A Transit Bottleneck Model for Optimal Control Strategies and its Use in Traffic Assignment in Paris

Jian Wen, Fabien Leurent, Xiaoyan Xie

Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2016) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

19th EURO Working Group on Transportation Meeting, EWGT2016, 5-7 September 2016, Istanbul, Turkey

A Transit Bottleneck Model for Optimal Control Strategies and its Use in Traffic Assignment in Paris

Jian Wen [a,b], Fabien Leurent [b,]*, Xiaoyan Xie [b]

[a] *École Polytechnique, Route de Saclay, 91120 Palaiseau, France*
[b] *LVMT, UMR-T 9403, École des Ponts, IFSTTAR, UPEM, UPE, Champs-sur-Marne, France*

Abstract

Control strategies are of crucial importance in operations of saturated transit lines. At a given bottleneck, delay propagates due to an excess of demand over supply. Operators therefore use optimal strategies (e.g. holding, fast-running, train-canceling etc.) in order to recover the level of service. While toolbox for train operation control system has thrived in past years, few traffic assignment models for large-scale transit networks on the planning side have touched on the impact of control dynamics: static macroscopic models perform quite well in simulating platform/vehicle saturation, but less so with bottleneck delay; dynamic models deal with dynamic passenger behavior under congestion, yet single vehicle runs are not explicitly considered either.
The objective of this paper is to (1) capture the dynamic control features in a bottlenecked rail system; (2) provide a computation-friendly simulation approach to optimize the control; and finally (3) fit the bottleneck model in a static capacitated assignment model to improve its performance. The paper first provides mathematical formulae to describe bottleneck phenomena in a high-demand high-frequency transit line. Formulae are then cast into a control problem of dynamic system, to which we apply dynamic programming to reduce the dimension of decision variables. A numerical application to the Line RER A in Paris is followed, in which control strategies are optimized at key stations in a rolling horizon; on-platform passenger activities are simulated accordingly. Results are compared with both static model and real-life data, showing the effectiveness of the control-based assignment.

Peer-review under responsibility of the Scientific Committee of EWGT2016.

Keywords: Transit Bottleneck; Traffic Assignment; Control Strategies; Delay Recovery; Dynamic Programming

* Corresponding author. Tel.: +33-1-81-66-88-54; fax: +33-1-81-66-88-00.
E-mail address: fabien.leurent@enpc.fr

2214-241X
Peer-review under responsibility of the Scientific Committee of EWGT2016.

1. Introduction

Bottleneck delay stems mainly from the excess of demand over supply, as a direct consequence of ordinary peak-hour saturation. At the source (“bottlenecked”) station, willing-to-board passengers delay doors’ closing and trains’ departure. Delay then accumulates through knock-on effect and propagates upwards and downwards along the transit line. Service frequency decreases; platform congestion grows: the service is caught in a vicious cycle. One typical example is the Line RER A in Paris, one of Europe’s busiest transit lines with over 1 200 000 passengers per day. During the morning peak, the service is severely delayed as a result of the huge demand. Report (STIF, 2012) shows that westbound service frequency is reduced by 19% on average, from 30 to 24.4 trains per hour.

Pender et al. (2012) find it an established practice to use rescheduling strategies (holding, fast-running and train-canceling) in order for operators to deal with bottleneck delay, with which the level of service can be effectively recovered with a relatively low control cost. However, past research on transit traffic assignment models has not totally explored it. Some assumed that frequency reduction effect does not propagate; many others adapted the reduced frequency uniformly to all stations. Both simplifications appear to be off from real-life practice, failing to simulate not only the spatial and temporal accumulation of delays, but also the efforts made by operators to apply dynamic controls for schedule recovery.

The paper aims at modeling the bottleneck delay along with the control strategies for its recovery, and finally adding dynamic operation features to the static assignment model. The rest of the paper is divided into four parts: Section 2 reviews different approaches of modeling bottleneck dynamics in the literature, linking the perspectives from both planning side and operation side. Section 3 develops the mathematical model, which describes operation conditions (rolling stock, passenger activity), system constraints (rail capacity, signaling system) and operators’ skillset in bottleneck recovery. The mathematical model is transformed into a dynamic programming problem, leading to an approximate solution which significantly reduces the computation time. In section 4, an application instance based on RER A is carried out, targeting a high-frequency rail transit system of both headway-based and schedule-based regulations. The results demonstrate how different strategies may influence the level of service. Section 4 also justifies the compatibility of this bottleneck model with the traditional traffic assignment models. Section 5 draws the conclusion of the study and indicates some directions for further research.

