q_{ms} + \sum_{a \in A_m^-} x_{as}$, $\forall s \in S, m \in N, m \neq s$ illustrating the conservation of flow by destination $s \in S$ for a node, other than the destination s. A_m^+ defines the set of outgoing arcs of node m (respectively A_m^-, for the set of ingoing arcs) and q_{ms} the origin – destination flow from node $m \neq s$ to destination $s \in S$.

Let the hyperpath set H_{ns} include the hyperpaths h from n to s. We define the hyperpath flow state $\mathbf{X}_{NS}$ as a linear combination of elementary flows along the hyperpaths h with coefficients q_{ns}^h:

$$\mathbf{X}_{NS} = [q_{ns}^h : s \in S, n \in N \setminus \{s\}, h \in H_{ns}] \tag{3}$$

Definition 5 – Feasible hyperpath flow state: *Considering the origin – destination trip matrix* $\mathbf{q}_{NS} = [q_{ns}]_{n \in N, s \in S}$*, the hyperpath flow state* X_{NS} *is feasible if it is non-negative and if the conservation of flow by origin destination pair* q_{ns} *is satisfied:*

$$q_{ns}^h \ge 0 \;\; \forall s \in S, n \in N \setminus \{s\}, h \in H_{ns} \tag{4a}$$

$$\sum_{h \in H_{ns}} q_{ns}^h = q_{ns}, \;\forall s \in S, n \in N, n \neq s \tag{4b}$$

The set of feasible hyperpath flow states is noted $\mathbf{E}_X$. An elementary path r is composed of a series of arcs from n to the destination s. The proportion of the flow from n to s of the hyperpath h at the elementary path r is noted $\hat{h}_r = \prod_{a \in r} \hat{h}_a$. The network flow state $\mathbf{x}_{AS} = A(\mathbf{X}_{NS})$ is linked to the hyperpath flow state via the arc flow state:

$$x_{as} = \sum_{n \in N} \sum_{h \in H_{ns}} q_{ns}^h \sum_{r \in R_{ns}(h)} \hat{h}_r 1_{\{a \in r\}} \tag{5}$$

5.2. Definition and characterization of the traffic equilibrium

Definition 6 – Traffic Equilibrium: *Given the OD trip matrix $\mathbf{q}_{NS} = [q_{ns}]_{n\in N, s\in S}$ and the arc-cost function, a hyperpath flow state $\mathbf{X}_{NS} = [q_{ns}^{h} : s\in S, n\in N\setminus\{s\}, h\in H_{ns}]$ with $\mathbf{x}_{AS} = A(\mathbf{X}_{NS})$ is a traffic equilibrium if there exists a cost matrix $\boldsymbol{\mu}_{NS} = [\mu_{ns} : s\in S, n\in N\setminus\{s\}]$, such that for all $s\in S$ and $n\in N\setminus\{s\}$ the following conditions apply:*

$$q_{ns}^{h} > 0\,,\ \forall h\in H_{ns} \quad (6a)$$

$$\sum\nolimits_{h\in H_{ns}} q_{ns}^{h} = q_{ns} \quad (6b)$$

$$C_{ns}^{\varepsilon}(h, x_{AS}) - \mu_{ns} \geq 0\,,\ \forall h\in H_{ns} \quad (6c)$$

$$q_{ns}^{h}\cdot[C_{ns}^{\varepsilon}(h, x_{AS}) - \mu_{ns}] = 0\,,\ \forall h\in H_{ns} \quad (6d)$$

The previous set of conditions can be defined as a non-linear complementarity problem (NCP) in the variable $(\mathbf{X}_{NS}, \boldsymbol{\mu}_{\mathrm{NS}})$. The associated cost function is:

$$(\mathbf{X}_{NS}, \boldsymbol{\mu}_{NS}) \mapsto [C_{ns}^{\varepsilon}(h, A(\mathbf{X}_{NS})) - \mu_{ns} : s\in S, n\in N\setminus\{s\}, h\in H_{ns}]$$

The characterization of traffic equilibrium is demonstrated through a Variational Inequality Problem (VIP).