2. Bibliographical Review

On the planning side, traffic prediction follows the sequential four-step process, traffic assignment being the last step. While early static transit assignments usually assumed free-flow boarding and running, many recent studies have been focused on congestion condition. Lam et al. (2002) propose a formulation for the capacity restraint assignment, in which line frequency is related to the passenger flows due to the fixed fleet size. Leurent and Liu (2009) and Schmöcker et al. (2011) study the seat capacity as another factor of congestion. Leurent et al. (2014) deal with a larger range of capacity phenomena and demonstrate that frequency also depends on track occupancy, since vehicle clearance time and safety margins have to be ensured. A static traffic assignment model is put forward to simulate the bottleneck: downstream frequencies are reduced by the same value; upstream links share the delay by increasing the running time. However, this model does not consider bottleneck phenomena as dynamic.

Dynamic assignment models, on the other hand, have been largely focused on passenger behavior. Nuzzolo et al. (2012) present a schedule-based model to investigate vehicle congestion, in which user behavior is affected by on-vehicle discomfort and operational delay. Schmöcker et al. (2008) propose a frequency-based dynamic approach, with attention being paid to platform congestion and passenger queuing caused by bottleneck. The “failing-to-board” probability is introduced in their study, thanks to which the simulated wait time gets much closer to the real case. However, it’s still underestimated due to the strongly simplifying assumptions about service regularity

On the operation side, near-capacity operation has been long tied to control strategies. Eberlein et al. (2001) analyze real-time control strategies with a general model. Their deterministic model tests the strategies of holding, stop-skipping and deadheading. Shen and Wilson (2001) add the delay to the system. The results show that holding strategies combined with short-turning can reduce the objective by about 10-60%. Sánchez-Martínez et al. (2015) formulate a holding optimizing model that reflects dynamic running times and demand. The model proves that control based on dynamic inputs outperforms its static equivalent in overcrowded cases and optimization-based control leads to better performance than target or mean headway strategies. Cats (2011) also evaluates different holding criteria and number and location of time point stops. Both greedy and genetic algorithms are used to get optimal strategies. Dynamic runs are also taken into consideration in path choice model for assignment, but a comprehensive assignment model is still to be tested.

Our bottleneck model is based on the static and macroscopic capacitated assignment model (static CapTA) of Leurent et al. (2012, 2014), while operation dynamics, having been long studied at the operation side, are introduced to describe the optimal controls. The following section will present the framework of our model, together with some mathematical process to make it precise, modularized and computation-friendly.

3. Bottleneck Model

Some key features such as mingled queuing and capacitated random boarding of the line model in static CapTA are maintained. However, the sub-track model is reformed throughout to simulate the bottleneck in a hybrid way, combining a macroscopic but time-dependent model of passenger demand with a microscopic model of vehicle runs. Given a bottleneck, we suppose that operators control trains to hold or run fast (stacking and holding) in order to avert knock-on effects and recover the schedule. The controllable variables include both dwell time and running time. Effects are evaluated on the basis of both regularity and punctuality. If the track occupancy exceeds its maximal value for a relatively long period, trains may also be cancelled (train-cancelling) with extra cost. All these suppositions constitute the backbone of our model: microscopic runs, macroscopic but non-constant arrivals as well as objective-based control.

The following part of this section is organized into five subsections to bring out our bottleneck model step by step. Subsection 3.1 analyzes the causes of bottleneck from both operator and passenger sides, taking RER A as a case study. Subsection 3.2 abstracts from the case the general settings of a high-demand high-frequency transit line, from which we start to build the mathematic model and solve it. System constraints and objective function are presented respectively in Subsection 3.3 and 3.4, and then assembled into a quadratic optimization program in Subsection 3.5. Subsection 3.5 also provides the formulation of control system for optimal strategies and the related mathematical process in order to reduce the complexity.

3.1. Passenger Arrivals and Vehicle Runs

The Line RER A in Paris provides a perfect introduction to bottleneck phenomena.

Most of the traffic of RER A takes place within the central trunk, between Vincennes and La Défense. During the morning peak (MPH, 7 a.m. to 9 a.m.) in a working day, up to 50,000 passengers travel through this segment every hour. RER A is obliged to run at full capacity, planning 30 westbound trains per hour. However, due to the huge demand at important transfer stations as well as constraints of infrastructure, the target frequency of 30 vehicles per hour (vph) is hardly achieved. The most problematic station is Châtelet - Les Halles, a transfer hub for 7 other transit lines. During MPH, trains are often delayed by long dwells because of vehicle-platform exchange.

To fully examine the operation of RER A, we use a Python bot to collect data from RATP real-time Passenger Information System (PIS), which provides information on train arrivals (time of next train, train approaching, train at platform etc.) for passengers via screens, webpages and smartphone applications. Working day operation data of RER A are recorded on a second-to-second basis during a two-week period between 7th and 18th, Dec 2015. The PIS data are then transformed to arrival / departure time at each station. Official Automatic Vehicle Location (AVL) data from RATP, the operator, validate the authenticity of our PIS data.

During MPH, RATP permits fulfilling the schedule with 10% flexibility, i.e. finishing the one-hour plan in 66 minutes. We focus therefore on a study period of 1.1 hour, starting from 8:10. All trains that arrive at Châtelet - Les Halles between 8:10 and 9:16 are taken into account. Observations from PIS suggest that passenger arrival experiences a significant “crest” (see Figure 1), during which the dwell time goes beyond the threshold of 60 s. Supposing that dwell time follow a normal distribution, we test and finally accept the hypotheses at a significance level of 0.05 that the dwell time mean during the “crest” (8:30 - 8:45) is larger than the mean during the rest of the study period (8:10 - 8:30 and 8:45 – 9:16). Statistics show that the average dwell time during the crest is 73.8 s, compared with 58.0 s during the rest.

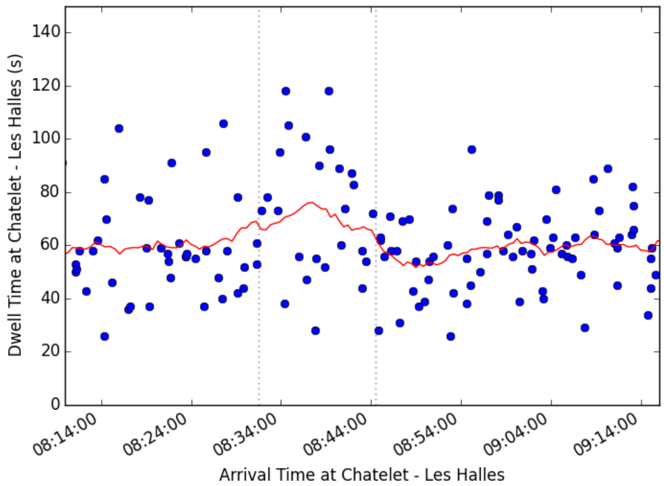


Fig. 1. Dwell time at Châtelet - Les Halles (the red line represents the average dwell time of 11 nearest neighbors).

Why is the threshold of 60-second dwelling important? To illustrate, we also need to investigate the operation situation of RER A. As mentioned, RATP targets a frequency of 30 vph, equivalent to the planned headway of 120 s, while the signaling system maintains a separation time of at least 60 s (including 30 s for entering / leaving the station plus 30 s as safety margin) between one departure and the next arrival. This fact limits the dwell time at each station to 60 s at most to avoid over-occupancy. Once the dwell is longer than this threshold, knock-on delay occurs inevitably and the bottleneck arises as a result. Therefore, having an "unacceptably long" dwell time, the "crest" should be highlighted in the study of bottleneck. We name this period as "bottleneck period". The rest of the study period is called "control period", since control strategies are used in response.

3.2. System Settings

The mathematical notations used in this paper can be found in Table 1.

Let us now consider a rail transit line that serves the urban area. This transit line captures some of the key features of the Line RER A in Paris: (1) highly trafficked in the central segment; (2) frequently saturated at peak hours though run at full capacity; (3) operated on the basis of both regularity and punctuality.

Line Settings. The studied segment consists of S stations, represented by the set $\mathcal{S}=\{1,2,\cdots,S\}$. Services are run independently in two directions (i.e. no constraints at terminals) so that we can focus on any of the two. Each direction provides one service mode, with V vehicles running one after another. $\mathcal{V}=\{1,2,\cdots,V\}$. Over-taking, stop-skipping and short-turning are not allowed in this system for the sake of simplicity.

Controllable Variables. The service runs with overall frequency f during the study period $\mathcal{H}$, keeping the arrival headway h. For a given vehicle $z\in\mathcal{V}$ at station $j\in\mathcal{S}$, the actual headway is denoted by $h_{z,j}$. Arrival time, separation time from the previous departure, dwell time, departure time and running time to next station $j+1$ are noted respectively as $t^a_{z,j}$, $t^s_{z,j}$, $t^w_{z,j}$, $t^d_{z,j}$, $t^r_{z,j}$. These variables are controllable in the operation. A bar applied to any notation indicates the planned value (e.g. $\bar{f}$ the planned frequency); a hat indicates the minimum limit due to system constraints (e.g. $\hat{t}^r_j$ minimum running time from j to $j+1$). The timing is deterministic if not otherwise specified.