Assuming that the arc travel cost functions C_a are continuous, we can prove that there exists an equilibrium state for the capacitated transit assignment model. Nevertheless, it is not proven that there is a unique solution.

5.3. Computation scheme

The solution of the traffic equilibrium problem of network assignment is guaranteed by applying a Method of Successive Averages (MSA), using a decreasing sequence of positive numbers $(\lambda_k)_{k\geq 0}$ with $\lambda_0 = 1$. An equilibrium state is reached with the following steps:

- Initialization: set $\mathbf{x}_A = 0$ and $k = 0$
- Cost formation: execute the line sub-model by line, based on $\mathbf{x}_A$. The ZIP algorithm loads the leg flows and enforces the capacity constraints (platform flowing, in-vehicle comfort and frequency modulation) and the UNZIP algorithm estimates the costs of the line legs.
- Auxiliary state: search of the shortest hyperpath for all origin – destination pairs and assign the OD flows to the network elements, yielding an auxiliary flow $\mathbf{y}_A$.
- Flow update: the previous and auxiliary state flows are combined, $\mathbf{x}'_A = \lambda_k y_A + (1-\lambda_k)\cdot\mathbf{x}_A$
- Convergence criterion: If the states $\mathbf{x}_A$ and $\mathbf{x}'_A$ are sufficiently close, break the loop with solution $\mathbf{x}'_A$. Else, increment k, set $\mathbf{x}_A = \mathbf{x}'_A$ and go to Cost formation.

The convergence is measured by the function: $GAP_{k,k-1} = \sum\nolimits_{o\in O, s\in S}(q_{os}\left\| c_{o(s)}^{k} - c_{o(s)}^{k-1}\right\| / c_{o(s)}^{k-1})$

6. An application instance

A classroom instance is used to demonstrate the effects of the model. That example lies in a simplified representation of the busiest railway line in the Paris Metropolitan area, the RER A. At the morning peak hour more than 50 000 passengers per hour pass through the busiest segment of the line. The nominal frequency of 30 trains/h is frequently decreases to 25 due to congestion at the central trunk stations.

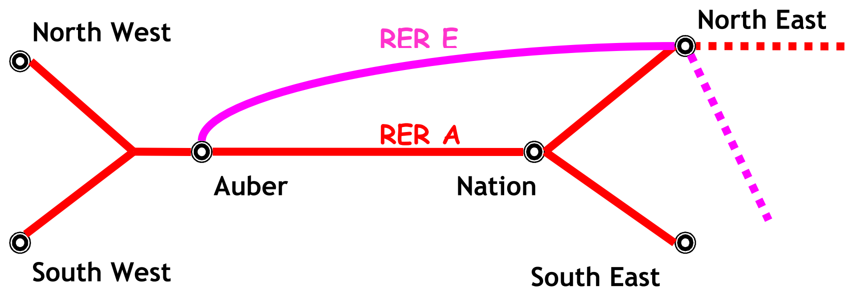

Figure 2. Abstraction of the selected network, emphasizing on line legs

The simplified network consists of two transit routes connecting the Northern (resp. Southern) branches to the central trunk, running from East to West. Another line, the RER E, competes with the RER A between the North East and Centre (Fig 2). Table 1 resumes the characteristics of the transit routes. The service network contains 6 stations as origin and destination zones with 50 arcs and 35 nodes. We have produces a truncated OD trip matrix to come as close as possible to the actual conditions at the central trunk, but in the same time emphasize on the effects produces by the capacity constraints.

Table 1. The characteristics of the transit routes

Route	Frequency (t/h)	Seat Capacity (per veh.)	Total Capacity (per veh.)
RER A north	18	432	1760
RER A south	12	600	1888
RER E	8	1100	2564

Table 2 contains the results after 500 iterations for three model variables; first without any capacity constraints, then taking account of waiting in platform and frequency modulation and thirdly with the three capacity effects modeled. There is a significant increase in the Generalized Time between the three model variables, especially as waiting and in-vehicle cost is concerned. Indeed we observe that by including in-vehicle sitting conditions, the in-vehicle cost explodes, but in the same time the waiting cost is reduced due to a larger number of passengers preferring the alternative RER E line from North East to Auber.