Signaling System. The transit line is equipped with a two-aspect fixed-block signaling system. All stations are blocked at two ends; interstation links are separated by n_j fixed blocks between stations j and $j+1$. Signaling system keeps a minimum safety separation $\hat{t}^s$ between two successive vehicles.

Table 1. Mathematical notations

Notation	Definition
$\mathcal{H}$	Study period of duration H . $\mathcal{H}_b$ the bottleneck period. $\mathcal{H}_c = \mathcal{H} \setminus \mathcal{H}_b$ the control period.
$\mathcal{S}$	Set of S stations along the line.
$\mathcal{V}$	Set of V vehicles running during $\mathcal{H}$.
f	Frequency. $\bar{f}$ the planned frequency. $f_{z,j}$ the actual frequency of vehicle $z \in \mathcal{V}$ at station $j \in \mathcal{S}$.
h	Arrival headway.
t^a	Arrival time. $\mathbf{T}_j^a$ the vector of arrival times of all trains at station j .
t^s	Arrival separation time with the previous departure. $\hat{t}^s$ the minimum separation time for safety.
t^w	Dwell time.
t^d	Departure time.
t^r	Running time to the next station.
n_j	Capacity of interstation block between stations j and $j+1$.
q_j	Passenger arrival rate at station j during $\mathcal{H}_c$.
α_j	Bottleneck period arrival rate ratio (compared with q_j) at station j .
ω_j	Average wait time at station j .

Bottleneck. At the bottlenecked station i , the demand exceeds the supply during the bottleneck period $\mathcal{H}_b$, when intensive boarding / alighting activity prolongs dwell time to at least $\hat{t}_i^{w,b}$. If the planned headway cannot be guaranteed due to the long dwell, delay accumulates ever since. During the rest of the study period (control period) $\mathcal{H}_c$, the dwell time can be reduced to $\bar{t}^w$.

3.3. System Constraints

Based on the settings made in Subsection 3.1, we are well-prepared to model this transit line. System constraints and variable interdependences will be clarified in this subsection.

Variable Interdependence. For any $z \in \mathcal{V}$ and $j \in \mathcal{S}$:

$$t_{z,j}^w = t_{z,j}^d - t_{z,j}^a \tag{1}$$

$$t_{z,j}^s = t_{z,j}^a - t_{z-1,j}^d \tag{2}$$

$$h_{z,j} = t_{z,j}^a - t_{z-1,j}^a \tag{3}$$

$$t_{z,j}^r = t_{z,j+1}^a - t_{z,j}^d \tag{4}$$

The equations (1) to (4) gives the definitions for dwell time, arrival separation, arrival headway and running time. These variables are interdependent and (1) to (4) should be respected at all times.

Signaling Constraints. The signaling system (see Subsection 3.2) plays the role to keep the traffic in order and avoid possible rear-end accidents. In this paper, we concentrate on bottleneck phenomena at platforms, while details on interstation runs would be out of scope. In that way, we suppose further that the interstation segment between stations j and $j+1$ (with n_j fixed blocks) be treated as one entire block ("interstation block"). This block can hold n_j vehicles at maximum. To put another way, the interstation block can be modeled with a first-in-first-out queue with capacity of n_j vehicles.

Under the suppositions made above, signaling constraints can be written as follows:

$$t^s_{z,j} \geq \hat{t}^s \tag{5}$$

$$t^d_{z,j} \geq t^a_{z-n_j,j+1} \tag{6}$$

Equation (5) presents the minimum separation. Equation (6) depicts the function of interstation block: the train z cannot depart from station j and enter the interstation block if trains from $z-n_j$ to $z-1$ are all in. Vehicle z has to hold until vehicle $z-n_j$ arrives at the next station $j+1$ and leaves a vacancy.

Regulation Constraints. These constraints reflect some basic principles of running:

$$t^r_{z,j} \geq \hat{t}^r_j \tag{7}$$

$$t^w_{z,j} \geq \hat{t}^{w,b}_i \text{ if } j = i \text{ and } t^a_{z,j} \in \mathcal{H}_b; \text{ otherwise } t^w_{z,j} \geq \bar{t}^w_j \tag{8}$$

Equation (7) states that trains cannot run faster than a speed limit. Equation (8) puts constraints on dwell time.