Table 2. The Generalized Time (in minutes) for each model variable

Model Variable	Optimal GT	Waiting cost	In-vehicle cost	Transfer cost	Access – Egress cost
Without Capacity	52,8	7,8	31	0	14
Without Comfort	67	22	31	0	14
With all capacity constraints	82,8	19,7	48,5	0	14

That origin destination pair demonstrates the behavior of passenger when comfort is included. Without capacity constraints only 27% of the OD flow chooses to RER E. Instead, when waiting constraints and in-vehicle comfort is considered, 86% of that OD flow prefer that route.

If we look at the disaggregated origin-destination costs between the non-capacitated and the capacitated model variables, the increase in the optimal generalized time varies from 16%, between North East and Auber – where an alternative service exists – to 145% between Auber and South West, due to the conditions of accessing the vehicles and the absence of seating available.

Let us focus on the results of CapTA model on the operation of the transit routes. The dwell time for both services increases at Nation due to the boarding and alighting passenger flows, as it is shown in table 3. Therefore a decrease in the vehicle frequency from 30 to 26,35 vehicles/h is observed at the downstream station of Auber. That contributes to a secondary effect of insufficient total capacity, causing a stock formation and an important increase in waiting time at Auber.

Table 3. Operation results of RER A under capacity constraints

Transit Route	Dwell Time at Nation (s)	Dwell Time at Auber (s)	Frequency at Auber	Expected Waiting time at Auber (min)
RER A north	61,6	40	15,80	60,3
RER A south	48,9	40	10,55	59,9

Conclusion

This paper develops the CapTA model, introduced in Leurent et al (2011). It addresses route choice at a network level, while including the effects of a wide range of traffic phenomena: in-vehicle passenger and seat capacity, access – egress capacity, the platform track capacity on the vehicle side. The model is based on a dual network representation of the service and the passenger layers. A line sub-model is used to estimate the generalized time between access and egress station within a line, according to the usual flows on each network element. A mathematical formulation of the traffic equilibrium is provided at the network level.

The CapTA model constitutes a framework for introducing capacity effects. In addition to the capacity effects already modeled, it would be straightforward to include corridor pedestrian capacity in a station, track capacity in vehicles. These features are expected to be useful in transit planning applications. On-going work, is focused on congestion assessment of the Paris metropolitan network.

Acknowledgement.

This research is supported by the Research and Education Chair on “The socio-economics and modeling of urban public transport”, operated by Ecole des Ponts ParisTech in partnership with the Transport Organizing Authority in the Paris Region (STIF), which we would like to thank for its support.

References

Lai, Wang, Jong (2011), *Development of Analytical Capacity Models for Commuter Rail Operations with Advanced Signaling Systems,* Paper #11-3424 presented at 90th Annual Transportation Research Board Meeting, Washington DC, January 2011.

Lam WHK, Gao ZY, Chan KS, Yang H (1999) *A stochastic user equilibrium assignment model for congested transit networks*, Transportation Research Part B, Vol. 33, pp. 351-368.

Leurent F (2010) On Seat Capacity in Traffic Assignment to a Transit Network, *Journal of Advanced Transportation*, 2010, Vol. 46, Issue 2, 112-138

Leurent F (2011a) *Transport capacity constraints on the mass transit system: a systemic analysis*, European Transport Research Review, Volume 3, Number 1, 11-21

Leurent F, Chandakas E (2012) *The transit bottleneck model*. Manuscript submitted for publication

Leurent F, Chandakas E, Poulhes A (2011) *User and service equilibrium in a structural model of traffic assignment to a transit network*, in Zak J (ed) The State of the Art in the European Quantitative Oriented Transportation and Logistics Research – 14th Euro Working Group on Transportation & 26th Mini Euro Conference & 1st European Scientific Conference on Air Transport. Elsevier Procedia Social and Behavioral Sciences, 20, 495-505

Nguyen S, Pallotino S (1988) *Equilibrium traffic assignment in large sxale transit networks*, European Journal of Operational Research, 37, 176-186

Ortuzar JD, Willumsen L (2004). *Modelling Transport*. Wiley, 3rd edition.