Schedule Constraints. The trains out of the study set $\mathcal{V}$ are supposed to fulfill perfectly the schedule. In this case, the arrival time of the first train and the departure time of the last train should be limited to the study period $\mathcal{H}$ for all stations $j \in \mathcal{S}$ so that there is no conflict with neighboring trains:

$$t^a_{1,j} \geq \bar{t}^a_{1,j} \text{ and } t^d_{V,j} \leq \bar{t}^d_{V,j} \tag{9}$$

3.4. Control Strategies and Objective Function

Vuchic (2005) mentions that when headway is regular and smaller than 4 minutes, no one consults the schedule any more, persuading transit operators to invest in headway-based regulation for high-frequency systems. However, schedule-based regulation still plays a part for systems sensitive to delays. For example, RER A, having two branches sharing tracks with other commuter lines, has to run under hybrid regulation: schedule-based operation in branches to avoid conflict with other trains while high-frequency headway-based operation in the busy central trunk.

Without the support of siding, crossover and spare vehicles, three control strategies are commonly used:

- Stacking: schedule-based, permitting all knocked-on trains to run as close as possible to the affected ones to catch the schedule (while running ahead of schedule is generally not permitted if no delay happens).
- Holding: headway-based, slowing down both preceding and following trains to regularize the headways.
- Train-cancelling: cancelling trains at terminus if the capacity of tracks is saturated.

The schedule-based control is measured by the punctuality. The objective function is therefore intuitive:

$$I_{punc} = \sum_{j \in \mathcal{S}} \sum_{z \in \mathcal{V}} (t^a_{z,j} - \bar{t}^a_{z,j})^2 \tag{10}$$

Supposing q_j the passenger arrival rate during $\mathcal{H}_c$ and $\alpha_j q_j$ during $\mathcal{H}_b$, the headway-based holding control, on the other side, can be evaluated by the regularity at each station weighted by the arrival flow.

$$I_{rglr} = \frac{\sum_{j\in\mathcal{S}} q_j \left[\alpha_j \cdot \sum_{z\in\mathcal{V}, t^a_{z,j}\in\mathcal{H}_b} (h_{z,j})^2 + \sum_{z\in\mathcal{V}, t^a_{z,j}\notin\mathcal{H}_b} (h_{z,j})^2 \right]}{\sum_{j\in\mathcal{S}} q_j \cdot H} \tag{11}$$

3.5. Optimization Program and Complexity Reduction

The general objective function evaluates both regularity and punctuality. The cost of cancelling trains will be taken into consideration as well.

$$I = \beta_1 I_{punc} + \beta_2 I_{rglr} + c_{cancel} \tag{12}$$

Based on Equation (12), we have finally the optimization problem:

$$\min_{\substack{t^w_{z,j}, t^r_{z,j} \\ \text{for } \forall z\in\mathcal{V}, j\in\mathcal{S}}} I, \text{ s.t. system constraints and variable interdependence} \tag{13}$$

This problem, quadratic, convex and linearly constrained, can be solved directly by the interior point algorithm (named “optimal solver”). However, the computation is costly in time, especially when a large control set is given. To simplify, we treat the entirety of the rolling stock as a discrete dynamic system, the states of which are characterized by the series of arrival times at each station. Consequently, we can rewrite the objective function as a cost functional for control problem: $\sum_{j\in\mathcal{S}} (\mathbf{T}^a_j)^{\mathrm{T}} \mathbf{G}_j \mathbf{T}^a_j$ (the constant c_{cancel} is omitted), where $\mathbf{T}^a_j$ is the vector of arrival times at station j; $\mathbf{G}_j$ is a positive definite matrix. The control problem can then be reformulated by using dynamic programming method according to Hamilton-Jacobi-Bellman equation, which breaks the decision apart into a sequence of smaller decisions at each station, $\mathbf{T}^w_j$ and $\mathbf{T}^r_j$ being control variables. The dynamic programming starts from the bottlenecked station i to upstream and downstream stations respectively:

$$V_{up}(\mathbf{T}^a_k, k) = \min_{\mathbf{T}^w_j, \mathbf{T}^r_j} \sum_{j=1}^{k} (\mathbf{T}^a_j)^{\mathrm{T}} \mathbf{G}_j \mathbf{T}^a_j = \min_{\mathbf{T}^w_k, \mathbf{T}^r_k} \left\{ (\mathbf{T}^a_k)^{\mathrm{T}} \mathbf{G}_k \mathbf{T}^a_k + V_{up}\left(\mathbf{T}^a_{k-1}, k-1\right) \right\} \tag{14}$$

and,

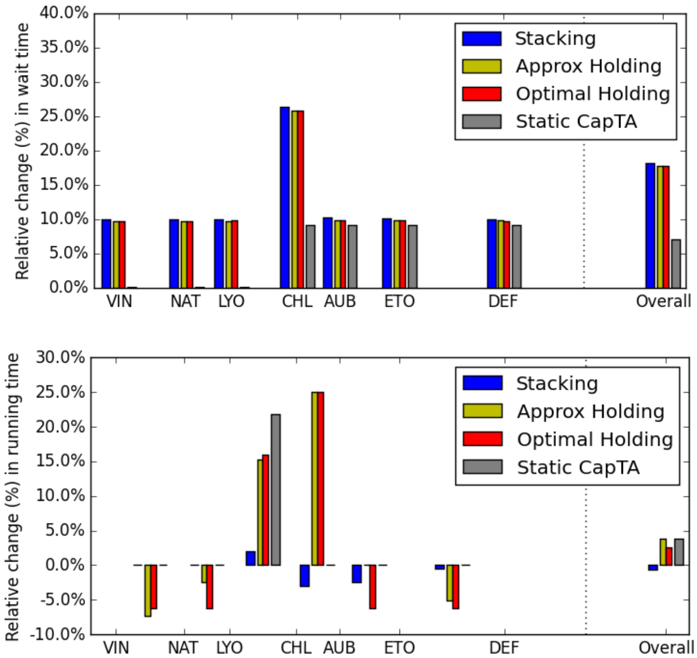
$$V_{down}(\mathbf{T}^a_k, k) = \min_{\mathbf{T}^w_j, \mathbf{T}^r_j} \sum_{j=k}^{S} (\mathbf{T}^a_j)^{\mathrm{T}} \mathbf{G}_j \mathbf{T}^a_j = \min_{\mathbf{T}^w_k, \mathbf{T}^r_k} \left\{ (\mathbf{T}^a_k)^{\mathrm{T}} \mathbf{G}_k \mathbf{T}^a_k + V_{down}\left(\mathbf{T}^a_{k+1}, k+1\right) \right\} \tag{15}$$

This dynamic control problem is much less demanding in time if we apply approximate algorithms (named “approximate solver”) to reduce the number of variables.

4. Application

The model is applied to simulate RER A.

Inputs. The application instance consists of 7 stations in the RER A central trunk, Vincennes to its easternmost, and La Défense to its westernmost. We set the prolonged dwell time $\hat{t}^{w,b}_i = 75$ s at Châtelet - Les Halles during $\mathcal{H}_b$, compared with the normal dwell time $\overline{t}^w_j = 60$ s. The bottleneck flow rate ratio $\alpha_j = 1.25$ for all stations. Other inputs (planned frequency, planned dwell time, planned running time, arrival rate, capacity etc.) remain identical to those used in the static CapTA.



(a)

(b)

Fig. 2. Relative change in dwell time and running time (the deterministic case).

Deterministic Case. In the first place, we simulate two scenarios, the stacking-only strategy and the holding-only strategy with the aforementioned settings. For holding, two solvers (the optimal and the approximate) mentioned in Subsection 3.5 are distinguished to demonstrate the performance of dynamic programming.

Figure 2 illustrates how different strategies may impact on wait time and running time. As first glance, we notice that holding enhances the robustness of transit systems when facing bottleneck (overall wait time 70.69 s by optimal holding vs. 70.95s by stacking), although this strategy has "side effects" of producing irregular running times as an immediate consequence of spacing. As a comparison, stacking has much less influence on running time, since trains are deprived of the ability to "cooperate" with other stakeholder trains. Results on wait time also justify the static CapTA's supposition that downstream stations suffer from a reduction in frequency. However, the unrealistic upstream assumption of static CapTA fails to model the knock-on effect, while the proposed bottleneck model successfully simulates the propagation along the line from a global perspective.

The most remarkable difference exists at Châtelet - Les Halles (CHL). During $\mathcal{H}_b$, the simulation implies that the boarding probability drops to 0.87 due to the limit capacity of vehicles, which therefore results in a significant increase in wait time. In contrast, the static CapTA, taking a constant arrival rate, could not predict this saturation.