Thomas R (1991) *Traffic Assignment Techniques*. Avebury Technical, Aldershot, England.

TRB (2003) *Transit Capacity and Quality of Service Manual*. On-line report prepared for the Transit Cooperative Research Program, available on-line at the following website address: http://gulliver.trb.org/publications/tcrp/tcrp_webdoc_6-a.pdf . First edition 1999.

Vuchic VR (2006) Urban Transit : Operations, Planning and Economics. Wiley.

Appendix A. Continuity of leg-cost flow relationship

Theorem 1 – Continuity of leg cost-flow relationship

Let the service leg costs G_{ij}^{z} of the transit services $z \in \ell$ be functions of the leg trip matrix, with $\mathbf{x}_{\ell} \geq 0$ as well as the expected waiting time for a line leg. If the travel cost functions and the waiting cost function are continuous with respect to $\mathbf{x}_{\ell}$, so it is the approximate average leg cost function $C_{i,j}^{\ell\ \varepsilon}$ defined previously, except at the points where $k'_{zi} = 0$ and $q_{ij}^{+} = 0$.

We develop 4 Lemmas for the demonstration of the Theorem.

Lemma 1 – Regularity of probability to board

The passenger stock for a transit service n_z is a function of the exogenous flow, the modulated frequency and the available capacity, $n_z = n_z(q_{ij}^{+}, f'_{zi}, k'_{zi})$, where $n_z = 0$ for $q_{ij}^{+} = 0$ and $n_z > 0$ for $q_{ij}^{+} > 0$. For $(k'_{\mathrm{zi}}, q_{ij}^{+}) \geq 0$ let $p(k'_{\mathrm{zi}}, q_{ij}^{+}) \equiv \min\{1; \dfrac{k'_{\mathrm{zi}}}{n_z(q_{ij}^{+}, f'_{zi}, k'_{zi})}\}$ if $k'_{zi} > 0$ and $q_{ij}^{+} > 0$, else $p(k'_{\mathrm{zi}}, q_{ij}^{+}) \equiv 0$ if $k'_{zi} = 0$, else $p(k'_{\mathrm{zi}}, q_{ij}^{+}) \equiv 1$ if $k'_{zi} > 0$ and $q_{ij}^{+} = 0$. Then the function p is continuous with respect to $(k'_{\mathrm{zi}}, q_{ij}^{+})$, except at $(0,0)$. It is continuously differentiable except along $\{k'_{zi} = n_z(q_{ij}^{+}, f'_{zi}, k'_{zi})\}$

<u>*Proof.*</u> Once we have defined the function of the passenger stock, the property is obvious for $k'_{\mathrm{zi}} > 0$, $q_{ij}^{+} > 0$ and $k'_{zi} \neq n_z(q_{ij}^{+}, f'_{zi}, k'_{zi})$ since the probability function will yield either $p = 1$ if $k'_{zi} > n_z(q_{ij}^{+}, f'_{zi}, k'_{zi})$, or $p = k'_{zi} / n_z(q_{ij}^{+}, f'_{zi}, k'_{zi})$ if $k'_{zi} < n_z(q_{ij}^{+}, f'_{zi}, k'_{zi})$ which is differentiable on each restricted domain. The two sub-domains of differentiability are separated by the line $k'_{zi} = n_z(q_{ij}^{+}, f'_{zi}, k'_{zi})$. If $k'_{\mathrm{zi}} > 0$ then the function p is right continuous at $q_{ij}^{+} = 0$, since $k'_{zi} / n_z(q_{ij}^{+}, f'_{zi}, k'_{zi}) \to \infty$ as $q_{ij}^{+} \to 0^{+}$ and $p = 1$. On the other hand, if $k'_{zi} = 0$ and $q_{ij}^{+} > 0$, then it holds for the function $p(0, q_{ij}^{+}) \to 0$ as $q_{ij}^{+} \to 0^{+}$ and therefore the function p is continuous at $(0,0)$.

Lemma 2 – Regularity of modulation factor

The platform temporal occupancy is a function of the in-vehicle flow vector $\mathbf{y}_z$ *and the transit service frequencies* f_z *for all the services in the set* Z_i^ℓ , $H_i' = H([\mathbf{y}_z]_{z\in\ell},[f_z]_{z\in\ell,})$. *The platform temporal occupancy function has a lower bound of* $H_i' \geq H_i^{\min}$ *and it is strictly increasing. Let the modulation factor be* $\eta_i = \min\{1; H/H'\}$. *Therefore, it is continuous with respect to* (y, f) *and continuously differentiable except along* $\{H = H'\}$.

__Proof.__ The proof follows the guidelines of the previous lemma. The property is obvious for $H \neq H'$, since the function yields $\eta_i = 1$ for $H > H'$ or $\eta_i = H/H'$ for $H < H'$, which is continuously differentiable on their restricted domain. The line $\{H = H'\}$ separates the two sub-domains of differentiability. The same applies for the modulated frequency.

Lemma 3 – Continuity throughout line loading

Along a transit service $z \in \ell$ *of a transit line, at every station* i , *the functions* k_{zi}' *and* y_{zi} *are continuous and sub-differentiable with respect to the vector* $\mathbf{x}_\ell$. *That also applies for the derived functions. The probability of the stock to board* π_{zi} *is regular for the domain, except if* $q_{ij}^+ = 0$.

__Proof.__ We use an inductive process for the proof of the Lemma property. From the origin station $i = 1$, we have $k_1' = k_z$ for a service and $y_{zi} = 0$. Let an Induction Assumption that the Lemma property holds for a station i and a transit service z . Let us consider the next station, $i+1$. If k_{zi}^- is the available capacity at the departure from station i , the number of passengers alighting at $i+1$, $y_{zi+1}^- = \sum_{j<i+1,\ k=i+1} y_{jk}^z$ and the available capacity $k_{zi+1}' = k_{zi}^- - y_{zi+1}^-$ are regular. Furthermore, in **Leurent (2011b)** it is demonstrated that the passenger stock function $v \mapsto n_{zi} = \sum_{z\in(i,j),\, j>i} v_j$ is continuous. Therefore the probability of the stock to board $\pi_{zi} = \min\{1, \dfrac{k_{zi}'}{n_{zi}}\}$ is continuous for $n_{zi} \geq 0$.

Lemma 4 – Continuity throughout line costing

If the perceived travel time G_{ij}^z *is continuous for each transit service* $z \in \ell$ *with respect to* $\mathbf{x}_\ell$, *then the approximate arc cost function* $c_{ij}^{\ell\varepsilon}$ *is also continuous for the line leg* (i, j) .

__Proof.__ In **Leurent (2010)** the author demonstrated that for a transit service the arc-cost function $G_{ij}^z(\mathbf{x}_\ell)$ is continuous. In addition, we determined in Lemmas 1 and 2 that the approximated probability for the stock to board a vehicle and the modulated frequency respectively are continuous functions. Therefore the approximate arc cost function is also continuous since continuity is maintained through the operators of addition and multiplication. It is also maintained through division since according to Lemma 1 it holds for the approximate function of the probability of the stock to board that $\pi_{zi}^e > 0$ with respect to $\mathbf{x}_\ell$.

Therefore the Lemmas 1, 2, 3 and 4 make Theorem 1 hold true.

Appendix B. Bundle Optimality

Theorem 2 - Optimal Bundle

By optimal bundle we call the bundle that minimizes the expected total travel time, including the waiting time from a node i *to the destination* s . *Therefore, it is optimal when further combination of a line cannot* *improve the expected travel time. The minimum expected travel time of the lines within the attractive set is inferior to the expected travel time.*

Proof. To prove the theorem, we first demonstrate that a bundle is optimal when it contains the lines whose minimum travel time is inferior to the expected travel time of the bundle. Conversely, we show that by withdrawing a line from that attractive set, the new bundle is not optimal.