Computational Performance. We can also examine from the simulation the effectiveness of the approximate solver. This solver produces almost identical results to those from the optimal solver (overall wait time 70.70 s vs. 70.69s, overall running time 132.2 s vs. 130.5 s). Its computation time on a personal computer with a 2.50 GHz processor is less than 10 ms, compared with 110 ms by its optimal counterpart.

This performance is fairly good, and application to large-scale assignment problems is thus feasible. The entire transit network of the Greater Paris Metropolitan Area has some 200 rail transit services as well as around 4000 bus services. Supposing that 20% of the rails service and 10% of the bus service are bottlenecked during the morning peak, an iteration of this model will require the supplementary computation time of 5 minutes. Currently, CapTA performs 20 iterations in 11 hours. The bottleneck model will then add approximately 100 minutes to an entire execution of the assignment model.

Stochastic Case. The deterministic case attacks the problem from an idealistic view, yet the real-life operation is not as predictable. The primary stochastic factor residing in the model is the dwell time, as Figure 1 has shown. For this reason, we suppose furthermore that the passenger arrival rate follows the Poisson distribution. The minimum dwell time, proportional to the number of boarding passengers, can be therefore described as a Poisson process, the mean of which remains $\lambda = \hat{t}_i^{w,b}$ at Châtelet - Les Halles during $\mathcal{H}_b$ and $\lambda = \overline{t}_j^{\,w}$ otherwise.

We simulate for 10 times the holding scenario with the approximate solver and compare the results with both deterministic case and PIS data as illustrated in Figure 3. The dwell time of stochastic holding shows consistent results with that of PIS. The simulated running time is not as good, providing a significant overestimation at Châtelet - Les Halles. This may imply that RATP does not put completely the headway-based holding into practice: the real strategy is rather a combination of holding and stacking. The difference also suggests that the vehicle run model is oversimplified to some extent, although such simplification is not a detriment to our conclusion.

As for the platform activity, the stochastic case gives an even longer wait time than the deterministic case does. This can be explained by the irregularity associated with the stochasticity. Such a long wait is considered to better reflect the reality of bottleneck congestion. The real-life wait time, however, still requires further exploration of the combination of AVL and Smart Card data.

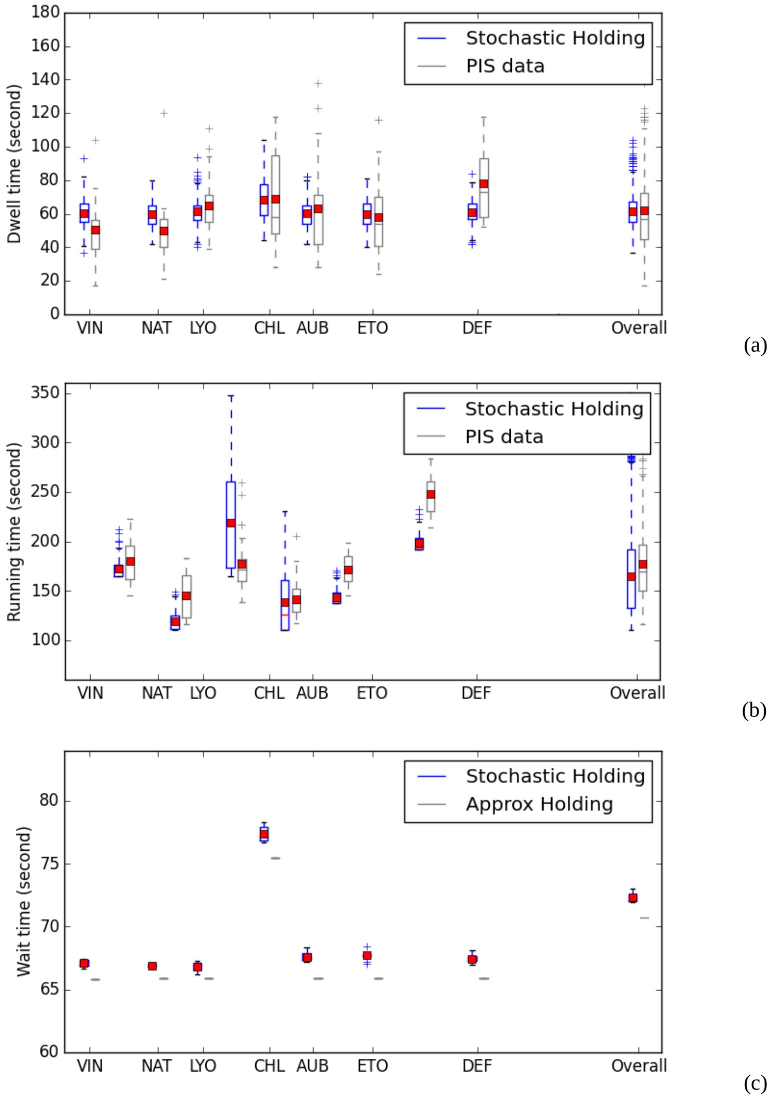


Fig. 3. Comparison of dwell time, running time and wait time (the stochastic case).

5. Conclusion

The bottleneck model establishes a framework to represent a wide range of strategies used by operators in bottleneck delay recovery process. The simulation proves that the bottleneck model outperforms static CapTA in predicting wait time and running time. It also demonstrates a good consistency with the running details in real-life practice. Furthermore, the concurrent accuracy and efficiency of the deliberately designed approximate solver promises future applications to large-scale assignment problems in planning and operation.

We have already focused our next step on integrating this bottleneck model to static CapTA and testing its performance in the transit network in Paris. Besides this test, many other developments can be also thought of:

- to model unpredictable incidents and to optimize the recovery;
- to formulate other strategies, e.g., stop-skipping, short-turning;
- to take into consideration the terminal constraints and rolling stock constraints; . . .

Each of the aspects is of great potential for research. With the development and wide application of automated data system, we believe that the research on this field will hopefully gather pace.

Acknowledgements

This research was supported by the Research and Education Chair on “Socio-economics and Modeling of Urban Transit”, operated by ENPC (Ecole des Ponts ParisTech) in partnership with STIF (Transport Organizing Authority in the Paris Region), to whom we are grateful. We would also like to give our thanks to Shaoqing Wang from LVMT for her suggestions and provision of the CapTA data, and to RATP (Autonomous Operator of Parisian Transports) for its AVL data.

References

Cats, O., 2011. Dynamic modelling of transit operations and passenger decisions.

Eberlein, X.J., Wilson, N.H. and Bernstein, D., 2001. The holding problem with real–time information available. Transportation science, 35(1), pp.1-18.

Lam, W.H., Zhou, J. and Sheng, Z.H., 2002. A capacity restraint transit assignment with elastic line frequency. Transportation Research Part B: Methodological, 36(10), pp.919-938.

Leurent, F, Liu, K. 2009. On seat congestion, passenger comfort and route choice in urban transit: a network equilibrium assignment model with application to Paris. Proc. TRB Annu. Congr. 2009, , Pap. #09-1784.

Leurent, F. and Chandakas, E., 2012. The transit bottleneck model. Procedia-Social and Behavioral Sciences, 54, pp.822-833.

Leurent, F., Chandakas, E. and Poulhès, A., 2014. A traffic assignment model for passenger transit on a capacitated network: Bi-layer framework, line sub-models and large-scale application. Transportation Research Part C: Emerging Technologies, 47, pp.3-27.

Nuzzolo, A., Crisalli, U. and Rosati, L., 2012. A schedule-based assignment model with explicit capacity constraints for congested transit networks. Transportation Research Part C: Emerging Technologies, 20(1), pp.16-33.

Pender, B., Currie, G., Delbosc, A. and Shiwakoti, N., 2012, September. Planning for the unplanned: An international review of current approaches to service disruption management of railways. In Australasian Transport Research Forum (ATRF), 35th

Sánchez-Martínez, G.E., Koutsopoulos, H.N. and Wilson, N.H.M., 2015. Event-Driven Holding Control for High-Frequency Transit. In Transportation Research Board 94th Annual Meeting (No. 15-4611).

Schmöcker, J.D., Bell, M.G. and Kurauchi, F., 2008. A quasi-dynamic capacity constrained frequency-based transit assignment model. Transportation Research Part B: Methodological, 42(10), pp.925-945.

Schmöcker, J.D., Fonzone, A., Shimamoto, H., Kurauchi, F. and Bell, M.G., 2011. Frequency-based transit assignment considering seat capacities. Transportation Research Part B: Methodological, 45(2), pp.392-408.

Shen, S. and Wilson, N.H., 2001. An optimal integrated real-time disruption control model for rail transit systems. In Computer-aided scheduling of public transport (pp. 335-363). Springer Berlin Heidelberg.

STIF, 2012. Schéma Directeur du RER A. Report prepared for the Syndicat des transports d'Île-de-France, available on-line at the following website address: http://www.stif.org/IMG/pdf/Deliberation_no2012-0163_relative_au_schema_directeur_du_RER_A.pdf.

Vuchic, V.R., 2005. Urban transit: operations, planning, and economics.