If B is a bundle and for line ℓ holds $c_{ij}^{\ell} + u_{js} \leq c_{i(s)}^{B} + \alpha / \hat{f}_B$, then bundle $B' = B + \{\ell\}$ is also optimal.

It suggests that if the minimum travel time of the line from the node to the destination is inferior to the expected travel time of the bundle, then the line is part of the attractive set. It is obvious that if we start from that inequality, and for $\hat{f}_B, \hat{f}_a \neq 0$, we get:

$$c_{i(s)}^{B'} + \alpha / \hat{f}_{B'} \leq c_{i(s)}^{B} + \alpha / \hat{f}_B$$

By induction it is obvious that a line bundle is not optimal, unless it contains all the lines whose minimum travel time is inferior to the bundle expected travel time.

Conversely, if B is an optimal bundle and for line ℓ, it holds $c_{ij}^{\ell} + u_{js} \leq c_{i(s)}^{B} + \alpha / \hat{f}_B$, then bundle $B' = B - \{\ell\}$ is not an optimal one.

We advance as previously, by showing that for $\hat{f}_B, \hat{f}_a \neq 0$, it holds:

$$c_{i(s)}^{B'} + \alpha / \hat{f}_{B'} \geq c_{i(s)}^{B} + \alpha / \hat{f}_B$$

Thus, we prove that the bundle is optimal if and only if it contains the lines where their minimum travel time to the destination is inferior to the expected travel time of the bundle.

A special case to be considered is that of the congested line, where the bundle is reduced to the congested line. Since a congested line is considered to provide a continuous service, where the revised frequency $\hat{f}_a \to \infty$, the arc resembles a private arc like the pedestrian arcs. Therefore if a line is congested, it forms a bundle of a single line, whose expected travel time cannot be further reduced by combining with other bundles.

Appendix C. Characterization of Traffic Equilibrium

Theorem 3 – Characterization of Traffic Equilibrium

Let us consider a hyperpath flow state $\mathbf{X}_{NS} = [q_{ns}^{h} : s \in S, n \in N \setminus \{s\}, h \in H_{ns}]$ which is feasible for the OD trip matrix $\mathbf{q}_{NS} = [q_{ns}]_{n \in N, s \in S}$. If $\boldsymbol{\chi}_{NS}(\mathbf{X}_{NS}) = [C_{ns}^{\varepsilon}(h, A(\mathbf{X}_{NS})) : s \in S, n \in N \setminus \{s\}, h \in H_{ns}]$, then $\mathbf{X}_{NS}$ corresponds to a traffic equilibrium if and only if, for any other feasible hyperpath flow state $Y_{NS} = [\xi_{ns}^{h} : s \in S, n \in N \setminus \{s\}, h \in H_{ns}]$, we have:

$$\boldsymbol{\chi}_{NS}(\mathbf{X}_{NS}) \cdot (\mathbf{Y}_{NS} - \mathbf{X}_{NS}) \geq 0 \qquad (7)$$

Proof. Firstly, while considering that $\mathbf{X}_{NS}$ is a traffic equilibrium state, we wish to prove that expression (7) stands. Therefore, for any feasible hyperpath flow state, we have $\xi_{ns}^{h} \geq 0$, for any $s \in S, n \in N \setminus \{s\}, h \in H_{ns}$. According to the definition of the traffic equilibrium, it holds for (8c):

$$\xi_{ns}^{h} \cdot [C_{ns}^{\varepsilon}(h, \mathbf{x}_{AS}) - \mu_{ns}] \geq 0, \forall h \in H_{ns}$$

Therefore, $\xi_{ns}^{h} \cdot C_{ns}^{\varepsilon}(h, \mathbf{x}_{AS}) \geq \xi_{ns}^{h} \cdot \mu_{ns}$.

For all the hyperpaths, $h \in H_{ns}$, we have: $\sum_{h \in H_{ns}} \xi_{ns}^{h} \cdot C_{ns}^{\varepsilon h} \geq \sum_{h \in H_{ns}} \xi_{ns}^{h} \cdot \mu_{ns}$. According to the feasible hyperpath flow state definition (6b) holds: $\sum_{h \in H_{ns}} q_{ns}^{h} = q_{ns}$. And:

$\sum_{h \in H_{ns}} \xi_{ns}^{h} \cdot C_{ns}^{\varepsilon h} \geq \sum_{h \in H_{ns}} \xi_{ns}^{h} \cdot \mu_{ns} = q_{ns} \cdot \mu_{ns}$ (a)

The sum over all the hyperpaths from equation (8d), gives respectively:

$\sum_{h \in H_{ns}} q_{ns}^{h} \cdot C_{ns}^{\varepsilon}(h, \mathbf{x}_{AS}) = \sum_{h \in H_{ns}} q_{ns}^{h} \cdot \mu_{ns} = q_{ns} \cdot \mu_{ns}$

If we combine that expression with (a), we get: $\sum_{h \in H_{ns}} C_{ns}^{\varepsilon h}(h, \mathbf{x}_{AS}) \cdot (\xi_{ns}^{h} - q_{ns}^{h}) \geq 0$.

By summing over all the nodes, $n \in N \setminus \{s\}$, and the destinations, $s \in S$, we build (7).

Secondly, while assuming that the expression holds for $\mathbf{X}_{NS}$, we prove that it is the hyperpath flow state corresponding to the traffic equilibrium. We assume a feasible hyperpath flow state, $\mathbf{Y}_{NS} = [\xi_{ns}^{h} : s \in S, n \in N \setminus \{s\}, h \in H_{ns}]$ almost equal to $\mathbf{X}_{NS}$, with only a difference on a node-destination pair (n, s). If the set of hyperpaths H_{ns} contains only one hyperpath, we can prove that by assuming $\mu_{ns} = C_{ns}^{\varepsilon}(h, \mathbf{x}_{AS})$ and by replacing it in (7), the expressions (6c) and (6d) hold. The flow state $\mathbf{X}_{NS}$ corresponds to traffic equilibrium. If there are more than one hyperpaths, we assume that for a small positive number θ, while $q_{ns}^{h} > 0$ and $\xi_{ns}^{h} \geq 0$, there is $\xi_{ns}^{h} = q_{ns}^{h} - \theta$ and $\xi_{ns}^{h'} = q_{ns}^{h'} - \theta$. Therefore, by applying to (7), we have:

$(C_{ns}^{\varepsilon h'}(h', \mathbf{x}_{AS}) - C_{ns}^{\varepsilon h}(h, \mathbf{x}_{AS})) \cdot \theta \geq 0$

According to the inequality, any hyperpath $h \in H_{ns}$ with a positive flow q_{ns}^{h} in the flow state $\mathbf{X}_{NS}$ has a minimum cost on H_{ns}. We further demonstrate that if we define $\mu_{ns} = \min_{h \in H_{ns}} C_{ns}^{\varepsilon}(h, \mathbf{x}_{AS})$, the (6c) and (6d) hold and the hyperpath flow state corresponds to a traffic equilibrium.

Appendix D. Existence of traffic equilibrium

Theorem 4 – Existence of traffic equilibrium

Assuming that the arc travel cost functions C_a are continuous, there exists an equilibrium state for the user and service equilibrium model.

Proof. The continuity of the travel cost functions guarantees that, according to the Theorems 1 and 3, the cost functions $\mathbf{x}_{AS} \mapsto C_{ns}^{\varepsilon}(h, \mathbf{x}_{AS})$ are continuous with respect to the network flow state. In addition, we have defined that there is a continuous function – corresponding to the combination of the hyperpath flows – such that $\mathbf{x}_{AS} = A(\mathbf{X}_{NS})$. Therefore, the composed functions, $\mathbf{X}_{NS} \mapsto C_{ns}^{\varepsilon}(h, A(\mathbf{X}_{NS}))$ and $\mathbf{X}_{NS} \mapsto \chi_{NS}(\mathbf{X}_{NS})$, are continuous. Since the set of feasible hyperpath flow states is convex and compact, there is at least one solution for the variational inequality, (7). Nevertheless, it is not proven that there is a unique solution.